EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

Using modern ICT for semi-automatic data collection of activity-travel patterns in the spatio-temporal context of ubiquitous urban environments: Experiences with tracing and data imputation

Using Modern ICT for Semi-Automatic Data Collection of Activity-Travel Patterns in Spatio-Temporal Context of Ubiquitous Urban Environments

Experiences with Tracing and Data Imputation

Anastasia Moiseeva, Joran Jessurun and Harry Timmermans
Eindhoven University of Technology
Faculty of Architecture, Urban Planning Group
5600 MB Eindhoven
The Netherlands
a.moiseeva@tue.nl

Keywords: GPS, activity-travel patterns, activity-based approach, imputation.

Abstract: The activity-based approach views the city as a collection of individual activities, actions, reaction and interactions in the spatiotemporal context of a large-scale environment. For understanding the choice behaviour of individuals and describing urban place in terms what is going on data on daily activity patterns of individuals should be collected. Advanced tracking technologies could facilitate and improve empirical research on human special behaviour. They provide researchers a far more accurate database than traditional means of data collection of a much higher degree of temporal and spatial resolution. Thus the purpose of this paper is twofold: first it aims to report an overview and experiences on data collection with advanced tracking technologies; secondly, it presents the theoretical background of the activity based approach and illustrates how resulting data obtained with GPS technologies can be used to represent the human activities in a spatiotemporal context according to the concepts of the activity-based approach. In addition, this paper presents a system TraceAnnotator for imputing different facets of activity-travel patterns from GPS tracers by using Bayesian belief networks. Activities and movement/travel patterns are represented using a GIS environment at the aggregated and individual levels.

1. INTRODUCTION: APPLICATION OF MODERN ICT IN BEHAVIOURAL RESEARCH

The rapid growth of information and communication technology (ICT) invading our daily lives in multiple ways. Mobile commuting has become possible by the creation of ubiquitous urban environments. It gives people great flexibility in staying in contact with their social network, work outside the office, check emails and internet whilst travelling, etc.

The other side of this coin is that modern ICT devices also allow others to trace people. In that sense, GPS devices, radio-frequency identification (RFID) tags and cellular phones offer new opportunities for data collection about how people move around in cities, either at the aggregate level or as individuals.

Asakura et al. (1999), and Asakura and Hato (2001) investigated the application of cellular phones to monitoring individual travel behaviour. Asakura and Iryo (2007) studied tourists' behaviour using tracking data collected with a mobile instrument. For the MIT project Real Time Rome (Real Time Rome, 2006) data were obtained from cell phones for understanding urban dynamics in real time: traces of information and communication networks, movement patterns of people and transportation systems, spatial and social usage of streets and neighbourhoods (http://senseable.mit.edu/realtimerome/ accessed 02 June 2010).

RFID tags find their application typically in large buildings of public spaces such as stations and shopping malls, such for example collecting data on pedestrian movement. The experiment on the application of RFID tags to travel data collection was conducted in a department store in Fukuoka as well as in a shopping street in Matsuyama (Hato and Kitamura, 2006).

Interest in the application of GPS systems started in the late 1990's, when the first pilot studies were conducted (e.g. Murakami and Wagner, 1999; Zito, 1995). Murakami and Wagner (1999) studied the results of a GPS based experiment for 100 households in the USA. Zito et al. (1995) applied GPS to travel time survey by floating vehicles. Early studies suggested that GPS-based data collection resulted in more accurate spatial and temporal data on travel behaviour than traditional data collection methods. In general, GPS was better able to detect trips of short distance that respondents often fail to report (e.g., Battelle, 1997; Hato and Asukari, 2001). These new opportunities have generated a rapidly increasing number first of pilot projects and now more serious projects on the use of these devices for tracing people in a variety of disciplines.

While the application of the tracking technologies has developed substantially in transportation science (e.g. Wolf, et al., 2003; Bohte and Maat, 2008) during the last decades, the use of GPS for tracking individuals has been recently explored in such disciplines and domains as urbanism (van der Spek, 2008) and spatial planning (Hovgesen et al., 2008). In the field of spatial planning and urban research, two different fields of research using GPS can be distinguished, namely travel surveys and activity patterns. The first field is related to the travel diary research – research in the field of travel choice behaviour, mostly on a regional or metropolitan scale. The second field covers the analysis of activity patterns on different scales. In this case, the relationship between activity and space-time is crucial.

Today it has become essential that for behavioural research, activity-based analysis and modelling needs we are able to impute from GPS tracers data on which activity is conducted (activity type/trip purpose), where (destination/location), when (start and end time) and for how long (duration), the transport mode(s) used, and possibly with whom the activity was conducted and the route taken (Timmermans et al, 2009). GPS traces do not provide in any straightforward manner information about trips and activities, hence they provide information about the distance covered, duration of travel, speed, acceleration and location coordinates. Nevertheless, the number of studies that include the registration of transportation mode, destination locations and destination types is growing in this field and they propose different methods of registration and identification (Wolf et al 2001; Wolf et al, 2003; Bohte and Maat, 2008; Clifford et al., 2008). The principles to interpret the GPS traces have been rather ad hoc. Moreover most of these studies have been limited in scope in the sense that data collection has been refined to a small number of days.

To explore the feasibility of using GPS devices, jointly with advanced data imputation models, we developed a system called Trace Annotator and conducted some small scale experiments to assess the feasibility of this approach. This paper will summarize our previous research efforts in this context. First, section 2 provides a brief discussion of the nature of human activities in time and space within the context of individual action space, activity spaces and time geography approach. In Section 3 we will outline the architecture of the system. This is followed by the data imputation used. In Section 4 we will discuss some of our experiences with the application of this technology. Finally, the last section draws some conclusions and concluding remarks are offered concerning interaction of individual activities in a hybrid physical-virtual space.

2. DAILY ACTIVITY-TRAVEL PATTERNS IN SPATIAL-TEMPORAL CONTEXT: THE INDIVIDUAL AND DISAGGREGATED LEVELS

Historically, human geographers and social scientist have given considerable attention to the study of locations aspects of human activities. The effects of space on human behaviour tended to treat time as an external factor, something that is relevant to understanding a given phenomenon, but not essential. Activity choices were seen being made in the context of distance alone, such as with the gravity model, and often these decisions were seen in an aggregate sense, with individual decisions viewed as minor variations of those of larger zonal-based groups (Golledge and Stimson, 1997).

However, for last three decades or so, the location of activities in time and as well in space has been given increasing significance in research. An innovative and instructive approach to the study of time, space, and human activities, known as the *Lund time geography approach*, has been developed by geographers at Lund University in Sweden. They attempted to develop a model of society in which constraints on behaviour (activity) may be formulated in the physical terms 'location in space, area extension and duration in time' (Hägerstrand, 1970: 11). This approach became widely known when in 1970 Torsten Hägerstrand published his paper *What about People in Regional Science?*.

At the heart of Hägerstrand's time geography as initially formulated was the notion that all of the actions and events that sequentially make up an individual's existence have both temporal and spatial attributes. For Hägerstrand the logical place to begin the study of human geography was with the individual. The approach suggested that there are relatively smaller number of primary factors in everyday life that imputing upon all individuals and constrain their freedom to occupy certain space and time locations. It was hypothesised that when these constraints are identified, it is possible to deduce reasons as to why a particular individual follows one path rather than another.

Thus Hägerstrand's time geography examines human activities under various constraints in a space–time context. Space and time are connected through the concept of *space–time path*, which depicts the sequence of an individual's activities at various locations over a time period (Hägerstrand, 1970). The possible locations that a person can visit within a given time window form a continuous space known as a space–time prism (Hägerstrand, 1970; Lenntorp, 1976). A space–time prism defines the spatial and temporal boundaries of an opportunity space available to an individual with respect to his/her constraints. Within a particular space-time prism, individuals need to trade off between travel time and activity duration. They may choose to travel further and spend less time at the destination to perform the activity, or to travel less and allocate more time to the activity. Another aspect of time use concerns the variation of time spent on various activities across the day of the week. There is a substantial amount of empirical evidence of such variation (e.g., Hanson and Huff, 1982; 1986; Huff and Hanson, 1986; Pas and Sundar, 1995).

During the 1970s behavioural geographers developed the concept of *action space* that describes individual's total interaction with and response to, his or het environment. Jakle et al. (1976) wrote that the action space: "specifically draws attention to the individual's relationship with his surroundings social and spatial environment and allows us to examine the patterns in which individuals interact in space. We can most effectively use the concept by dividing it into meaningful components – movement and communications" (p.94). Thus in a broad context, action space provides a framework within which individual or group spatial interaction can be viewed.

The movement component of an action space may be termed an *activity space*, which is defined as the subset of all locations within which an individual has direct contact as a result of his or her day-to-day activities. Thus activity spaces represent direct contact between individuals and their social and physical elements, whereas communication channels, such as newspapers, radio, television, internet and other communication meanings/tools, are indirect links with their environment. All human activities occur coincidentally in *time* and *space*. Hence it is necessary to identify the *temporal* and the *spatial* aspects of activity space. The temporal aspect concerns the frequency and regularity with which an individual chooses to participate in a specified activity and the duration time of this activity. The spatial aspect of activity spaces involves both the location of activity and pattern of movements.

The time-geographic and the action-space and activity-space approaches to the study of human spatial behaviour in a spatiotemporal context of large-scale environments have been integrated and extended in what is referred as an activity approach to the analysis of individual and household activities and travel behaviour (Golledge and Stimson, 1997). For example, Kurani and Lee-Gosselin, (Kurani and Lee-Gosselin, 1997) have attributed 'the intellectual roots of activity analysis' to fundamental contributions from Hägerstrand (1970), Chapin (1974), and Fried et al. (1977). Hägerstrand forwarded the time geographic approach that delineated systems of constraints on activity participation in time-space. Chapin identified patterns of behavior across time and space. Fried, Havens, and Thall addressed social structure and the question of why people participate in activities. *The* *motivation of the activity approach is that travel decisions are activity based, and that any understanding of travel behavior is secondary to a fundamental understanding of activity behavior*.

