Processing of National Travel Survey GPS Pilot Data

A technical report prepared for the Department for Transport

Eindhoven University of Technology

By Tao Feng, Anastasia Moiseeva and Professor Harry Timmermans

September 2011

Contents

Contents ... 2

1. INTRODUCTION ... 3

2. BACKGROUND ... 5
   CONTEXT ... 5
   BENCHMARKS ... 7

3. BASIC PRINCIPLES UNDERLYING TRACEANNOTATOR ... 9
   Basics ... 9
   Architecture ... 10

4. NTS GPS PILOT STUDY DATA AND PROCESS DESCRIPTION ... 14
   Overview ... 14
   Training data ... 15
   GPS and accelerometer data ... 17
   Person and household information ... 22
   GIS data ... 23
   Process description ... 25

5. USE OF ACCELEROMETER DATA ... 27
   Background ... 27
   Analyses and results ... 28
   Assessment of imputation models ... 32
   Conclusions and discussion ... 39

6. APPLICATION TO NTS GPS PILOT DATA ... 40

7. VALIDATION ... 48
   Context ... 48
   Framework and rules ... 48
   Results ... 49

8. IDENTIFICATION OF SPECIAL ACTIVITIES ... 50

References ... 53

Appendix A: glossary of terms ... 55

Appendix B: rule for merging fragments of activity or travel episodes ... 57

Appendix C: rules for validation of inferred mode and purpose of trip ... 58

Appendix D: summary of validation results ... 62

Appendix E: questions for NTS GPS processing ... 64

1. INTRODUCTION

1.1 In November 2010, the National Centre for Social Research (NatCen) contracted Eindhoven University of Technology to undertake GPS data processing for a pilot of the National Travel Survey (NTS) for Great Britain, which used accelerometer equipped Global Positioning System (GPS) devices to collect personal travel data to replace the paper travel diary.

1.2 This report presents the background, technical details and application results of the data processing stage of the NTS GPS pilot project. It documents the development and application of a tool, called TraceAnnotator - developed by the team to process (semi-)automatically multi-day GPS traces - which was then applied to the data collected by NatCen. During this work we further improved our algorithms by increasing complexity and identifying specific conditions, or even by visual inspection and manual correction.

1.3 Details on the data collection and further background to this pilot project conducted for the Department for Transport (Great Britain) can be found in the *National Travel Survey 2011 GPS Pilot Field Report*, by Josi Rofique, Alun Humphrey and Caroline Killpack (NatCen, August 2011)[1].

1.4 The key requirements of the tasks described in this report were to:

- Input into designing new questions for the NTS placement/pick-up interviews to aid data processing;
- Process data – including the matching of GPS data to interview data and Geographic Information System (GIS) data, and
- Technical documentation of the data processing.

1.5 We were required to clean and process the data into trip and trip stages and the infer mode and purpose of journey. Outputs were also to include the journey start and end point and the length of the journey (distance and time).

1.6 The GPS data were collected for 874 respondents aged 12 or more during the seven day travel week that followed the NTS pilot survey, alongside additional information collected during the CAPI placement and pick-up interviews to assist data processing. Data was collected using the MGEData Mobitest GSL accelerometer- equipped GPS device[2]. Questions were added to the standard NTS interviews on personal points of interest (schools, work, gym, supermarkets etc),

[1] Unpublished at time of writing.

[2] Specifically, an offline version of the device at: http://www.mgedata.com/en/hw-and-sw-products/hw/mobitest/mobitest-gsl

details of when respondents failed to charge or carry their devices, and where atypical, the respondent’s hours of work.

1.7 A number of amendments were made to the existing TraceAnnotator system to process the NTS GPS data. Development work was done to use the accelerometer traces to infer transport modes and link the resulting algorithm to the TraceAnnotator system. Some key variables had to be estimated as devices were not set-up correctly; requiring further adjustment and the nature of some of the training data also required additional adjustments.

2. BACKGROUND

CONTEXT

2.1 The 2011 GPS NTS pilot survey is the culmination of a review of new technologies[3] and a small-scale feasibility project[4] which concluded that GPS devices are the most suitable technology option to deliver affordable and practical improvement to improve the quality and reliability of NTS diary data and that GPS technology has real promise without any fundamental barriers of feasibility or public acceptability. The replacement of the current travel diary with GPS devices could substantially reduce NTS respondent burden, offer long-term cost savings and improve data accuracy. The pilot study was designed to set out how the NTS could work in a real life situation and to collect data on a scale large enough that the findings would be statistically robust.

2.2 The use of GPS has been examined in several small scale pilot projects to date. A GPS trace consists of detailed time, latitude and longitude information recorded at a regular interval ranging from once per second. More advanced devices can also record speed, distance and height. The information on speed and distance, and other indicators that can be derived from such GPS data can be used to infer trip (stages) and transport modes.

2.3 In general, the aim of any imputation or inference model is to find the variables, functions and/or conditions that discriminate most between the transport modes and thus results in the most accurate inference (i.e. the highest percentage of correctly classified transport modes). The success of the approach will depend on (i) the accuracy of the measurements and generated statistics such as distance and speed of the GPS device, (ii) the (absence of) variability in the relevant GPS measurements across respondents, time and spatial context, (iii) the inherent differences between transport modes in specific settings, and (iv) the ability of the inference system to detect these critical differences between transport modes in terms of the chosen variables, perhaps taking temporal and spatial context into account.

2.4 Two approaches have been employed in transport research for inferring transport modes from GPS traces. First, and most common, researchers by inspecting GPS traces have formulated ah hoc rules to associate scores on particular variables with a specific transport mode. For example, Bohte and Maat

[3] Review of the Potential Role of 'New Technologies' in the National Travel Survey, Wolf et al., 2006 http://webarchive.nationalarchives.gov.uk/+/http:/www.dft.gov.uk/pgr/statistics/datatablespublications/personal/methodology/ntsreports/ntsreviewtechnologies.pdf

[4] National Travel Survey GPS Feasibility Study, Anderson et al., December 2009, http://webarchive.nationalarchives.gov.uk/+/http://www.dft.gov.uk/pgr/statistics/datatablespublications/personal/methodology/ntsreports/ntsgpsstudy.pdf

2.5 These rules are neither mutually exclusive nor exhaustive and such deterministic rules tend not to capture the stochastic nature of the GPS data. The second, less frequently applied, approach is based on formal data mining and/or statistical models, which can be deterministic or probabilistic. This approach tends to involve more rigour in the sense that inference mechanisms are more closely derived from the data as opposed to the researcher, defining the mechanisms.

2.6 Regardless of a formal or ad-hoc approach, a key problem of any inference system is that at the start of the process, there is no evidence as to the correctness of the imputed transport modes. Some researchers have relied on their personal inspection of (a sub-sample of) GPS traces. This approach can be criticised in the sense that there is no confirmation of correctness by the respondents. More commonly, other researchers have used so-called prompted recall surveys. These surveys involve asking respondents to check and if necessary correct the data imputed from the GPS traces, sometimes combined with the invitation to provide additional information that cannot be captured by GPS or other technology. In terms of reducing respondent error, one might claim that the combined use of GPS and a prompted recall requires less effort because respondents only need to check the imputed data and correct it, rather than provide full (out-of-home) activity-travel diaries. On the other hand, correcting errors in imputed schedules may sometimes be even more demanding.

2.7 In terms of assessing the quality of the inference process, information about the frequency and nature of corrections of the imputed journeys, stages and travel purposes (identified by differentiating between travel and activity episodes), can be used. The proportion of cases for which the mode of transport was imputed correctly and other facets of the schedule provide information about the validity of the adopted approach.

