# Using Bluetooth and WiFi to unravel real-world slow mode activity travel behaviour

van den Heuvel, J.P.A.; Ton, Danique; Hoogendoorn, S.P.; Daamen, W.

**Publication date**
2015

**Document Version**
Final published version

**Citation (APA)**
van den Heuvel, J. P. A., Ton, D., Hoogendoorn, S. P., & Daamen, W. (2015). *Using Bluetooth and WiFi to unravel real-world slow mode activity travel behaviour*. Paper presented at 14th International Conference on Travel Behaviour Research, Windsor, United Kingdom.

# Using Bluetooth and WiFi to unravel real-world slow mode activity travel behaviour

Winnie Daamen (w.daamen@tudelft.nl); corresponding
Jeroen van den Heuvel (j.p.a.vandenheuvel@tudelft.nl / jeroen.vandenheuvel@nsstations.nl)
Danique Ton (danique.ton@nsstations.nl)
Serge Hoogendoorn (s.p.hoogendoorn@tudelft.nl)

## Abstract

Slow modes have an increasing share in urban mobility. The lack of accurate revealed data has so far hampered scientific research aimed at unravelling slow mode mobility.

Multiple types of data can be collected to gain better insight into slow mode transport and traffic operations, such as counts on specific locations (cross-sections), distributions of flows over the network and dynamics thereof. Typical data collection techniques for vehicular traffic, such as induction loops, cannot be applied, among other things due to the fact that slow modes are not restricted to lanes. Therefore, other, non-intrusive, ways to collect these data need to be investigated.

In our paper we look at the applicability of Bluetooth (BT) and WiFi sensors to collect data on pedestrian and cycle flows, using two case studies. The first case study covers the data filtering process, to come from the raw sensor data to the information necessary for behavioural research. It describes the application of 9 sensors in the inner city of Amsterdam. The second case study deals with a BT/WiFi sensor network, installed in the station of Utrecht, the Netherlands. Using these data, we have successfully estimated choice models for the route choice and activity choice behaviour of departing train travellers, showing the potential use of BT/WiFi as a (revealed) data source for modelling travel behaviour in a station.

### Slow modes in urban mobility

Walking and cycling - “slow modes” - are important in urban mobility (Ravalet, 2013, Harms et al, 2014). In The Netherlands, bikes are used for over one out of four trips (27%), while approximately one out of six trips consists of walking (16%). In the cities of Amsterdam and Utrecht, the modal share of bikes is 35-40% (Ministerie van Infrastructuur en Milieu, 2013). In the last decades, growth for bike traffic has concentrated in the urban areas, while rural areas have shown a decrease (CROW-Fietsberaad, 2014). For example, in the city of Amsterdam the share of cyclists increased from 33% to 53%, while at the same time the shares of cars (from 39% to 24%) and public transport (from 27% to 21%) decreased (numbers extracted from the yearly national survey on mobility). Similar trends are seen in other traditional biking countries such as Denmark and Germany, with shares ranging between 10 and 20%, but also in countries where cycling is less popular (Harms et al, 2014).

Although it is excellent for urban quality, sustainability and accessibility of city centres, growth of slow modes cause various challenges when they are too successful: (over)crowded city streets, traffic jams, and bicycle parking problems. For example, bicycle traffic jams in Utrecht city centre are reported frequently on social media, for example on 24 April 2014 (Massa, 2014). The main shopping street of Amsterdam city centre (Kalverstraat) has been closed several times due to overcrowding. Pedestrian traffic in this street has grown spectacularly by 50%, from 50,000 on an average Saturday in 2012 to almost 75,000 in 2014 (Parool, 2014). Similar challenges occur at large train stations in The Netherlands. Examples are the closure of platforms at Amsterdam Central station on 5 December 2013 due to service disruptions caused by a storm (RTL Nieuws, 2013). More recently, platforms of the train station of Amsterdam Airport Schiphol were overcrowded due to a major power grid failure. See Figure 1 for some impressions.

**Figure 1: Clockwise, starting from upper left picture: Bicycle traffic jams in Utrecht city centre (24-04-2014), crowding at Kalverstraat in Amsterdam City centre (28-12-2013), closed platforms at Amsterdam Central station (5-12-2013), and crowding at Amsterdam Airport Schiphol train station (7-4-2015).**

Similar challenges are reported in other countries. A famous example is the historical inner city of Venice, Italy. The city centre consists of a network of waterways and footways, often limited in width. On busy days, tens of thousands of tourists visit the city and create – together with regular city traffic – a large pressure on the urban infrastructure (Mamoli et al, 2012). Another example is the annual Festivities in the Belgium city of Ghent, which attract approximately 1.5 million visitors to the historical city centre during the 10-day event (Versichele et al, 2012).

The main question is when do these challenges evolve into problems (so far, no major incidents have been reported) and how large will these problems really be, and what can be done to solve them. To this end, a better understanding of the underlying activity and travel behaviour, as well as the resulting traffic dynamics is required. However, the lack of accurate, revealed data has so far hampered scientific research aimed at unravelling slow mode mobility. Due to recent technological developments, new methods of data collection in traffic have become available. Some of these new technologies have found their way to slow mode traffic research. Examples of these new techniques are Bluetooth and WiFi sensors, sensors detecting the number of devices with an active Bluetooth and / or WiFi connection in their field of view. In this paper, we will give an overview of existing data collection techniques and show the possibilities of applying Bluetooth and WiFi sensors for slow modes.

This paper starts with an overview of traditional (mostly manual) data collection methods for slow modes, followed by an overview of automated data collection methods. After that, we give a technical description of Bluetooth and WiFi sensors, resulting in a comparison of traditional data collection techniques and these Bluetooth and WiFi sensors. To show the practical applicability of Bluetooth and WiFi sensors we introduce two applications: the first case showing the filtering process deriving data on trips in the city centre of Amsterdam, while the second case covers Utrecht Central Station focusing on the estimation of choice models, while the second case. We end with conclusions and recommendations for further research.

**Traditional data collection methods for vehicular traffic**

Real-world (non-laboratory) traffic data collection methods can be deployed for local and global traffic flow measurements. With “local”, we refer to the traffic dynamics at a specific location in the network, for example a junction, square or an entrance. In contrast, “global” refers to the traffic flows in the network, for example a ring road, a city centre or a pedestrian facility. The traffic measurements themselves can be classified into microscopic and macroscopic perspectives. The microscopic perspective consists of data of movements of individual persons, i.e. trajectories, routes or travel times. The macroscopic perspective consists of data of movements of traffic flows, in which individuals cannot be distinguished. Examples are flows and densities. An overview of the data for each category is shown in the table below. This table holds both for vehicular traffic and for slow modes.

**Table 1: Overview of data for the distinguished categories.**

| | | **Measurement objective** | |
|---|---|---|---|
| | | **Local** | **Global** |
| **Measurement perspective** | **Microscopic** | Trajectories | Routes<br>Travel times |
| | **Macroscopic** | Counts<br>Speed distribution<br>Flow patterns over time | Densities (distribution) over network<br>OD matrix<br>Travel times |

An overview of the traditional vehicular data collection methods is given in Table 2.

**Table 2: Classification of traditional vehicular data collection methods.**

| | | Measurement objective | |
|---|---|---|---|
| | | **Local** | **Global** |
| **Measurement perspective** | **Microscopic** | Video cameras<br>Probe vehicles | License plate recognition<br>Video cameras<br>GPS<br>RFID<br>Surveys |
| | **Macroscopic** | Inductive loops<br>Pneumatic tube<br>Radar<br>Laser<br>Video cameras<br>Probe vehicles<br>Infrared | |

In the following, we give an overview for the data collection methods used for slow modes. Here, we distinguish between manual and automated data collection, as the first ones are the more traditional techniques, while the second ones are more recent innovations and not yet widely used in slow mode research.