The activity approach explicitly recognizes and addresses the inability of trip-based models to reflect underlying behavior and, therefore, the inability of such models to be responsive to evolving policies oriented toward management versus expansion of transportation infrastructure and services (McNally and Rindt, 2007).

King and Golledge (1978) noted that an activity approach to urban analysis provides insights into the functioning of urban areas and to their spatial structure. Analyst of urban spatial behaviour views the city as a collection of individual activities, actions, reaction, and interactions. They described urban places in terms of what is going on instead of in terms of quantities of land use of various types. In adopting this approach, the rationale is that if we know: (1) how people actually use an urban area, (2) how they respond in choice situations, (3) how they sequence their activities and the duration of their activities, and (4) the relationship of each of these things to changes in their own circumstances, then we will be in a better position to evaluate policies designed to change urban environments. And we will be in a better position to describe the city as it is used be the people living in it.

A number of general conclusions may be drawn concerning the nature of human activities in time and space within the context of individual action space, activity spaces and time geography approach. For every individual there appear to be three components of daily activity patterns. These are: (1) the time of an activity; (2) the space over which the activity takes place; and (3) the type of activity (Golledge and Stimson, 1997). All of these are highly interconnected.

Time is taken into consideration in two ways: first, in terms of the duration of each activity; and second, in terms of the time of occurrence of the activity. The duration of each activity is a basic ingredient in the account of a day's activities. But not enough is known about this aspect of activity patterns. The time of day when activities occur also is important and may be critical in determining whether or not an activity will take place together with others that are needed during the daytime.

The distribution of facilities in space for particular activities appears to be of prime importance in determining whether or not a given activity will be performed and, perhaps more so, in determining the frequency with which an activity will be performed. We may hypothesize that activities will have a longer duration when access is close and easy, rather than when it is distant and/or difficult. Further, the distribution of potential places of activities is viewed by the individual from a certain perspective - that is, from where he or she is located. This means there is a need to relate activity patterns to the overall cognitive map that people have of a particular urban environment and, in particular, to identify orientation nodes.

According with the time-geographic approach there exist three basic relationships of space–time paths between different individuals: (1) co-location in time, (2) co-location in space, (3) co-existence. Co-location in time represents activities in different space–time paths that interact with each other within a common time window. Co-location in space occurs when activities in different space–time paths occupy the same location in different time windows. Co-existence describes the cases when activities take place at the same location and within a common time window.

Another term, space–time bundle, is used for more general cases of co-existence relationship of space–time paths. It represents situations where space–time paths are in close spatial and temporal proximities. Miller (2005b, p. 386) defines a station as ‘‘a location in space where paths can bundle for some activity”. Common examples of station include homes, offices, and shopping malls where individuals converge to participate in particular activities. Depending on the spatial resolution level for stationary bundles, a station can be a building, a neighborhood, or a city. A space–time bundle can happen during movements in situations such as carpooling or riding public transit (Miller, 2004), which is called a mobile bundle as opposed to a stationary bundle at a fixed location.

As it was pointed out the availability of travel and activity data allows researchers to examine particular facets/components of daily activity-travel patterns in space-time context and the relationships of activity-patterns between different individuals. Next, we will describe how necessary data on travel and activity from GPS tracers might be obtained with the imputation system TraceAnnotator; these results will be displayed with Geographic Information System TransCad.

3. TRACE ANNOTATOR

The general purpose system, called ‘TraceAnnotator’, has been developed to process automatically multi-day or multi-week GPS traces. For the purpose of this project, the TraceAnnotator has been configured in the way that the data processing could be relatively divided into two main processes:

1. imputation of transportation modes and activity episodes, where data imputation is established by using a Bayesian belief network.

2. imputation of activity type, where GPS data are fused with GIS land use data and personalised land use data.

The system has been designed according to the following goals:

- that it can process arbitrarily large datasets;
- that it can process the data fast;
- that it can handle GIS calculations;
- that it can be used without programming;
- that it can be easily extended.

TraceAnnotator has been developed using Java and it uses the following technologies:

- Spring for configuration using xml files (http://www.springsource.org/).
- GeoTools for the GIS based components (http://geotools.codehaus.org/).
- Netica software for the Bayesian Network component used in the implementation of the ClassifierFilter (http://www.norsys.com/).

The main two classes of TraceAnnotator are Sample and Filter (see **Figure 1**). A sample is one measurement from a GPS trace. It contains attributes like date, time, latitude and longitude. A list of samples is called a sample trace or just trace. A filter processes a trace of samples. Most filters will manipulate each sample. For example, it can add new attributes or change existing attribute values, but a filter could also write derived data into an external file. Multiple filters can be chained together and each filter can make some changes to the sample or do some data processing and then send the sample to the next filter. By using these filters as building blocks more complex processing can be done without having to program new filters.

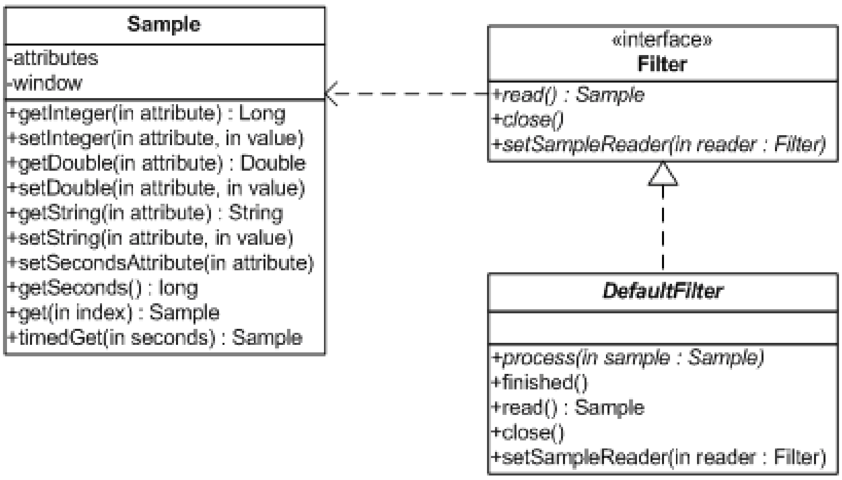

Figure 1 Sample and Filter are the main two classes of the TraceAnnotator

Filters can be chained together to do more complex processing, but normally only one sample is processed at a time. To be able to perform calculations across multiple samples, the concept of a sample window is introduced. A sample can have a sample window, meaning that samples before or after the current sample can be accessed. By also adding a seconds attribute to the sample it is possible to get a window of samples specified by time. Using this, requests like give me all samples between -60 seconds to 60 seconds (in the future) can be resolved. Normally samples do not have an initialized window. This can be done by adding WindowFilter into the filter chain. All filters that appear after that WindowFilter are allowed to use the window of a sample. But if an attribute of a sample is added or changed, accessed using a window, another WindowFilter needs to be added. This is because the WindowFilter will have to buffer the future sample values.

3.1 Filters

The different transport modes and activity episodes cannot be distinguished without additional information as average acceleration, maximum acceleration, maximum speed and other factors, including errors in the GPS device itself. The basic idea underlying the concept of a filter is to provide the chain of necessary calculations in order to derive additional information. The filters can be divided into eight categories:

1. Simple Filters
2. GIS Filters
3. Windowed Filters
4. Classifying Filters
5. Custom Filters
6. Input Filters
7. Output Filters
8. Misc Filters

Simple Filters - Simple filters just process individual samples. Some examples of Simple Filters are:

- ConvertToUnit – to convert a value from one unit type to the other. For example, this filter could be used to convert the value of the attribute *distance* from km/h to m/s.
- TimeZoneFilter – to convert a given date and time from one time zone to another time zone. This filter also takes daylight saving time into account.

GIS Filters – Filters that uses GIS information (like shapefiles).

- GISDistanceFilter – Calculates the distance from the position stored in the longitude and latitude attributes to the closest feature in a given shapefile.

  For example, this filter calculates the distance between the given position and the railway track (GIS information about railway track is stored as shapefile).
- DropBadGISLoggerSamplesFilter – Checks if the latitude and longitude attributes have a valid value. This filter will only return samples for which those values are valid.

Windowed Filters – Filters that make use of a window of samples. A window is used to lookup samples that are before or after the current sample. Most windowed filters also use the specialized seconds attribute. The seconds attribute and the window must first be initialized using the WindowFilter before using these filters. For example, for the calculation of average speed and average acceleration we define a sample window between -60 seconds to 60 seconds; for the calculation of the accumulated distance during every 3 minutes – a sample window is between -240 seconds to 0 seconds. Examples of Windowed Filters are:

- WindowFilter – this filter will initialize the window and the seconds attribute on the samples. All windowed filters must be placed after an instance of this window filter.

- AverageFilter – to calculate the average value of an attribute in the given window.
- MinMaxFilter – to calculate the minimum, maximum and range of an attribute in the given window.
- SumFilter – Calculates the sum of an attribute in the given window.
- TimeDerivativeFilter – can be used to calculate the speed from a distance attribute or the acceleration from a speed attribute.
- ModeFilter – Calculates the value of an attribute with the highest frequency in the given window.

Classifying Filters – Filters that can be used to classify a sample.

