ROUTE CHOICE ANALYSIS USING A GPS-BASED DATA COLLECTION

Alessandro Vacca, University of Cagliari, Italy, alevacca@unica.it

Italo Meloni, University of Cagliari, Italy, imeloni@unica.it

ABSTRACT

This paper aims to provide greater insight into the factors underlying route choice behaviour mechanisms, examining actual user route choice decisions. The information has been drawn from a database containing information collected using a smartphone with built-in GPS tracking device, for recording both the spatial-temporal characteristics of the routes chosen and the activities carried out by the users during the survey. We monitored a sample comprising 109 volunteers, who travelled 626 routes during morning peak hours. The main objective was to compare the observed choices with the least cost paths in terms of time and distance using both GIS and a specialised software for simulating transport systems (CUBE, Citilabs). The results show that for systematic habitual travel individuals tend to place more importance on time and distance attributes and thus to choose routes similar to those simulated using the model (least cost paths). In addition, for repetitive trips between the same origin and destination, users travel the same route 70% of the time, the remaining 30% involving two or more routes.

Keywords: GPS, Route Choice, Personal probe system, simulation based routes

INTRODUCTION

Traffic congestion problems are on the rise worldwide, especially in urban areas and to an even greater extent in the developing countries, and likewise the negative effects congestion produces. At the same time, however, there is greater environmental awareness about the deteriorating quality of life and the effects of climate change. What is more, with the current economic downturn fewer resources are available for large transport infrastructure projects that in any case require long time frames for completion. Actions that endeavour to overcome these problems, include, especially in the short term, measures aimed at more efficient urban traffic management. From this viewpoint, one area where researchers could focus their attention is in gaining a deeper insight into the processes underlying drivers’ route choice decisions, in an attempt to encourage them to use the road network more efficiently through a variety of measures and target actions (information campaigns, marketing, advanced communication systems with ITS technologies and ATMS - advanced traffic management systems, pricing policies). The choice of route is that part of the trip most influenced by these types of measures and systems. The behavioural component strongly influences the choice mechanism. Thus, for example, adequately informed and sensitized individuals, by simply changing their travel choices, may decide to use the proposed alternatives and consequently affect the characteristics and operating level of the transportation system.

Thus, improving our understanding of route choice is important for a number of reasons. First of all the complexity of the decision making process underlying route choice. Of all travel choice decisions this is one of the most complicated to interpret both as to the number and role of the different aspects and attributes, and the inherent dynamics of the choice itself. Travel decision making involves both the individual characteristics of the user (social, economic, personal) and the topological features of transport supply (route characteristics). User choice is the result of the dynamic interaction of these two components. Compared to other choice dimensions, route choice is affected to a greater extent by habit, characterizing it as repetitive and relatively unchanged over time and space.

Because of the complexity of this issue researchers have traditionally focused their attention on the simpler components related to choice, namely the directly measurable objective attributes such as time, distance and cost, in any case all those attributes more or less directly associated with the operating and topological characteristics of the transport network. Following this approach models and simulation software have been developed that, though on the one hand suitably describe the global network load conditions, on the other do not adequately reproduce individual behaviour dynamics, (analysis of behaviour al attributes at the individual level). The reason for this also lies in the fact that the data used for calibrating these tools consist of aggregated vehicle counts in a variety of representative sections, but that only provide quantitative information on traffic volumes.

Recently, research in this area has focused on gaining greater insight into those aspects directly related to behaviour. A number of theories have been developed, aimed at improving the interpretation of route choice by analysing and factoring in subjective characteristics of the individual. Some authors have developed discrete choice models further, calibrating models based on revealed preference data (Ramming, 2001) (Nielsen, 2004) (Wolf, et al., 2004) (Bekhor, et al., 2006). Others have incorporated psychological factors into the mathematical structure of the models, using latent variables (McFadden, 2001) (Prato & Bekhor, 2006) (Prato, et al., 2012). Another branch of research are the so-called learning models, based on the assumption that actual route choice is influenced by routes previously travelled. These models include those similar to the FL model (Fudenberg & Levine, 1998) as well as those based on Thorndike's law of effect such as REL (Erev, et al., 1999) (the good results obtained adopting a particular strategy are such that the same choice can be repeated). Another research line concerns experiments based on the prospect theory (Kahneman & Tversky, 1979) (Tversky & Kahneman, 1992), in particular its extension the CPT (cumulative prospect theory). Experimental studies have been conducted aimed at evaluating choice behaviour as a function of a particular reference point, evaluating user risk propensity under conditions of uncertainty (Avineri & Prashker, 2003) (Avineri & Bovy, 2008) (Bogers, et al., 2005) (Xu, et al., 2011). The problem of implementing the CPT is that no in-depth studies have yet been conducted on real cases.

The main problem, however, common to all research addressing the behavioural aspects of route choice, lies in data acquisition, from the data that identify the choice actually made by the individual to the data characterizing his/her behavioural aspects (socio-economic, cognitive aspects, experience, activities engaged in at destination, etc). In fact, in the traditional methodology for constructing route choice and traffic assignment models the actual routes chosen by each individual are not known, but only the aggregated values of flows along routes throughout the network, distributed on the basis of the objective characteristics of the network. This certainly hinders any detailed analysis of the behavioural attributes at the individual level.

To overcome these shortcomings for some years now the attention has been focused on collecting detailed information about the spatial-temporal configuration of routes actually chosen, associating them with the behavioural characteristics of the decision maker/driver travelling those routes between origin and destination. This has been made possible also thanks to the advances in GPS technology. A number of recent studies have employed GPS based data collection, mostly through in-vehicle surveys (Li, et al., 2005) (Zhu & Levinson, 2010). A detailed description of GPS based travel data collection techniques can be found in (Wolf, 2006).

This paper falls under the latter research line. We have conducted an in-depth analysis of a database containing information on actual route choices of a sample of users travelling within an urban area. The data were collected using a smartphone with built-in GPS tracking system able to record, by means of an appropriate interactive application, not only the spatial-temporal evolution of the route travelled during the two-week survey, but also trip start and end times, activities carried out at destination, travel mode, number of passengers in vehicle, etc. Information concerning personal characteristics have also been collected, through a questionnaire administered to each individual monitored.

The aim of this work is therefore to analyse the travel behaviour of this sample of users first determining the extent of variation in choice with respect to the routes simulated with a standard approach using a traffic macrosimulation model implemented in Cube software (Citilabs). This variation is evaluated using the following indicators:

1. Coverage (Ramming, 2001), that represents the percentage of overlap between actual routes and the least cost paths;

2. Time/Distance Deviation Index (Spissu, et al., 2011), that indicates the percent difference between journey times of actual and simulated routes.

Moreover, because of the nature of the data, it was possible to determine the characteristics associated with repetitive routes between the same destination and origin, evaluating the degree of intravariability of the choice for a single user and its characteristic aspects. The remainder of the paper is organized as follows: section 2 provides a brief review of existing studies on route choice using GPS-based data collection systems, Section 3 describes the data and implementation methodology while Section 4 describes the analyses conducted. Lastly Section 5 draws conclusions.