**Manual data collection methods for slow modes**

Compared to car traffic, it is relatively complex to observe slow modes with the traditional traffic data collection methods. Most typical data collection techniques for vehicular traffic (see Table 2) cannot be applied, since slow modes are not restricted to lanes and they do not have a visible unique identifier (license plate). We first focus on the manual data collection techniques that have been applied for slow modes. Global macroscopic measurements, for example aerial observations, require a view at slow mode traffic flows from high altitude and without interference of the line of sight. A classic pedestrian study which is (partially) based on aerial observations is the work of Pushkarev and Zupan (1975). Global macroscopic measurements are a challenge in urban areas and cannot be applied in indoor environments, for example in train stations. Local macroscopic measurements require many observations, often at several sites depending on the attribute(s) the researcher is interested in. An example is the estimation of the maximum flow or critical density in a bottleneck, for example stairways or escalators. A classic pedestrian study in this field is the work of Fruin (1987). When aiming at a significant set of observations, with a sufficient range of flow conditions, local macroscopic measurements are labour-intensive time-consuming activities, and therefore expensive to implement. Moreover, human observers tend to underestimate the real number of pedestrians in more complex situations (Timmermans, 2009).

Manual measurement methods also cause challenges when deployed for microscopic measurements. Local microscopic measurements have to deal with the many lateral movements which are common in the movement of pedestrians and cyclists. These lateral movements are caused by the nature of walking and cycling, which is less stable than car driving. Moreover, compared to car traffic, pedestrians and cyclists have more degrees of freedom to (instantly) change directions, also due to the fact that slow modes are not restricted to lanes. Therefore, slow mode traffic flows consist of more complex, chaotic patterns, which cause difficulties in keeping track of individual movements. Traditional local microscopic measurements consist of post-processing time-lapse photography or video recordings by human observers, since it is impossible for human observers to accurately track individuals real-time in a dense flow of pedestrians or cyclists.

Global microscopic movements of pedestrians and cyclists pose a different kind of challenge. In contrast to cars, slow traffic is not equipped with a license plate, which is the unique identifier which can be used to match observations at multiple sites in the network. Frequently used manual methods are stalking and questionnaires (Bovy and Stern,1990). In stalking, the observer follows the traveller in an unnoticed way, while recording temporal and geographical characteristics of the movement. In questionnaires, the respondent is asked questions about a specific trip or is asked to record several trips and trip attributes in a travel diary. Both methods have their own specific disadvantages. Stalking is extremely labour-intensive, since the observation time of one movement is equal to the movement time the traveller. Questionnaires are more efficient, but rely on the degree in which the respondent is able to recall sufficient information about his/her movement. Several studies have indicated that this is difficult for respondents, and results in a low reliability of collected data. Both methods are mostly suitable to collect data on routes and global paths of pedestrians and cyclists. Detailed operational behaviour (e.g. trajectories) is not covered.

In Table 3 manual data collection methods are classified, and the previously mentioned examples are given. In a slow traffic environment these methods have in common that their low productivity causes the measurement costs to be relatively high when comparing to vehicular traffic. The low productivity is caused by the need of a large number of human observer hours in the data collection and analysis process. Moreover, many data collection methods have specific disadvantages which cause low data reliability and/or observation biases. In in-door situations, some measurement methods are not applicable at all. Therefore, other ways of real-world data collection need to be investigated, being preferably non-intrusive in order not to interfere with the normal walking or cycling behaviour.

**Table 3: Classification of manual data collection methods.**

| | | **Measurement objective** | |
|---|---|---|---|
| | | **Local** | **Global** |
| **Measurement perspective** | **Microscopic** | Manual, post-processing of time-lapse or video | Stalking<br>Questionnaires |
| | **Macroscopic** | Manual counts<br>Qualitative description of traffic phenomena | Aerial observations |

**Automated data collection methods for slow modes**

As with vehicular traffic, emerging technologies increasingly enable automated data collection for slow modes. For local microscopic and macroscopic pedestrian measurements, several technologies are used. Image based sensors (time-lapse and video) are most frequently used, but infrared or laser sensors have also been applied for these objectives (Timmermans, 2009; Voskamp, 2012). These technologies have in common that occlusion of pedestrians results in measurement errors. Occlusion occurs at high traffic densities and/or suboptimal positioning of the sensors, and makes it difficult for the sensor to detect and distinguish individual pedestrians. In addition, this will lead to underestimation of large flows and high densities. Moreover, sensor movements cause problems in image processing (Duives, 2012). Specifically with respect to video and time-lapse technologies, privacy issues can pose limitations on their usability. Privacy regulations in many countries require strict limitation of access to data which potentially can be used to directly or indirectly infer the identity of a person who is linked to a registration of movement in the dataset (Van den Heuvel et al, 2013). In this context, video or time-lapse photography based sensors need to be capable of on-board (“embedded”) processing of images. In this situation the sensor itself generates the data files which describe the traffic conditions. To our knowledge, there are no embedded sensors which can be deployed in dense traffic situations. The alternative is that the sensors transmit their recordings to an off-site server, which converts the visual recordings to traffic data files. Such a server would require a secure environment with restricted access, and high bandwidth connectivity with the on-site sensors. This poses many organisational challenges and high costs. Moreover, state-of-the art technologies are not capable of processing high-density pedestrian flows fully automatically, as shown by Duives (2012) at aerial measurements of the annual music event Lowlands in the Dutch town Biddinghuizen.

With respect to global measurements, technological progress has resulted in several types of automated data collection methods. In the earlier mentioned study in Venice the researchers have used a combination of traditional on-site counts, traditional post-recording manual processing of time-lapse images, and an automated data collection method by GPS (Mamoli et al, 2012). In this study the route choice and the occurrence of bottlenecks were registered by following a subset of the inner city's visitors using GPS-recorders. The objective was to test this technology against the traditional methods. A clear disadvantage of traffic measurements by GPS is intrusiveness, since travellers have to be equipped with sensors before entering the network. This limits the scalability of the measurements which results in small samples, and might introduce observer biases.

A second source of automated measurement data is the mining of mobile phone data of telecom operators – also referred to as "mobile data analytics" – which recently has found its way to transport studies (Keij, 2014). A large study in the context of urban travel in Boston, USA and Rio de Janeiro, Brazil has been presented recently (Çolak et al, 2015). In this study, time and location registration of call records are converted to trip data. The paper does not report any registration of the mode of transport. In a recent paper, Alvarez and Leeson (2015) state that mobile phone data has proven to be valuable in vehicular traffic, but argue that it is very difficult to deploy for recording pedestrian movement due to spatial resolution limitations. The authors conclude that the development of a new, more detailed and accurate technology (i.e. 5G) is required. If these limitations are overcome, the mobile phone data of telecom operators will provide a vast amount of urban traffic data, potentially at extremely low costs per unit of measurement.

Finally, Bluetooth and WiFi technologies allow tracking of individual travellers through a network by the registration of radio devices such as mobile phones, tablets and laptops. An early study in this field has been presented by O'Neill et al (2006). As an alternative for deploying GPS receivers, the researchers have installed several Bluetooth sensors in Bath (United Kingdom), both at the university campus and in the city centre. These sensors were used to detect the presence of Bluetooth-enabled mobile phones which were carried by pedestrians. The objective was to determine the correlation between Bluetooth measurements and manual gate counts by human observers. By matching the unique identifier of the detected Bluetooth devices of several measurements in time (single sensor) and space (multiple sensors) the researchers respectively induced the time spent at one site or the direction of movement between sites. Versichele et al (2012) have used Bluetooth measurements to observe human movement at the Ghent Festivities. This study focussed at the movements behaviour at the festival, as well as the transportation mode to the festival. In the latter, only public transport modes were considered. Bikes, cars and walking as access and egress modes were out of scope of the measurements.