2.8 The NTS GPS pilot was designed such that it did not include a prompted validation survey using processed trip data, but instead used manual inspection of mapped traces against the imputed mode and purpose. There are several reasons why it was not appropriate to include a prompted validation survey:

- It is not currently feasible to do a validation survey at the NTS pick-up interview because the data will require some time to be downloaded and processed into such a system. If data could be downloaded and processed in time for use in the pick up interview, a face to face survey would not be feasible as it would require all respondents to be present (proxy responses would not be possible without obtaining permission from the absent respondent(s) and proxies are also unlikely to be able to provide reliable information about all journeys made by all absent household members).

- One of the most important quality measures on any survey is the response rate. The response rate for the GB NTS is currently approximately 60 per cent. Increasing respondent burden would be likely to have a negative impact on this rate and thereby data quality.

- To date, GPS travel survey studies that have employed prompted recall validation surveys have typically been small-scale academic studies: a web based post-validation survey is not going to be representative or practicable in a general population survey. Some 9.2 million UK adults have never used the internet and 27 per cent of UK households do not have an internet connection[5]. Introducing an additional follow-up survey using telephone methods would be prohibitively expensive. The annual sample size for the GB NTS is approximately 20,000 individuals, therefore any minor additions to survey length can quickly sum to major cost implications.

BENCHMARKS

2.9 In the following text, we provide an overview of results of previous research on this issue. These should serve as a benchmark for the results of the present study, although differences between studies do not only depend on the device used and the inference method, but are also influenced by the kind of environment in which observations were made. High rise, densely populated areas (sometimes referred to as urban canyons) are more likely to cause GPS devices to lose their satellite signal and thereby higher inaccuracies in the data recorded. The accuracy of inferred results also depends on the distribution of trips by transport modes because some transport modes are more difficult to infer than others.

2.10 Global Positioning Systems use ‘triangulation’ to three or more satellites orbiting the earth to determine the location of a device. In principle this technology can be adopted anywhere. The accuracy of the position increases with the number of satellites that are used to locate the device.

[5] Internet Access - Households and Individuals, 2010, Office for National Statistics, http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/historical-internet-access/internet-access-2010-households-and-individuals.pdf

2.11 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 in traditional surveys (e.g., Battelle, 1997; Hato and Asukari, 2001). Studies focusing most on the technical aspects of GPS devices (e.g., Stopher et al., 2003; Ohmori et al., 2006) drew attention to potential technology issues, such as the time required for the GPS device to locate the satellites, limited battery life, and errors due to human behaviour.

2.12 The number of trips has most commonly been identified by defining some arbitrarily set period of no movement, typically ranging from 30 seconds to 2 minutes (e.g., Schönfelder et al., 2002; Chung and Shalaby, 2005; Wolf, et al., 2004; Forrest and Pearson, 2005; Li and Shalaby, 2008; Bohte and Maat, 2008). If this time is too short, a stop at a traffic light, for example, may have been falsely detected as an activity episode; if it is too long short activities such as dropping off a child at school may have gone undetected. Reported accuracies regarding the number of trips typically vary between 85%-90%.

2.13 Some studies have also imputed transport mode and trip purpose. Most of these have relied on ad hoc rules, relating to speed and distance, while trip purpose has been typically imputed on the basis of (detailed) geo-spatial information systems (e.g., Wolf, et al., 2001). Unfortunately, geo-coded land use databases often lack accuracy as they are not well maintained. Even if the data are up to date, it may not be possible to differentiate between trip purposes. For example, some locations may be classified as having mixed use; that is, having commercial and residential uses. Consequently, some authors have instead used fuzzy approaches (e.g., Tsui and Shalaby (2006); Schuessler and Axhausen (2008). Results suggest that the success of the imputation of mode is higher for trains and cars, and lower for walking, running and bicycling. The imputation of trip purpose is typically regarded as being more difficult than mode allocation. Reported overall the proportion of correctly inferred trips (‘hit ratios’) for transport mode typically vary around 80%. However, hit ratios for car trips, may be as high as 95%; cycling and walking trip ratios vary between 65%-80%, with an exceptional 98% for walking trips reported by Tsui and Shalaby (2006). Reported hit ratios for trip purpose are rarely much higher than 70% and more commonly around 40%.

3. BASIC PRINCIPLES UNDERLYING TRACEANNOTATOR

Basics

3.1 The state-of-the-art system that was developed and used in the present study is based on the following principles:

> i Fuzzy logic or probabilistic systems are preferable to deterministic system to account for the inherent stochastic nature of the studied process, and
>
> ii Errors in the measurements should be considered in addition to commonly used variables.

3.2 The architecture of TraceAnnotator was originally designed for a scenario where respondents would ‘upload’ their GPS traces and be invited to participate in a web-based prompted recall to verify the imputed schedules. It was therefore important that they should receive the imputed schedules as quickly as possible. Moreover, because geo-coded data in the Netherlands are not widely available or very expensive, the basic system is based on the data that the GPS traces provide. The system was developed for data collection over a longer period of time than seven days. Therefore, a learning algorithm was implemented which uses the results of the prompted recall surveys as an input to training the system’s parameters, and over time the structure of the classification system. Because respondents are asked to report activity type and trip purpose at different locations, in principle a user-based geographic information base can be developed over time to further improve the accuracy of the trip purpose/activity type inference.

3.3 The TraceAnnotator system is a Bayesian Belief Network (BBN) or Bayesian classifier system, which replaces commonly ad hoc rules with a dynamic structure, leading to improved classification if consistent evidence is obtained over time from more samples (more traces). A BBN is a model for reasoning about uncertainty, which 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. The network is a graphical representation of probabilistic causal information through a directed acyclic graph and sets of probability tables behind them. The graph consists of nodes and arcs which represent discrete or continuous variables and causal/influential relationships between variables, respectively.

Architecture

3.1 TraceAnnotator has been configured such that data processing is divided into two main processes:

i The imputation of transport modes (and corresponding trips and stages) and activity episodes, where data imputation is established by using a BBN.

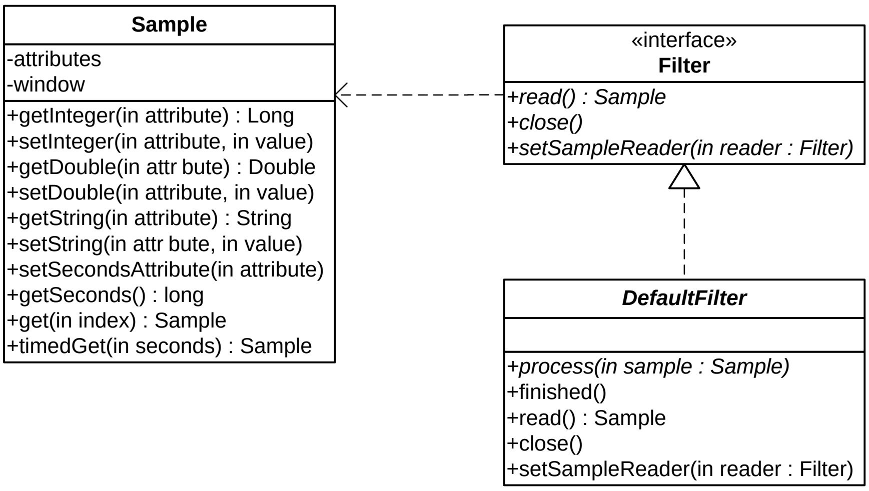
ii The imputation of activity type, where GPS data are fused with GIS and personalized land use data.

3.2 TraceAnnotator is written in Java and 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 BBN component used in the implementation of the ClassifierFilter (http://www.norsys.com).