- NeticaFilter – This filter uses the Bayesian belief network library Netica to classify samples. It is possible to specify the following: a Netica network, which sample attribute values must be entered into which node of the Bayesian belief network and into which attribute the result should be stored.

Custom Filters – Filters developed for a project to perform some very specific tasks.

- ActivityCalendarFilter – This filter extracts activities and trips from a trace. Then this filter merges extracted activities and trips as defined by the customised merging rules. These rules are based on the type of activities and trips and on the time threshold value of activities and trips. The defined threshold value of time is 3 minutes. All trips and activities, which are less than 3 minutes, are merged with other trips or activities. This Filter keeps a relation with the original GPS traces. Thus, for trips the route is stored and for an activity the location is stored. After that the reverseGeocode operation is performed for all derived activities. This operation provides an address for the activities from the known latitude and longitude attributes of the activities. This reverse geocoding is a powerful component of the system as it saves an enormous amount of work for the researcher and is a powerful link in the data fusion process.

Remained groups of Filters are Input and Output Filters.

Input Filters – Input filters are at the beginning of the filter chain and are used to create new samples from files or other sources.

Output Filters – Filters that can be used to save sample back to a file or another resource.

- SampleOutputStreamFilter – This filter can be used to write a sample back to a comma separated file.

- SplitSampleOutputStreamFilter – This filter writes samples to a comma separated file. The filename will contain the date of the sample (derived from the seconds attribute). For each day a new file will be generated that will contain the samples of that day.

Misc Filters include Filters that do not have their own category yet and do not fit well in the other categories. One of the design goals of the system is to make it very simple to implement new filters. The DefaultFilter class can be used to implement most filters.

3.2 Application of the system ‘TraceAnnotator’

Figure 2 shows a simplified version of the configuration of TraceAnnotator. At the beginning of the chain is the SampleInput reader that reads samples from the logger.csv file (data format from a GPS device). The next filter will remove bad samples that contain invalid locations. For this reason it needs to know the attribute names of the attributes containing the latitude and longitude coordinates. Samples that are invalid, missing samples or samples that have not allowed values, will be removed. Next the ConvertToType filter will remove the unit from the DISTANCE attribute and convert it to a Double datatype, making further processing quicker. The next filter will use the DATE and TIME attribute to construct a SECS attribute. When no timezone is given, it is assumed to be universal time (UTC). Then the sample arrives at the GISDistanceFilter. This filter will take the latitude and longitude and search in the given shapefile for the closest feature within the search distance. This distance will then be stored in the destin attribute. In this example, the distance between a point and the closest railroad is calculated and stored in the RRDIST attribute. After that, the sample will arrive at the WindowFilter. This filter will assign a window to the sample that references 50 samples in the future and 50 samples in the past. The WindowFilter constructs this window by processing all samples that passes this filter and only sending a sample to the next filter when it has processed its complete window. Now the sample has a window and it can be processed by windowed filters like the Average Filter. This filter calculates the average speed and deviation for the average speed over the window of -60 seconds to 60 seconds. It will put the results in AVGSPEED and STDDEVSPEED attributes.

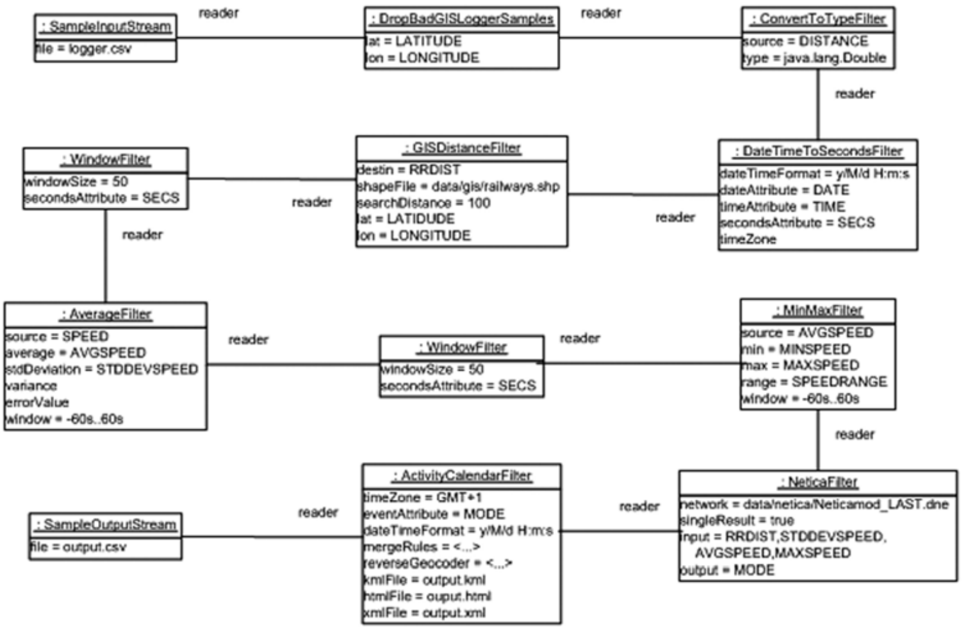

Figure 2 A simplified version of the configuration of the TraceAnnotator

The MinMaxFitler will calculate the minimum and maximum values of the AVGSPEED attribute, but that attribute is not yet available in the window of the sample. To make it available we have to put the samples through another WindowFilter. After that, the MinMaxFilter can do its calculations.

In this way, many derived data can be generated that can be used by a classification filter. In our system we use the Netica library for classification using a Bayesian Belief Network. The filter needs a trained network. It will put the input attributes into the input nodes of the BBN and then get the classification result from the output node and put it into the output attribute. In this case, the network is used to determine the transportation mode of a certain sample. It can also determine if the sample is part of an activity.

The ActivityCalendarFilter is the only specialized filter for this project. It will look for parts in the trace with the same MODE and generate trips and activity episodes from that. Then, it will use some merge rules, which can be specified in the configuration of the filter, to merge some trips or episodes together (see **Filters**). For example, a 1 minute walk after using a bike can be merged into the bike trip. This way some errors and trips or episodes that are too detailed can be removed.

The last step of the ActivityCalendarFilter is to reverse geocode all activity episodes to give its location a textual representation. This is done by reverse geocoding all points in the episodes and then choose the location name with the highest frequency of address records. The results of this filter will be stored in different files that can be used by other software for analysis (html, kml, csv files).

The SampleOutputStream filter gets its samples from the ActivtyCalendarFilter and will store them in an output file. This can be used to inspect all generated attributes.

3.3 Spatial data and geo-referencing

The GISDistanceFilter can determine the distance between a point and the shapes in a shapefile. The filter uses the Geotools library to implement this functionality. For the project also a reverse geocoder has been implemented. A reverse geocoder can find a textual representation for a certain location from a given latitude and longitude point. This reverse gecoder is used by the ActivityCalendarFilter to determine the locations where the activities take place. It can use a road shapefile to find addresses, but also kml files can be used to identify points. Kml is the format used by Google to represent geographical data.

3.4 Internet Based Prompted Recall Survey

For the validation process of modes and activities episodes generated by TraceAnnotator it was decided to develop the Internet-based prompted recall instrument. A specially designed web application allows survey participants to upload multi-days GPS tracers. Because the first application of the system concerns a study on the relationship between characteristics of the built environment, individuals' space-time behaviour and formation of mental maps, the daily travel patterns of the survey participants are displayed only in a tabular format (see Figure 3). There is no map display, as this might influence individuals' mental maps. However, it is straightforward to add a map.

After uploading a multi-day GPS tracer on the web application, TraceAnnotator processes GPS tracer data. As a result, Activity Agendas, arranged by the date (yy-mm-dd), are generated for every day.

Respondents are requested to check information for every Agenda, make necessary changes, save these changes and confirm the Agenda. After confirmation, the Agenda will move to the link 'Confirmed Agendas'. In the link 'Confirmed Agendas' participants can only view Agendas but they cannot edit Agendas (see **Figure 3**). The following changes can be made:

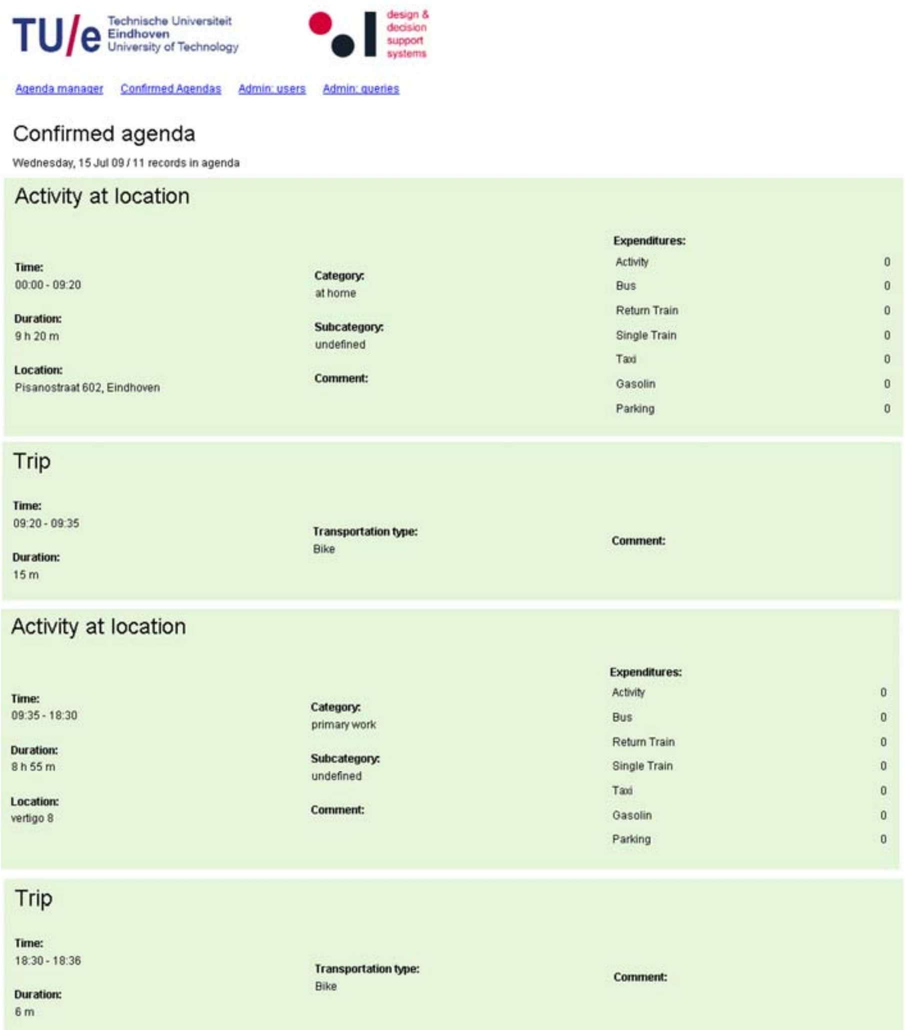

Figure 3 The Screenshot of the Confirmed Agenda after validation process
(a part of the Agenda)