A third cluster of studies with Bluetooth measurements of pedestrians has been reported by Voskamp (2012), Ton (2014) and Van den Heuvel et al (2015). The researchers have used Bluetooth measurements to analyse behaviour of passengers at the main train station of Utrecht, which is the largest train station in The Netherlands with approximately 250,000 train passengers per average work day. These studies focused on the route choice behaviour at the train station by arriving and departing train passengers. The main disadvantage of Bluetooth measurements is the relatively low penetration rate: 7-8% by O'Neill (2006), 5-10% by Voskamp (2012), 11% by Versichele et al (2012). Versichele et al (2012) argue that a low penetration rate can result in a sampling bias due to over and underrepresentation of specific groups of people. Since the sensors used by Ton (2014) were capable of measuring both Bluetooth and WiFi signals, this study reported a WiFi penetration rate of over 20%, which is significantly higher than Bluetooth. The main advantage of Bluetooth and WiFi sensors is the low cost of measurements, especially when these measurements are being deployed for longer time frames.

Similar to the previous section, Table 4 presents the various examples of automated data collection methods which have been presented in this section. In contrast to the traditional methods, a clear distinction between the two measurement perspectives cannot be made, since the technologies are or can be used for both perspectives. However, the classification according to the measurement objective has been made. Video, time-laps, infrared and laser technologies are used to local measurements, while GPS, Bluetooth, WiFi and mobile phone data are used for global measurements.

**Table 4: Classification of automated data collection methods.**

| | | **Measurement objective** | |
|---|---|---|---|
| | | **Local** | **Global** |
| **Measurement perspective** | **Microscopic** | Video<br>Time-lapse<br>Infrared<br>Laser | GPS<br>Bluetooth, WiFi<br>Mobile phone data |
| | **Macroscopic** | | |

We can conclude that recent slow mode traffic automated measurements haven been deployed in studies of pedestrian traffic in a broad set of traffic conditions: from low traffic volumes in Bath to extremely high volumes in Ghent and Biddinghuizen; in several travel motives, from commuting and travelling in Utrecht, shopping in Bath and to leisure in Venice, Ghent and Biddinghuizen. We have found no studies with automated measurements of bicycle traffic. With respect to the technology we can conclude that Bluetooth/WiFi are currently the most suitable technologies for measuring slow mode traffic flows in an urban network, especially to derive the distribution of flows and route choice in the network. In the next sections we will further explore the technology and its potential. Firstly by defining the technology capabilities in the context of traffic research in general, and secondly by illustrating two applications in automated slow mode traffic measurements in urban mobility.

**Bluetooth/WiFi technology for automated slow mode urban traffic measurements**
Bluetooth and WiFi are radio modules embedded in many devices which are carried by people, both when staying at one location and while travelling. Bluetooth and WiFi allow its parent device to communicate with other devices in the vicinity without being physically connected by a cable. Commonly used examples are mobile phones, tablets and laptops. In this context, the Bluetooth functionality is mostly used to pair with accessories, for example headsets, car kits and input devices (keyboard, mouse, etc.). Invented in 1994, Bluetooth has broadly become available since the ‘00’s. For mobile phones, a penetration rate of 90% has been reported for 2014 (Bluetooth Interest Group, 2015). The WiFi functionality is used to connect to a wireless network. Like Bluetooth, WiFi found its way to the market in the ‘00’s after being invented in 1997. Initially, its main application was at home, to share one internet connection over multiple computers without cables (The Economist, 2004). Currently, WiFi is also used to offer wireless networks in public spaces used by many kinds of devices, in many cases used to connect to the internet. For example, smartphones are commonly used devices. It is estimated that currently about half of the adult population on earth owns a smartphone. By 2020 this share will increase to 80% (The Economist, 2015). The maximum range of a single unit of both technologies is about 100 meters, but is often less due to local radio interference (The Economist, 2004; Bluetooth Interest Group, 2015). The areas of

applications of both Bluetooth and WiFi are increasing rapidly, under what is often referred to as “the Internet of Things”.

As the wide-spread application of both technologies is relatively recent, its use in the context of traffic measurements is even more recent. Due to the widespread use of these technologies by travellers, potential sample sizes have become attractive for traffic research. Moreover, an increase in suppliers combined with the continuous decrease of cost of computer chips, has driven cost of large-scale deployment of Bluetooth and WiFi sensors down. This has made the technical, automated measurement solution competitive to traditional, labour intensive methods. Thirdly, the reports on the first use cases in traffic research, both in practice and in academics, show promising results and invite others to exchange ideas and experience. This contributes to the generation of new use cases and applications. Fourthly and finally, the privacy of travellers (as carrier of the mobile device) can be respected relatively easy, at low cost. Despite the ongoing debate about this topic (Abbord-Jard et al, 2013), the authors expect this advantage to contribute to a public acceptance of the deployment of radio based technologies for traffic research.

For traffic measurements, Bluetooth and WiFi technology have three fundamental, technical characteristics in common:
1. **Detection**. The first is the periodic, low-interval search of the mobile device for other devices, which in other research are referred to as “swipes” or “inquiries”. Bluetooth inquiries may take up to 10 seconds due to the large number of frequencies to be scanned. WiFi inquiries potentially have a much lower interval time of 8 milliseconds (Abbord-Jard et al, 2013). Since inquiries by radio modules consume energy, device manufacturers set the inquiry interval at a rate which balances the time to establish a connection and the battery life. For traffic research, every inquiry which is detected by a sensor, confirms the presence of a mobile device within scanning range of the sensor (location) at a specific moment (time). In this concept, a mobile device acts as a proxy for a traveller;
2. **Identification**. Secondly, the the Bluetooth of WiFi module can be identified by a Media Access Control Address. This MAC address is a unique, but anonymous identifier which is used for establishing the link between a (mobile) device and other devices in the network (Abbord-Jard et al, 2013). For traffic research, the MAC address enables the identification of identical mobile devices (and the traveller to which it belongs) at different sensors in the network. Routes can be generated by combining detections of the same mobile device by different sensors;
3. **Location**. Thirdly, the strength of the wireless signal – received signal strength indicator, or RSSI - provides information about the quality of the connection, either of a short swipe or of an established connection. The higher the signal strength, the faster and more stable the connection tends to be (Versichele et al, 2012). For traffic research, signal strength data provides some information about the position of the mobile device to the sensor. This position information is relative to other detections. Information about the exact location or direction of movement cannot be derived by the RSSI-data only.

Combining these technical characteristics of Bluetooth and WiFi generates a powerful source of traffic data. Firstly, routes of travellers through the network can be reconstructed by combining detections of MAC addresses by multiple sensors in the transport network. This reconstruction is based on the logical order of the MAC address detections at the different locations, which is directly linked to the sequence of the movement by the traveller. Signal strength (RSSI) data can be used to improve validity, by filtering out potential noise (i.e. traffic at adjacent links). Moreover, RSSI data can used to determine the order of detections in time in facilities with a dense grid of sensors which causes mobile devices to be scanned by multiple sensors at once. Secondly, travel times can be derived from the difference between start and end time of trajectories. Thirdly, transport network occupancy can be estimated by presenting the total number of detections at each moment in time.

To illustrate the potential of Bluetooth and WiFi sensors for slow mode traffic measurements, two cases are presented in the next two sections. The first case is about the walking patterns and flows in the main shopping street of Amsterdam, The Netherlands. Delft University of Technology has performed temporary measurements of pedestrian flows in the Kalverstraat. This case study shows the data filtering necessary to come from raw data to travel behaviour information. Particularly challenging was the separation of pedestrians from cyclists in the sensor data. Secondly, the traveller operations in the main train station of the city of Utrecht is presented. Utrecht is the fourth largest city in The Netherlands, and its main station is the largest station in terms of daily train passenger volumes. To measure pedestrian flows inside the facility as part of the SMART Station program, train operator and station manager Netherlands Railways (NS) has deployed a large number of Bluetooth and WiFi sensors in the station hall since 2012.