3.3 The main two classes of TraceAnnotator are Sample and Filter (see Figure 3.1). A sample is one measurement from a GPS trace. It contains attributes such as date, time, latitude and longitude, distance, speed, etc. 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 3.1: Sample and Filter are the main two classes of the TraceAnnotator


Sample
-attributes
-window
+getInteger(in attribute) : Long
+setInteger(in attribute, in value)
+getDouble(in attr bute) : Double
+setDouble(in attribute, in value)
+getString(in attribute) : String
+setString(in attr bute, in value)
+setSecondsAttribute(in attribute)
+getSeconds() : long
+get(in index) : Sample
+timedGet(in seconds) : Sample
«interface»
Filter
+read() : Sample
+close()
+setSampleReader(in reader : Filter)
DefaultFilter
+process(in sample : Sample)
+finished()
+read() : Sample
+close()
+setSampleReader(in reader : Filter)

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 used. A sample can have a sample window, meaning that samples before or after the current sample can be accessed. By adding a second's attribute to the sample it is possible to get a window of samples specified by time. Using this, requests like 'all samples between -60 seconds to 60 seconds (in the future)' can be resolved.
3.4 The different transport modes and activity episodes cannot be distinguished without additional information such 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. 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. The filters can be divided into eight categories:

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

- ConvertToUnit converts 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, and
- TimeZoneFilter converts a given date and time from one time zone to another time zone. This filter also takes daylight saving time into account.

ii GIS Filters – filters that uses GIS information (like shape files):

- GISDistanceFilter calculates the distance from the position stored in the longitude and latitude attributes to the closest feature in a given shape file. This filter can for example calculate the distance between the given position and the railway track (GIS information about railway track is stored as a shape file).

- DropBadGISLoggerSamplesFilter checks if the latitude and longitude attributes have a valid value. This filter will only return samples for which those values are valid.

iii Windowed Filters – makes use of a window of samples. A window is used to lookup samples that are before or after the current sample. 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 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 calculates the average value of an attribute in the given window;
- MinMaxFilter calculates 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, and
- ModeFilter calculates the value of an attribute with the highest frequency in the given window.

iv Classifying Filters –can be used to classify a sample. For example:

- the NeticaFilter uses the BBN library Netica to classify samples. It is possible to specify which sample attribute values must be entered into which node of the BBN and into which attribute the result should be stored.

v Custom Filters – were developed to perform some very specific tasks:

- The ActivityCalendarFilter extracts activities and trips from a trace. This then 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 reverse Geo-code 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 geo-coding 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.

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

vii Output Filters – can be used to save sample back to a file or another resource. For example:

- The SampleOutputStreamFilter can be used to write a sample back to a comma separated file, and
- The SplitSampleOutputStreamFilter writes samples to a comma separated file. The filename will contain the date of the sample (derived from the seconds attribute). A new file will be generated containing the samples for each day.

viii Miscellaneous Filters - do not have their own category and do not fit well in the other categories.

3.5 Having described the functionality of the system, it may be helpful to emphasize some key differences between TraceAnnotator and other approaches. In particular, it should be evident that the system operates on two processes. The first process works at the sample or epoch level, whereby each sample is “Annotated” in terms of transport mode, activity episode, activity type, etc. in probabilistic terms using the BBN. Because these epoch are relatively sensitive to the adopted temporal resolution, errors in the GPS will impact the results. Therefore, these annotated epoch data are aggregated in a second process to derive journeys, stages and travel purposes.

4. NTS GPS PILOT STUDY DATA AND PROCESS DESCRIPTION

Overview

4.1 Four different kinds of data were collected and were therefore available for this project:

i GPS and accelerometer training data recorded for a non-random selection of trips using the GPS device and summarised in a supplementary record sheet;

ii GPS and accelerometer traces recorded during a seven-day travel week by NTS GPS pilot respondents;

iii Interview data for these pilot respondents, related to personal and household statistics and self-reported aspects of a set of activities, and

iv Several data sets about networks and land use in geo-coded formats.

Table 4.1: Overview of available data

• *Training data*	
Description	Samples of GPS traces recorded for different transport modes, plus corresponding trip diary
Contents	- 26 Excel sheets which were downloaded through the MGE tool, and 1 file recording the related diary - 1 dataset (in format of .MTA, .MTS, and .MTD), and 1 file recording the related diary
Source	Collected and sent by DfT and NatCen staff.
• *GPS and accelerometer traces*	
Description	GPS and accelerometer traces which need to be processed
Contents	- 874 datasets (in format of .MTA, .MTS, and .MTD) which relate to the traces of 874 people, one person can have multiple days of data (10 of them did not really include the trace data).
Source	Collected and sent by NatCen
• *Interview data*	
Description	Personal and household information
Contents	Data are organized and delivered in two waves, named as Feb and Mar, respectively. For each wave, there are data of: - Personal information - Household data - Home address - All addresses (except for home) - Vehicle information - Postcode
Source	Collected and delivered by NatCen.
• *GIS data*	
Description	Data of line and nodes of road, underground, public transport mode, land use etc.
Contents	Some data are in the format of matching GIS requirement, some are indirectly transferable. The list of data are: - ITN Layer (OS MasterMap Integrated Transport Network Layer, Road network and road routing information) - ITN PATHS GB (Reference of link and nodes) - London Underground (Line and stations of underground transport in London) - The National Public Transport Data Repository (NPTDR) SQL Backup (National Public Transport Data Repository, snapshot of public transport journey,) - OS MasterMap Topography Layer as GML (road, land use, etc.) - InterestMap GB (information of land use data) - National Public Transport Access Node (NaPTAN) (stations and stops of public transportation) - Meridian-2 (track line and stations of road, rail, etc.)
Source	Most were supplied by DfT, some were downloaded from online websites

Training data

4.2 The training data set related to 39 travel days of an opportunity sample. The data consisted of GPS and embedded accelerometer data, for particular modes of transport and types of journeys to create a broad specimen of trips. These included multi-mode commutes, long distance national rail trips to visit family and friends, short urban rail trips for shopping purposes, cycling and motorcycle trips, examples of people who run between home and work, long distance coach trips, and trips involving Newcastle Metro, London Underground, Docklands Light rail, and Sheffield trams to build up the light rail profile. In addition, a written description of the trips recorded in these traces was provided. Table 4.2 illustrates of such a travel record.

Table 4.2: example of a manual record kept for GPS device test data

Date	*Trip no*	*Start*		*Finish*		*Mode of trip stage*	*Purpose*
		Origin	*Time*	*Destination*	*Time*		
11/03	1	Private address (Home)	0930	Carshalton train station	0935	Car	Visit friends & family
		Carshalton train station	0940	Kings Cross St.Pancras train station	10.30	Train	
		Kings Cross St.Pancras train station	11.00	Newcastle Central Station	14.00	Train	
		Newcastle Central station	14.05	Gateshead Stadium Metro station	14.15	Tyne & Wear Metro	
		Gateshead Stadium Metro station	14.15	Private address	14.20	Walk	
	2	Private address	15.00	Metro Centre, Gateshead, NE11	15.15	Car	Shopping
	3	Metro Centre, Gateshead, NE11	16.40	Private address	16.55	Car	Visit friends & family
	4	Private address	17.00	Private address	17.25	Car	Visit friends & family

Note: postcodes were supplied for private addresses but these are not presented here for reasons of confidentiality.

