Using wearable GPS devices in travel surveys: A case study in the Greater Copenhagen Area

Thomas Kjær Rasmussen*, Jesper Bláfoss Ingvardson, Katrín Halldórsdóttir, Otto Anker Nielsen

{tkra, jbin, katha, oan}@transport.dtu.dk

DTU Transport, Technical University of Denmark

* Corresponding author

Abstract

GPS data collection has become an important means of investigating travel behaviour, as it ideally provides far more detailed information than traditional travel survey methods. While setting fewer requirements to the respondents, it however sets high requirements to the post processing of the data collected. This study proposes a combined fuzzy logic- and GIS-based algorithm to process raw GPS data. The algorithm is applied to GPS data collected in the highly complex Greater Copenhagen Area network and detects trip legs and distinguishes between five modes of transport. The algorithm shows promising results by (i) identifying trip legs for 82% of the reported trip legs, (ii) not classifying non-trips such as scatter around activities as trip legs and (iii) identifying the correct mode of transport in more than 90% of trip legs for which corresponding observed modes are available.

1. Introduction

Over the last 20 years Global Positioning Systems (GPS) have been applied in various investigations of transport-related issues. These applications include, among other, (i) evaluation of system performance, such as measuring historical and real-time congestion and flow levels (Quiroga and Bullock, 1998; Quiroga, 2000; Li et al., 2004; Herrera et al., 2010), (ii) analysis of travel behaviour, such as response to road pricing schemes (Nielsen, 2004), and (iii) estimation of route choice parameters for use with route choice models (Rich and Nielsen, 2007; Chen et al., 2013). In recent years much effort has been given to investigate the use of GPS devices as the data source for travel surveys (Wolf, 2000; Gong et al., 2011; Bolbol et al., 2012; Stopher et al., 2005, etc.). When compared to traditional travel diaries collecting data via GPS devices ideally provides the investigator with far more detailed information on travel times, routes used and locations of activities. Another advantage of using GPS data is that it is not dependent on individuals' (possibly mis-) perception of travel time, travel distance and departure time. In traditional travel diaries there is often a common problem of underreporting of trips (e.g., Stopher et al., 2007; Forrest and Pearson, 2005). This problem is likely to be reduced when using GPS as all movements of participants are logged (Stopher et al., 2008). Additionally, the data collection sets far less requirements to the respondents as answering time-consuming questionnaires regarding route choices are avoided. This enables larger sample sizes, but also sets higher requirements to work done in the post-processing, where detailed travel information are derived from the raw data.

Originally, GPS units required external power supply and were therefore limited to being installed in vehicles. Over the years the equipment has evolved and the data collection is no longer restricted to being vehicle-based. Today GPS units are sufficiently accurate, lightweight and have long enough battery-life to make multi-day person-based data collection for all conducted trips possible (e.g. Stopher and Shen, 2011; Gong et al., 2011; Bolbol et al., 2012). Multi-day person-based data collection facilitates a complete analysis and better understanding of individuals' travel patterns including choice of mode of transport, combination of modes, route choices in multi-modal transport and day-to-day variations. Hence, ultimately, it is possible to obtain detailed travel information similar to or even more detailed than those obtained from traditional travel surveys at lower costs in terms of time usage spent by respondents and interviewers.

Collecting GPS data generates a lot of raw data and sets high requirements to the post-processing of the data. Various approaches to do so has been proposed in literature, and this paper describes the results obtained by applying an existing method to a multi-day person-based GPS data set collected among families in the Greater Copenhagen Area. The cleaning of GPS points and identification of trip legs has been done using POSDAP (2012) developed by Schüssler and Axhausen (2009). This paper extends the method by Schüssler and Axhausen (2009) by proposing a method that utilises geographical information systems (GIS) to identify mode choice. Furthermore, additional algorithms to detect illogical mode changes have been

developed and applied to the data. In addition to the GPS data, corresponding traditional interview-based travel survey data were collected for one of the days in the survey period. This made it possible to validate the results of the trip and mode identification algorithms.

The remainder of this paper is structured as follows. Section 2 reviews the existing literature focused on using GPS as a travel survey collection method. Section 3 introduces the case study, while Section 4 introduces the method used to post-process the GPS data. Section 5 reports the results obtained by applying the proposed method to the case study. A discussion and comparison to results found in similar studies are done in Section 6, and Section 7 concludes the findings.

2. Literature review

The literature review is divided into two parts with section 2.1 focusing on how GPS devices have been used in travel surveys, whereas section 2.2 focuses on the proposed approaches for post-processing the raw GPS data.

2.1. GPS in travel surveys

GPS technology limited the first travel surveys using GPS devices as data collection method to only being vehicle-based, as the devices were large and power consumption was high (Wagner et al., 1996; Yalamanchili et al., 1999). These early studies sought mainly to supplement telephone-based travel surveys by collecting additional data to identify e.g. detailed route choices as well as to verify exact time of day, and detect unreported trips (Wolf, 2000). The additional trip information was specified by the respondents, e.g. trip time and trip purpose, when starting a trip (Yalamanchili et al., 1999; Du and Aultman-Hall, 2007). This was often done on a connected personal digital assistant (PDA).

The first study to expand the method to support several modes of transport was Draijer et al. (2000) in which respondents were asked to wear a GPS and a PDA device on all trips. Due to the size and weight of the devices (approx. 2kg) there was however a consistent underreporting of walking, cycling and public transport trips as well as trips with purpose of shopping and visiting friends. Additionally, as the respondents were asked to turn the device on/off when starting/ending a trip, the survey design demanded constant effort from the participants.

Several studies combine GPS traces with additional information gathered by a travel survey questionnaire. Among these are the studies by (de Jong and Mensonides, 2003; Bohte and Maat, 2009; Tsui and Shalaby, 2006), which were internet-based surveys where respondents needed to confirm the trips identified by the trip identification algorithm. As GPS devices have become smaller and lighter, multi-modal GPS based travel surveys have become extensively applied as travel survey method. As part of the evolvement of the survey method, much has been done to reduce the effort needed by the respondents, and many studies today do

therefore not ask participants to provide trip information *en-route* (Schüssler and Axhausen, 2009; Stopher and Shen, 2011). This however sets higher requirements to the post-processing algorithms to identify trip legs and mode from raw GPS data consisting solely of time and space information.

Later studies, including Schüssler and Axhausen (2009) and Bolbol et al. (2012), have analysed the correctness of fully automatically processed GPS data not including any questionnaire data in the trip and mode detection. In Schüssler and Axhausen (2009) GPS data from 4.882 participants wearing the GPS devices for 6.65 days on average was processed without any further information, and the results were compared to the existing (national) travel survey. The results showed that in aggregate figures the trip and mode identification only deviates slightly from that of the census data. However, no disaggregate comparison of individual data was performed. This was done in Bolbol et al. (2012), where 81 respondents wore a GPS device for 2 weeks but also answered a travel diary questionnaire. Based on speed and acceleration only, the study designated each trip to one of six different modes. When comparing to the travel diary it was found that most modes could be inferred correctly. Some modes however had very similar speed and acceleration profiles, making it harder to distinguish between bus and metro, and between bus and bicycle.

Alternative approaches utilising information on local spatial information has been proposed to try and overcome this, including map matching to mode-specific networks by use of GIS-software(e.g., Chen et al., 2010; Chung and Shalaby, 2005; Tsui and Shalaby, 2006; Bohte and Maat, 2009). Another approach incorporating local spatial information was proposed by Stopher et al. (2005), which used an elimination method where walking segments were first identified followed by public transport and private car. For each mode a set of rules was developed, e.g. for public transport the trip had to follow the transit network and had to include regular stops that did not coincide with intersections.

All the methods proposed in the above mentioned studies have shortcomings in either (i) not including modes which are important in an application to the Greater Copenhagen Area (rail is not included in Chung and Shalaby (2005), and bicycles are not included in Gong et al. (2011)), (ii) relying on prompted recall surveys where participants need to verify their trips (Bohte and Maat, 2009; Stopher et al., 2008), and/or (iii) including a very small sample of participants (only 9 participants in Tsui and Shalaby (2006)). These shortcomings will be addressed in this present study by including (i) the five most dominant modes in the Greater Copenhagen Area covering in total 97.5% of all trips undertaken[1], and (ii) 183 participants totalling 644 person days of travel, creating a sufficiently large sample to validate the algorithms.