EARLIER APPROACHES

In the past, information on route choice was collected on the basis of question and answer surveys, either by phone, mail or one-to-one interviews, participants providing information on the route travelled (Ben-Akiva, et al., 1984), (Mahmassani, et al., 1993), (Ramming, 2001), (Prato, et al., 2005). As these surveys rely on the interviewees' memory, the data gathered were not always correct and reliable (Murakami & Wagner, 1999), (Barbeau, et al., 2009). Data collection has benefited substantially from the widespread use of information systems and the rapid progress of modern GPS devices, as data can now be recorded automatically in electronic form and above all with greater precision than personal interviews (Murakami & Wagner, 1999), (Nakazato, et al., 2006), (Hato, et al., 2006), (Bricka, et al., 2009) (Barbeau, et al., 2009).

However, the quality of GPS data depends not only on the efficiency of the device and the satellite system but also on the situation in which the survey is conducted. In addition, data collected in the form of coordinates should undergo map-matching (Christopher, et al., 2000) (Zhou & Golledge, 2006) (Quddus, et al., 2007) (Marchal, et al., 2005) (Schuessler & Axhausen, 2009) so as to be able to track complete routes as a series of arcs on the network. Notwithstanding the above limitations, over the last ten years data collected using GPS devices or more advanced integrated systems (PDA; smartphone, etc.) have been utilized to gather travel information of both a spatial and temporal nature. In particular, these data have been used to analyse activity-travel choice behaviour (Asakura & Iryo, 2007), (Stopher & Greaves, 2007), to study the dynamic processes of activity scheduling and re-scheduling in more detail (Zhou & Golledge, 2007) to gain a better understanding of spatial behaviour (Schönfelder & Axhausen, 2004), to quantify private vehicle emissions (Beckx, et al., 2010), to calibrate the route set generation (Fafieanie, 2009) and finally to study the effect of personal cap and trade mileage policies on individual activity travel patterns (Meloni, et al., 2011). GPS technology has also been exploited in an endeavour to improve our understanding of those attributes governing route choice (Jan, et al., 2000), (Li, et al., 2005), (Parkany, et al., 2006), (Papinski, et al., 2009), (Papinski & Scott, 2010), (Zhu & Levinson, 2010), (Spissu, et al., 2011).

(Jan, et al., 2000) aimed to understand how objective route choice data can help in grasping the differences between actual behaviour (collected with in-vehicle GPS) and those determined using traffic assignment models, from both the practical and theoretical viewpoints. For this purpose, they used a sub-sample from a database containing around 3000 trips collected during a one-week survey involving 100 households (216 drivers) in Lexington, Kentucky. The information concerns the network attributes such as distance, speed and capacity, while for each route recorded (denoted by a user ID) times and distances have been drawn from the GIS implementation. The database does not contain any information regarding traffic lights or delays due to turning manoeuvres. They have calculated the following: (a) objective journey times, assuming flow conditions reduced by 15% of free-flow speed; (b) journey times recorded with the GPS; (c) the shortest journey time between the same OD, obtained in the same way as (a). The percentage of arcs shared by two routes with the same OD are then calculated. The authors found that for most pairs shared routes accounted for around 60% or just over. The PDI (path deviation index), has also been calculated as the ratio between the area delimited by the two routes and the length of the shortest path, to determine the distance between two paths which, for very small distances, allow one to take the paths between two pairs with low PDI as being equal. Users usually tend to travel the same route for the same trip, the PDI increasing with the distance between the OD (distance deviation index), while journey times are similar to the shortest paths. In the same paper the authors observe that only limited information could be collected with the GPS and hope that further technological advances in the tools currently available will make it possible to gather different types of data.

(Li, et al., 2005) claim that an understanding of route choice behaviour can contribute to improving assignment algorithms, based on more realistic assumptions. They used the data collected during an in-vehicle GPS survey called “Commute Atlanta”, choosing a sub-sample of 182 drivers from 138 households over a 10-day period for whom the socio-economic characteristics were known. For each route recorded, start and end time, journey time and distance are known. Examining trips made between 05:00 and 10:00 on weekdays, they found that roughly 40% of commuters only uses one route for their trips, while 60% used at least two. Considering as routine using the same route at least twice during the survey period, around 33% travelled routine routes, more than the 15.5% reported by Abdel-Aty et al. (Abdel-Aty, et al., 1994). The authors develop a binary logit model for route choice, taking as dependent variable the use of one or more routes between the same OD. For specifying the model the socio-economic characteristics of both the individual and the household, as well as information on the trips and the characteristics of the most used route, come into play. The results show the strong explanatory power of flexible working hours and of trip chaining on the dependent variable to choose more than one route, compared to the socio-demographic characteristics of the commuters and the relative route attributes. The authors observe that this type of study can gain precision when detailed information on the activities carried out and the trips made over long periods of time are available.

(Parkany, et al., 2006) The aim of their research is to investigate the deviation between route choice and the shortest path, referring to peak and off-peak hours and between revealed and stated data. They use data gathered between March 2002 and July 2003 at Lexington (Kentucky), by means of an in-vehicle GPS for each of the 276 households involved in the survey, monitored over a period of 10 days. The study also envisaged a post GPS collection survey, e-mailing questionnaires to a sub-sample of 106 drivers (so as to obtain good correspondence with the mailback interview). The post-interview gathered data on travel habits, route preferences in peak and off-peak hours, factors relating to attitudes. The deviation is calculated as the ratio between journey time of the chosen route and the shortest path (time deviation index), indicating as “diverters” those commuters for whom the ratio exceeds 10%. The reasons for the deviation lie in the fact that factors other than time come into play in route choice and that commuters may not have sufficient information or the ability to minimise travel time, confirming the theory that more than route is used between the same OD. Models are estimated to evaluate whether the routes indicated in the stated preferences are consistent with the revealed preferences, showing that variables such as habit, attitudes and route characteristics are all associated with route choice.

(Papinski, et al., 2009) examine how commuters schedule their trips and the planning sequence. They used data for a sample of 31 commuters and selected 21 home-to-work routes recorded with GPS, so as to evaluate the deviation with planned trips (as the home-to-work trip is a programmed activity, it is assumed it is planned beforehand). They use a three-step method: pre-interview, data collection and post-interview. They identify 4 types of route planning: main roads first, sequential, mixed or no planning. The authors show that 20% of the participants make diversions with respect to the planned route (for various reasons, from the desire to try a different route to mood change). In spatial terms it was found that short routes lend themselves better to diversions, as commuters are more familiar with the local streets. The outcomes, however, could be criticized because in the study the users are encouraged to think explicitly about their routes, and this could cause bias in the subsequent phase of choice, altering the decision-making process.