- Change type of transportation mode and edit location of the activity.
- Change Activity at Location to the Trip and vice versa.
- Insert missing Trip or Activity using free time slot. Free time slots occur as a result of a missing signal or COLDSTART.

Respondents also have to indicate for every activity the type of activity which has been conducted, if it is incorrect or not identified. The activity types are listed by categories (work, servicers, healthcare, etc). Activity types such as at home, primary work, grocery shopping is automatically identified by the system on the base of the personal spatial data. It was decided that respondents are not allowed to adjust time attributes (beginning, end and duration). Observations and corrections, which have been indicated by respondents in the prompted recall survey, are then used to update the conditional probabilities of the variables in the Bayesian belief network.

4. DATA IMPUTATION

As discussed in the previous section, a Bayesian network makes up the core of the system and can be viewed as the engine that replaces ad hoc rules with a dynamic structure, leading to improved classification if consistent evidence is obtained over time from more samples (more traces). A Bayesian Belief Network, belief network or BBN is a model for reasoning about uncertainty. A BBN represents all factors deemed potentially relevant for observing a particular outcome and thus can be used to predict the conditional probability of observing a particular outcome. BBNs are a graphical representation of probabilistic causal information based on two components: a directed acyclic graph and a probability distribution (set of probability tables). The graph consists of nodes and arcs as shown in **Figure 4**. The nodes represent variables, which can be discrete or continuous. The arcs represent causal/influential relationships between variables. This is an effective way to describe the overall dependency structure of a large number of variables.

In our case, the Bayesian belief network represents the multiple relationships between different spatial, temporal and other factors, including errors in the technology itself and the facets of activity-travel patterns that we wish to impute from the GPS traces. We use a Bayesian belief network to impute automatically the type of transportation mode, activity stops, COLDSTART and URBANCANYON from the GPS tracers (11 possible states of the node ‘Mode’, see Figure 4). BBN makes explicit the dependencies between different variables. We included the following variables (look **Transport Mode and Activity Episode Identification Algorithm)**: distance to the railway track, average and maximum acceleration, average speed, max speed, deviation from the average speed, accumulated distance during every 3 minutes, personal information concerning possession of car, bike and auto, the number of satellites that GPS device used for records (usedsat), the number of satellites which are available at the moment (viewsat), position accuracy of 3d coordinate (PDOP) and horizontal accuracy of 2d coordinate (HDOP), data fixed or not (valid). While modeling arcs between nodes of the network, we took into account all possible factors. If there may be direct dependencies between variables, then they are conditionally dependent. Since a type of 'Mode' determines how the other nodes will be changed, we model this relationship by drawing an arc from the node 'Mode' (the parent node) to all the other nodes. The arcs represent causal relationships between variables.

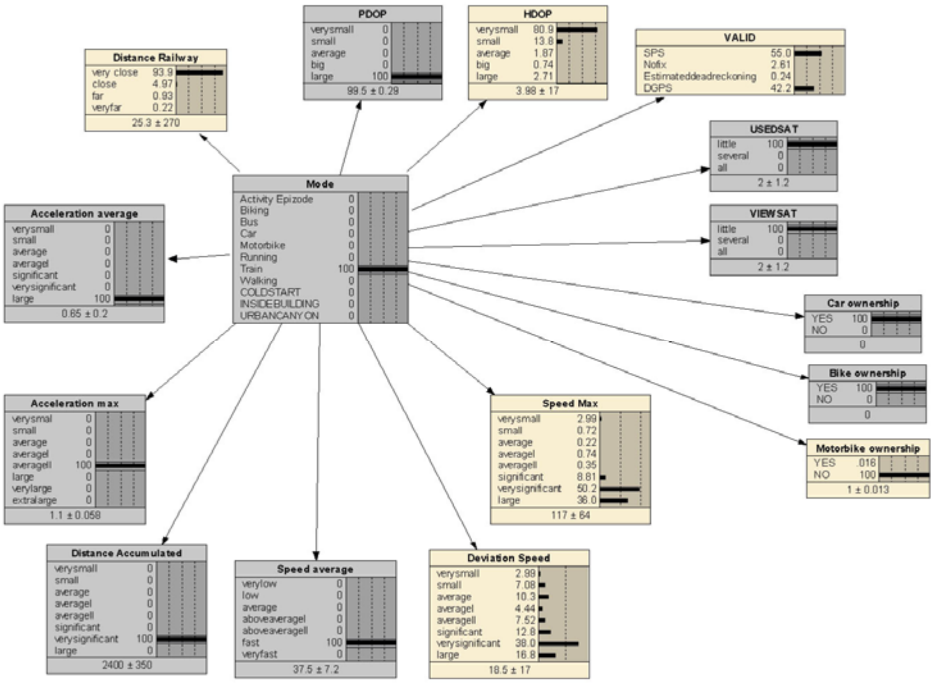

Figure 4 Bayesian belief network: Transportation Mode and Activity Episode Identification

In the system, the BBN is a computational object able to represent compactly joint probability distributions, which denote dependencies and independencies among the variables as well as the conditional probability distributions of each variable, given its parents in the graph. As an example, **Figure 5** shows the conditional probability of the variable 'Distance Railway' given the variable 'Mode'. The probabilities provided for each combination of events are based on the real data received from the participants in the prompted recall and thus either are the confirmed interpretations of the GPS traces or the corrected interpretations.

From **Figure 5**, we can see that the conditional probability is highest (93.878%) for the train when distance to the railways is very close (less then 25 m). It means that it is very likely that the used transport mode reflected in the GPS trace is the train because the distance to the railway track is less than 25 meters.

Node: RRDIST Apply Okay

Chance % Probability Reset Close

Mode	very close	close	far	veryfar
Activity Epizode	1.377	2.406	1.145	95.071
Biking	0.367	0.140	0.140	99.353
Bus	0.834	1.147	1.721	96.298
Car	1.658	1.154	1.024	96.164
Motorbike	0.373	0.187	0.187	99.253
Running	0.0889	0.0889	0.178	99.644
Train	93.878	4.973	0.929	0.220
Walking	1.374	3.453	6.276	88.897
COLDSTART	5.621	0.0549	0.165	94.160
INSIDEBUILDING	0.0984	0.0890	0.0703	99.742
URBANCANYON	35.197	24.017	6.004	34.783

Figure 5 Conditional probability table of the variable 'Distance to the Railway' given the variable 'Mode' (screenshot of the Netica network)

4.1 Pilot Studies

The results of pilot study conducted by Moiseeva, Jessurun and Timmermans (Moiseeva et al, 2010a, 2010b) for testing the system 'TraceAnnotator' showed promising results. The aim of the first pilot (Moiseeva et al, 2010a) was, first, to show the success of the imputation accuracy of trips, transportation modes and activities from the multi-days GPS tracers, secondly, to illustrate that Bayesian belief network can learn over time reducing respondent's burden. The imputation accuracy for various facets of activity-travel patterns was considerable high and varied between 85-99%, in average 92% of all transport modes was correctly identified (train (99%), bus (98.4%), walking (94%), car (93.3%), bike (85.5%). The study also showed that system can learn over time, for instance, the accuracy of correctly identifying biking trips increased from 85.5 % up to 93%, while this percentage for walking trips increased from 94% up to 97%.

The aim of the second pilot study (Moiseeva et al, 2010b) was to show two alternative approaches to impute transport modes and activities from multi-week activity travel diaries. The first approach, the personal histories approach, is based on *creating personalised Bayesian belief network* for every participant and then learning the conditional probability table of Bayesian network with individual's data obtained from previously processed traces. The second approach, the aggregated histories approach, is based on updated conditional probability table of *common BBN* using previous multidimensional sequences of activity-travel patterns aggregated across respondents.

Although the initial Bayesian belief network performed quite well, the results of this study showed that the imputation accuracy improves faster over time when learning is based on the aggregated histories of the respondents. In average for all transport modes the imputation accuracy in case of personal histories learning reached max 96% whereas for the aggregated histories learning it reached 99 -100% of the accuracy.

Initially 96% of activities were correctly identified by the system. After learning network on the base of personal histories and aggregated histories correspondently in average 96% and 99% activities were identified.