4.3 This data set was used to extract some key statistics that were used in the imputation process to (partially) train the BBN. It also served to examine the inference quality of the network: how successful could the system classify/predict the transport modes observed in this data (see chapter 5). Because this data set did not contain explicit information about activity episodes, it could not be used to examine inference quality with regard to the non-travel related facets of activity-travel patterns. To compensate for this, a small sample of activity episodes from the NTS sample around the start and end time of a trip was extracted to improve the learning of the BNN. Moreover, the manual identification process also considered other important variables, like the number of satellites used to estimate a position, horizontal accuracy of the position estimate (metres), date and time, distance travelled since last point (metres), etc.

4.4 The BBN uses a small number of variables collected in the main pilot during the individual interview which flag ownership or regular access to cars, motorcycles and bicycles. The DfT and NatCen staff who provided training data were not interviewed; therefore these data were not available. To circumvent this, input data were set to be consistent with the training data. For example, if the transport mode recorded was motorcycle, then the dummy variable for ‘MOTORCYCLE’ was set equal to YES. To match those exceptions that someone who may travel by taxi or a hired car but does not really own a car, we randomly reversed the ownership characteristic for a small percentage of that mode. For instance, we assume 85% of car trips are conducted by a car driver (“CAR”=YES), while the other 15% of car trips relate to people who don’t have a car (“CAR”=NO).

GPS and accelerometer data

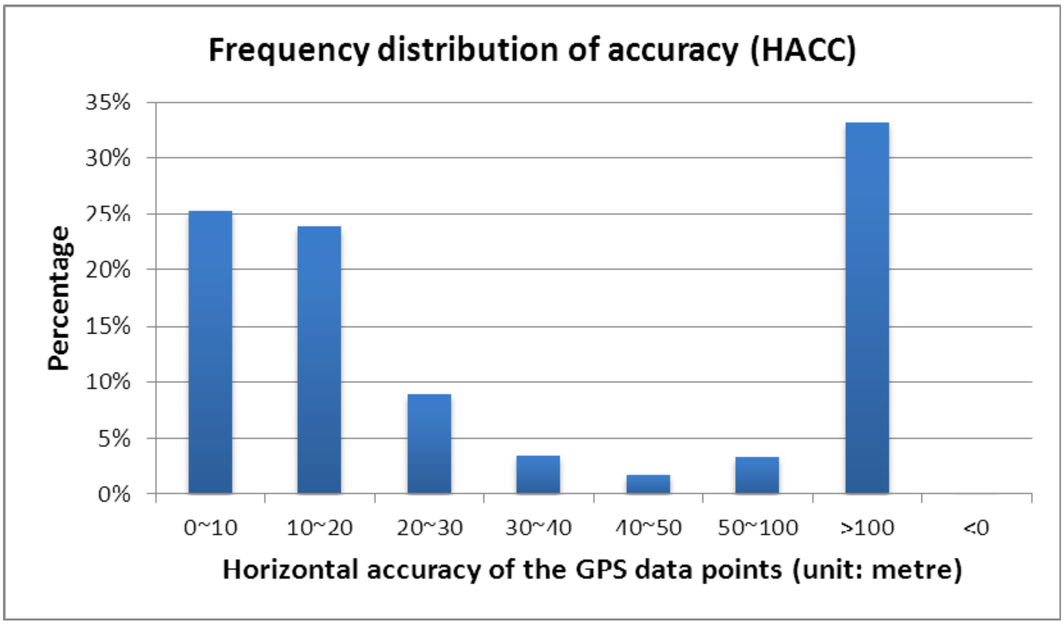
4.5 Data for the training sample and the main pilot was collected using an MGE Mobitest GSL accelerometer - equipped GPS. The devices were set to record GPS and accelerometer data once every second. These recording make up the so-called epoch level data.

Problems identified regarding the GPS data

Erroneous GPS points

4.6 Most studies filter out erroneous GPS measurements a priori by setting some filter. TraceAnnotator does not do this because such erroneous data may sometimes provide useful information as well. For example, a longer period of missing data may indicate travel by metro. Rather than filtering these out from the start, erroneous measurements are therefore "annotated" in the first phase of processing the GPS data. The GPS devices also record a number of measures which quantify the quality of the conditions under which the GPS information was calculated, the number of satellites used, and the measurement accuracy in horizontal and vertical dimensions are included in the BBN. It can happen frequently if there is no received good signals, such as urban canyon, inside building which are included to identify epochs where classifying travel episodes or activity episodes seem problematic. Depending on the length of such states over time, at the stage of merging consecutive epochs into episodes, the problem that epoch data cannot be classified due to missing signals may disappear. Figure 4.1 portrays the frequency of inaccurate recordings. It also shows that negative values were recorded in 39 traces .

Figure 4.1: horizontal accuracy of the GPS data points



Base: 103,848,420

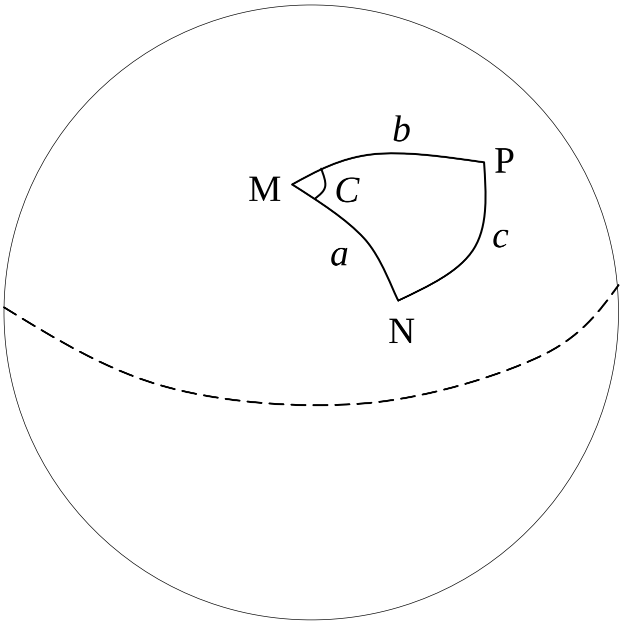
Missing distance and speed data

4.7 During the main fieldwork stage it was discovered that the default setting for the device meant that distance since the last position was only calculated by the device when the threshold for the horizontal accuracy measure (HACC) was within a range of 10 metres. For this reason, many zeros were found, making the distance measurements provided by the device not very useful. Similarly, it was also discovered that contrary to the order specification, the devices supplied were not programmed to record speed. Consequently, distance and speed, which are essential inputs to the inference system, had to be estimated. Because inaccurate GPS data often produced highly unrealistic distance and speed estimates, and the project time table did not allow for developing any advanced method, after trying and comparing several alternatives, the approaches described in the following paragraphs were used. However, some inaccuracies had to be remedied in the application of TraceAnnotator, as will be described later (Chapter 6)

4.8 Because of the surface of the earth, the distance between two points cannot be simply calculated as in Euclidean space. A more feasible algorithm which considers the circle of the sphere is required. We developed an algorithm based on Haversine formula (Sinnott, 1984) for calculating great-circle distances between two points on a sphere from their longitudes and latitudes. The Haversine equation is often used in navigation, giving great-circle distances between two points on a sphere from their longitudes and latitudes. It is a special case of a more general formula in spherical trigonometry, the law of Haversines, relating the sides and angles of spherical "triangles". Haversine' formula calculates the shortest distance over the earth's surface between the points, ignoring any hills, etc. Figure 4.2 shows these inherent relations: the three points (M, N and P) are connected by the great circle (a for M~N, b for M~P and c for N~P). The law of Harversines is formulated as follows:

$$haversin(c) = haversin(a - b) + sin(a)\ sin(b)\ haversin(C)$$

Figure 4.2 The law of the Haversines



Based on the law of Haversines states, the equations to calculate the distance between two points are:

$$d = R \cdot haver\sin^{-1}(h) = 2R \cdot \arcsin(\sqrt{h})$$

$$haver\sin(\theta) = \sin^2(\theta / 2) = (1 - \cos(\theta)) / 2$$

$$h = haver\sin(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2) haver\sin(\Delta\lambda)$$

where,
haversin is the Haversin function,
d is the distance between the two points;
R is the radius of the sphere,
φ$_1$ and *φ*$_2$ are the latitude of point 1 and point 2, and
Δλ is the longitude separation.
R was set as 6,371 km. Substituting Equation (2) into Equation (3), we get

$$h = \sin^2(\Delta\varphi / 2) + \cos(\varphi_1)\cos(\varphi_2)\sin^2(\Delta\lambda / 2)$$

where *Δφ* is the altitude separation.