---

[1] Retrieved from the Danish National Travel Survey (Christiansen, 2012)

2.2. Post-processing GPS data

Post-processing of raw GPS data typically involves four steps, namely (i) GPS data cleaning, (ii) trip and activity identification, (iii) trip segmentation into single-mode trip legs, and (iv) mode identification. The approach varies slightly between studies, e.g. in Stopher et al. (2005) step ii) and iii) are performed jointly. After performing these four steps, some analyses apply additional steps. Chen et al. (2010), Stopher et al. (2005) and others infer the purpose of the trips identified, while e.g. Schüssler and Axhausen (2009) match the identified trip legs onto the corresponding modal networks.

Most analyses sets off with a cleaning and filtering step, where systematic and random errors are removed from the data. This is often conducted by use of the number of satellites visible and the horizontal dilution of position (HDOP) (Stopher et al., 2005). The latter determines how the satellites are dispersed, and the most accurate positioning is obtained when the satellites are well dispersed equalling a low HDOP value. Random errors can be dealt with by including a data smoothing algorithm (Schüssler and Axhausen, 2009).

Dependent on the functionality of the GPS device used, trip end points (activity points) are often identified at points when the device has been stationary for a period of time and/or if the density of observations has been high for a period of time indicating that an activity has occurred (Schüssler and Axhausen, 2009; Stopher et al., 2005). The result is a number of trips which are defined as being from one activity point to the next. This approach is evaluated on a subsample of trips in Schüssler (2010), which finds that the algorithm correctly detected 97% of stated activities without detecting any false activities. Most studies further split trips into trip segments, defined by a change of mode. Correct trip segmentation is crucial for the subsequent identification of the mode of travel of the trip segments. Consequently, de Jong and Mensonides (2003) divide trips into trip segments whenever the speed drops to 0 km/t, with the option to combine segments again if no mode change occurred. Schüssler (2010) and Tsui and Shalaby (2006) identifies walking segments if speed and acceleration are low, as all other modes are assumed to be preceded or followed by such walking segments (or by time gaps).

Several studies find that most modes can be identified by only using the speed and acceleration profiles gathered by the GPS device. Moreover, Bolbol (2012) found that, while using the acceleration rather than the speed profile induces better results when distinguishing between modes, the best results are obtained if combining the two profiles. While this is an easy and efficient approach for some modes, it is often not sufficient to enable a clear distinction between certain modes. For example Bolbol (2012) found that bus and bicycle trips in the Greater London Area have similar speed and acceleration profiles, and Tsui and Shalaby (2006) found that bus characteristics overlap with characteristics of several other modes. Other techniques have been proposed to improve the mode detection, including GIS analyses for detection of public transport trips based on a tram, metro, rail, and bus network (Tsui and Shalaby, 2006). In Gong et al. (2011) rail and bus trips are identified based on the proximity of origins and destinations to rail stations and

bus stops. A similar approach for bus trips is proposed in Schüssler (2010). Using the proximity to bus stops of origins and destinations to identify bus trips seems not sufficient in urban areas in which the bus network is extensive; trip legs starting and ending near bus stops might have been done by e.g. bicycle rather than by bus. Another way of improving the distinguishing between modes is to utilise information about the respondents implicitly in the identification of modes. In Stopher et al. (2008) this is done by only allowing the algorithm to assign car or bicycle as mode for a trip only if the household has a car or bicycle at disposal, respectively. However, these approaches have limitations if comparing to a typical Scandinavian city where the bus network is dense and the ownership of bicycles is relatively high.

3. Case study: Greater Copenhagen Area

The study area covers the Greater Copenhagen Area in which approximately 2 million people live, and the case study utilises data collected as part of the ongoing research project "Analyses of activity-based travel chains and sustainable mobility" (ACTUM). The dataset includes 53 households, corresponding to 183 persons in the range from 6 to 58 years of age. The households were sampled from the Danish National Travel Survey (Christiansen, 2012), and all participants were asked to bring a GPS device on all trips undertaken within a period of 3-5 days. Additionally, each respondent were asked to fill in an internet-based travel diary corresponding to one of the days for which GPS data was also collected. This enables a validation of the proposed fully automatic trip- and mode detection algorithm.

Linking the travel diaries to the recorded GPS observations generated a database containing travel diaries with corresponding GPS data for 101 person days. Consequently, there were 82 persons for which data could not be linked, and an analysis identified that this was due to one of the following three reasons; (i) the respondent failed to answer the survey for a day where she also carried a GPS device , (ii) no or only a little GPS data was collected for the day where the survey was filled in, or (iii) there was a large difference between the number of trips reported and what could be seen in the GPS-traces.

The GPS device used for the data collection was the wearable KVM BTT08M (KVM, 2013). This device logs data every second, thereby facilitating a high level of accuracy for the identification of en-route travel choices and trip ends. In total (i.e. not only the GPS data on days for which a travel diary was available), the dataset contains 6,419,441 collected GPS points (observations), corresponding to 1,783 hours of travel (including stationary and error data), and was collected on approximately 644 person days of travelling.

The proposed method applies GIS-analyses in various steps. These GIS-based analyses utilise a detailed digital representation of the road and public transport networks of the Greater Copenhagen Area. The road network used is based on the road network of NAVTEQ (2010) and is in a format that allows for a complex map matching algorithm to be run (Nielsen and Jørgensen, 2004). The public transport network used for

the mode identification of rail trip legs is a digital representation of the rail line alignment in the Greater Copenhagen Area. The analysis distinguishing between bus and car utilises a disaggregate network representation containing bus route alignment and stop locations for all bus-lines and bus-line variants in the Greater Copenhagen Area.

4. Method

The study proposes a fully automatic method to post-process GPS data. Without requiring any information about the GPS carrier, the method performs, and iterates between, a series of steps identifying activities (trip ends), trip legs and the most probable mode chosen. The method proposed is based on the fully automatic trip and mode detection algorithms developed in Schüssler and Axhausen (2009). The method proposed in Schüssler and Axhausen (2009) are modified in order to improve the results by e.g. utilising the availability of a highly disaggregate representation of the public transport network covering the Greater Copenhagen Area. Moreover, the proposed method extends the approach of Schüssler and Axhausen (2009) by (i) accompanying the use of speed and acceleration distributions of the trip segments identified with GIS analyses to better distinguish between modes with similar speed and acceleration characteristics, and (ii) incorporating advanced feedback loops between steps, allowing inconsistent mode-sequences to feedback into and alter the trip leg detection algorithm. The proposed method contains the traditional 4-step process (Figure 1), and a detailed description of the steps of the algorithm is presented in the following subsections.
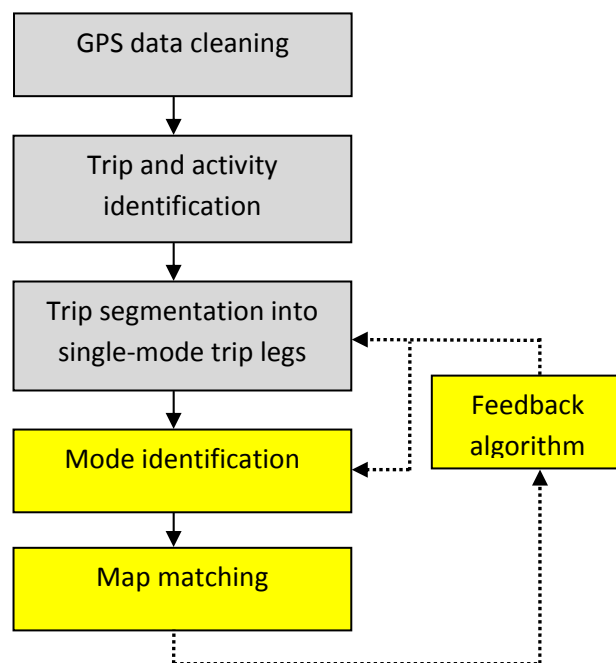

Figure 1: Approach used in this study. Boxes highlighted in grey denote steps that are similar to corresponding steps in Schüssler and Axhausen (2009) whereas the boxes highlighted in yellow are steps where this paper contributes with new, alternative methods.