(Papinski & Scott, 2010) maintain that to best support route choice analysis tools able to process the data efficiently and with few operations are required. For this purpose they describe the development and application of route choice analysis (RCA) kit that automatically calculates more than 40 variables relating to the route characteristics, making it possible to compare observed with the shortest routes (calculated using an appropriate module integrated with the RCA). They use GPS recorded data for a sample of 237 drivers, who participated in a two-day survey for home-to-work trips, comprising three phases: 1) completion of a paper trip diary (memory jogger), 2) GPS tracking (2 days), 3) post-survey interview. Each commuter was provided with a PDA incorporating the GPS device. When inconsistencies existed between the stated information and the GPS data, or when data were missing, the participant was contacted to complete the information.. The RCA calculates the shortest path with cost functions based on the distance and time: these are then compared using the variables that the kit calculates, with those actually observed. It was found that the actual routes are substantially longer in terms of time and distance than the shortest travel time paths, identifying the greatest differences in directness, characteristics of the longest stretch, number of turns, speed and road type. Considering the shortest path in terms of distance covered, these include roughly 34% of expressways and highways, against 44% of the observed paths, that also have fewer turns. The RCA also calculates the RDI, route directness index, a measure of the degree of route circuity. The shortest paths were found to have an RDI of 1.38 and 1.34 (the decimal points indicate the percent distance greater than the distance in a straight line between the origin and destination examined) against 1.53 of the observed paths. They conclude maintaining that cost functions based simply on time or distance are not able to adequately reproduce the dynamics of actual trips.

(Zhu, et al., 2010) explored the phenomenon that, contrary to the user-equilibrium theory, makes individuals choose more than one route for repetitive trips between the same OD. They used in-vehicle GPS based data collection for a sample of 657 home-to-work trips in the Minneapolis area, Saint Paul, Minnesota (Twin Cities), made by 95 individuals. The recorded data were processed for mapping onto a network implemented in GIS and then analysed. Setting as different those routes that deviated by at least 30% in terms of distance, they found that around 40% of the commuters monitored travelled along more than route. Then, by comparing the average simulated travel times and the standard deviation with those recorded using the GPS, they showed that in most cases their model underestimated the time (likely because of the fact that the model poorly reproduces the delays at junctions that by contrast the GPS records correctly). In conclusion, they maintain that the commuter tends to minimize travel times though remaining within a certain route reliability threshold.

(Spissu, et al., 2011) developed a methodology for analysing GPS data for actually observed routes, in order to evaluate the degree of individuals' intra/intervariability and to compare actual routes chosen with the least cost paths. They monitored a sample of 26 students from the University of Cagliari, over a two-week period using a personal probe system called Activity Locator (Meloni, et al., 2011). This device not only tracked the routes travelled but also recorded the different activities the participants engaged in on the survey days. They found that for a specific OD pair, in 92% of the cases the same person used just one route (intravariability); Extending the analysis to all the students, (for trips between the same OD pair) they found that in around 72% of cases one route was used, 24% two routes and for the remainder three routes or more (intervariability). Calculating the overlapping index, for the percentage of arcs common to both the actual and simulated routes, they found that roughly 70% of the logged data had values of between 0 and 0.6, while less than 30% exhibited higher values and hence more similar characteristics to the respective least cost path. They then evaluated the time deviation index and the distance deviation index (percent variations of travel time and distance of the observed versus the simulated route) and found that the participants generally tended to minimise the distance for short trips (around 4 km), while time was considered more important for longer trips (more than 7 km).

In almost all the examples described, the GPS data were collected through in-vehicle surveys that make it possible to gather objective information on trips (travel times and distances). Pre- and post-analysis interviews were conducted to gather information about the subjective characteristics of the individuals. GIS platforms were used to study the routes. In the present study, which is a follow up to the survey conducted by (Spissu, et al., 2011), the data were collected using an integrated system able to also record the activities conducted, along with all the characteristics associated thereto. In this way a complete database was created containing all the information (objective and not) concerning the trips.
For comparisons with the objectively most convenient paths, then, was used a static macrosimulation model (implemented in CUBE, Citilabs Ltd.) of the entire study area, which reproduces the network characteristics actually encountered by the drivers referring to the data used. It seems that, in the works illustrated above, for the calculation of the minimum cost paths they have been used exclusively GIS techniques, without a calibrated reference simulation model, which necessarily affects the quality of the data used for the comparison.

DATA DESCRIPTION

In April 2011, 109 participants from Cagliari (Italy) metropolitan area were recruited for a two-week program to be conducted in seven separate waves from May 2011 to April 2012. Each participant was asked to carry a smart phone with built in Gps in which an application called “Activity Locator” – implemented by CIREM – was pre-installed. This device permitted users to send activity and travel attributes to a dedicated server in real time, together with automatic positioning points (timing, latitude, longitude). At the end of the survey all the information gathered was reworked into a single activity-travel database.
The “Activity Locator” system comprises (a) a client software installed in a portable Gps-integrated device, (b) a server software that transmits and receives information to/from each client, and (c) an Internet connection. The client software is a Java application that can be installed in any smart phone (Symbian or Android platform) with built-in GPS currently available on the market. The application enables to track individual daily routes and the

collection of all activity-travel related information through a sequence of pull-down menus that reproduces the classical activity diaries, but in real time. The server software collects the information sent by each client; each user can be identified in real time on a map (powered by Google Maps) by a symbol containing all the user information (i.e. spatial, temporal, and activity information). The data are immediately available in database formats (i.e. xls, csv, xml) and downloadable onto any desktop or laptop computer. The data are transferred by each client to the server and vice versa via an Internet connection. The activities recorded with the Activity Locators provided to the participants included more than thirty disaggregated activity in- and out- of-home types (i.e. personal care, household care, meals, work, sport, recreational, *etc*.), as well as company involved (i.e. alone, with family members only, with family and others, *etc*.). Also a number of travel attributes were included related to trip mode, vehicle occupancy, and ticket paid. In particular, considering travel modes, trips can be classified into: (a) Car as driver, (b) Car as passenger, (c) Motorcycle, (d) Walking, (e) Bus, (f) Train, (g) Light Railway. In this paper, only the Car as driver mode has been analysed.

During the survey weeks a total of 8831 trips were recorded by 109 individuals. Each route was mapped onto the network using a map matching technique (Corona, et al., 2012), so as to obtain a base that could be edited and analysed using commercial GIS software. As we envisaged analysing the observed routes using traffic macrosimulation software, the routes were mapped as a series of nodes, coordinates and numbering corresponding with the model network graph. A subset of routes was drawn from the database for trips made between 07:00 and 10:00 in the morning for car as driver mode only making a total of 626 routes, so as to use data consistent with the peak hour for which the model is calibrated.

METHODOLOGY

As mentioned, this work resumes the research conducted by (Spissu, et al., 2011). They carried out a pilot study using the same devices on a sample of 26 users compared to the 109 examined here that differ not only in quantity but also in composition. A group of students from the University of Cagliari was selected for the first survey, so the data did not have the heterogeneous features of the data used in the present work.