**Case 1: City Centre of Amsterdam**

The municipality of Amsterdam increasingly deals with (over)crowding by pedestrians and cyclists in its inner city, not only during special events (e.g. Kingsday), but also during regular operations (e.g. during peak hours). In order to quantify the potential problems with respect to safety and capacity, insight is needed into the magnitudes of pedestrian and bicyclist flows at particular locations as well as the relations between the different parts of the inner city. To investigate the latter, the Delft University of Technology has collected data through hybrid Bluetooth / WiFi sensors at several locations in the inner city of Amsterdam (Baelde, 2015). In the following, we first introduce the setup of the data collection. Then, we give some details on the filtering process. From the filtered data links between observed locations have been identified. We show how sensitive the results are for filtering and we discuss the practical issues in such data collection efforts.

**Data collection setup**

As indicated before, the data collection has taken place in the inner city of Amsterdam, see the figure below. We have focused on the main shopping street in the area, the Kalverstraat, observed by three sensors, and an intersection, Muntplein, with 6 sensors. The sensors used for the data collection are combined Bluetooth / WiFi sensors. These have two long-range antennas, with a range of about 70 meters, thus each sensor has a range in the shape of an ellipse with a length of 140 meters, a width of 30 meters and the location of the sensor as its centre point. In the figure the field of view of each sensor has been indicated by the semi-transparent red areas, where the black cross indicates the position of the sensor. Here, it can be seen that the sensor view is blocked by the buildings alongside the roads, leading to irregularly shaped (approximations of) fields of view. Each sensor has been attached at a height of 4m above street level. The data collection has taken place on Friday 7 November 2014 and Saturday 8 November 2014. The duration was limited since the sensors are capable of functioning stand-alone during 40 hours.

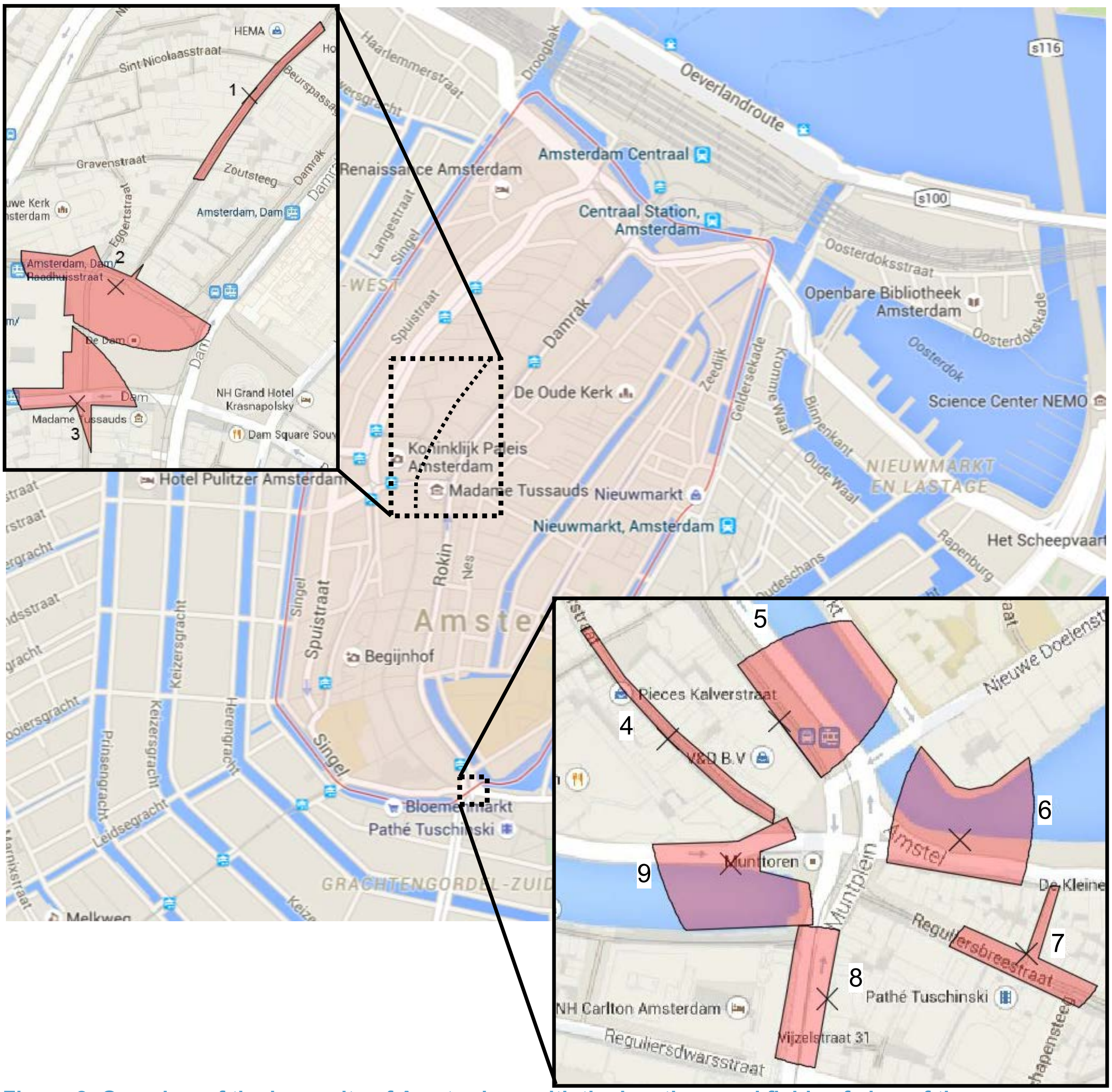



**Figure 2: Overview of the inner city of Amsterdam, with the locations and fields of view of the sensors.**

Table 5 shows the characteristics of each sensor, being the name of the street, the type of traffic observed by the sensor, the start and end time of the observations, and the number of raw data points for both the Bluetooth and the WiFi sensor. The number of WiFi-data points is larger than the number of Bluetooth-data points. This is according to our expectation (and literature), as more people have WiFi enabled on their devices. However, the ratio between Bluetooth- and WiFi-data points differs for the different sensors, with a minimum of 1.52 for the sensor at Amstel and a maximum of 5.14 for the sensor at the Kalverstraat South. There seems to be a relation between this ratio and the type of traffic observed: high ratios hold for sensors with only pedestrians, while sensors also observing cars have a much lower ratio. This can be expected, as modern cars often have Bluetooth equipment installed (e.g. car communication kits), implying a relatively high probability that a car is detected by the Bluetooth sensor.