4.1. GPS data cleaning

The parameter values used for data cleaning varies across studies (Stopher et al., 2005; Schüssler and Axhausen, 2009; Tsui and Shalaby,2006; Stopher, 2005). However, most studies require, in order to get coordinates in three dimensions, four satellites to be visible (Stopher et al., 2005; Schüssler and Axhausen, 2009). Furthermore in order to get accurate positions it is often required that the satellites are sufficiently dispersed corresponding to a HDOP-value of less than 4-5 (Tsui and Shalaby,2006; Schüssler and Axhausen, 2009). This study adopts the values from Schüssler and Axhausen (2009) which requires that a minimum of four satellites were visible and that the horizontal dilution of position (HDOP) was less than four. In addition, due to the small variation in altitude in Denmark, only observations with altitude levels between -37 meters and 201 meters were regarded as acceptable. This corresponds to the altitude range in Denmark +/- 30 meters[2]. Lastly, the data were smoothed using a Gauss kernel smoothing approach to remove systematic errors and perform data smoothing as suggested by Schüssler and Axhausen (2009).

4.2. Trip and activity identification

The trip and activity identification algorithm developed in Schüssler and Axhausen (2009) was applied to identify trips. The activities (trip ends) are identified by locations where the bearer of the GPS is stationary for a period of time of a length being an optimal compromise between identifying short stops (e.g. picking up persons) and falsely detecting activities (e.g., when driving in congested traffic or waiting at traffic signals) (Chen et al., 2010; Tsui and Shalaby, 2006; Stopher et al., 2005; Wolf, 2000). Consequently this current study adopts the method put forward by Schüssler and Axhausen (2009), and defines an activity point if one of three criteria is met; (i) if there is a time gap between consecutive observations of 120 seconds or more, (ii) if the speed has been lower than 0.01 m/s for at least 60 seconds, or (iii) if the location of the GPS device is within a limited area for at least 60 seconds. The first situation occurs when the device has been stationary for a while, which causes the unit to turn off to save battery, or if the GPS signal is lost during a trip leg. If the signal is lost the last observation will be flagged as "beginning of (time) gap", and the first observation when the signal is re-established will be flagged as "end of (time) gap". The last criterion is analysed by use of *clouds* or *stop points* which is the number of observations within a 15 meter radius of the respective observation. If this number exceeds 30 observations for more than 60 seconds (60 observations as the GPS device used logs every second) an activity is flagged. When walking at a normal pace the number of observations within a 15 meter radius is typically 20-25, whereas the corresponding number for travel by car typically is below 5.

4.3. Trip segmentation into trip legs

A trip between two activities might involve several trip legs with different modes of transport or changing between vehicles of the same mode (e.g. changing between train lines). An example of a trip involving

---

[2] +/- 30 meters is regarded as the standard variation of measurement for the GPS devices

several trip legs is a trip where the traveller bikes from home to a train station, walks from the bicycle to the platform at the train station, rides on the train and then walks from the departing train station to the destination. Such a trip involved four trip legs, and in order to get a disaggregate representation of the travel patterns of the travellers, the trips identified needs to be split into trip legs. This is done by applying the approach of Schüssler and Axhausen (2009). Trip legs are identified by assuming that a *short* walking stage is needed between modes, e.g. from bike to train or from bus to another bus, similarly to in Tsui and Shalaby (2006). The walking segments are identified by means of the unique characteristics of walking (low acceleration and low speed). If an identified walking trip segment is *long* (90 seconds or longer), it is defined as a separate trip leg.

The current study specifies a new trip leg if 3 criteria are fulfilled. The first criterion requires the speed in the mode change (or walking stage) to never exceed 2.00 m/s and the acceleration to never exceed 0.1 m/s^2. The second criterion requires the length of the derived trip legs to be at least 90 seconds for walking trip legs and at least 120 seconds for trip legs of all other modes. The last criterion requires that the short walking section between two trip legs must include observations that are flagged as a "start of walk" (or "end of gap") and an "end of walk" (or "beginning of gap") (see section 4.2). Though not identical, these criteria are similar to the ones applied in Schüssler and Axhausen (2009)[3].

4.4. Mode identification

As part of the GPS data post-processing, each trip leg is associated with the most probable transport mode used. The identification process is partly based on speed and acceleration profiles, similar to the approach used in Schüssler and Axhausen (2009) and in Bolbol et al. (2012). The driving conditions in the Greater Copenhagen Area however ranges from being slow moving traffic through congested urban areas to fast moving traffic on motorways. Additionally, in urban areas it is hard to distinguish between whether driving in a bus or following behind it in a car, or even biking next to it. Consequently, it is hard to distinguish between modes solely based on acceleration and speed profiles.

The present study therefore proposes the three-step mode identification process illustrated in Figure 22. This process is based on analyses using both the speed and acceleration profiles as well as more advanced analyses conducted in GIS-software. The steps are explained further in the following subsections.

---

[3] Schüssler and Axhausen (2009) used a maximum speed of 2.78m/s and a minimum duration of walking trip legs of 60 seconds.
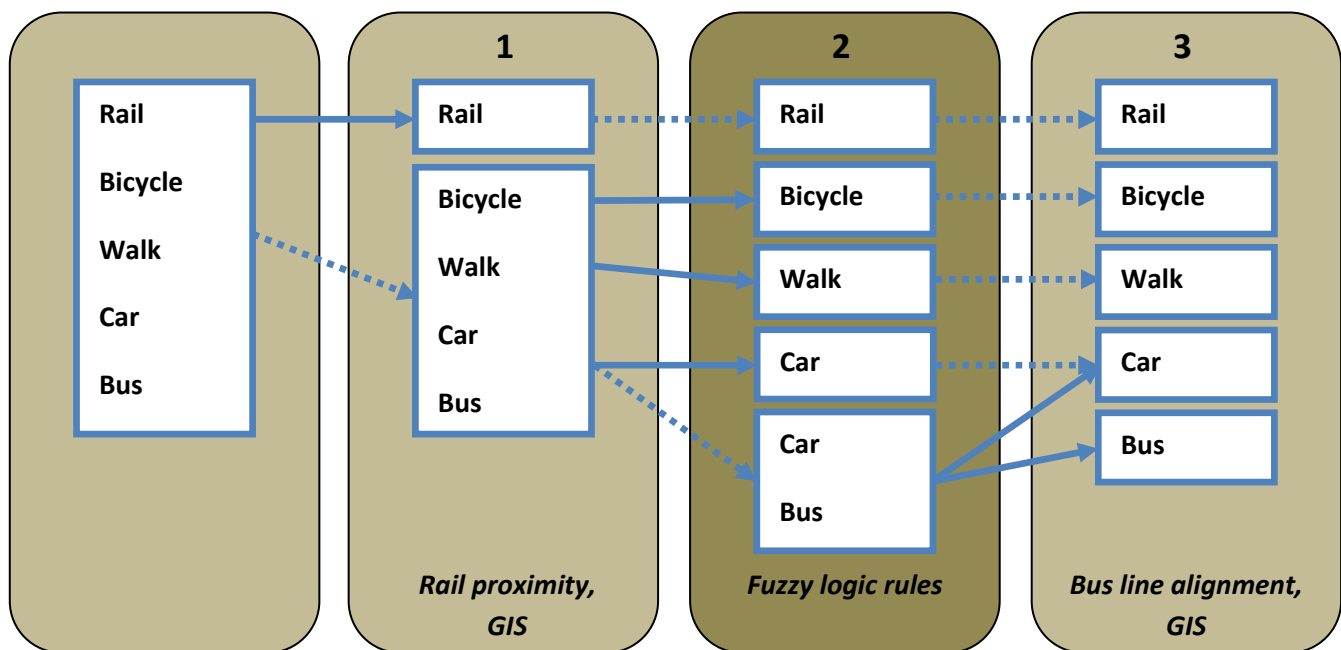
Figure 2: The step-wise mode classification algorithm. Continuous arrows denote mode classification whereas dotted arrows denote no change from previous step. Step 2 is directly adopted from Schüssler and Axhausen (2009), but with adapted fuzzy logic rules.