The objective of the presented paper is to compare the paths recorded using the GPS system with those mathematically considered as being of minimum cost. For this purpose, it was necessary to define an operating methodology that, starting with the data collected using the Activity Locator, generated two comparable databases (one for the real paths and one for the simulated ones). Once this stage was completed, it was possible to proceed with the comparative analysis. The entire process can be divided into four main sections, described below:

1. Calibration of the simulation model;
2. Creation of the real paths database;
3. Simulated paths construction;
4. Comparison and analysis of the results.

Model of the Cagliari metropolitan area

As mentioned above, the analyses were performed with the support of a specialized traffic simulation software. CRiMM (Centre for Research Models and Mobility) has implemented in CUBE (Citilabs), a suite of specialized programs for transport planning, the macrosimulation model of the Cagliari metropolitan area .
The network graph consists of 606 zones, with a total of 12594 nodes and 22552 links, each characterized by attributes such as distance, free flow speed, free flow travel time, capacity, functional classification. Many intersections have also been modelled, including types of traffic lights and geometric characteristics. For the current application we used morning peak hour matrices, with a total of 92570 assigned trips (equivalent vehicles). Figure 1 shows a schematic representation of the model calibration and validation processes.

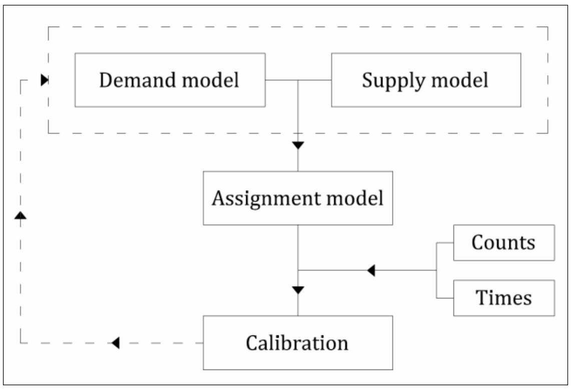

Figure 1 - Calibration process

The model has been successfully calibrated and validated, obtaining the following indices: R2 = 0.924, RMSE = 22.9%, MAPE = 17.6%. The route choice model is based on the classic method of minimizing path cost, which is calculated using Dijkstra's algorithm (1959). The implemented function is a linear combination of the time and distance attributes, with the following specification:
Ck = a + b * time * distance
where alpha is equal to 1, time is equal to the travel time calculated on the basis of link flow speed (like BPR), the distance is the link length and beta is a parameter calculated as β = $(C_e*P_c)*60/V_t$
where, C_e is the operating cost (0.195 €/km), P_c is the perception of operating cost (0.4) and V_t is the value of time (15 €/h). Therefore β is equal to 0.312.

Observed routes database construction

The construction of the observed paths database was particularly important and challenging as, in order to be implemented in the CUBE network, the data needed to be encoded in a particular manner. Route information collected using the Activity Locator is recorded in a server in a point shape file form, one for each user, in which each path is identified by a user ID and an ordered sequence of GPS points. For each of these, information about the start and end times, the spatial coordinates and other attributes related to the activities associated thereto, are known. However, this information cannot be used directly in CUBE because the software requires the path to be encoded by a two-dimensional vector, the first corresponding to the A node and the second to the B node (the AB sequence also determines link direction). This vector is composed of as many rows as there are links in the path. So, because it is necessary to have a database containing information about the links, it was decided to use map matching techniques to associate each GPS point to a network link, using the CUBE graph as the basis in order to ensure consistency with the node numbering. The result is that each "mapped" path consists of a consecutive nodes vector, identified by a number that corresponds to the "physical" node in the graph. By subsequent processing it was then possible to convert this vector into an AB sequence, which enables the identification of the path links order. Once the link sequence has been constructed, it is possible to associate to each one the corresponding characteristics, including functional classification, length, free flow speed, free flow travel time and capacity. In the present study, only the time and distance characteristics have been associated to each path, postponing the allocation of qualitative characteristics to a later stage. However, for correct implementation in the simulation model, each mapped path had to be manually checked and reconstructed where necessary. Because the network graph is particularly dense, we found that during map matching a GPS point could often be associated to a link adjacent to the one it should actually belong to, obtaining a "visually" acceptable result but that cannot however be used for implementation in the simulation model. An example may clarify this concept: imagine a route having origin node 1 and destination 2, formed by arcs 1-3, 3-7, 7-4, 4-2, and the opposite path with origin 2 and destination 1, formed by arcs 2-5, 5-8, 8-6, 6-1.

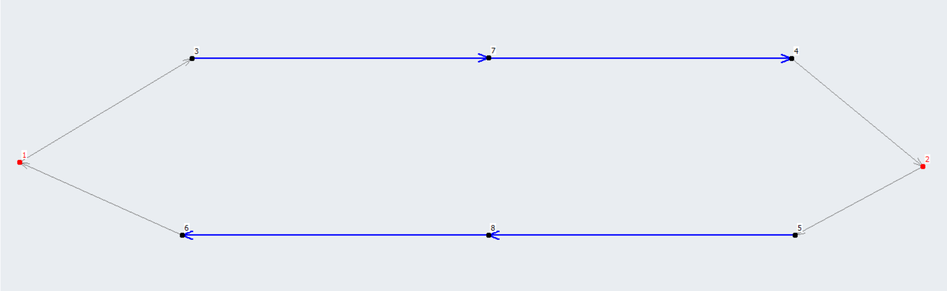

Figure 2 - Cube routes example

During map-matching an incorrect link association could occur, such that the path 1→2 comprises the sequence of nodes (1,6,8,5,2). This means that, when constructing the 1-2 comparison path, the model can never reproduce what is mapped (that represents the route actually chosen by the user), because the reference nodes used to delineate the path will not be considered, as they belong to a path running in the opposite direction, thus producing completely erroneous results. The whole process can be described schematically as follows:

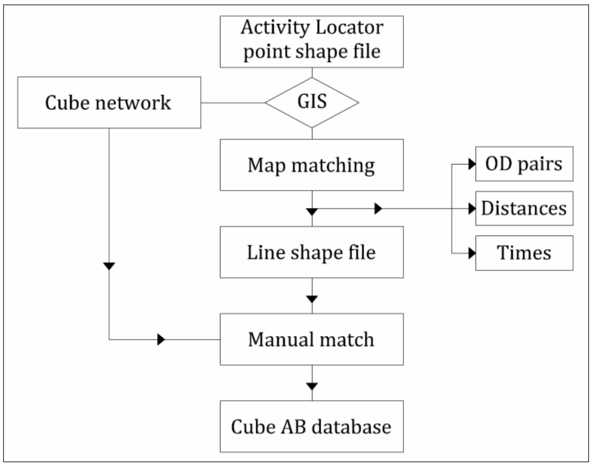

Figure 3 - Observed routes database creation

Simulated paths database construction

Once the observed database had been created, the next step was to programme in CUBE an application for extracting from each route, the corresponding least cost path in terms of travel time and distance as well as that calculated by the model for each iteration performed during assignment. The reason for this was that had we focused solely on the equilibrium route this would have caused an error as all the routes calculated by the model need to be taken into account for loading the traffic volumes into the network.
In the case at hand, the model reached convergence after 70 iterations. Thus for each OD pair it was possible to extract 72 simulated path of which 1 had the shortest distance, 1 the shortest travel time and 70 the least cost (whose function has been described above). Many of these are repeated, as it often happens that the model considers the same route as the least cost path more than once.