**Table 5: Characteristics of each sensor.**

| Sensor id | Name | Traffic type | Start time | End time | Number of raw BT data points | Number of raw WiFi data points | BT/WiFi ratio |
|---|---|---|---|---|---|---|---|
| 1 | Nieuwendijk | Pedestrians | 07-11-2014 15:08 | 08-11-2014 22:21 | 51,824 | 247,579 | 4.78 |
| 2 | Dam | Pedestrians, trams | 07-11-2014 14:22 | 09-11-2014 00:59 / 01:43 | 80,561 | 130,098 | 1.61 |
| 3 | Kalverstraat North | Pedestrians, bicycles, cars | 07-11-2014 14:41 | 09-11-2014 00:27 | 60,556 | 105,092 | 1.74 |
| 4 | Kalverstraat South | Pedestrians | 07-11-2014 11:46 | 08-11-2014 22:26 | 96,692 | 496,605 | 5.14 |
| 5 | Rokin | Pedestrians, bicycles, cars, trams | 07-11-2014 12:10 | 09-11-2014 00:23 | 87,452 | 195,981 | 2.24 |
| 6 | Amstel | Pedestrians, bicycles, cars | 07-11-2014 12:33 | 08-11-2014 22:43 | 66,462 | 101,057 | 1.52 |
| 7 | Reguliersbree-straat | Pedestrians, trams | 07-11-2014 13:00 | 08-11-2014 22:07 | 96,248 | 417,300 | 4.34 |
| 8 | Vijzelstraat | Pedestrians, bicycles, cars, trams | 07-11-2014 10:42 | 08-11-2014 22:35 | 193,631 | 411,707 | 2.13 |
| 9 | Singel | Pedestrians, bicycles, cars | 07-11-2014 09:38 | 08-11-2014 22:38 | 70,445 | 265,754 | 3.77 |

**Data filtering**

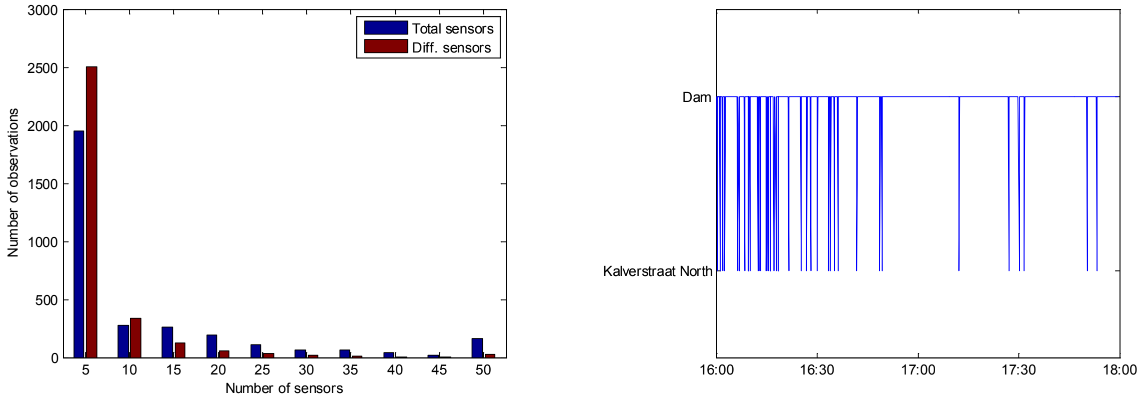
The first task in the data analysis is the filtering of the data, as the sensors do not only detect traffic participants (pedestrians, bicyclists, drivers), but also static devices with Bluetooth or WiFi connections. As the filtering methodology does not depend on the time period and the type of data, we will focus on the data filtering on the Friday evening peak from 16:00 – 18:00 for the Bluetooth data. For this time period, the number of raw Bluetooth data points is 57,584.

The filtering consists of the following steps:

1. Remove overlap in detection period for a single sensor.
2. Remove devices with many observations.
3. Identify pedestrians.

The first filtering step consists of removing data points with overlap in the observed time period. So if one data point indicates that a device has been detected at a certain sensor between 16:00:44 and 16:01:10 and the second data point indicates a detection period between 16:00:59 and 16:01:34 at the same sensor, then these two data points are combined into one data point, with the combined detection period. This resulted in a reduction of about 20% to 46,427 data points.

The second filtering step consists of removing data points of devices that have a large amount of observations at different sensors. We expect that pedestrians will walk directly from their origin to their destination, passing a limited number of sensors. Even tourists or shoppers are not expected to walk back and forth through the area for too many times. Figure 2a shows the number of observations at different sensors for each device, with a difference in the total amount of observations in blue and the number of observations with difference in sensors in red. Where we would expect a limited amount of sensors for a single trip, we found that up to 500 observations have been included for this individual (implying someone walking around in circles). When looking into detail for one (example) device, these many observations are limited to two sensors, that is, during the 2 hour period, this device moves from the one sensor to the other sensor and back, or sometimes leaves the sensor area for a short period of time and then returns (see Figure 3b). These are obviously not logical movement patterns for pedestrians, so the devices with more than 100 observations at sensors are filtered. This leads to 29,903 remaining data points.




a. Number of observations per device.

b. Sensors where a device with more than 500 observations has been recorded.

**Figure 3: Characteristics of observations at sensors for each device.**

The third filtering step relates to identifying pedestrians, and removing other traffic participants from the data set. We do this based on the walking speed. Typically, the average pedestrian speed appears to be 1.34 m/s (Buchmueller and Weidmann, 2006), with a standard deviation of 0.37 m/s. As upper boundary, we use a speed of 2 m/s: all devices moving faster than this speed will be removed from the data. Here, some of the technical properties of the sensors need to be considered, in particular the delay of the observation by the sensors. Previous experiments show that these delay are in the range of 3-10 seconds (Kostakos, 2008; O'Neill et al., 2006). Based on the distance between sensors and the maximum speed, we calculate the minimum travel time between two sensors and remove the detection delay (which is the worst case scenario). The resulting threshold values are given in Table 6. If the delay between two detections is shorter than this minimum travel time, the data point is removed from the data set. This leads to a data set of 712 data points. Based on this data set, we will identify the flows between the different locations in the inner city of Amsterdam.

**Table 6: Minimum walking times between Bluetooth sensors in seconds.**

| | Nwd | Dam | KvsN | KvsS | Rokin | Amstel | Rbs | Vijzels | Singel |
|---|---|---|---|---|---|---|---|---|---|
| Nwd | X | 25.8 | 65.8 | 359.3 | 425.1 | 455.9 | 474.2 | 409.4 | 349.5 |
| Dam | 25.8 | X | 0 | 291 | 356.8 | 387.6 | 405.9 | 344.4 | 284.5 |
| KvsN | 65.8 | 0 | X | 253.5 | 319.3 | 350.1 | 368.4 | 329.9 | 270 |
| KvsS | 359.3 | 291 | 253.5 | X | 0.8 | 24.1 | 42.4 | 27.8 | 13.0 |
| Rokin | 425.1 | 356.8 | 319.3 | 0.8 | X | 16.5 | 35.1 | 20.5 | 14.7 |
| Amstel | 455.9 | 387.6 | 350.1 | 24.1 | 16.5 | X | 27.8 | 13.0 | 47.1 |
| Rbs | 474.2 | 405.9 | 368.4 | 42.4 | 35.1 | 27.8 | X | 24.9 | 65.1 |
| Vijzels | 409.4 | 344.4 | 329.9 | 27.8 | 20.5 | 13.0 | 24.9 | X | 14.9 |
| Singel | 349.5 | 284.5 | 270 | 13.0 | 14.7 | 47.1 | 65.1 | 14.9 | X |

## Identifying flows between locations

From the resulting data set, the number of pedestrians travelling directly from one sensor to another sensor has been derived, see Table 7. We can clearly see the different flow sizes between the sensor combinations, identifying the main connections (e.g. from the Dam to the

Kalverstraat North) and the smaller connections, typically between sensors not directly connected (indicated by the yellow in the table).
Table 8 shows the total amount of observations at one sensor, which is the sum of all arrivals and departures shown in Table 7 and the number of pedestrians that have only been observed at the sensor (thus not walking through the network). As we do not have ground truth measurements, we can only give a first impression of the quality of the data set. We would expect about 2500 pedestrians per hour passing the main sensors (Dam, Kalverstraat Noord, Kalverstraat South). From the table, it can be derived that we have observed at maximum 164 pedestrians during 2 hours, leading to a penetration rate of 3-4%, which is slightly lower than we would expect. We can thus conclude that Bluetooth / WiFi sensors are a promising technique to be applied, but the filtering does have a major effect on the results.