5. EXPERIENCES

In this section we will illustrate different facets of activity-travel patterns collected with a GPS device and processed with the TraceAnnotator. 5 participants were involved in data collection carrying the GPS logger Bluetooth A+. The participants of the study were required to be residents of Eindhoven (the Netherlands) and perform the majority of their every day activities in Eindhoven. The survey participants have been asked to download their GPS data at least once per week, upload traces to the web application and confirm their Activity Agendas. Data on activity-travel patterns were collected during 10 weeks.

Confirmed activity-travel data have been used for illustration purposes.Following facets of individual daily activity patterns have been represented: (1) movement patterns (trips) and transport modes; (2) the space over which the activity takes place; (3) the type of activity and (4) day and the time of activities and trips. The participants of the study are PhD candidates from the Eindhoven University of Technology and live in the northern part of the city. Small sample size and homogeneous in work-home trips partially explains their activity patterns at the city scale and that a significant part of their daily trips was conducted by foot or by bicycle. GIS TransCad offers an operational environment to manage, query, analyse, and visualise complex and dynamic activity and travel data collected at the individual and aggregated levels.

5.1 Activity Space: Spatial Component

The spatial component of individual activity spaces can be represented in terms of the location of activity and pattern movements.

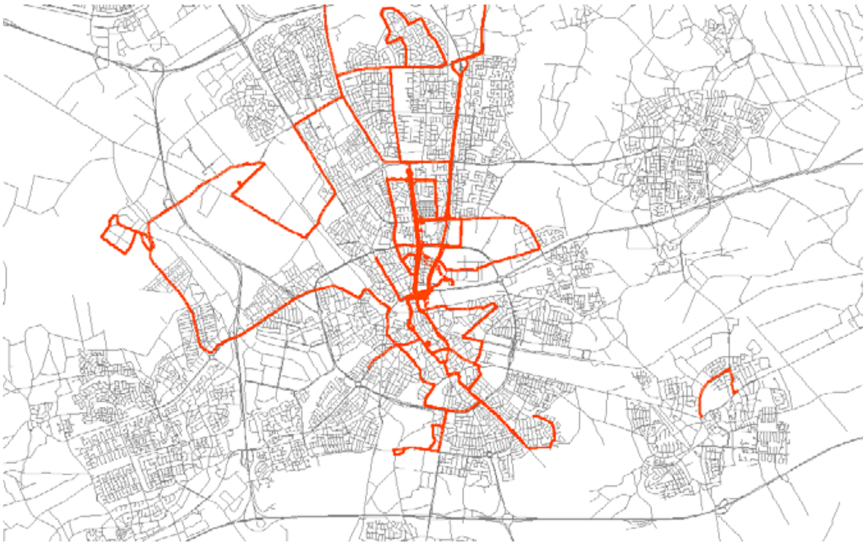

Figure 6 Movement Patterns in Activity Space: Bus Trips Aggregated between Participants

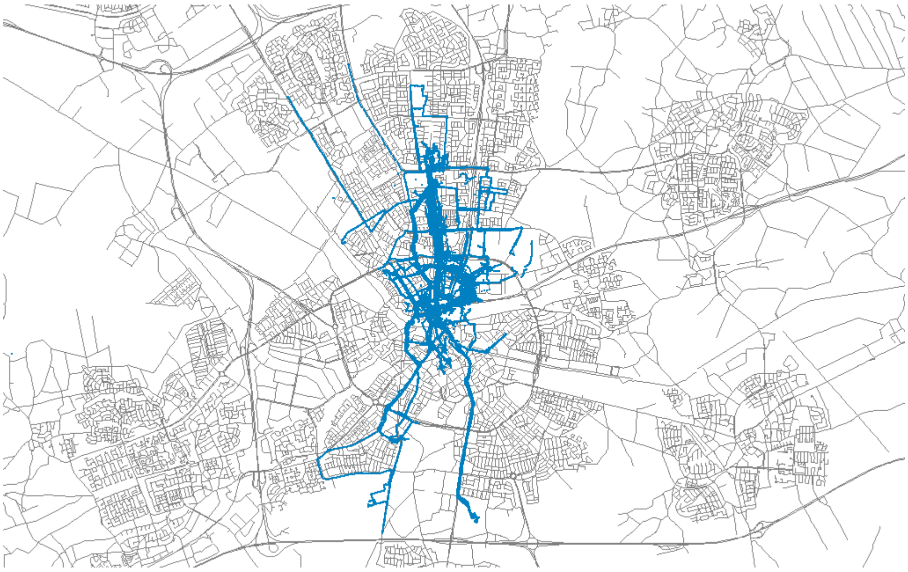

Figure 7 Movement Patterns in Activity Space: Bike Trips Aggregated between Participants

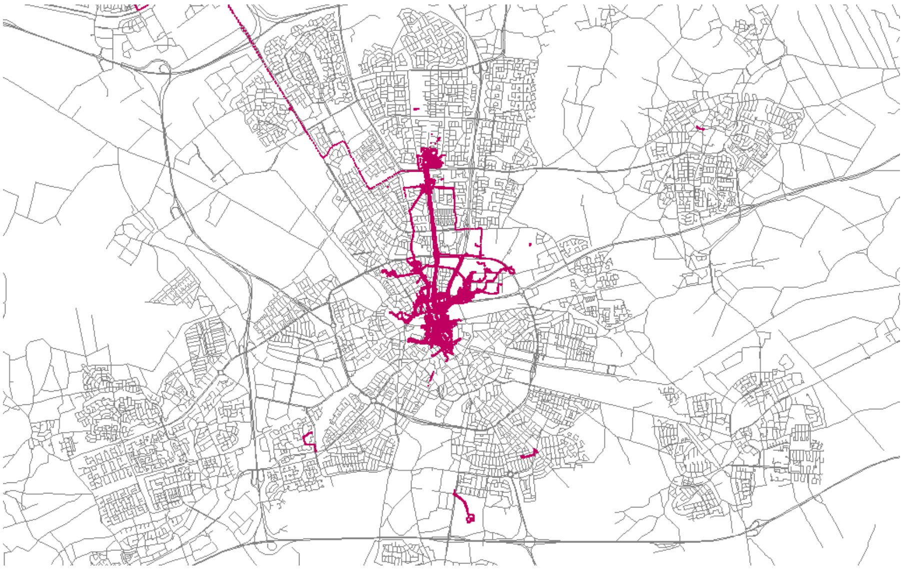

Figure 8 Movement Patterns in Activity Space: Walking Trips
Aggregated between Participants

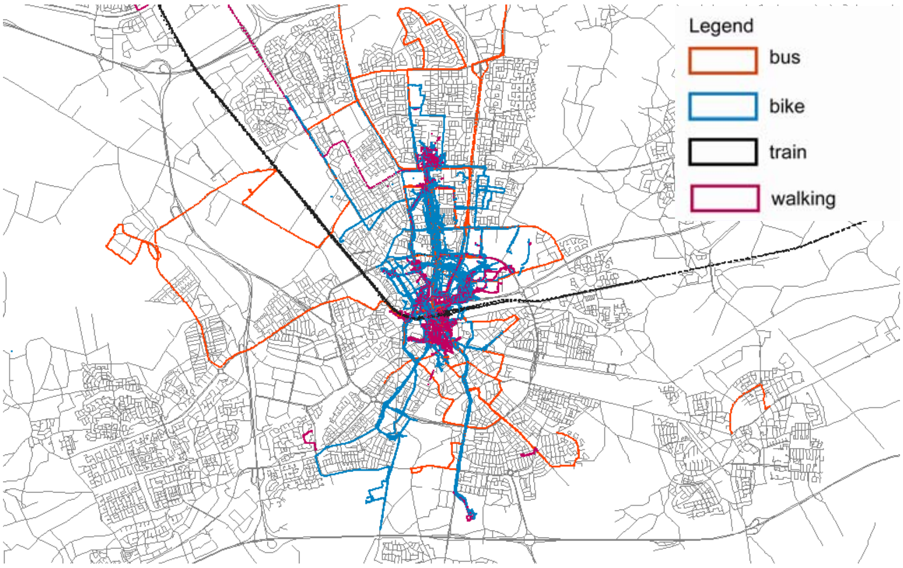

Figure 9 Movement Patterns in Activity Space: Trips Aggregated between Participants

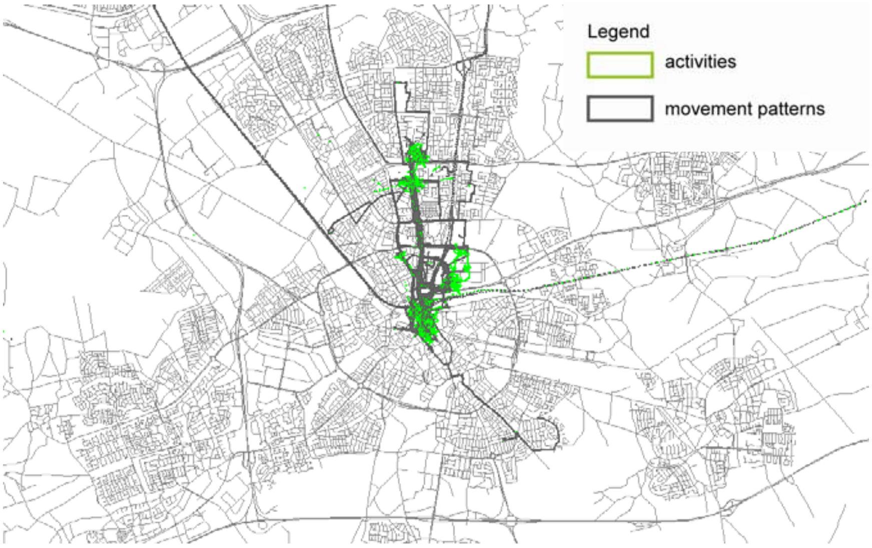

Figure 10 Activities Locations in an Individual Activity Space (participant 1)