Data analysis

For analysing the data a routine was programmed in CUBE that compares simulated and observed routes in spatial terms. It operates as follows: once the actual route has been input (reconstructed as described in the previous paragraph) as well as the origin and destination nodes, the routine identifies, from an internal database created beforehand, all the paths the model has calculated between that specific OD pair At this juncture a database is generated composed of 70 lines, each referring to a single path to which the cost, time and distance attributes have been associated.. Two other modules calculate “directly” (without the need for a preexisting database) the shortest paths in terms of travel time and distance. Once all the necessary information has been extracted, the last step consists in comparing in spatial terms the simulated and observed paths, the final result being the length in km at which the observed and simulated paths match perfectly, a necessary prerequisite for calculating coverage (Ramming, 2001)).

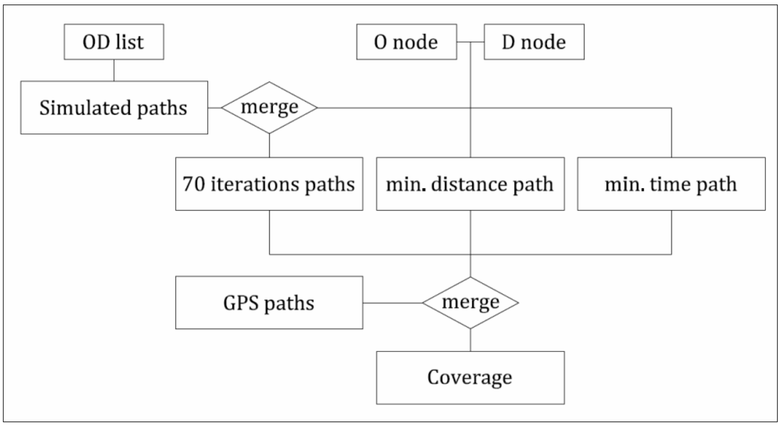

Figure 4 - Data processing

Using the Activity Locator it was possible to collect data for all the trips made by the participants monitored during the two weeks of the survey. For the purposes of this study, we selected from among a total of 8831 trips, only those made on weekdays at between 7:00 and 10:00 in the morning for the car as driver mode. This was necessary because as described in Section 4.1 the model has been calibrated for a matrix for trips made during morning peak hours.. For this reason the database should contain information on traffic conditions identical to those used in the simulation which clearly influence calculation of the least cost path. Not all the 109 drivers monitored had travelled in car as driver mode in the time interval chosen. Thus 93 drivers were selected with the necessary requirements. The total number of trips for this sub-sample is 626, with average travel time of 18’42’’, average distance of 7.41 km, and average speed of 23.8 km/h. For each trip we identified the type of activity that the driver engaged in at destination. In this way it was possible to distinguish between home-to-work trips and trips for discretionary activities. Table 1 shows the percent share of the activities associated with the 626 trips selected:

Table 1 - Activity percentages

Activity	Percentage
Work/study	73.56%
Shopping/errands	9.13%
Pick up/drop-off passengers	7.37%
Recreational activities /sport	2.72%
Meals/meal preparation	2.24%
Coffee break	1.76%
Household chores /looking after family	1.12%
Personal care	0.48%
Waiting for public transport/lift	0.48%
Voluntary work	0.32%
Going for a walk	0.32%
Leisure (tv, books, internet)	0.16%
Doctor	0.16%
Sleeping/relaxing	0.16%

As mentioned above, with the Activity Locator it is also possible to determine the distance of the trips recorded by interpolating the GPS points. Note that, though the actual routes were reconstructed manually in the simulation model, the two average distance values only differ by 6.1%, demonstrating the soundness of the analysis techniques adopted. In fact, comparing the average distance of the paths obtained directly from map-matching techniques (thus without the post-reconstruction described above) the difference with the average distance recorded using the GPS is 16.3%, a further indication of the difficulties encountered in data processing.

Coverage and spatial-temporal deviation

From the first comparison between actual and simulated paths it was possible to understand whether the drivers chose routes mathematically considered as least cost paths, as a function simply of travel time and distance attributes. Thus, where necessary, the coverage was calculated for each path with the corresponding least cost determined by the model, using the following formula:

$$COVE_k^i = \frac{l_{sim}^i}{l_k}$$

where *i* denotes the type of simulated path (least cost, shortest distance, shortest travel time) and k identifies the trip, l_{sim}^i is the length of shared path along the two routes, while l_k is the total path length recorded with the GPS.

Because of the nature of the data, the same route may have been used several times by the driver during the survey period. Therefore a further selection was performed so as to consider, for comparative purposes, only routes that differed from one another. For example, if a driver travels the same route for 10 trips then this is considered only once. Thus the sample was reduced to 298 different routes (moving from a trips database to a routes). The graph of Figure 5 shows that 49% of the routes have a coverage greater than or equal to 90%, while the remainder take lower values. This further confirms the findings reported in the literature, i.e. that drivers do not necessarily use the least cost route for their trips.

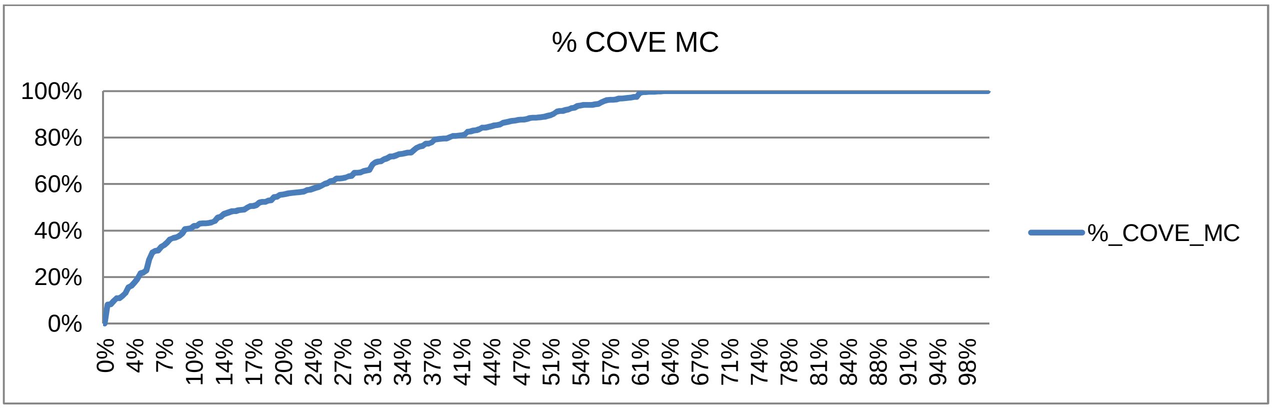

Figure 5 - Entire sample coverage

With the recorded travel times and distances it was also possible to calculate indexes that provide information about the percent difference between the actual distances and travel times and the corresponding least cost path. For this purpose we used the formulae suggested by (Spissu, et al., 2011) for the Time Deviation Index and Distance Deviation Index.