4.9 The performance of this algorithm was examined using specific examples. Results suggest that the algorithm is quite robust in estimating distances from GPS data. Nevertheless, when the GPS data recorded are inaccurate, distances calculated this way may still be (highly) unrealistic. To correct for such values, any unrealistic distance calculation, based on typical speed of the transport mode, was in the merge stage imputed by calculating the distance of the last reliable epochs before the epochs with unrealistic distances/speeds.:

4.10 Speed was estimated using the following approach. As trace data was recorded for every second, we designed an algorithm to calculate the speed for any particular time period as it is not immediately clear what the best epoch would be. Based on reported experiences (mainly on the Internet) and our own experimentation, (average) speed was calculated for every three seconds based on accumulated distance. Results indicated that the three second epoch seems a good choice to compensate to some extent for deviations from real instantaneous velocity. The relevant equations for the calculation of speed are:

$$v_i = accudistance_i / (time_i - time_{i-3})$$

where,

v_i is the speed at time point i,
$acudistance_i$ is the accumulated distance from point i-3 to i.
$time_i$ and $time_{i-3}$ are the time at point i and point $i-3$. Thus,

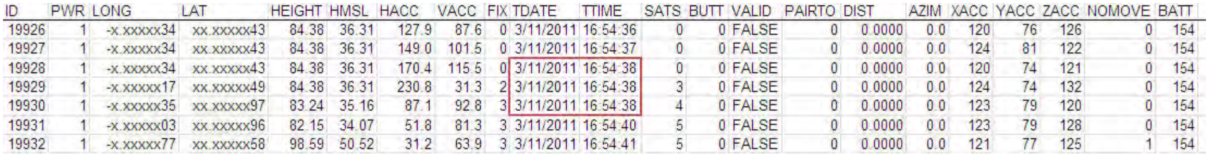
$$accudistance_i = dist_{i,i-1} + dist_{i-1,i-2} + dist_{i-2,i-3}$$

where $dist_{i,i-1}$, $dist_{i-1,i-2}$ and $dist_{i-2,i-3}$ are distances between two points in consecutive time.

Problems regarding the continuity of time recorded

4.11 Although the device is supposed to record the information on a second-by second basis, this was not always the case, as shown in Figure 4.3.

Figure 4.3: sample of raw data recorded by GPS device



ID	PWR	LONG	LAT	HEIGHT	HMSL	HACC	VACC	FIX	TDATE	TTIME	SATS	BUTT	VALID	PAIRTO	DIST	AZIM	XACC	YACC	ZACC	NOMOVE	BATT
19926	1	-x.xxxxx34	xx.xxxxx43	84.38	36.31	127.9	87.6	0	3/11/2011	16:54:36	0	0	FALSE	0	0.0000	0.0	120	76	126	0	154
19927	1	-x.xxxxx34	xx.xxxxx43	84.38	36.31	149.0	101.5	0	3/11/2011	16:54:37	0	0	FALSE	0	0.0000	0.0	124	81	122	0	154
19928	1	-x.xxxxx34	xx.xxxxx43	84.38	36.31	170.4	115.5	0	3/11/2011	16:54:38	0	0	FALSE	0	0.0000	0.0	120	74	121	0	154
19929	1	-x.xxxxx17	xx.xxxxx49	84.38	36.31	230.8	31.3	2	3/11/2011	16:54:38	3	0	FALSE	0	0.0000	0.0	124	74	132	0	154
19930	1	-x.xxxxx35	xx.xxxxx97	83.24	35.16	87.1	92.8	3	3/11/2011	16:54:38	4	0	FALSE	0	0.0000	0.0	123	79	120	0	154
19931	1	-x.xxxxx03	xx.xxxxx96	82.15	34.07	51.8	81.3	3	3/11/2011	16:54:40	5	0	FALSE	0	0.0000	0.0	123	79	128	0	154
19932	1	-x.xxxxx77	xx.xxxxx58	98.59	50.52	31.2	63.9	3	3/11/2011	16:54:41	5	0	FALSE	0	0.0000	0.0	121	77	125	1	154

4.12 Figure 4.3 shows that three identical times are recorded (16:54:38), but with different coordinate locations. When processing the traces with this problem, the first time stamp was assumed to be the correct one. For example, in calculating the average speed for one minute, we picked only the first case if there are multiple records with the same time slot.

4.13 Based on our experience with other devices, the share of missing information was substantial, not only in well-documented cases, but also for example when travel involved the train, when often there was not a good signal..

Accuracy of the recorded time

4.14 The accuracy of recorded time will influence how accurately the imputation model is able to predict mode of transport. Before we applied the training data to the BBN model, we inspected the raw data traces and matched the exact transport mode used at specific times according to the record sheet. When the training data was collected, the original raw trace data was recorded each second by the device, but the record sheet was typically completed after the journey, sometimes using approximate rather than exact times, so there were some deviations in the accuracy of the time. This is a general problem in self-reported data, because people may not be able to remember exactly what the start/end time of the travel is. Therefore, when matching the exact time with certain transport mode, in addition to the record sheet, we carefully checked other reasonable variables which helped us to identify the time and transport mode, i.e. instantaneous speed, distance, the accuracy of horizontal measurement (HACC), The number of satellites used to position the device (SATS), and the variable related to the accelerometer reading that determines whether the device is moving (NOMOVE).

Wrong date recorded

4.15 There were several instances where the GPS devices had recorded the wrong date. In total, 118,285 data points from a total of 103,848,420 epoch data (0.1 per cent), distributed among 842 trace files (there are 899 files) had wrong dates. This relatively often happened at the start of a trace. Examples of such dates are 1/7/1980, 1/10/1980, 1/7/2055 and so on. If the trace file includes multiple days of logged data in one file, it would normally include multiple data with wrong dates. It seems this problem tends to happen the first time the device is turned on, but this is not always the case. Such dates were corrected to maintain the consistency of the imputed data. .

Large longitude values

4.16 The values of the longitude and latitude should be in certain range and not zero for Great Britain. However, sometimes data contained unreasonable large values for longitude. More precisely, there were 181,089 unrealistic longitudes (0.2 per cent) in 98 trace files.To cope with this problem, we added the functionality in our software that before we process the data, we drop any unreasonable data from the resource. A condition was added, for example, if the value of longitude is larger than 100, this epoch was not processed. And if the values of longitude and latitude are both zeros, the epoch was not processed as well in the Bayesian network.

Missing recordings

4.17 The GPS traces showed a large proportion of epochs for which no data were recorded. This is in part due to the fact the accelerometer goes to sleep after 600 seconds of no movement. The lack of GPS data could be attributed to cold start, urban canyon etc. If the number of seconds of no recordings was larger than 500, the corresponding time was classified as an activity episode, used to detect travel purpose. Otherwise, the consecutive epochs of missing data were treated by merge rules. This happened in 12 per cent of the trips.