Figure 11 Activities Locations in an Individual Activity Space (participant 2)

Figure 12 Activities Locations in an Individual Activity Space (participant 3)

Figure 13 Activities Locations in an Individual Activity Space (participant 4)

Figure 14 Activities Locations in an Individual Activity Space (participant 5)

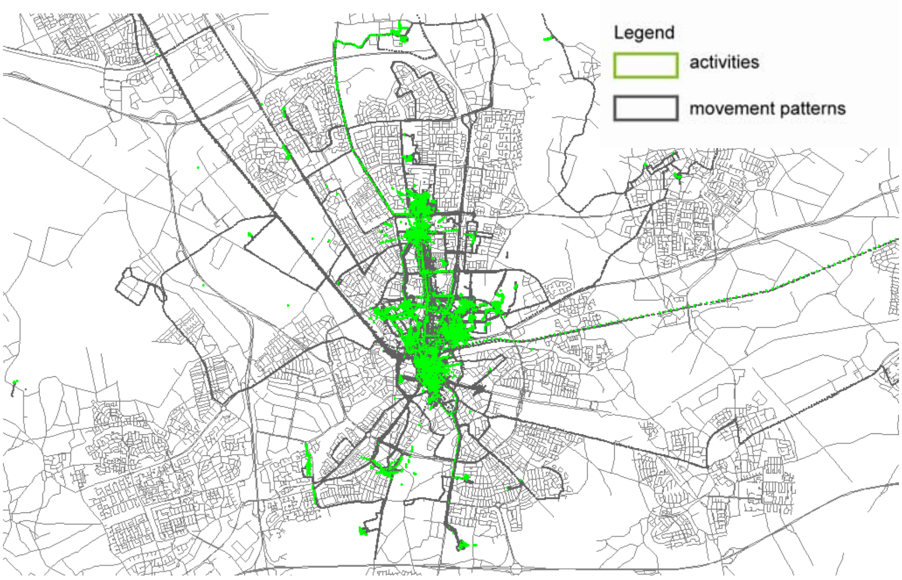

Figure 15 Activities Locations in the city scale (aggregated between participants)

Figure 16 Activities Locations and Movement Patterns in the city scale (aggregated between participants)

5.2 Activity Approach: The Use of an Urban Area

Activity Approach views the city as a collection of individual activities, action, and reactions. Collection of individual activities represents the way how people use an urban area. Figure 17 shows that majority of activities are concentrated in the following areas (I) Shopping Area Winkel Centrum Woensel, (II) Technical University Eindhoven; (III) City Centre and (IV) shopping street with specialised shops and different cafés and restaurants (Kruistraat).

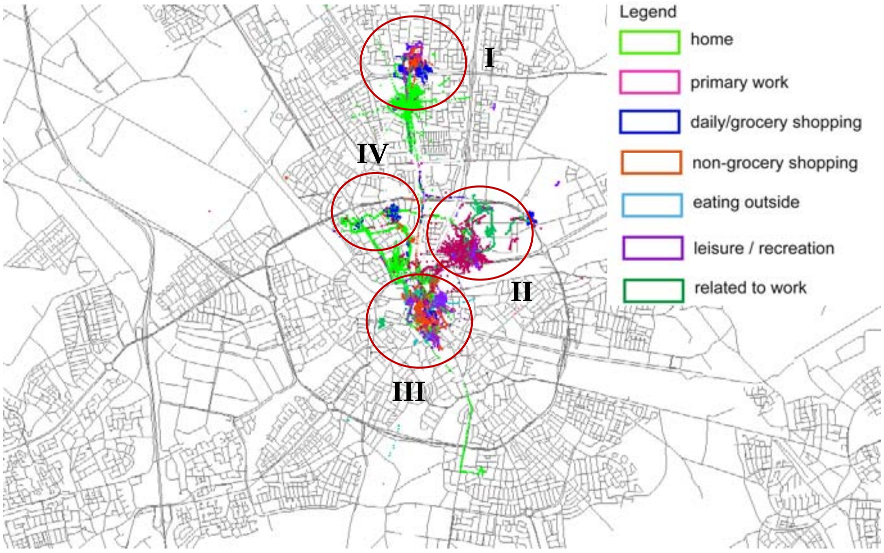

Figure 17 Activity Types Aggregated between Respondents (zoom out)

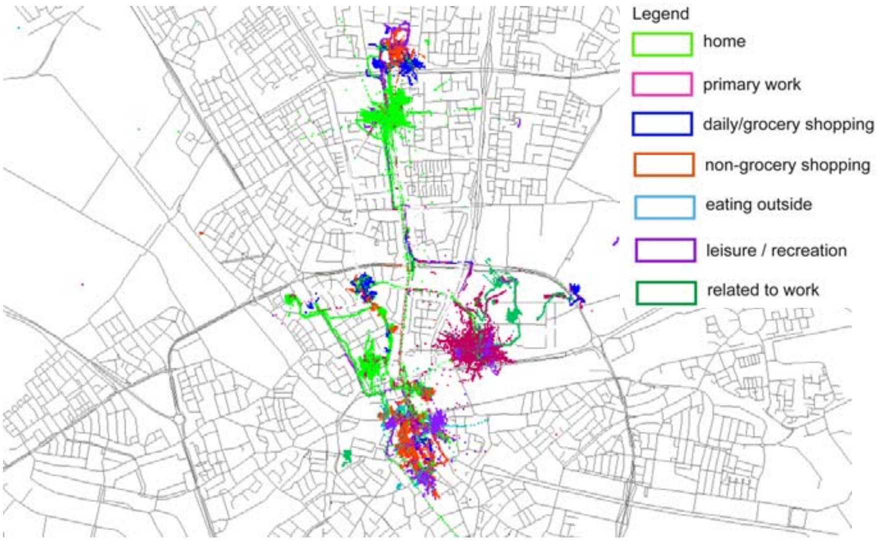

Figure 18 Activity Types Aggregated between Respondents (zoom in)

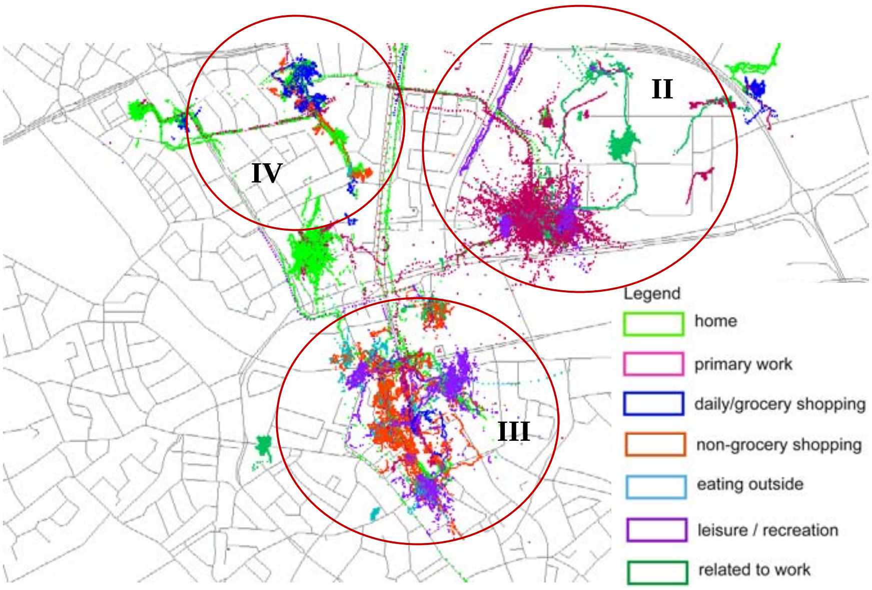

Figure 19 Collection of individual activities: city centre (III), Technical University Eindhoven (II), Kruistraat (IV) (zoom in)

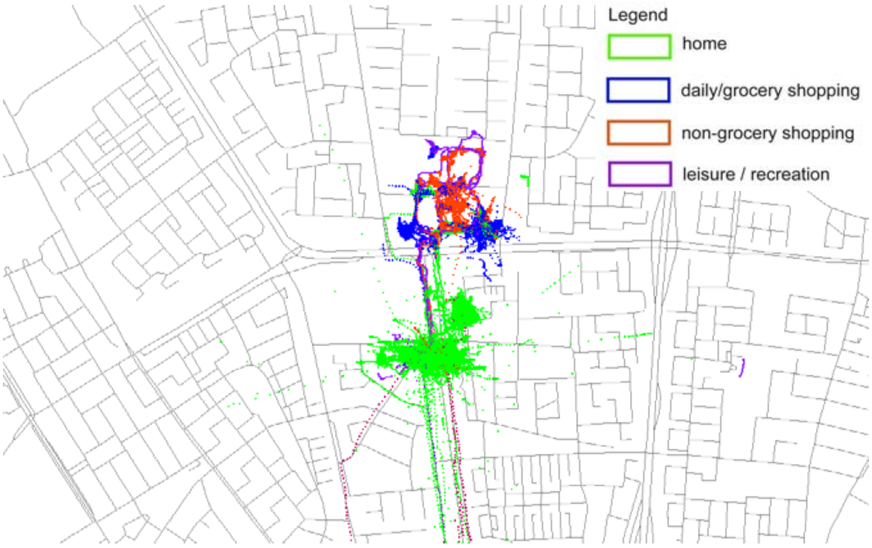

Figure 20 Collection of individual activities: Winkel Centrum Woensel (zoom in)

5.3 Co-location in Space, Time and Co-existence

According to the time-geographic approach there exist three basic relationships of space–time paths between different individuals: (1) co-location in space, (2) co-location in time, (3) co-existence (co-location both in time and space). Figures 21-23 illustrate co-location of the non-grocery shopping activities: participant 1 and 3shopped mostly in the city centre (III) and in the Shopping Center Winkel Centrum Woensel (I). In turn, all participants shopped in the Kruistraat (IV). Participant 3 shopped at all 3 locations, while others only at 2 locations.