**Table 7: Number of movements for each sensor pair. In yellow the sensors located next to each other.**

| | Nwd | Dam | KvsN | KvsS | Rokin | Amstel | Rbs | Vijzels | Singel |
|---|---|---|---|---|---|---|---|---|---|
| Nwd | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| Dam | 7 | 0 | 25 | 2 | 2 | 1 | 0 | 1 | 2 |
| KvsN | 1 | 39 | 0 | 1 | 3 | 1 | 0 | 1 | 1 |
| KvsS | 0 | 2 | 3 | 0 | 2 | 0 | 1 | 0 | 4 |
| Rokin | 0 | 2 | 0 | 6 | 0 | 5 | 0 | 4 | 8 |
| Amstel | 0 | 3 | 5 | 0 | 6 | 0 | 4 | 9 | 2 |
| Rbs | 0 | 3 | 1 | 0 | 0 | 3 | 0 | 0 | 2 |
| Vijzels | 1 | 2 | 0 | 1 | 0 | 7 | 3 | 0 | 23 |
| Singel | 0 | 7 | 4 | 3 | 1 | 5 | 0 | 5 | 0 |

**Table 8: Total observations of pedestrians per sensor.**

| | Nwd | Dam | KvsN | KvsS | Rokin | Amstel | Rbs | Vijzels | Singel |
|---|---|---|---|---|---|---|---|---|---|
| Arrivals | 9 | 63 | 40 | 13 | 14 | 22 | 8 | 20 | 42 |
| Departures | 7 | 40 | 47 | 12 | 25 | 29 | 9 | 37 | 25 |
| Single obs | 4 | 61 | 6 | 57 | 3 | 4 | 3 | 3 | 2 |
| Total | 20 | 164 | 93 | 82 | 42 | 55 | 20 | 60 | 69 |

### Case 2: Utrecht Central Station

At an average workday Utrecht Central Station serves over 170,000 arriving and departing train passengers, and almost 60,000 passengers who transfer between trains. The station consists of seven central platforms. Every platform is connected to the station hall by two escalator and one stairway. The station hall is located on top of the central section of the platforms. In the train station, the vertical infrastructure – escalators and stairs – are common bottlenecks in the pedestrian network. This is particularly the case when train arrivals coincide at both sides of one central platform, or at the two platform sections - A and B – of one platform side. This occurs frequently, since the station acts as train hub in the Dutch railway network with arrivals and departures of approximately 40 trains per hour.

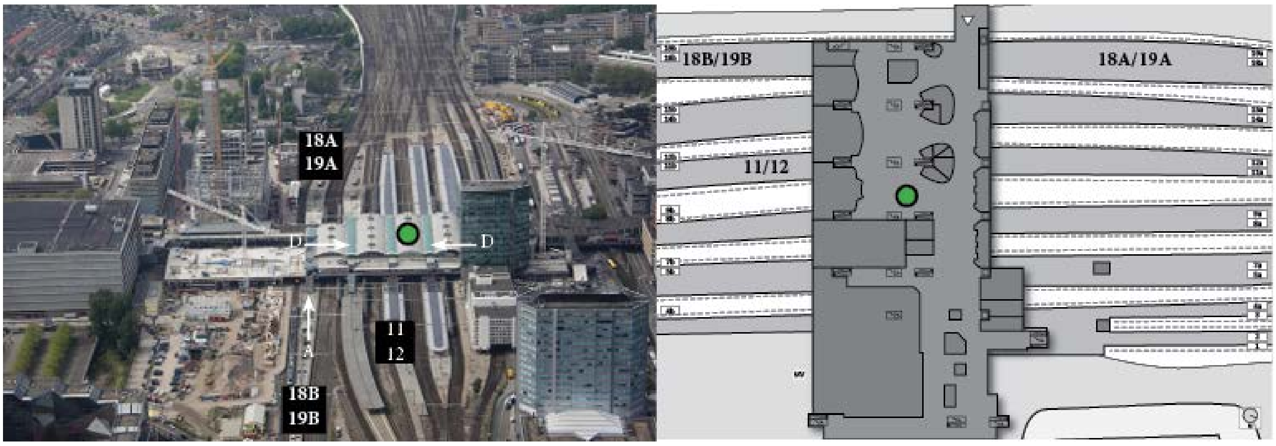



Figure 4: Overview of Utrecht Central station.

At Utrecht Central station, Bluetooth and WiFi sensors have been used to measure pedestrian flows since 2012. As the stations and real estate division of Netherlands Railways, NS Stations has been implementing new technologies for pedestrian flow measurements under the SMART Station program. The main objective of this program is to develop new sources of pedestrian behaviour data inside train stations. This data is being used for station (re)design, management and operation. Currently, there are two other SMART Stations in The Netherlands: Leiden Central station and Amsterdam Airport Schiphol train station.

Since the start of the program, the SMART Station data has supported a large number of studies, both practical and scientific. In this paper we will present two cases which are the result of the graduation projects of two master students. Both studies were aimed at determining factors which contribute to route choice of train passengers inside Utrecht Central station.

The first case study has been performed in 2012 by analysing the impact of congestion at bottlenecks on route choice behaviour by arriving train passengers (Voskamp, 2012 and Van den Heuvel et al, 2015). For this study a limited number of nine Bluetooth sensors has been installed at platform 18/19, both A and B section. The research objective was to determine the factors which are relevant for the choice of vertical infrastructure to exit the platform (Figure 5). Three potential factors have been tested: 1. the type of vertical infrastructure – escalators or stairs -, 2. the waiting time due to queueing for the congested escalator or stairs and 3. the destination of the pedestrians inside or outside the station hall.

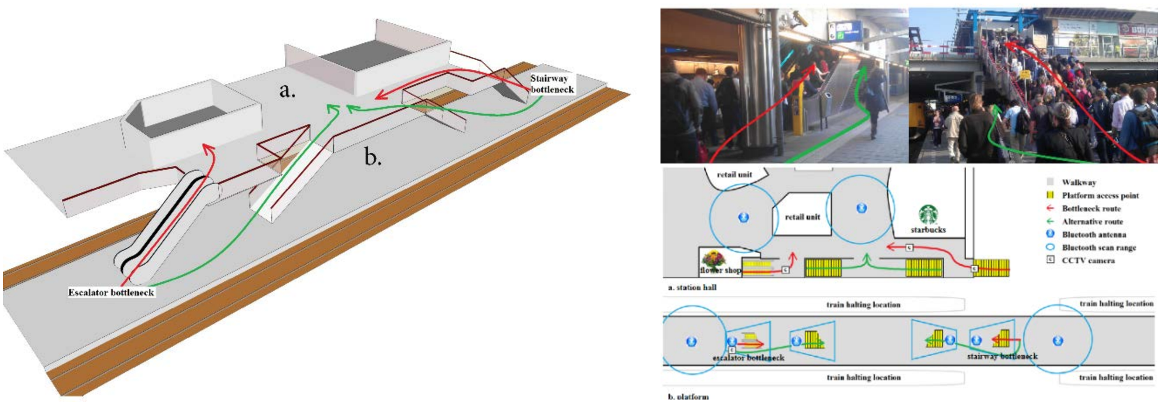



Figure 5: Route choice behaviour by arriving passengers.

In the second case study the route choice behaviour in the station hall by departing train passengers has been analysed (Ton, 2014 and Ton et al, 2014). This study has used data collected by over 30 hybrid Bluetooth/WiFi sensors which were already present in the station hall (Figure 6). One of the research objectives was to determine the factors which contribute to the choice of escalator or stairway to leave the station hall and access the platform. Seven potential factors have been tested: 1. walking distance to the escalator or stairway in the station hall, 2. walking time to the escalator or stairway in the station hall, 3. the train stop location alongside the platform, 4. occurrence of train delays, 5. visibility of the escalator or stairway in the station hall, 6. orientation of the escalator or stairway relative to the direction of approach of the passenger and 7. peak or off-peak time of the day as a proxy for familiarity at the station. Because this platform offered the best situation for this study, data of pedestrian traffic to platform 11/12 was used.