Person and household information

4.18 During the placement interview, data on person and household characteristics and frequent activities were collected. The most important person and household data used in the imputation concern the availability and use of various transport modes: car, motorcycle and bicycle. A script was written to reference the person/household and vehicle data via the household serial number and the person number.

4.19 To identify children, age was extracted from person data file. Respondents in the [12-15] age category were considered to be children.

4.20 Address data was collected for activities such as school, work, shopping etc. The data were then used for the imputation of travel purpose. More specifically, addresses for frequently visited locations of a variety of activities were used. Because the interview data only provided postcode and location name, geo-coding was needed to find the associated longitudes and latitudes. The ESRI shape format which provides the spatial data of postcodes for the whole United Kingdom was used for this.

GIS data

4.21 The imputation of activity type (trip purpose) also drew upon several GIS data bases which include data on various types of land use and points of interests. GIS data were also used, complementary to the GPS and accelerometer data, for inferring transport modes, such as train, light rail, bus and metro. Distance to stations and stops were used in the classification of transport mode. In particular, the following data were used:

i *Transport Infrastructure Data*

(a) Road network - Mastermap (http://www.ordnancesurvey.co.uk/oswebsite/products/os-mastermap)
This data set contains the Integrated Transport Network (ITN) which is essentially a line down the middle of all roads broken into toids with information on the type of road which can be used in conjunction with the other three layers to know more about the area around the road.

(b) Railway track + stations – Meridian 2 (http://www.ordnancesurvey.co.uk/oswebsite/products/meridian2)
This was considered to be the best available source for this kind of information, even though it is not entirely complete, there are occasionally some minor gaps in the route and a few stations are not present.

(c) Light rail + stations/stops - Meridian and Points of Interest database (see ii below for more on this).

(d) Underground - a track line and an underground station file for London Underground (NB does not include underground outside London). .

(e) Bus route + stops - NaPTAN and NPTDR. NaPTAN is better for bus stops but NPTDR also contains information on the order of the route between stops. Considering that bus lines coincide with roads lines, a decision was made to use only the bus stops data in the imputation. Therefore, distances to bus stations were calculated for each activity location. This derived data was also needed to detect and calculate waiting times,

(f) Ferry/boat route +stations - These data are available in both the NaPTAN and Meridian 2 datasets, but we used the NaPTAN data.

ii *Land Use and Points of interest data*

InterestMap GB (derived from the Ordnance Survey Points of Interest database) was used for this purpose.

4.22 Because these data sources were supplied in a range of different formats, these were first converted into a consistent format (ESRI Shape files). A common feature of GIS such data sets, is the uncertainty of the extent to which these data are up-to-date and/or complete.

4.23 Several of these sources use different coordinate systems. The MGEData GPS device uses the default coordinate system WGS84 which uses longitude and latitude information, while almost all the GIS data supplied used OSGB1936 or other similar GB coordinate systems, which use easting and northing measurements to label the spatial location. Thus we also had to transform all the GIS data into the WGS84 system. POI data were used to identify activity type. To that effect these data were prepared into shape files according to the list of activity purpose. Table 4.3 specifies which POI data were used for what specific activity type/trip purpose.

Table 4.3: use of Points of Interest data and personal profile data to code purpose of journey

Purpose from/ to	*Label*	*Purpose*	*Data Availability*	*Data Source*
1	Home	Home	Home address	PP
2	Work	Work	Work address	PP
3	Education	Education/Education escort	School, centre/association, etc.	POI: Education and health PP
4	Shoppinggroc	Food/grocery shopping	Food, drink and multi item retail, etc.	POI: Retail PP
5	shoppingother	All other types of shopping	Clothing and accessories, household, office, leisure and garden, motoring, etc.	POI: Retail PP
6	personalmedical	Personal business – medical	Hospital, dental surgeries, physical therapy, clinics, etc.	POI: Education and health PP
7	Personalother	Personal business – other	Consultancies, employment and career agencies, hire service, advertising, etc.	POI: Commercial services PP
8	Eatdrink	Eat/drink	Cafe, bar, restaurant, etc.	POI: Accommodation, eating and drinking PP
9	Entertainment	Entertainment/public social activities	Bodies of water, botanical and zoological, historical and cultural, landscape features, recreational, tourism	POI: Attractions PP
10	Sports	Sports	Gambling, outdoor pursuits, sport and entertainment support services, sports complex, venues, stage and screen.	POI: Sport and entertainment PP
11	Holiday	Holiday base	Camping, caravanning, mobile homes, holiday parks and centres, etc.	PP & POI: Accommodation, eating and drinking
12	Friend	Visiting family and friends	Friend address	PP
13	Other	Other type of purpose apart from purposes of 1 ~ 10		POI
14	Waiting	Wait at stations or bus stops	-	
999	Unknown	The purpose is unknown	-	

Note: PP: personal profile data; POI: point of interest data

4.24 In 73 per cent of the cases where purpose was allocated, the identification was based on the personal profile data; in the remaining 27 per cent, POI were needed.

Process description

4.25 The project involved the following steps:

i Data preparation, including

- converting all GIS data to the same format;
- estimating distance and speed data;
- estimating missing data for person and activity episodes;
- geo-coding of personal address data;

- selecting relevant travel days;
- linking person and GPS data sets, and
- processing all data to create input for the annotation.

ii Developing Bayesian belief model using the accelerometer data and comparing its performance to a model including both accelerometer and GPS data.

iii Updating and testing the BBN for inferred travel modes based on the training data set.

iv Inferring trip purpose for the same data set using land use and points of interest data.

v Preparing the data of personal profiles and personal addresses based on the interview data.

vi Applying the updated version of TraceAnnotator to infer transfer modes, trips and stages for the NTS pilot sample.

vii Developing and applying a framework to assess the validity/feasibility of the processes traces of the NTS pilot sample.

viii Post-processing of traces to identify (sub)sequences that show evidence of

- Waiting activity;
- Series of calls;
- Off-network travel;
- Round trip, and
- Children playing.

ix Converting results into requested data formats for transfer back to NatCen.

5. USE OF ACCELEROMETER DATA

Background

5.1 . It has been well-documented that problems arise when using GPS devices due to signal loss (for example, when travelling underground, under tree canopies or inside buildings), and urban canyons (highly uncertain data due to interception of signals, e.g. high rise buildings). Moreover, no imputation algorithm can be improved beyond the inherent variability in the variables used to differentiate between transport modes or activity types. For example, if in a particular environment a bus and car travel at exactly the same speed, and make exactly the same number of stops at exactly the same locations, an algorithm based on speed and acceleration will not be able to differentiate between these two transport modes.

5.2 In domains other than transport, accelerometer data have been used to identify the type of people's physical activity (e.g. Bao and Intille, 2004; Ravi et al., 2005). An accelerometer is a sensor that returns a real valued estimate of acceleration in relation to the device along the x, y and z axes. Accelerometers have been primarily used as motion detectors and for body-position and posture sensing (Ravi et al., 2005), especially in the context of identifying different types of physical activities (i.e., walking, running, sitting & relaxing, watching TV, scrubbing, brushing teeth, climbing, etc.). Studies have reported accuracies up to 80-90%, although most studies have been conducted in laboratory settings. Bao and Intille (2004) conducted an experiment using five accelerometers placed at different places of the body instantaneously to check the sensitivity of the accelerometer device for 20 activities. They found that the accelerometer attached to the thigh and wrist produced stable results. Accuracy ranged between 41 and 89% for activities involving movement.