4.4.1. Step 1: Rail proximity

Rail networks are typically characterized by not having the same spatial location as the street- and path network, with the exception of on-street light rail and tram lines. This is the case in the Greater Copenhagen Area, and trip legs using rail can easily be distinguished from others by their close proximity to the alignment of the rail network. Consequently, the first step identifies rail trip legs based on the proximity of observations to the rail network. If more than 75% of observations in one trip leg are located less than 25 meters from the rail network, the trip segment is classified as being a rail trip. Additionally, to avoid classifying walking trips on railway station platforms as rail trips, the length of a rail trip leg is required to be at least 250 meters. This is less than the shortest distance between railway stops in the Greater Copenhagen Area, but longer than most within-platform walking trips. An example of a successfully identified rail trip leg is shown in Figure 33 where almost 98% of observations for a trip leg are located within 25 meters of the rail network.

Trip leg information

Date:	09-11-2011
Starting time:	16:07:01
Ending time:	16:22:53
Trip length:	15 min 52 sec
Number of obs.:	947
Number of obs. within 25m:	924
Percentage rail:	97.6%

Figure 3: Example of a rail trip leg on the Danish S-train ring line (Ringbanen) identified by the rail proximity algorithm.

4.4.2. Step 2: Fuzzy logic rules

Having identified rail trip legs, the next step is to apply the fuzzy logic method by Schüssler and Axhausen (2009) to determine the mode of travel of the remaining trip legs. The distinction between walk, bicycle, car and bus is done by applying certain logic rules to the speed and acceleration profiles of the trip legs. The modes are best distinguished if using the median speed together with peak values of speed and acceleration. Most studies represent the peak values using 75-95 percentiles of the speed and acceleration to take into account outliers (Stopher et al., 2005; Gong et al., 2011; Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009). As in Schüssler and Axhausen (2009) this study uses the 95th percentiles of speed and acceleration in addition to the median speed.

The set of rules are based on dividing the distribution of the median speed as well as the distribution of the 95th percentiles of speed and acceleration into intervals as proposed by Schüssler and Axhausen (2009). The distributions were divided into four (possibly overlapping) intervals, i.e. very low, low, medium and high, based on an empirical analysis of the sample of trip legs for which the mode was known (Figure 44). Combining these defined intervals across the profiles and applying certain fuzzy logic rules facilitates the mode identification.
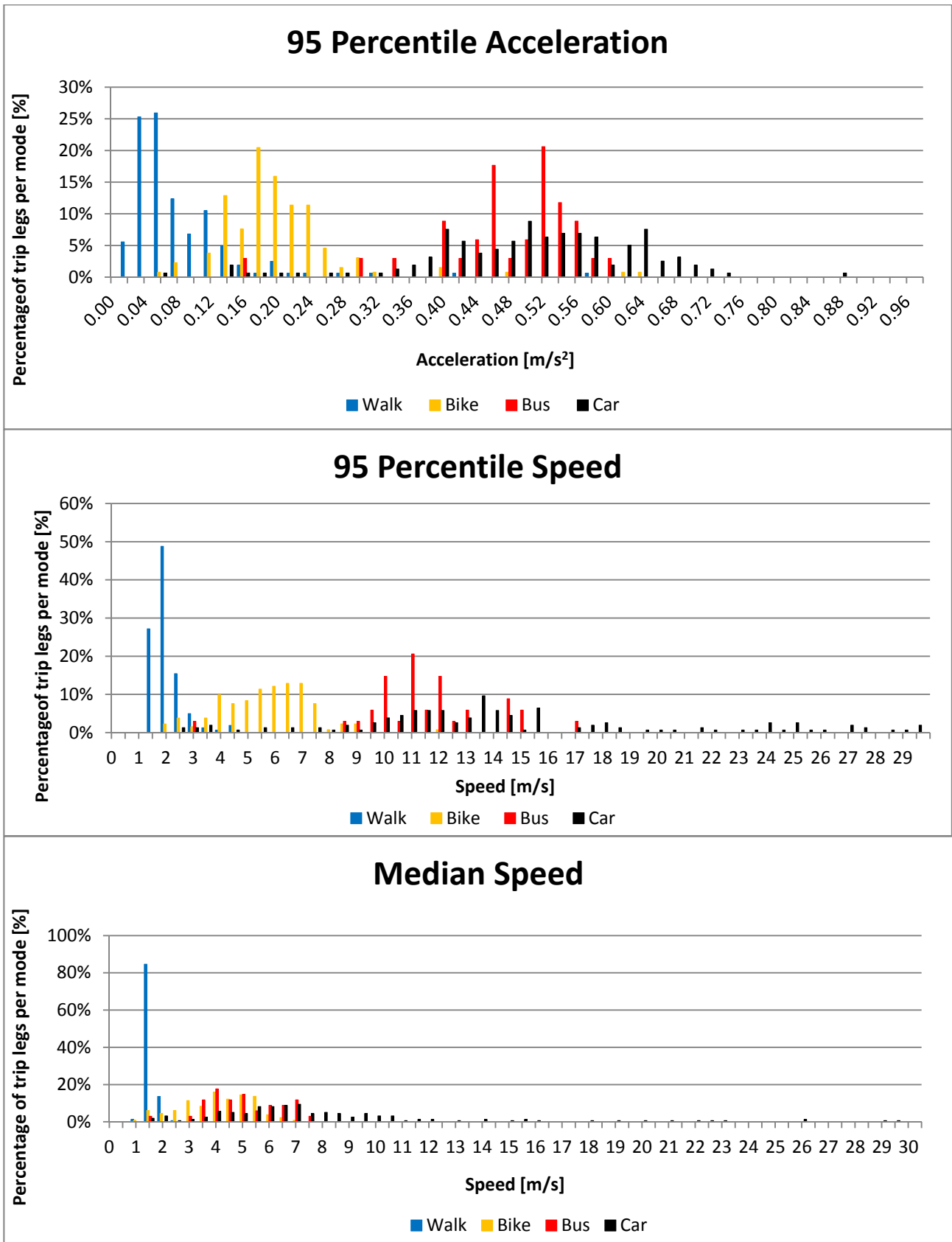
Figure 4: The distributions of the 95ᵗʰ percentiles of speed and acceleration and the median speed for the subset of trip legs for which the mode is known (rail excluded).

In Figure 4 it can be seen that even though most trip legs are easily distinguishable based on a combination of speed and acceleration, this method does not uniquely separate modes. Walk and bicycle are the two

modes which can, based on the profiles, be identified with a high chance of success as these have the least overlaps with the other modes due to the consistently low maximum speed. When combining the profiles, the two modes can also be distinguished from each other.

Trip legs undertaken by car and bus are difficult to distinguish from each other based on the profiles. This is due to the somewhat similar acceleration and speed profiles of the two modes in urban areas where trip legs range from slow moving traffic in dense city centre traffic to fast moving motorway traffic with very different speed and acceleration characteristics. In order to facilitate the distinction between bus and car trip legs, this study proposes a two-fold approach to identify bus trips. Initially a separation is done based on the profiles, clearly distinguishing between trips which are assumed to be definitely car trips and trips which are either car or bus trips. Hence, all trip legs which confer with the speed and acceleration intervals for bus trips are classified as potential bus trips. Through this, this sample of *initially* classified bus trips includes all actual bus trips and a large subset of car trips. Subsequently the identification of *actual* bus trips among this set of *potential* bus trips is done in step 3. The union of the set of trip legs initially identified as car trips and the trip legs not classified as *actual* bus trips among the set of *potential* bus trips constitutes the set of car trips.