The *Distance Deviation Index* (DDI) is calculated as:

$$DDI = \frac{d_{obs} - d_{sim}}{d_{sim}}$$

where d_{obs} is the distance of the observed route, and d_{sim} is the distance of simulated route. On the same way, the *Time Deviation Index* (TDI) is calculated as following:

$$TDI = \frac{t_{obs} - t_{sim}}{t_{sim}}$$

where, t_{obs} is the travel time of the observed route, and t_{sim} is the travel time of simulated route.

The graph of Figure 6 shows the indexes calculated for the routes travelled to participate in work/study and discretionary activities.

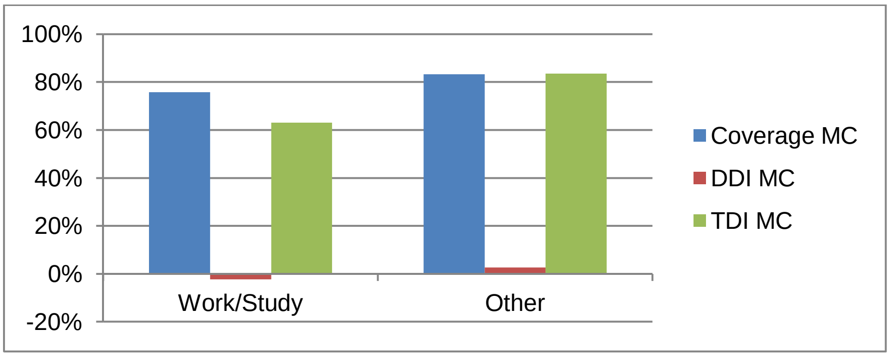

Figure 6 - Coverage, DDI and TDI divided by purpose

As can be seen the distance deviations are very small, indicating that the drivers have good spatial perception of their trips and the transportation network. Even if the average coverage suggests that the paths are not necessarily the least cost ones (and hence, mathematically, the most convenient), the decision to take a different route does not result in excessive differences in terms of length. Interestingly, careful examination of the time deviation index (TDI) shows that, while the average coverage for the two types of routes is almost comparable (with deviations of a few percentage points), the differences in terms of TDI are greater.

There is a simple explanation for this phenomenon. The figure shows that, in both cases, users choose the least cost paths for about 79% of the trips. The lower values of TDI for the home-to-work trips indicate that, as these trips are typically systematic, often repetitive and follow a route that users are familiar with, when deciding which route to take, they tend to minimize not only the distance but also the time, contrary to what happens for discretionary activities trips. Figure 7 shows the relationship between coverage, TDI and DDI versus trip length. The DDI index remains more or less steady around 3%. Looking at the trend of TDI, we can observe that it takes particularly high values for trips shorter than 10 km, while for trips of over 10 km the deviations do not exceed 10%.

This phenomenon can be explained by the fact that, for short routes, differences of even a few minutes lead to deviations greater than in the case of longer trips, in which the differences between the travel times are compensated by total travel time (for example, a 5-minute difference in a path of 10 minutes produces TDI=50%, while in a path of 25 minutes it is TDI = 20%).

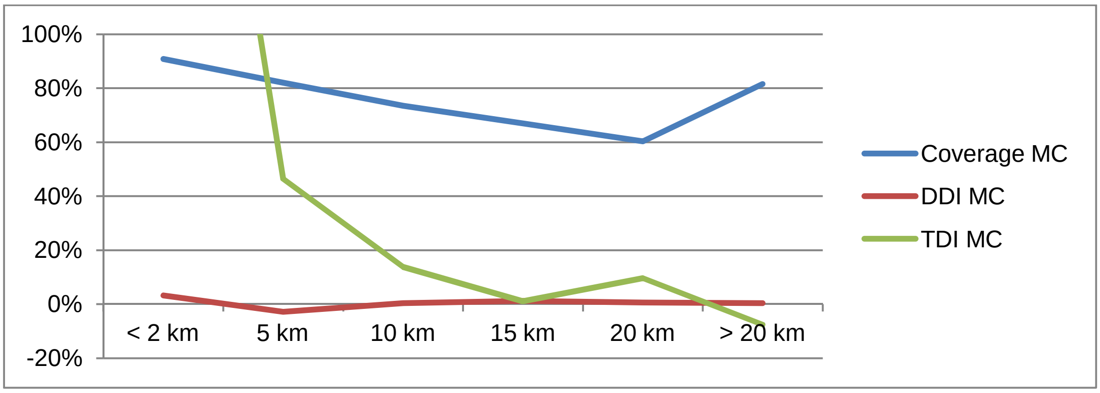

Figure 7 - Trip lenght analysis

Interestingly, if instead of considering simply a single trip between an OD pair we take the whole sample into account, thus including repetitions of the same path, the indicators follow the same trend described above, except for the TDI of the work/study trips, which takes a value of 20%. This further substantiates what has already emerged, i.e. routine trips and familiarity leads drivers to make route choice decisions that place more importance on the time and distance attributes than on the other factors.

Intravariability

We then turned out attention to repetitive trips made by the same driver between the same OD, thus allowing to assess the degree of intravariability of the routes chosen . Out of a total of 626 records, 381, approximately 60.9%, had the required characteristics. Figure 8 shows the percentage of the sample that repeated the path between the same OD for the number of times indicated on the X axis:

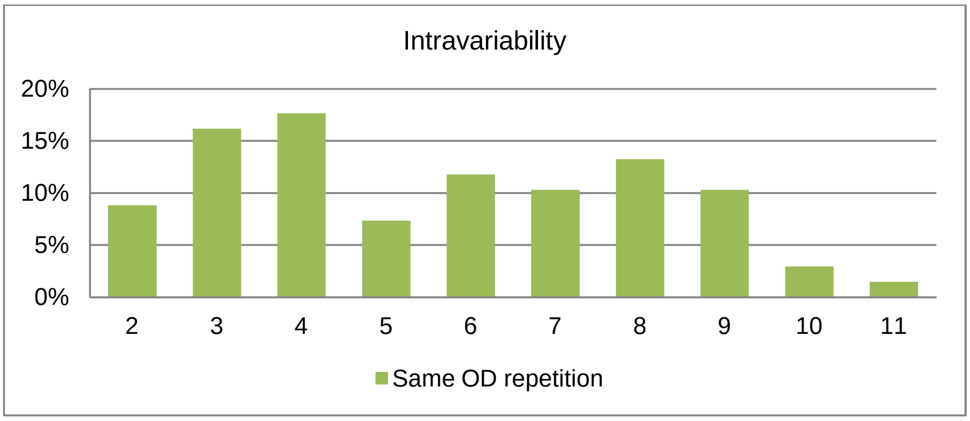

Figure 8 - Same OD repetition

As can be seen 69.1% of the drivers took the same route, 22.1% took at least two while the remaining 8.8% chose three or more routes. During the two survey weeks, the single route drivers made on average 5.28 trips between the same OD, against 6.29 for drivers choosing more than one route. Focusing on this portion of the sample, drivers choose a prevalent route for 68% of the trips.