Given these promising results of using accelerometer data for recognizing people's physical activities, some researchers realized their potential in the context of identifying transport modes. For example, Troped et al. (2008) used a combination of GPS and accelerometer data to discriminate between four types of what was called activity modes, but what would be called transport modes in the transport community (walking, jogging/running, bicycling, inline skating or driving a car). They found that imputation based on accelerometer data was correct in 89% of their cases, and this was increased to 93% when both types of data were used. However, only 61% of the driving minutes were correctly classified, and the combined used of these two data did not always result in better predictions, suggesting that the advantage of adding GPS to accelerometer monitoring, and vice versa, may depend on the type of analysis conducted and/or the purpose of the study. Similarly, Cooper et al. (2010) combined accelerometer and GPS data to investigate the level and location of physical activity of children walking to school. The mean values of the accelerometer data were used to identify differences among three transport modes (walking, car and bus). In a more recent study, Oliver et al. (2010) argued the merits of combining GPS, GIS and accelerometer data.

5.3 The combined use of GPS and accelerometer data might be especially relevant because the accelerometer does not depend on any signals from satellites, and hence might provide more stable data recordings. Moreover, accelerometer data might reinforce GPS data or be used in a complementary fashion when the accuracy of GPS data is substantially reduced. Accelerometer data has already been used in large-scale commercial applications to derive detailed information about individual's behaviour when travelling. For example, Ipsos MORI have used MGEData GPS devices to collect accelerometer data for the travel survey they undertake on behalf of Postar and subsequently investigated predicting features such as mode of travel[6].

5.4 In this section of the report we explain how the TraceAnnotator was used to compare the performance of three approaches based on the training data: using only GPS data, using only accelerometer data and using combined GPS and accelerometer data to identify transport modes. We then conclude which approach was most accurate in predicting the correct mode of travel and whether augmenting GPS data with accelerometer data improves the accuracy of such inferences. .

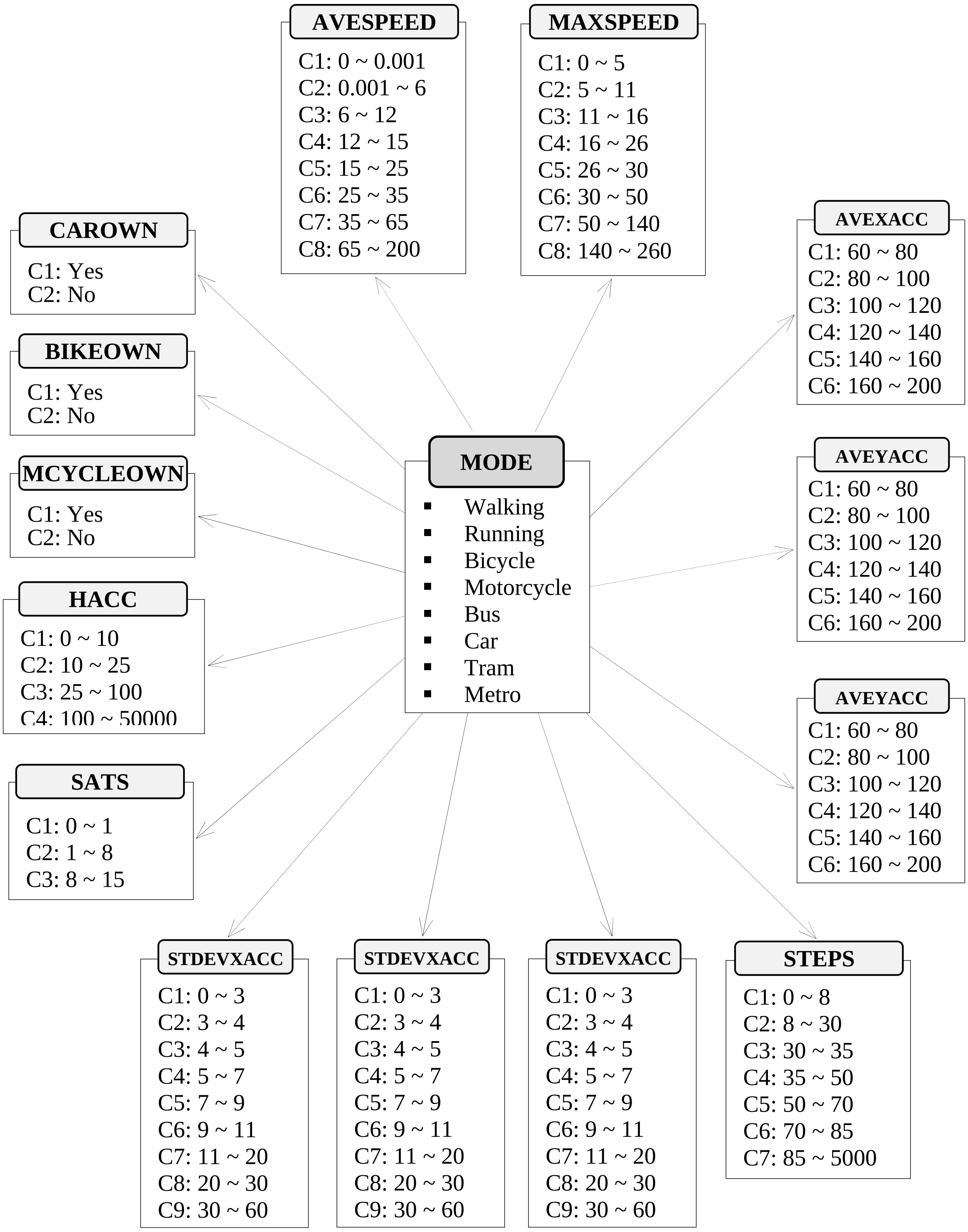
5.5 To better understand the potential of this approach; first the ability of accelerometer data to detect differences in transport modes was examined. To that effect we only incorporated a limited number of transport modes. In particular train journeys were excluded from this work as the GPS data was missing or incomplete for the majority of training data examples. This is likely to be due to overcrowding, the device being placed in a difficult position such as in a handbag under a table, the metal based coating applied to certain Train Operator's carriage windows which is known to block GPS signals. Next, a more full fledged approach which also included inference of the purpose of the journey was completed.

Analyses and results

5.6 Figure 5.1 shows the network structure that was used to infer the type of transport mode from the GPS traces. The output is the conditional probability that a particular type of transport mode has been used as a function of the states of the variables included in the BBN. The node MODE considers 8 different transport modes: walking, running, bicycle, motorcycle, bus, car, tram and metro.

[6] http://ipsos-rsl.com/researchspecialisms/ipsosmediact/whatwedo/postar.aspx

Figure 5.1: Model structure and classification settings for the inference of mode of transport in the comparative exercise



CAROWN: Yes if the respondent has a car, otherwise No; BYKEOWN: Yes if the respondent has a bicycle, otherwise No; MCYCLEOWN: Yes if the respondent has a motorcycle, otherwise No; HACC: estimated horizontal measurement error, *m*; SATS: number of satellites used for position calculation; AVEXACC: average value of X-axis acceleration change; AVEYACC: average value of Y-axis acceleration change; AVEZACC: average value of Z-axis acceleration change; STDEVXACC: standard deviation of X-axis acceleration change; STDEVYACC: standard deviation of Y-axis acceleration change; STDEVZACC: standard deviation of Z-axis acceleration change; STEPS: the average time duration of the device not moving in one minute; AVESPEED: the average speed in every three minutes, *km*/*h*; MAXSPEED: the maximum average speed, *km*/*h*.