4.4.3. Step 3: Bus line alignment

The present study proposes a new approach for separating car- and bus trips. The new method is based on a thorough analysis of coherence between GPS carrier stopping locations and bus line bus stops; the subset of *possible* bus trip legs are analysed to identify whether they follow the stopping pattern of any certain bus line. The initial step of the identification specifies that if at least 15 GPS-observations are located less than 25 meters from a bus stop the algorithm will flag the trip leg as stopping at the bus stop. Next, if the GPS carrier stops at at least 60% of potential bus stops[4] on any bus line, the trip will be flagged as a *probable* bus trip (on bus lines fulfilling this criterion). The rather low percentage of 60% is applied to take into account bus routes with few passengers where the bus often does not stop at all bus stops. For high level-of-service bus lines, the threshold is set at 80%.

Subsequently, the sample of *probable* bus trips are analysed with regards to the origin and destination of the trip legs. The start point *and* end point of trip legs classified as *probable* bus trip legs are analysed to see if they are located less than 100 meters from a bus stop on any of the bus line(s) identified previously. If this is the case then the trip leg is classified as a bus trip. Otherwise the trip leg will be classified as a car trip. An additional benefit of applying this method is that the most probable actual bus line used is identified.

---

[4] Between boarding and alighting stops

Initial results showed that due to sometimes long stopping times, the trip leg identification algorithm in some cases splits one actual bus trip leg into several trip legs. To take this into account, the bus line alignment algorithm, for every flagged trip leg, will search for previous and subsequent trip legs within a timeframe of +/- 300 seconds which are classified as either car or bus. If such trip legs exist, these will also be flagged as *probable* bus trip legs assuming that within a short timeframe it is more probable that two consecutive trip legs with similar speed and acceleration characteristics are of the same mode rather than a shift from car to bus or vice versa.

Two examples of the application of the proposed method are shown in Figure 55. The example to the left is an actual bus trip, whereas the example shown to the right is an actual car trip. The stop analyses for the two trip legs showed that there are clusters of observations at a large percentage of stops associated to several bus lines. The second part of the analysis determines whether the trip leg start and end in the close proximity of any of the bus stops on the bus lines identified; in the example to the left the GPS carrier stops at eight out of eleven bus stops of bus line 168 *and* the trip leg also begins and ends close to bus stops on this line. This causes the trip leg to be correctly classified as a bus trip leg. In the example to the right, the GPS carrier does stop several places along bus line 161, but the trip leg is correctly classified as a car trip leg as the trip leg starts and ends more than 100 meters away from a bus stop served by bus line 161.
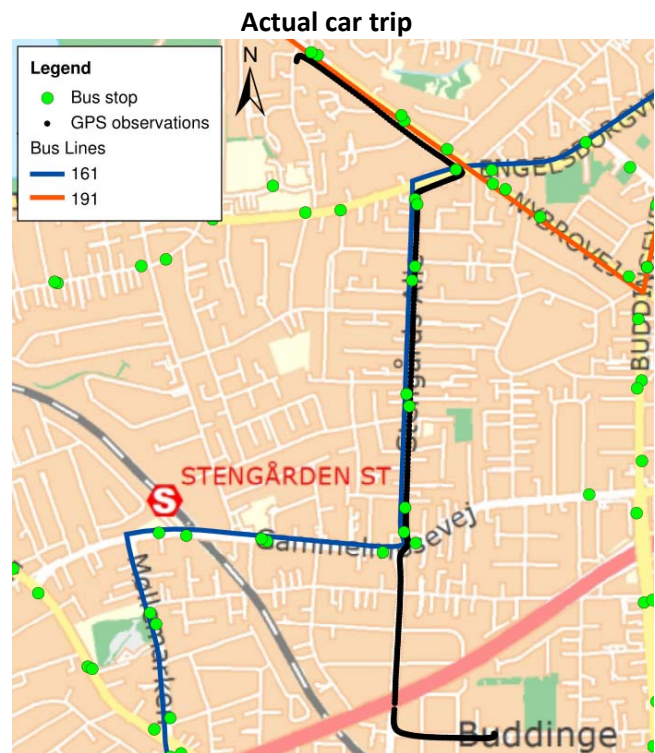
Actual bus trip

Actual car trip

Trip information

Date:	27-10-2011
Starting time:	08:40:38
Ending time:	08:49:37
Trip length:	8 min 59 sec

Bus line stop overlap	# stops	Hit%
145	4/5	80%
161	3/4	75%
167	8/11	72%
168	8/11	72%
81N	3/4	75%

Origin stop:	168
Destination stop:	67/145/161/167/168
Classified as:	Bus trip
Actual bus line:	168

Trip information

Date:	09-11-2011
Starting time:	06:42:26
Ending time:	06:49:07
Trip length:	6 min 41 sec

Bus line stop overlap	# stops	Hit%
161	3/5	60%

Origin stop:	None
Destination stop:	None
Classified as:	Car trip
Actual bus line:	-

Figure 5: Example of results from the bus stop algorithm.

4.5. Algorithmic feedback

Having assigned a most probable mode to all identified trip legs, the next step is a feedback algorithm which seeks to improve the results of the trip leg- and mode identification algorithm by identifying and correcting irregular mode shift patterns. This is done to avoid wrong modal classification caused by irregular changes in speed and/or acceleration for a trip leg, e.g. when queuing in congested traffic. The feedback algorithm uses simple logic rules to analyse and identify irregular mode shifts and is based on a tree of probable mode transfers. For example, it is likely that a bicycle stage follows or precedes a bus stage

as some passengers might bicycle to and from the bus stop, but it is not very likely that a bicycle stage is followed by a car stage with only a short time gap between the trip legs.

Specifically, the algorithm searches for consecutive trip legs ending and beginning within 300 seconds and 50 meters of each other for which the mode identification method described above have detected a *non-likely* mode change. For such cases all trip legs concerned are merged into one single leg, and the mode for this merged trip leg is classified as the most probable mode for the merged trip leg. The *non-likely* mode changes are bicycle to/from car and bus to/from car.

The feedback algorithm additionally searches for sets of three consecutive trip legs where the first and third stage is identified as a car trip, but the second is identified as another mode. If the time between the end of the first trip leg and the beginning of the third is less than 300 seconds *and* the distance between the end point of the first trip leg and the start point of the third trip leg is more than 30 meters apart (more than three times the standard deviation of measurement), the three trip legs are assumed to be the same car trip. Hence, they are connected to constitute one single trip leg with the mode being set to car. This is assumed as there is a net movement from the end point of the first trip leg to the starting point of the third trip leg done within a short period of time. This indicates that the trip legs are most likely one single actual trip leg;, and the approach then ensures trip legs that are mistakenly split due to congestion to be successfully connected.

4.6. Map matching

In the last part of the algorithm, all trip legs (except trip legs identified as rail trips) are map matched to the NAVTEQ road network (NAVTEQ, 2010) using a map matching algorithm developed at DTU Transport (Nielsen and Jørgensen, 2004). This has two purposes; (i) to detect and correct trip legs which are wrongly split due to congestion on motorways, and (ii) to remove non-trips, such as short trip legs generated as a consequence of the GPS device being turned on when no trip is actually undertaken. The first case is identified as the case where the matched end link and start link of two consecutive trip legs are either a motorway or ramp. In such a case the two trip legs are merged into one with the mode being set to car or bus depending on the most probable mode of the merged trip legs. In the latter case non-trips are defined as trip legs for which less than half of the mapped route is generated as a consequence of mapping of actual GPS observations[5]. Such non-trips are discarded. This should ensure short non-trips to be successfully removed, but might also cause the removal of many actual walking trips, e.g. short walking trips through parks, etc.

---

[5] In cases where only a part of the observed route can be map-matched, the map matching algorithm generates the shortest path between links to which observations can be mapped.