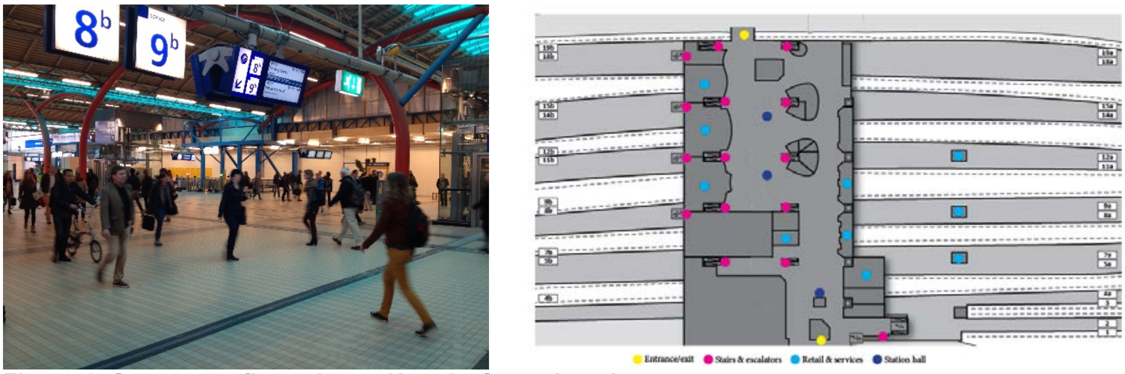



**Figure 6: Sensor configuration at Utrecht Central station**

For both studies, the concept of “disaggregate travel demand models based on discrete choice analysis” has been used (Ben Akiva and Lerman, 1985). In these “discrete choice models” a traveller is assumed to make choices from sets of alternative options for his/her trip. Each set of alternative options has to comply with the MECE-principle: mutually exclusive and collectively exhaustive. Consistent with the model assumption of utility maximization, it is assumed that the traveller choses the (one) option which delivers the highest utility at the time the choice is made. Like many other applications in transport research, route choices by passengers inside a train station can be considered as discrete choices of routes with different attributes.

The 2012-study of route choice of arriving train passengers has revealed that waiting time at congested vertical infrastructure has a significant impact on route choice. To illustrate the waiting time effect, a sample of “raw” Bluetooth measurements is projected in the left graph of Figure 4. Each dot in this graph represents the detection of a Bluetooth-device, which is carried by a passenger at the platform, directly after the arrival of train 8827 at 10 April 2012 during morning peak hour. Red dots represent route choices for the escalator, which is the primary route between train and station hall. Blue dots represent passengers who walked to the station hall by the stairway which becomes the main alternative when the primary route (escalator) gets congested. The green marked dots indicate that passengers tend to choose the primary route when no congestion has formed yet. After about 30 seconds after the train arrival, the primary route gets congested, and a share of arriving passengers starts to use the alternative route (indicated by the yellow marked dots). The red marked dot is an error, caused by a detection of a person who was already at the platform (just) before the train arrived.

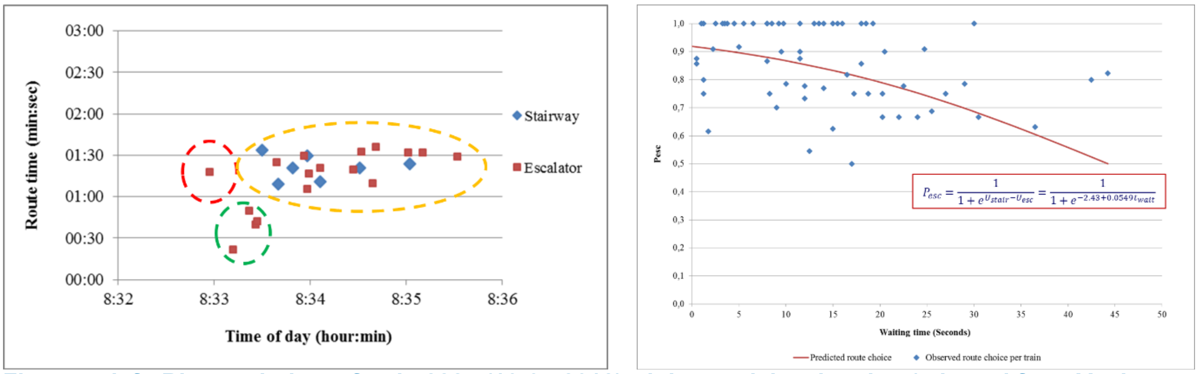



Figure 7: left: Bluetooth data of train 8827 (10-04-2012); right: model estimation (adapted from Voskamp, 2012).

The right graph of Figure 7 shows an aggregation of observations at the escalator bottleneck for 269 train arrivals. Each dot represents one train, and gives the share of arriving passengers who chose the bottleneck route and the waiting time due to congestion. The red graph and the equation represent the model which has the best fit with the observations. The model shows at a waiting time of approximately 45 seconds, about half of the passengers chose the alternative, non-congested route to the station hall. Under non-congested conditions about 90% takes the primary route to leave the platform.

The 2014-study of route choice of departing train passengers has revealed that route time is also an important attribute for departing passengers (left side in Figure 8). The relation has an inverse nature: the longer a route takes, the less travellers tend to choose this route. Distance has a similar, but less strong impact. Escalator or stairway orientation relative to the direction of approach by the passenger also proved to be a route attribute with significant impact. The better the orientation attribute of a route alternative is, the more people tend to choose for it. The study results indicate that travellers tend to favour right over left turns when leaving the station hall to enter the platform. Finally, the train location is a significant factor. Trains at Utrecht Central station can stop at three locations alongside the platform: A-section, B-section or central platform section (A+B). This result indicates that departing passengers, when choosing their route in the station hall, also include expectations about the walk at the platform before boarding the train.

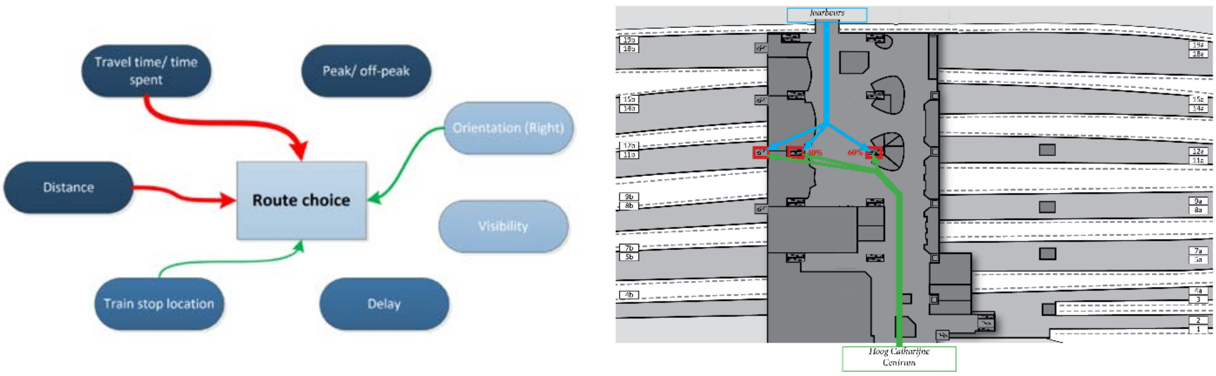



Figure 8: departing passengers; route choice factors (left; Ton, 2014) and route distribution (right; Ton et al, 2014).