Figure 21 Co-location in space: non-grocery shopping (aggregated between respondents) zoom out

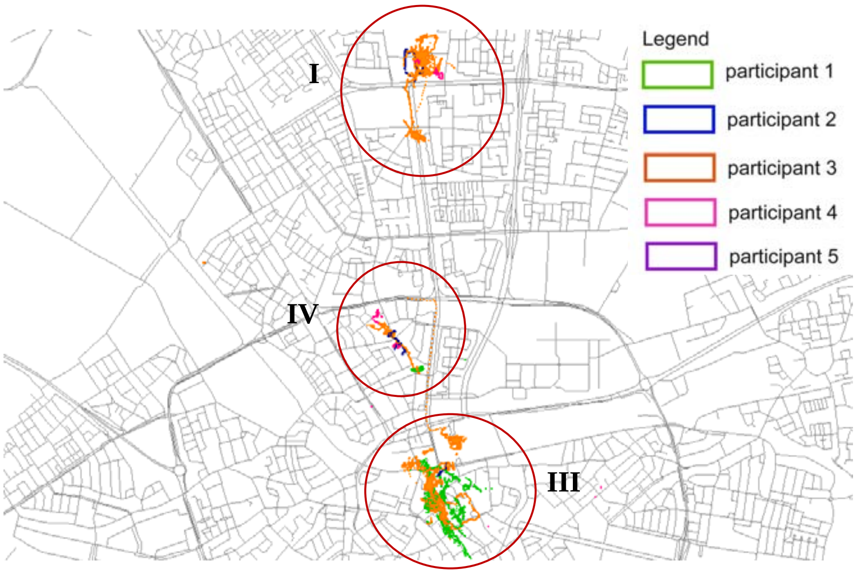

Figure 22 Co-location in space: non-grocery shopping (aggregated between respondents) zoom in

Daily grocery shopping activities (Figures 23, 24) occurred as well in the city centre (III), the shopping centre Winkel Centum Woensel (I) and in the Kruistraat, where open market takes place every Saturday (IV). The difference with non-grocery shopping is that the 2 participants (2 and 4) tend to buy grocery at the same location I. Participants 1and 5 shopped for grocery at locations III and IV, probably due to the closeness of the both locations, while participants 3 visited locations I and IV.

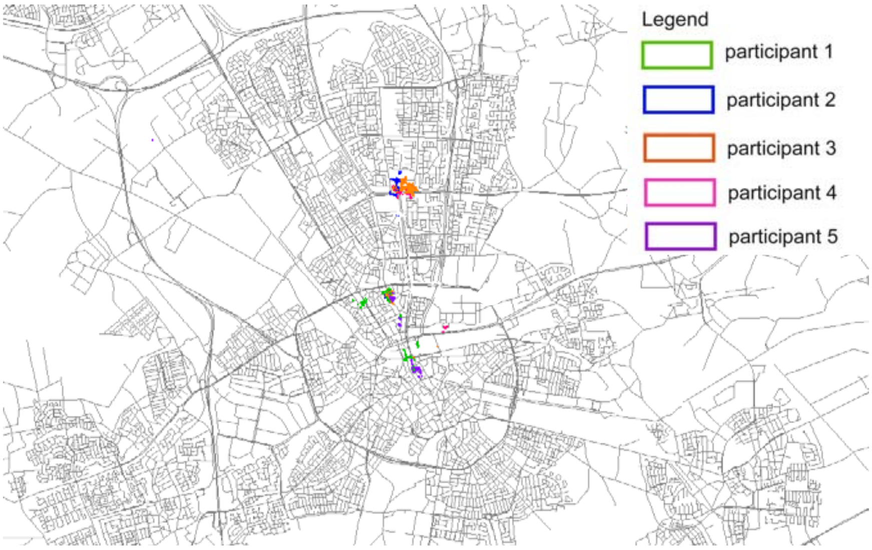

Figure 23 Co-location in space: daily/grocery shopping (aggregated between respondents) zoom out

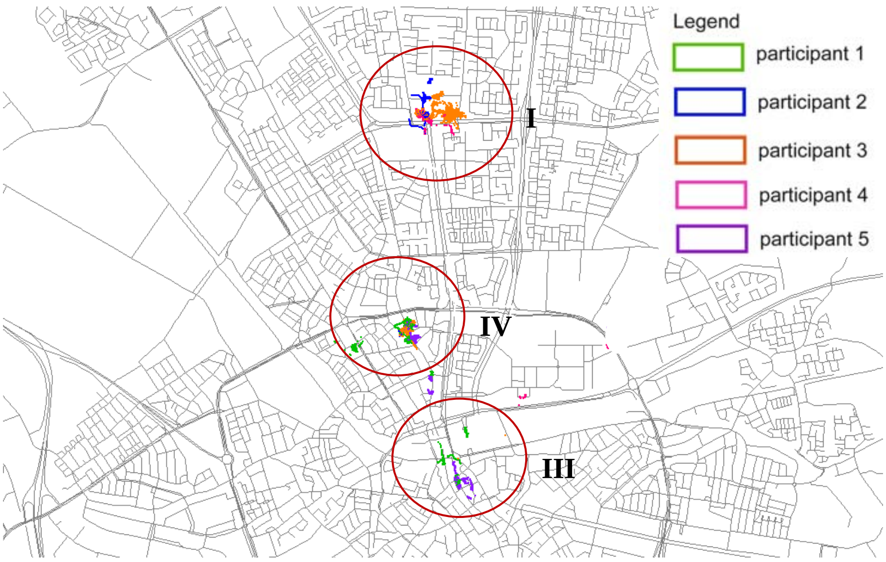

Figure 24 Co-location in space: daily/grocery shopping (aggregated between respondents) zoom in

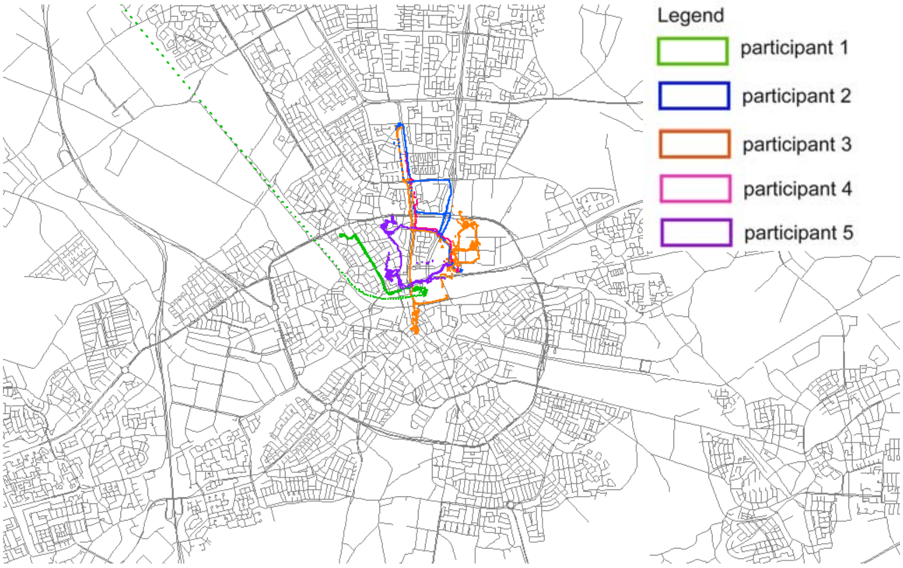

Figure 25 Activity travel patterns: co-location in time and coexistence (zoom out)

Figures 25-26 illustrate co-location in time of the activity-travel patterns of participants during one day on 19 June 2009 (Friday) from 8.00 a.m. till 8.0 p.m. During the day participant 2-5 located at the Technical University Eindhoven. For these participants Technical University Eindhoven represents a space–time bundle where activities coexist in time and space. After working participants 2 and 4 went directly home, while participant 3 went to the city centre, participant 5 to the Kruistraat and participant 1 took a train.

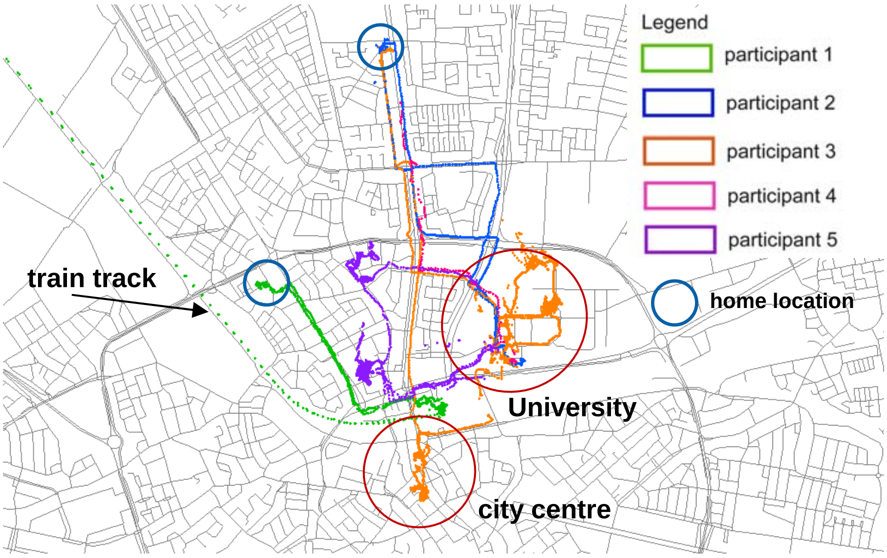

Figure 26 Activity travel patterns: co-location in time and coexistence (zoom in)

6. CONCLUSIONS

This paper has provided further insight into the advantages of advanced tracking technologies such as GPS for collecting travel behaviour of individuals in time and space and for different applications in the activity-based approach.