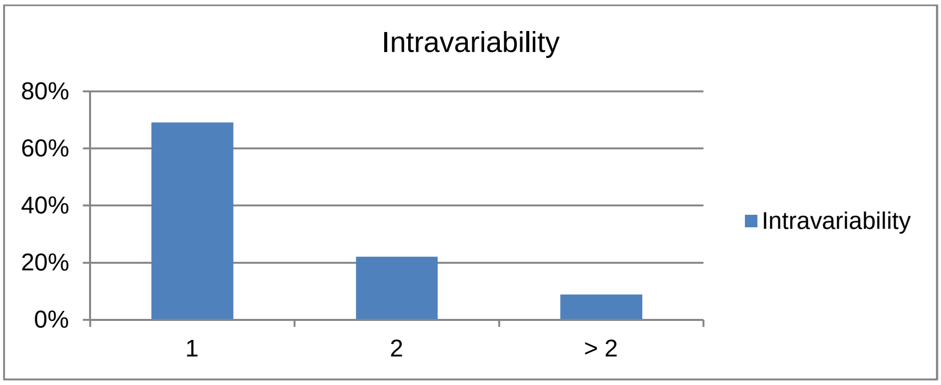

Figure 9 – Intravariability

It is therefore useful to examine how and to what extent the most used route differs from the others, using as the basis of comparison the indices calculated for the analyses conducted for the least cost path. The results show that, where the most used route (49 trips) has greater coverage, in 98% of cases the departure time is later than the second best, on average by around 13 minutes (8:29 against 8:16). Leaving later, the network is less congested. Between 8:00 and 8:30 traffic is heavier due to students travelling to school and to the increased frequency of buses, most of which share the same lane as private vehicles, holding up the traffic. In the 20 cases for which the most frequently chosen route deviates more than the least cost path, departure times do not exhibit a regular pattern. Practically the same characteristic is never repeated which, as in the previous case, could be represented by the fact that the trips having greater coverage always start at a later time. In the future it will be possible to calculate:

1. Path Deviation Index: represents the area delimited by the actual path and the least cost path;
2. Number of turns (and thus the directness index);
3. Percentage of “main” arteries;
4. Number of signalized intersections along the route.

CONCLUSIONS

In this work we describe a methodology for analysing GPS-collected data, aimed at identifying the differences between actual route choice and the least cost path calculated with a static macrosimulation model. The results show that, as a general rule, the model reproduces around 78% of the routes actually chosen by the driver, thus in keeping with the findings reported in the literature (Ortuzar & Willumsen, 2001), which indicate that cost functions based on time and distance are able to explain between 60%-80% of routes observed in practice. We found that distance is the attribute most likely to be minimized, indicating good spatial perception, while the same cannot be said of journey times. For habitual home-to-work travel, where the driver is familiar with the route and traffic conditions, route choice most closely resembles the so-called rational choice, the deviation indexes having lower values than trips made for discretionary activities. It was also observed that when more than one route is taken for repetitive trips between the same OD, the most frequently used (with characteristics most closely resembling the least cost path calculated with the simulation model) shows a later departure time than the second best routes, thus in less congested traffic conditions. It is interesting to compare the results of the pilot study conducted previously (Spissu, et al., 2011) with the present findings. In the pilot study, for repetitive trips between the same OD 92.7% followed one route, the remainder two or more. The difference of around 23% compared to the present study can be attributed to the fact that we only considered trips made during morning peak hours for a heterogeneous sample of drivers, while the pilot study covered the whole day. However, other studies addressing this topic also report different percentages; for example (Jan, et al., 2000) found 34% to take the same route or (Li, et al., 2005) 40% to take the same route, against 60% at least two. (Zhu & Levinson, 2010) maintain on the other hand that around 40% of the drivers they monitored use more than one route, a difference of around 10% compared to our results (31%), though this difference is to be attributed solely to the characteristics of the sample. Further research will extend the analysis to all the routes covered by car as driver recorded using the Activity Locator, so as to also account for the differences based on reference period, trip purpose, etc.

Future developments, in particular extrapolation of additional spatial attributes of the observed routes will aim to estimate and test route choice models incorporating the individuals' behavioural factors to be gleaned from the available data. It would be interesting to test specifications that include spatial-temporal restrictions, along with variables directly related to daily activity scheduling that could contribute to a better understanding of user attitudes towards travel.

REFERENCES

Asakura, Y. & Iryo, T., 2007. Analysis of tourist behaviour based on the tracking data collected using a mobile communication instrument. *Transportation Research Part A: Policy and Practice,* p. 684–690.

Avineri, E. & Bovy, P. H., 2008. Identification of Parameters for Prospect Theory Model for Travel Choice Analysis. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 2082, pp. 141-147.

Avineri, E. & Prashker, J. N., 2003. Sensitivity to Uncertainty: The Need for a Paradigm Shift. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1854, pp. 90-98.

Bamberg, S., Ajzen, I. & Schmidt, P., 2003. Choice of Travel Mode in the Theory of Planned Behavior: The Roles of Past Behavior, Habit, and Reasoned Action. *Basic and Applied Social Psychology,* 25(3), pp. 175-187.

Barbeau, S. et al., 2009. TRAC-IT – A Software Architecture Supporting Simultaneous Travel Behavior Data Collection & Real-time Location-Based Services for GPS-Enabled Mobile Phones. *Paper presented at the 88th Annual Meeting of the Transportation Research ,* January.

Beckx, C., Panis, L., Janssens, D. & Wets, G., 2010. Applying activity-traveldata for the assessment of vehicle exhaust emissions: Application of a GPS-enhanced data collection tool. *Transportation Research Part D: Transport and Environment,* 15(2), p. 117–122.

Bekhor, S., Ben-Akiva, M. & Ramming, S., 2006. Evaluation of choice set generation algorithms for route choice models. *Annals of operation research,* 144(1), pp. 235-247.

Ben-Akiva, M., Bergman, M., Daly, A. & Ramaswamy, R., 1984. Modeling inter urban route choice behaviour. *Proceedings of the 9th International Symposium on Transportation and Traffic Theory, VNU Science Press,* pp. 299-330.

Bogers, E. A., Viti, F. & Hoogendoorn, S. P., 2005. Joint Modeling of Advanced Travel Information Service, Habit, and Learning Impacts on Route Choice by Laboratory Simulator Experiments. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1926, pp. 189-197.

Bovy, P. & Stern, E., 1990. Route Choice: Wayfinding in Transport Networks.