5. Results

In order to allow evaluating the effect of the new components proposed in this current study, two configurations of the proposed method (section 3) was tested on the available dataset; (i) an *Algorithm 1* including trip leg and mode identification as well as feedback algorithm (sections 4.1-4.5), but excluding the map matching algorithm, and (ii) an *Algorithm 2* which includes *Algorithm 1* and the map matching algorithm (sections 4.1-4.6). An algorithm similar to *Algorithm 1* but without the feedback algorithm was also evaluated. The results of this evaluation are however not reported, as only 11 trip legs were connected by the feedback algorithm, thereby inducing the results of the two to be almost identical. To also facilitate the evaluation of the effect of the proposed mode identification algorithm, a *Baseline algorithm* was also tested, and the results of this are also reported in the following. The *Baseline algorithm* includes trip leg and mode identification as proposed by Schüssler and Axhausen (2009), i.e. with the mode identification step based on fuzzy logic rules. For each of the three algorithms, different configurations of intervals as well as rules were tested, and the following reports the results for the configuration that, for each algorithm, induced the best overall success rate.

While the above algorithms have been run on the full dataset consisting of approximately 664 person days, the following only presents results of a subset of this data, namely the trip legs for which the travel mode etc. can be known for certain. This includes trips that are directly connected to the travel diary data supplied by the respondents as well as trips where in-depth investigation made it possible to, with a low risk of error, deduct the travel information manually. The results of the mode identification are presented using two assessment measures, namely (i) the *success rate* which denotes the number of correctly identified trip legs by the algorithm as percentage of the number of *observed* trip legs of that mode in the sample; and (ii) the *confidence rate* which denotes the number of correctly identified trip legs by the algorithm as percentage of the number of trip legs of that mode identified by the algorithm. The latter refers to the percentage of trip legs where the mode is correctly identified in the sample of trip legs from the output of the algorithm. Hence, the first measure relates to the *observed* travel survey trip legs whereas the second measure relates to the trip legs in the output of the *algorithm* which also includes non-trips (see section 4.6).

5.1. Trip leg identification

The total number of trip legs identified is 754, 744 and 464 if using the *Baseline algorithm*, *Algorithm 1* and *Algorithm 2,* respectively. This compares to the total number of trip legs in the travel survey of 521, and in the evaluation it is important to bear in mind that three sources of error influence the numbers; firstly, there are trip legs in the travel survey where no corresponding GPS trip legs could be identified. This could be due to either the respondent not wearing the GPS, the GPS device not being able to get an acceptable signal or the device did not function properly. Secondly, some trip legs were identified by the algorithm

even though no corresponding trip information was reported by the respondents in the diary. This error is due to underreporting by the respondents, which has also been observed in other studies including Stopher et al. (2007) and Wolf et al. (2003). Lastly, the algorithm detects trip legs based on a dwell time threshold which might cause actual trip legs to be wrongly split. For example, one actual trip leg might be separated by the algorithm into several trip legs due to long dwell times while travelling, e.g. stop-and-go in congested traffic. The opposite might also happen, in that several actual trip legs might be connected into one trip leg by the algorithm if the dwell time between trip legs is very low. The latter could happen when a fast cyclist transfers to a local train or bus without any waiting time at the station.

Figure 6 illustrates the results of a comparison between the trip legs generated by the proposed methods and the trip legs reported by the respondents.
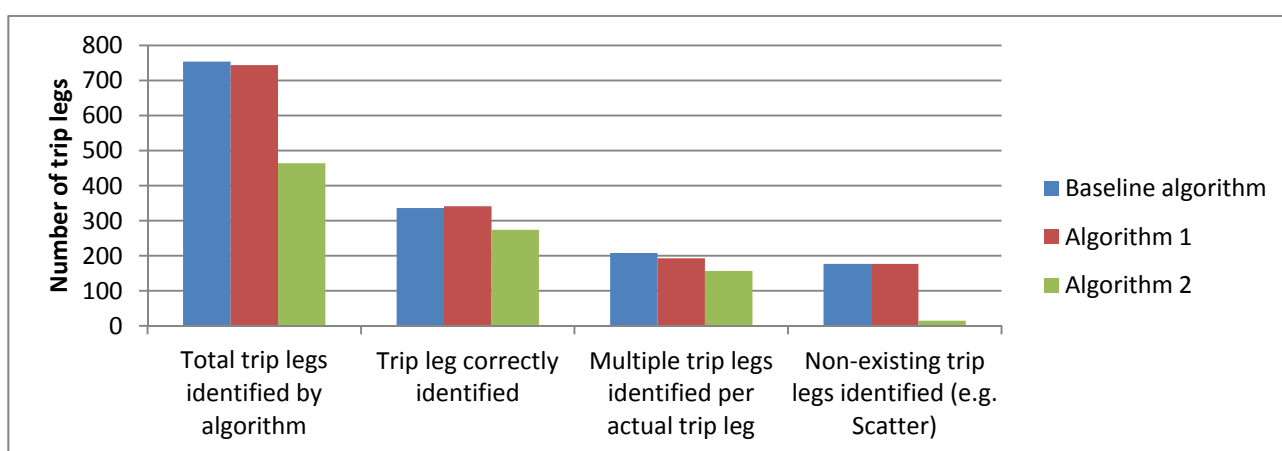

Figure 6: Classification of trip legs identified by the algorithms.

Looking at the *Baseline algorithm*, it can be seen to generate trip legs with the correct origin and destination for 45% of the identified trip legs, while 28% of the identified trip legs are partial trip legs, i.e. one reported trip leg is identified as two or more trip legs by the algorithm. A further 24% of the trip legs identified are non-trips[6] which should not have been detected as a trip leg. The remaining 4% represent trip legs that either include several actual trip legs (not split correctly), short walking trip legs at an activity point (e.g. walking 50 meters to the car or walking within an office building), or trips where observations are too low quality for general usage.

Applying the proposed mode-order and feedback algorithm improves the results (*Algorithm 1*). Fewer trip legs are identified and more trip legs are correctly identified. Additionally, the feedback algorithm causes fewer actual trip legs to be wrongly split into several trip legs, as eleven partial trip legs were successfully connected into actual complete trip legs. If also including the map matching algorithm of *Algorithm 2* a further nine trip legs were successfully connected into four actual trip legs and 143 trip legs were correctly

---

[6] Non-trips include random scatter and short trip legs which are not actual trips , e.g. walking around at the work place, etc.

removed from the sample, cf. Figure 66. However further analysis showed that some trip legs which should be merged remain, and more effort should be put on connecting these trip legs in a further study. Overall, the best results are achieved when using *Algorithm 2* as this identifies the entire actual trip leg as one trip leg in 59% of the cases, and the entire actual trip leg as one or several trip legs in 93% of the cases. Additionally, the percentage of trip legs identified which should not have been identified drops to a very low level (3%).

The *Baseline algorithm* and *Algorithm 1* identifies many very short trip legs (non-trips), and a manual verification has shown that these are successfully removed by the map matching algorithm of *Algorithm 2*. This can also be seen by the distribution of trip lengths for the identified GPS trip legs, as this corresponds better to the distribution of the length of the trip legs reported in the diary (Figure 77).
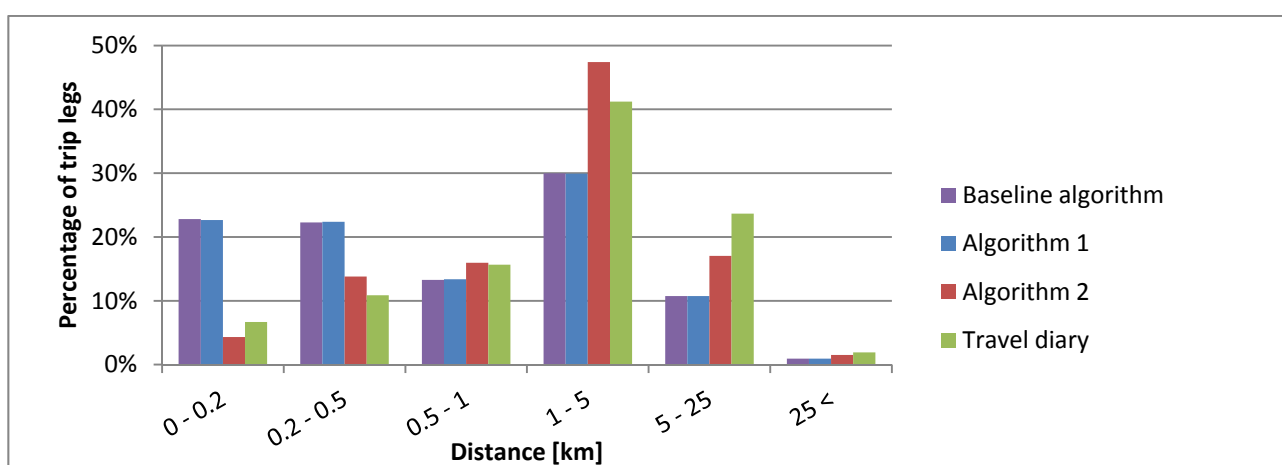

Figure 7: Trip length for identified GPS trip legs compared to stated travel survey trip lengths.