## Conclusions

In this paper we have identified various ways to observe slow modes. In addition to the traditional manual data collection techniques, recently developed automated data collection techniques have been presented. One of the new sensor classes, hybrid Bluetooth / WiFi sensors, has been applied in the inner city of Amsterdam to observe flow patterns and in a large train station to observe route choice. Both cases illustrate the potential of these sensors for slow mode traffic research. As individual (but anonymous) movements of pedestrians and cyclists can be automatically measured for a long time period, the collected datasets are extremely rich and useful for many applications in traffic research for slow traffic modes. To the best knowledge of the authors, this type of data cannot be collected by the traditional methods.

The cases also have illustrated that the researcher should be aware of several issues to avoid invalid research conclusions. Firstly, not all devices detected by the sensors are pedestrians, cyclists or other traffic participants in scope of the research. This means that it is always required to filter the collected data in a thorough, extensive and well-defined way. In other words: the filter itself is an important determinant for data quality and therefore for the validity of the research results. Secondly, the sensors themselves have a detection delay due to the design of the technology. Research design and filtering procedures have to be able to cope with this issue. Thirdly, and finally the penetration rate is rather low, especially for Bluetooth sensors (~5-10%). This might introduce a selection bias. Although to a lesser extent, the same issue occurs with WiFi sensors (~ 20% penetration rate). Currently, it is not clear whether these ratios will increase or decrease in the future. A downward trend is caused by (semi-)automatic off and on switching of Bluetooth or WiFi due to privacy concerns and/or the extension of battery life of the mobile device. An upward trend is caused by the increasing number of wearables with Bluetooth and or WiFi (i.e. smart watches). Only the future will tell which trend will be dominant. We will continue our research in testing the suitability of automated data collection methods in our slow mode research.

**References**

Abbott-Jard, M.; Shah, H. & Bhaskar, A. (2013), Empirical evaluation of Bluetooth and Wifi scanning for road transport, *in* 'Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia', pp. 14.

Alvarez, P. & Leeson, A. (2015), 'Big data helps pedestrian planning take a big step forward', *Transportation Professional* **April**, 26-27.

Baelde, F. (2015), ‘Monitoring Crowds in Amsterdam’, Additional Thesis work, Delft University of Technology.

Ben-Akiva, M. E. & Lerman, S. R. (1985), *Discrete choice analysis: theory and application to travel demand*, Vol. 9, MIT press.

Bluetooth Interest Group (2015), 'History of the Bluetooth Special Interest Group', http://www.bluetooth.com/Pages/HistoryofBluetooth.aspx [Last accessed at 6 May 2015]

Bovy, P. H. & Stern, E. (1990), *Route choice. Wayfinding in transport networks.*, Kluwer Academic Publishers.

Buchmueller, S.; Weidmann, U. (2006), Parameters of pedestrians, pedestrian traffic and walking facilities, Schriftenreihe 132, Institute for Transport Planning and Systems.

Çolak, S.; Alexander, L. P.; Alvim, B. G.; Mehndiretta, S. R. & Gonzalez, M. C. (2015), Analyzing cell phone location data for urban travel: current 2 methods, limitations and opportunities 3, *in* 'Transportation Research Board 94th Annual Meeting'.

CROW-Fietsberaad (2014), *Fietsen in Nederland: patronen, trends en beleid*.

Duives, D. C. (2012), 'Analysis of pedestrian crowd movements at Lowlands', Master's thesis, TU Delft, Delft University of Technology.

Fruin, J. J. (1987), *Pedestrian planning and design (Revised Edition)*, Elevator World, Inc..

Harms, L.; Bertolini, L. & te Brömmelstroet, M. (2014), 'Spatial and social variations in cycling patterns in a mature cycling country exploring differences and trends', *Journal of Transport & Health* **1**(4), 232--242.

Keij, J.E. (2014). Smart phone counting: Location Based Aplications using mobile phone location data. Master thesis. Delft University of Technology.

Kostakos, V. (2008). Using Bluetooth to capture passenger trips on public transport buses. University of Madeira.

Ministerie van Infrastructuur & Milieu (2013), *Mobiliteitsbalans 2013*, Kennisinstituut voor Mobiliteitsbeleid | KiM. [in Dutch]

Mamoli, M.; Michieletto, P.; Bazzani, A. & Giorgini, B. (2012), Venice as pedestrian city and tourist magnet: mass events and ordinary life, *in* 'Conference of the International Forum on Urbanism'.

Massa (2014), 'Fobie voor Utrecht Centraal', http://massa-media.nl/2014/04/fobie-voor-utrecht-centraal [Last accessed at 1 April 2015; in Dutch]

O'neill, E., Kostakos, V., Kindberg, T., Fatah, A., Penn, A., Fraser, D., et al. (2006). Instrumenting the city: developing methods for observing and understanding the digital cityscape. Bartlett School.

Parool (2014), 'De maximale drukte is bereikt voor de Kalverstraat'. http://www.parool.nl/parool/nl/4/AMSTERDAM/article/detail/3741390/2014/09/06/De-maximale-drukte-is-bereikt-voor-de-Kalverstraat.dhtml [Last accessed at 1 April 2015; in Dutch]

Peterson, B. S.; Baldwin, R. O. & Kharoufeh, J. P. (2006), 'Bluetooth inquiry time characterization and selection', *IEEE Transactions on Mobile Computing* **5**(9), 1173-1187.

Pushkarev, B. & Zupan, J. (1975), *Urban space for pedestrians: a report of the Regional Plan Association*, MIT Press.

Ravalet, E.; Christie, D.; Munafт, S. & Kaufmann, V. (2014), Analysis of walking in five Swiss cities: a quantitative and spatial approach, *in* '14th Swiss Transport Research Conference'.

RTL Nieuws (2013), 'Treinen Amsterdam rijden met vertraging', http://www.rtlnieuws.nl/nieuws/binnenland/treinenamsterdamrijdenmetvertraging [Last accessed at 1 April 2015; in Dutch]

The Economist (2004), 'CASE HISTORY - A brief history of Wi-Fi', http://www.economist.com/node/2724397/print [Last accessed at 6 May 2015]

The Economist. (2015), 'Smartphones - Planet of the phones', http://www.economist.com/node/21645180/print [Last accessed at 6 May 2015]

Timmermans, H. J. P. (2009), *Pedestrian Behavior: Models, Data Collection and Applications*, Emerald Group Publishing Limited.

Ton, D. (2014), 'Navistation. A study into the route and activity location choice behavior of departing pedestrians in train stations.', Master's thesis, Delft University of Technology.

Ton, D.; van den Heuvel, J. & Hoogendoorn, S. (2014), 'Zo kiest de treinreiziger zijn trap of roltrap. Wetenschappelijk onderzoek naar routeroute in het station.', *Verkeerskunde* **7**, 16-17 [in Dutch]. http://www.verkeerskunde.nl/routekeuzeopstations [Last accessed at 10 May 2015]

Van den Heuvel, J.; Thiellier, E. & van Gerwen, N. (2013), 'Privacy by design bij reizigersmetingen op stations', *Privacy & Compliance* **3**, 17-21. [in Dutch]. http://repository.tudelft.nl/assets/uuid:c50a8044-97b1-4474-89ef-6d97cf87d800/297298.pdf [Last accessed at 10 May 2015]

Van den Heuvel, J.; Voskamp, A.; Daamen, W. & Hoogendoorn, S. (2015), Using Bluetooth to estimate the impact of congestion on pedestrian route choice at train stations, *in* 'Traffic and Granular Flow '13'.

Versichele, M.; Neutens, T.; Delafontaine, M. & Van de Weghe, N. (2012), 'The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities', *Applied Geography* **32**(2), 208--220.

Voskamp, A. (2012), 'Measuring the influence of congested bottlenecks on route choice behavior of pedestrians at Utrecht Centraal', Master's thesis, Delft University of Technology.