The design of a system a system 'TraceAnnotator' developed by authors is illustrated. This system processes semi-automatically multi-days GPS traces. The process of imputing transportation modes, activity episodes and other facets of activity-travel patterns is based on a learning Bayesian Belief network, which represents the multiple relationships between spatial, temporal and other factors, including errors in the technology itself. This technology is combined with prompted recall data. A series of pilot provide evidence to the potential of these devices and the developed algorithm to collect data on activity-travel patterns. Respondent burden for these passive data collection method is relatively low.

Data obtained from the GPS are the valuable source to analyse complex spatial-temporal relationships among activities and interactions in the physical spaces. Some illustrative examples have been provided in this paper.

In conclusion, it seems that on one side due to the future expansion of communication devices, collecting data of individual activity-travel behaviour is less demanding. On the other side the widespread adaptation of the information and communication technologies has introduced additional changes to human activity and travel patterns that have significant implications to our everyday life and the human organization of space. Modern ICT allows us with additional flexibility to choose the ways that an activity is performed and allows to carry several activities at the same time (multitasking). We now have more choice of conducting activities either in physical space and have a direct interaction with others (working in a office), in virtual space (e.g. teleworking and communication with colleges) or even combination of both virtual and physical activities, which can result in multitasking (travel in a train and have a business teleconference).

Although physical and virtual spaces are different from each other in their nature, they are not independent from each other. Instead, they often in different complex manners influence and interact. In this respect data on activities conducted in the physical spaces are not sufficient enough to understand the complex nature of the contemporary activities, additional data on virtual activities and multitasking should be collected to examine interaction between physical and virtual activities.

For a futures research GPS technologies, enabled to collect data on activities in at physical locations, might be combined with a web prompted recall, which will serve as a tool to collect data about the virtual activities and multitasking. For instance, if a trip by the train has been identified participants should fill additional fields related to the activities which occurred during the travelling; working at home can be combined with reporting virtual activities. These data can be used to better understand the dynamics of space-time behavior, allowing a considerably more rigorous and detailed analysis of the interdependencies between time, space and choice behavior. It should improve the effectiveness of our decision support tools that assess and/or predict the impact of urban planning and design on user response.

7. REFERENCES

Asakura, Y., Hato, E., Nishibe, Y., Daito, T., Tanabe, J., and H. Koshima, 1999, Monitoring Travel Behaviour using PHS based Location Positioning Service System, *Proceedings of the 6th ITS Conference in Toronto* (CD-ROM).

Asakura, Y. and E. Hato, 2001, *Analysis of Travel Behaviour using Positioning Function of Mobile Communication Devices*. Travel Behaviour Research; the Leading Edge, Hensher, D. (ed.) Pergamon, pp.885-899.

Asakura, Y. and T. Iryo, 2007, Analysis of tourist behaviour based on the tracking data collected using a mobile communication. *Transportation Research Part A*, 41, pp. 684-690.

Battelle, 1997, *Lexington Area Travel Data Collection Test*. Final Report. Prepared for Federal Highway Administration, September, 1997.

Bohte, W. and K. Maat, 2008, Deriving and validating trip destinations and modes for multi-day GPS-based travel surveys: An application in the Netherlands. *Proceedings 87th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Chapin, F.S, 1974, *Human activity patterns in the city*. New York: Wiley.

Clifford, E., Zhang, G. and P. Stopher, 2008, Determining trip information using GPS data, *Working paper (Institute of Transport and Logistics Studies)*; ITLS-WP-08-01, http://www.itls.usyd.edu.au/downloads/working_papers/itls-wp-08-01.pdf

Fried, M., J. Havens, and M. Thall, 1977, *Travel Behavior – A Synthesized Theory*, NCHRP, Transportation Research Board, Washington, Final Report.

Golledge, R.G. and R.J. Stimson, 1997, *Spatial behaviour: a geographic perspective*. The Guilford Press: New York

Hägerstrand, T, 1970, What about people in regional science?, *Papers of the Regional Science Association*, 24:7-21.

Hanson, S. and Huff, J. O.: 1982, Assessing day-to-day variability in complex travel patterns, *Transportation Research Record,* 891, pp. 18–24.

Hanson, S. and J. Huff, 1986, Classification issues in the analysis of complex travel behaviour. *Transportation*, 13, 271-293

Asakura, Y. and E. Hato, 2001, Analysis of Travel Behaviour using Positioning Function of Mobile Communication Devices. *Travel Behaviour Research: the Leading Edge,* Hensher, D. (ed.) Pergamon, pp.885-899.

Hato, E. and R. Kitamura, 2006, Data Oriented Travel Behavior Analysis Based on Probe Person Systems, *TRB Conference: Innovations In Travel Demand Modelling* in Austin, Working Paper.

Hovgesen, H., Bro, R. and N. Tradisauskas, 2008, Tracking visitors in Public Parks – Experience with GPS in Denmark. *Urbanism on Track (Volume 1),* IOS Press, the Netherlands, pp 65-76.

Huff, J.O. and S. Hanson, 1986, Repetition and variability in urban travel. *Geographical Analysis 18(2)*, pp. 97-113.

Jakle, J. A., Brunn, S. and S.C. Roseman, 1976, *Human Spatial Behaviour*. North Scituate, MA: Duxbury Press.

King, L. J. and R. G. Golledge, 1978, *Cities, Space and behaviour*. Englewood Cliffs, NJ: Prentice-Hall.

Kurani, K. S. and M. Lee-Gosselin, 1997, Synthesis of Past Activity Analysis Applications, in *Activity-based travel forecasting conference*, Washington, DC: U.S. Department of Transportation, Report DOT-97-17.

Lenntorp, B, 1976, Paths in Space–time Environments: A Time Geographic Study of Movement Possibilities of Individuals. Lund Studies in Geography B: Human Geography. CWK Gleerup, Lund, Sweden.

McNally, M. G. and C. R. Rindt, 2007, The Activity-Based Approach. Center for Activity Systems Analysis, Institute of Transportation Studies, UC Irvine, UCI-ITS-AS-WP-07-1, http://escholarship.org/uc/item/86h7f5v0

Miller, H, 2004, Activities in space and time. In: Hensher, D., Button, K., Haynes, K., Stopher, P. (Eds.), *Handbook of Transport 5: Transport Geography and Spatial Systems*. Elsevier Science, London, pp. 647–660.

Miller, H, 2005, Necessary space–time conditions for human interaction. *Environment and Planning B*, 32, 381–401.

Moiseeva, A., Jessuran, J and H.J.P.Timmermans, 2010a, Semi-automatic imputation of activity-travel diaries using GPS traces, prompted recall and context-sensitive learning algorithms, *To appear in Transportation Research Record: Journal of the Transportation Research Board* http://pubsindex.trb.org/view.aspx?id=909867

Moiseeva, A., Jessuran, J and H.J.P. Timmermans, 2010b, Semi-Automatic Imputation of Long-Term Activity-Travel Diaries Using GPS Tracers: Personal Versus Aggregated Histories. *12th World Conference of Transport Research*, Lisbon, Portugal, July 2010

Murakami, E. and D.P. Wagner, 1999, Can using Global Positioning System (GPS) improve trip reporting? *Transportation Research C*, **7**, pp. 149-165.

Pas, E.I. and S. Sundar, 1995, Intra-personal variability in daily urban travel behavior: Some additional evidence. *Transportation 22*, pp. 135-150.

Spek, S van der, 2008, Spatial Metro – Tracking Pedestrians in Historic City Centres. *Urbanism on Track (Volume 1)*, IOS Press, the Netherlands, pp 79-100.

Timmermans, H.J.P., Moiseeva, A. and J. Jessurun, 2009, Collecting Activity-Travel Data Using Modern ICT: Recent developments and Prospects, *Proceedings 14th HKSTS International Conference*

Wolf, J., Guensler, R. and W. Bachman, 2001, Elimination of the travel survey diary – experiment to derive trip purpose from Global Positioning System travel data. *Transportation Research Record*, 1768, pp. 125-134.

Wolf, J., Guensler, R., Frank, L. and J. Ogly, 2001, The use of electronic travel diaries and vehicle instrumentation packages in the year 2000 Atlanta regional household travel survey: test results, package configurations, and deployment plans. *Proceedings 9th International Association of Travel Behaviour Research Conference*, Gold Coast, Queensland, Australia.

Wolf, J., Loechl, M., Myers, J. and C. Arce, 2001, Trip rate analysis in GPS-enhanced personal travel surveys. *International Conference on Transport Survey Quality and Innovation Kruger Park, South Africa*. August 2001.

Wolf, J., Loechl, M., Thompson, M. and C. Arce, 2003, Trip Rate Analysis in GPS-Enhanced Personal Travel Surveys, in P. Stopher and P.M. Jones (eds.) *Transport Survey Quality and Innovations*, Pergmon, Oxford, pp. 483-498

Zito, R., D'Este, G. M. and M.A.P.Taylor, 1995, Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems? *Transportation Research C*. Vol.3, No.4, pp.193-209.

Internet Resources

SERENE Project, http://www.dcs.qmul.ac.uk/~norman/BBNs/BBNs.htm
Real Time Rome Project, 2006, http://senseable.mit.edu/realtimerome/