5.2. Mode identification

This section presents the results of an evaluation of the proposed methods' capability to identify the correct mode of the trip legs. This is done by comparing the mode identified by the algorithm to the actual mode chosen as reported by the respondents, for each trip leg. Table 1 reports the results of such a comparison for the *Baseline algorithm*. As can be seen, approximately 82% of the trip legs are assigned the correct mode of transport when only considering trip legs which are actual trip legs. Table 2 reports the results of an analysis of *Algorithm 1*, and the corresponding success rate is 90%. Consequently, including the mode-order analysis and feedback algorithm proposed improves the results considerable. Especially the method proposed to identify rail trips are very efficient – using the fuzzy logic rules caused 24% of the rail trip legs to be correctly identified, whereas the corresponding number for *Algorithm 1* is 97%. Applying the proposed method to identify car and bus also improves the results considerably, as the success rate for bus rises from 38% to 73% while the success rate for car rises from 82% to 93%. The success rates for walking and bicycling reduces slightly for *Algorithm 1* when compared to the *Baseline algorithm*, however the overall success rate of *Algorithm 1* is considerably better.

Table 1 and Table 2 however also highlights a weakness of the two approaches, namely the identification of a lot of trip legs which are not reported in the diary (non-trips generated due to e.g. scatter). This induces the confidence rates to be 62% and 69% respectively. Including the mode order analysis thus classifies generated trip legs considerably better, especially for bicycle, bus (no generated bus trip legs are wrongly classified) and rail. Consequently, while the proposed mode-order and feedback algorithm improves the mode classification of actual trip legs considerably, *Algorithm 1* detects – as also found in section 5.1 – too many trip legs which are not part of any actual trip, e.g. walking around at the work place.

Observed / Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips	Confidence rate
Walk	184	12	2	6	-	111	58.4%
Bicycle	9	121	-	13	-	52	62.1%
Bus	-	1	14	9	-	2	53.8%
Car	-	4	21	143	25	12	69.8%
Rail	-	-	-	3	8	2	61.5%
Other	-	-	-	-	-	1	-
Total	193	138	37	174	33	180	**62.3%**
Success rate	95.3%	87.7%	37.8%	82.2%	24.2%	-	**81.7%**

Table 1: The results of the mode identification when using *Baseline algorithm*.

Observed / Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips	Confidence rate
Walk	180	11	2	2	-	111	58.6%
Bicycle	2	114	-	6	-	15	83.2%
Bus	-	-	27	-	-	-	100.0%
Car	4	8	8	156	1	48	69.3%
Rail	3	-	-	-	33	4	82.5%
Other	3	1	-	3	-	2	-
Total	192	134	37	167	34	180	**68.5%**
Success rate	93.8%	85.1%	73.0%	93.4%	97.1%	-	**90.4%**

Table 2: The results of the mode identification when using *Algorithm 1*.

As mentioned in section 5.1, adding the map matching algorithm of *Algorithm 2* removes many of such non-trips. This improves the overall confidence rate from 69% to 85%, which is a considerable improvement (Table 2). The improvement in confidence rate is however at the cost of reducing the amount of observed trip legs for which a corresponding trip leg is generated. This is because the map matching algorithm in addition to removing a lot of non-trips also removes a large number of generated trip legs for which a corresponding observed trip leg exists. Especially, many trip legs undertaken by foot or bicycle are discarded by the algorithm as the map matching was conducted on a road network. The row denoted by "Success rate (all)" highlights this by representing the share of the *total* number of observed trip legs for which a generated trip leg with the correct mode has been identified.

All trip legs identified as bus by the algorithm are correct. However, looking at the success rates of 73-77% for bus, these are the lowest success rates obtained across modes. A disaggregate analysis has identified two primary reasons for these lower percentages, namely (i) problems associated with the trip leg identification algorithm, i.e. an actual bus trip mistakenly being split into several trip legs due to congestion, longer dwell times, etc.; or (ii) in some cases buses skip a large percentage of stops, e.g. during evening hours where fewer passengers board the bus.

Observed / Algorithm	Walk	Bicycle	Bus	Car	Rail	Non-trips	Confidence rate
Walk	75	6	1	1	-	13	78.1%
Bicycle	1	104	-	5	-	3	92.0%
Bus	-	-	27	-	-	-	100.0%
Car	1	7	7	152	1	19	81.3%
Rail	-	-	-	-	33	2	94.3%
Other	1	1	-	1	-	1	-
Total	78	118	35	159	34	38	**84.6%**
Total (all)	192	134	37	167	34	180	
Success rate	96.2%	88.1%	77.1%	95.6%	97.1%	-	**92.2%**
Success rate (all)	39.1%	77.6%	73.0%	91.0%	97.1%	-	**(69.3%)**

Table 3: The results of the mode identification when using *Algorithm 2.*

6. Discussion

Data allowing a disaggregate digital representation of the infrastructure has become more and more available in recent years. In addition to this, GPS data has become widely used as a means to analyse peoples travel behaviour. Consequently, as more and more GPS data are available and as GPS data collection typically involves large datasets to be analysed on disaggregate network representations, the availability of automated post-processing procedures are essential. In this study such a fully automated and disaggregate method to process raw GPS data is proposed. Moreover, the study proposes a fully automatic, combined fuzzy logic- and GIS-based method to process raw GPS data collected in a person-based survey. With modifications, the proposed method should be applicable to all case studies where data is collected as individual-based GPS-traces and where detailed information on the local infrastructure is available. This study has applied the proposed method to GPS data collected in the Greater Copenhagen Area, for which a disaggregate digital representation of the infrastructure is available. Though the process is automated, it is important to notice that the parameters used in e.g. the segmentation of the speed- and acceleration profiles has to be adapted to fit the characteristics of the case study. The study found that even small changes to the parameters changes the results considerably. In this process of selecting the parameters which induces the best results, it is very useful to have available corresponding stated information of trips undertaken by the respondents (e.g. trip start and end time and location, mode chosen etc.).

When considering only observed trips for which one or more corresponding trip legs exist, the proposed algorithms *Algorithm 1* and *Algorithm 2* both produce success rates above 90%. Comparing to the overall success rates obtained in other studies, these results seems promising. Gong et al. (2011) obtained a success rate of 82.6%, Chen et al. (2010) obtained a success rate of 79.1%, Bolbol et al. (2012) obtained a success rate of 87.4%[7] and Chung and Shalaby (2005) obtained a success rate of 91.6% in their study including 60 trips. Considering the success rates at the level of the mode, it is seen that especially the success rates of 77% for bus and 97% for rail are high; Gong et al. (2011) and Bolbol et al. (2012) obtained success rates of 35.7% and 84.1% for rail, and 62.5% and 58.29% for bus. The high success rates of the current study are obtained as a result of utilising the availability of disaggregate network data. This is supported by the current study's application of the method proposed by Schüssler and Axhausen (2009) on the same dataset – success rates of 24% and 38% were obtained for rail and bus respectively.