Bricka, S., Zmud, J., Wolf, J. & Freedman, J., 2009. Household Travel Surveys with GPS: An Experiment. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 2105, pp. 51-56.

Christopher, W. E., Bernstein, D. & Kornhauser, A. L., 2000. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies,* 8(1-6), pp. 91-108.

Corona, C., Sanjust, B., Simbula, M. & Trogu, D., 2012. On Implementing a two-step map matching process of Gps-based data on a GIS platform. *Paper presented at the 7th International Conference on Informatics and Urban and Regional Planning.*

Erev, I., Bereby-Meyer, Y. & Roth, A., 1999. The Effect of Adding a Constant to all Payoffs: Experimental Investigation and Implications for Reinforcement Learning Models. *Journal of Economic Behavior and Organization,* Volume 39, pp. 111-128.

Fafieanie, M., 2009. Calibrating route set generation by map matching GPS data. *Master Thesis, University of Twente, Enschede - The Nederlands.*

Fudenberg, D. & Levine, D., 1998. The Theory of Learning in Games. *MIT press.*

Hato, E., Shinji, I. & Mitani, T., 2006. Development of MoALs (Mobile Activity Loggers supported by gps-phones) for travel behavior analysis. *Paper presented at the 85th Annual Meeting of the Transportation Research .*

Horowitz, J. L., 1984. The stability of stochastic equilibrium in a two-link transportation network. *Transportation Research Part B: Methodological,* 18(1), pp. 13-28.

Jan, O., Horowitz, A. & Peng, Z.-R., 2000. Using GPS data to understand variations in path choice. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1725, pp. 37-44.

Kahneman, D. & Tversky, A., 1979. Prospect Theory: An Analysis of Decisions under Risk. *Econometrica,* 47(2), pp. 263-291.

Katsikopoulos, K., Duse Anthony, Y., Fisher, D. & Duffy, S., 2002. Risk attitude reversals in drivers' route choice when range of travel time information is provided. *Human factors,* 3(44).

Li, H., Guensler, R. & Ogle, J., 2005. An Analysis of Morning Commute Route Choice Patterns Using GPS Based Vehicle. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1926, pp. 162-170.

Mahmassani, H., Joseph, T. & Jou, R., 1993. Survey approach for study of urban commuter choice dynamics. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1412, pp. 80-89.

Marchal, F., Hackney, J. & Axhausen, K. W., 2005. Efficient map-matching of large GPS data sets - Tests on a speed monitoring experiment in Zurich. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1935, pp. 93-100.

McFadden, D., 2001. Economic choices. *The American Economic Review,* 91(3), pp. 351-378.

Meloni, I., Spissu, E. & Bhat, C. R., 2011. The effect of personal cap and trade mileage policies on individual activity travel patterns: the Activity locator project. *Transportation Letters: The International Journal of Transportation Research,* 3(4), pp. 293-307.

Murakami, E. & Wagner, D., 1999. Can using Global Positioning System (GPS) improve trip reporting?. *Transportation Research Part C: Emerging Technologies,* 7(2-3), pp. 149-165.

Nakazato, M. et al., 2006. Internet GIS-Based Activity-Travel Simulator for Investigating Alternative Activity-Travel Patterns. *11th International Conference on Travel Behaviour Research Kyoto.*

Nielsen, O., 2004. Behavioural responses to pricing schemes: description of the Danish AKTA experiment. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* 4(8), pp. 233-251.

Ortuzar, J. & Willumsen, L. G., 2001. *Modeling Transport.* s.l.:John wiley & sons.

Papinski, D. & Scott, D. M., 2010. A GIS-based toolokit for route choice analysis. *Journal of Transportation Geography.*

Papinski, D., Scott, D. M. & Doherty, S. T., 2009. Exploring the route choice decision-making process: a comparison. *Transportation Research Part F,* 12(4), pp. 347-358.

Parkany, E., Du, J., Aultman-Hall, L. & Gallagher, R., 2006. Modeling Stated and Revealed Route Choice: Consideration of Consistency, Diversion, and Attitudinal Variables. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1985, pp. 29-39.

Prato, C. G. & Bekhor, S., 2006. Applying branch and bound technique to route choice set generation. *Transportation research record: Journal of the Transportation Research Board,* Issue 1985, pp. 19-28.

Prato, C. G., Bekhor, S. & Pronello, C., 2005. Methodology for Exploratory Analysis of Latent Factors Influencing Drivers' Behavior. *Transportation Research Record,* Issue 1926, pp. 115-125.

Prato, C. G., Bekhor, S. & Pronello, C., 2012. Latent variables and route choice behavior. *Transportation,* 39(2), pp. 299-319.

Quddus, M. A., Ochieng, W. Y. & Noland, R. B., 2007. Current map-matchingalgorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies,* 15(5), pp. 312-328.

Ramming, M. S., 2001. *Network Knowledge and Route Choice.* s.l.:Massachusetts Institute of Technology.

Schönfelder, S. & Axhausen, K. W., 2004. Structure and innovation of human activity spaces. *Arbeitsbericht Verkehrs- und Raumplanung 258,Zurich: IVT, ETH.*

Schuessler, N. & Axhausen, K. W., 2009. Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT). *Arbeitsberichte Verkehrs- und Raumplanung, 568, IVT, ETH Zürich,* Issue 568.

Simon, H., 1955. A behavioral model of rational choice. *The quarterly journal of economics,* 69(1), pp. 99-118.

Spissu, E., Meloni, I. & Sanjust, B., 2011. Behavioral Analysis of Choice of Daily Route with Data from Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 2230, pp. 96-103.

Stern, E., Holm, E. & Maarseven, M., 1993. Information and Commuters Behaviour: A comparative analysis. *Europe on the Move*, pp. 350-369.

Stopher, P. R. & Greaves, S. P., 2007. Household travel surveys: Where are we going?. *Transportation Research Part A 41,* p. 367–381.

Tversky, A. & Kahneman, D., 1992. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty,* Volume 9, pp. 195-230.

Wolf, J., 2006. Applications of new technologies in travel surveys. *Travel Survey Methods—Standards and Future Directions,* pp. 531-544.

Wolf, J. et al., 2004. Eighty weeks of global positioning system traces: approaches to enriching trip information. *Transportation Research Record: Journal of the Transportation Research Board,* Issue 1870, pp. 46-54.

Xu, H., Zhou, J. & Xu, W., 2011. A decision-making rule for modeling travelers' route choice behavior based on cumulative prospect theory. *Transportation Research Part C,* Volume 19, pp. 218-228.

Zhou, J. (. & Golledge, R., 2007. Real-time tracking of activity scheduling/schedule execution within a unified data collection framework. *Transportation Research Part A 41,* p. 444–463.

Zhou, J. & Golledge, R., 2006. A Three-step General Map Matching Method in the GIS Environment: Travel/Transportation Study Perspective. *International Journal of Geographical Information System.*

Zhu, S. & Levinson, D., 2010. *Do people use the shortest path? An empirical test of Wardrop's first principle.* Minneapolis, s.n.