The present study analysed all generated trip legs and found that many of these do not have a corresponding observed trip leg reported in the travel diary. This can partly be because of underreporting, but the analysis found that many non-trips were identified around activity locations. While the method proposed in this present study aims at removing such non-trips (through the map matching step), the other studies reviewed does not seem to explicitly deal with such non-trips. Clearly, it is however important for an algorithm to minimise the number of non-trips identified, especially in applications where there is no corresponding travel diaries available for verification purposes.

While generating promising results when applied to data collected in the Greater Copenhagen Area, the proposed method could be further improved in several ways. Firstly, the trip leg identification algorithm was found to wrongly split trips into several trip legs, e.g. when queuing at intersections. While identifying and connecting some of these trip legs, many are not caught by the rules of the feedback algorithm. Further research could go into how to improve the feedback algorithm by using the very disaggregate network representation available. It should be noted that the mode identification algorithm will also induce higher success rates if the trip leg identification does not split trip legs wrongly; analyses showed that several trip legs were misclassified due to trip legs being wrongly split. Secondly, the trip leg identification algorithm identifies many non-trips. The map matching algorithm succeeds in removing most of the non-trips, however at the cost of also removing many trip legs which are actually performed. These wrongly removed trip legs are primarily walk and bicycle trips. The wrong removal of these can be partly explained as a consequence of using the street network for the map matching, and further research could test whether expanding the network to also include paths would improve the results.

---

[7] Success rate has been calculated as the rate is not presented in the paper.

7. Conclusions

In this study a combined fuzzy logic- and GIS-based method to process raw GPS data collected in a person-based survey is proposed. The method relies upon the availability of a disaggregate representation of the transport network. Moreover, it separates the raw GPS data into trip legs and assigns the most probable mode, distinguishing between car, bus, rail, bicycle and walk. The method proposed also considers the identification, and removal of, non-realistic trip legs generated from e.g. scatter (non-trips). This is an issue not found addressed in other studies. Applied to data collected in the highly complex Greater Copenhagen Area, the algorithm produces good results. Especially the methods proposed for the identification of rail and bus trip legs seem to generate very good results when compared to success rates obtained in other studies.

8. Acknowledgements

9. References

Bohte, W. and K. Maat. 2009. "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: Emerging Technologies*, 17(3), pp. 285-297.

Bolbol, A., T. Cheng, I. Tsapakis, and J. Haworth. 2012. "Inferring hybrid transportation modes from sparse GPS data using moving window SVM classification." *Computers, Environment and Urban Systems*, 36(6), pp. 526–537.

Chen, C., H. Gong, C. T. Lawson, and E. Bialostozky. 2010. "Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study." *Transportation Research Part A: Policy and Practice*, 44(10), pp. 830-840.

Chen, J., M. Bierlaire, and J. Newman. 2013. "A probabilistic map matching method for smartphone GPS data." *Transportation Research Part C: Emerging Technologies*, 26, pp. 78-98.

Christiansen, H. 2012. "Documentation of the Danish National Travel Survey." DTU Transport, Department of Transport, Kgs. Lyngby, Denmark.

Chung, E. and A. Shalaby. 2005. "A Trip Reconstruction Tool for GPS-based Personal Travel Surveys." *Transportation Planning and Technology*, 28(5), pp. 381-401.

de Jong, R. and W. Mensonides. 2003. "Wearable GPS device as a data collection method for travel research." Working Paper, ITS-WP-03-02, The University of Sydney, Sydney, Australia.

Draijer, G., N. Kalfs, and J. Perdok. 2000. "Global Positioning System as Data Collection Method for Travel Research." *Transportation Research Record*, 1719, pp. 147-153.

Du. J, and L. Aultman-Hall. 2007. "Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues." *Transportation Research Part A: Policy and Practice*, 41(3), pp. 220-232.

Forrest, T. and D. Pearson. 2005. "Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by GPS." *Transportation Research Record*, 1917, pp. 63–71.

Gong, H., C. Chen, E. Bialostozky, and C. T. Lawson. 2011. "A GPS/GIS method for travel mode detection in New York City." *Computers, Environment and Urban Systems*, 36(2), pp. 131-139.

Herrera, J.C., D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, A. M. Bayen. 2010. "Evaluation of traffic data obtained via GPS-enabled mobile phones: The *Mobile Century* field experiment." *Transportation Research Part C: Emerging Technologies*, 18(4), pp. 568-583.

KVM, 2013. "GPS-BTT08M". "Product description." Webpage, http://www.kvm.com.au/store/pdf/GPS-BTT-08M.pdf.

Li, H., R. Guensler, J. Ogle, and J. Wang. 2004. "Using Global Positioning Systems Data to Understand Day-to-Day Dynamics of Morning Commute Behavior." *Transportation Research Record*, 1895, pp. 78–84.

NAVTEQ (2010). NAVSTREETS Street Data Reference Manual v2.8.

Nielsen, O.A. 2004. "Behavioral Responses to Road Pricing Schemes: Description of the Danish AKTA Experiment." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 8(4), pp. 233-251.

Nielsen, O. A. and R. M. Jørgensen. 2004. "Map-Matching Algorithms for GPS Data – Methodology and Test on data from the AKTA Roadpricing Experiment in Copenhagen." *Presented at 5th TRIannual Symposium on Transportation ANalysis (TRISTAN V)*, Le Gosier, Guadeloupe.

POSDAP. 2012. "Position Data Processing." Webpage, http://sourceforge.net/projects/posdap/.

Quiroga, C.A. and Bullock, D. 1998. "Travel time studies with global positioning and geographic information systems: an integrated methodology." *Transportation Research Part C: Emerging Technologies*, 6(1-2), pp. 101-127.

Quiroga, C.A. 2000. "Performance measures and data requirements for congestion management systems." *Transportation Research Part C: Emerging Technologies*, 8(1-6), pp. 287-306.

Rich, J. and O. A. Nielsen. 2007. "A socio-economic assessment of proposed road user charging schemes in Copenhagen." *Transport Policy*, 14(4), pp. 330-345.

Schüssler, N. and K. W. Axhausen. 2009. "Processing GPS raw data without additional information." *Transportation Research Record*, 2105, pp. 28–36.

Schüssler, N. 2010. "Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behaviour." Ph.D. thesis. ETH Zurich, Switzerland.

Stopher, P. R., Q. Jiang, and C. Fitzgerald. 2005. "Processing GPS Data from Travel Surveys." Presented at 2nd International Colloquium on Behavioral Foundations of Integrated Land-Use and Transportation Models: Frameworks, Models and Applications, Toronto, Ontario, Canada.

Stopher. P, C. Fitzgerald, and M. Xu. 2007. "Assessing the accuracy of the Sydney Household Travel Survey with GPS." *Transportation*, 34(6), pp. 723-741.

Stopher, P., E. Clifford, J. Zhang, and C. Fitzgerald. 2008. "Deducing Mode and Purpose from GPS Data." Working Paper, ITLS-WP-08-06, The University of Sydney, Sydney, Australia.

Stopher, P. and L. Shen. 2011. "In-Depth Comparison of Global Positioning System and Diary Records." *Transportation Research Record*, 2246, pp. 32-37.

Tsui, S. Y. A. and A. S. Shalaby. 2006. "Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems." *Transportation Research Record*, 1972, pp. 38–45.

Wagner, D. P., D. M. Neumeister, and E. Murakami. 1996. "Global Positioning System for personal travel surveys." FHWA, U. S. Department of Transportation.

Wolf, J. 2000. "Using GPS Data Loggers To Replace Travel Diaries In The Collection of Tracel Data." Ph.D. thesis, Georgia Institute of Technology.

Wolf, J., M. Oliveira, and M. Thompson. 2003. "Impact of underreporting on Mileage and Travel Time
Estimates: Results from Global Positioning System-Enhanced Household Travel Survey."
Transportation Research Record, 1854, pp. 189–198.

Yalamanchili, L., R. M. Pendyala, N. Prabaharan, and P. Chakravarthy. 1999. "Analysis of Global Positioning
System-based data collection methods for capturing multistep trip-chaining behavior."
Transportation Research Record, 1660, pp. 58-65.