5.7 The variables (nodes) included in the network are derived from the raw files extracted from the GPS device. These raw files consist of two types of data: GPS data and accelerometer data. The GPS data provide basic information about coordinates, date and time, accuracy measurements of the device and other information such as distance at every second (although as previously noted, this was only recorded where horizontal accuracy of a position was within a range of 10 metres). The accelerometer data provides information about the change in acceleration on three-axis with respect to the device, moving or non-moving of the device, and state of the device (turned off, sleeping, etc.). To feed the imputation model, we first generated some statistical variables (averaged) on the scale of three minutes based on the second-by-second data. We used two variables for measuring the speed pattern, AVESPEED and MAXSPEED which are the average speed and maximum speed for the three minutes epoch. The additionally generated variables related to the accelerometer data are non-moving time duration (STEPS), average value and standard deviation of the three-axis acceleration change (AVEXACC, AVEYACC, AVEZACC, STDEVXACC, STDEVYACC and STDEVZACC).

5.8 Since the accelerometer records the information about movements by an inherent variable (NOMOVE) in the device, if the device does not sense movement, the value is automatically increased by one per second. This information is considered important to differentiate traveling activities with similar speeds but different motions, i.e., running and cycling. To use this variable properly, we created a variable equal to the average value of time duration (STEPS). The higher the STEPS value is, the more random the motion of traveling becomes. Seven states were specified after carefully examining the distributional frequencies of the sample, ranging from C1 (not much random movement) to C7 (high levels of random movement).

5.9 In addition to the data extracted from the GPS device, the network also incorporates the effects of personal characteristics on the imputation of transport mode. Three variables (CAROWN, BIKEOWN and MCYCLEOWN) were used to describe whether the respondent has a car, a bicycle or a motorcycle. On case of the training data, where no person information was collected, these data were set consistent with the transport mode data. In case of the large sample, these input variables were explicitly collected from the personal profiles.

5.10 In addition, two precision related variables of the device were included: HACC and SATS. HACC is the key variable of the device for accuracy control. That means specific data such as distance from last recorded GPS point (DIST) will not be recorded if the measured HACC value is greater than the threshold. In our data collection, the threshold value was set as 10 meters by default.

5.11 The BBN also requires the states of the input and output variables to be classified. In this case, the initial conditional probability tables and the assumed relationships between the variables are based on either expert judgments or on the processing of a small set of GPS traces. Special emphasis on the setting of classifications is also given to process these 'invalid' data (the number of satellites, quite large value of HACC, etc.). For instance, the samples without any satellites should in general have a zero value for speed, but still have the instantaneous accelerometer data.

5.12 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. The example shown in Table 5.1 is the conditional probability of the variable 'Mode' by the seven levels of the variable 'STEPS'. One can see from Table 5.1 that, running behavior has the highest score (96%) associated with the top level 'C7' - which indicates that the person carrying the device is moving in a highly random or 'jerky' fashion - followed by walking (70%) and cycling (66%). This is understandable in that running involves continuous movement with many random variations. On the other hand, the tram has the highest probability at the minimum level of movement 'C1' (49%), followed by metro (28%) and bus (11%). These are notably modes of transport where the individual is most likely to be sitting passively. It should be noted that the distribution of car mode seems flat. This might be accounted for by the different places where people put their device.

Table 5.1: Conditional probability table of STEPS variable

%	*C1*	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*
Walking	0%	0%	0%	2%	9%	18%	70%
Cycling	0%	1%	0%	1%	6%	26%	66%
Running	0%	0%	0%	0%	0%	3%	96%
Motorcycle	0%	0%	0%	7%	34%	53%	6%
Bus	11%	37%	8%	16%	9%	17%	2%
Car	4%	10%	3%	18%	35%	23%	6%
Metro	28%	46%	7%	9%	7%	2%	1%
Tram	50%	19%	2%	12%	17%	1%	0%

Base = 52,421 training data

Assessment of imputation models

5.13 Table 5. 2 sets out the variables used in each of the three models being assessed. The GPS-only model includes the nodes only related to the GPS traces (AVESPEED and MAXSPEED), while the Accelerometer-only model excludes the speed related data and only incorporates accelerometer variables. To assess the model performance with combined data, all related variables are combined in the last model.

Table 5.2: Model structure and input variables for the three models

Models	*Contents*	*Variables*
1	GPS-only	AVESPEED, MAXSPEED, CAROWN, BIKEOWN, MCYCLEOWN, HACC, SATS, MODE
2	Accelerometer-only	STEPS, STDEVXACC, STDEVYACC, STDEVZACC, AVEXACC, AVEYACC, AVEZACC, CAROWN, BIKEOWN, MCYCLEOWN, HACC, SATS, MODE
3	GPS-and-Accelerometer	AVESPEED, MAXSPEED, STEPS, STDEVXACC, STDEVYACC, STDEVZACC, AVEXACC, AVEYACC, AVEZACC, CAROWN, BIKEOWN, MCYCLEOWN, HACC, SATS, MODE

5.14 The data used in this section to assess the three models comprises of the small training data set described in paragraphs 4.2 - 4.4. After processing the data by excluding the trips for which modes were not included in the comparative model (for example, rail), the data consisted of 80,670 data records. This dataset was further divided into 65% and 35% respectively for the purpose of model calibration and validation. Thus, the number of data records used for calibration and validation were respectively 52,424 and 27,630. The performance of the models was assessed in terms of accuracy (hit ratio, i.e. the percentage of corrected predicted classes). In addition, a confusion matrix, describing how misclassified cases were assigned, was constructed. Thus, the main diagonal of this matrix lists the proportion of correctly classified cases, while off-diagonal elements describe the proportion of incorrect imputations. It should be noted that the comparison required some operational decisions. For example, typically the start and end times in the verbal descriptions of the training data often were rounded-off. Comparisons were based on most closely matching times.

5.15 Table 5.3 present the results for the first comparison, which focuses on transport modes only and excludes rail travel as there were too many missing GPS data for that mode. It illustrates that the accuracy of all models for the calibration data is higher than for the validation data. Taking the calibration based models as an example; the accelerometer-only model achieves a higher precision (96%) than the GPS-only model (81%). When comparing the GPS and accelerometer model to the accelerometer-only model the accuracy is increased by less than one percentage point for the calibration data models and three percentage points for the less accurate validation data based models.

Table 5.3: Results for hit ratios of the three models (proportion of epochs for which mode was correctly identified)

Model	*Calibration data*	*Validation data*
GPS-only	81%	75%
Accelerometer-only	96%	82%
GPS-and-Accelerometer	96%	85%

5.16 Table 5.4 shows the confusion matrix for the model based on Accelerometer-data-only. It shows that the percentage of correctly predicted transport mode is higher than 90% for all modes except for bus (89%) and tram (84%). In case of the bus, eight per cent of the cases (epochs) is misclassified as walking and two per cent as tram. In case of the tram six per cent and ten per cent of tram modes are misclassified as bus and walking.

Table 5.4: Confusion matrix of model Accelerometer-only for validation data

	Predicted							
Actual	*Walking*	*Bicycle*	*Running*	*Motorcycle*	*Bus*	*Car*	*Metro*	*Tram*
Walking	96%	0%	0%	0%	1%	3%	1%	0%
Bicycle	2%	94%	0%	0%	0%	4%	0%	0%
Running	0%	2%	98%	0%	0%	0%	0%	0%
Motorcycle	0%	0%	0%	100%	0%	0%	0%	0%
Bus	8%	0%	0%	0%	89%	0%	1%	2%
Car	2%	1%	0%	0%	0%	95%	1%	0%
Metro	2%	0%	0%	0%	0%	0%	97%	0%
Tram	10%	0%	0%	0%	6%	0%	0%	84%

5.17 The results of the combined accelerometer-GPS data model are reported in Table 5.5. The main difference from the previous model is the inclusion of speed related data into the model. It is expected that speed information may allow further discrimination between transport modes. This seems to be supported by the results: Table 5.5 shows a substantial improvement in hit rates for all transport modes.

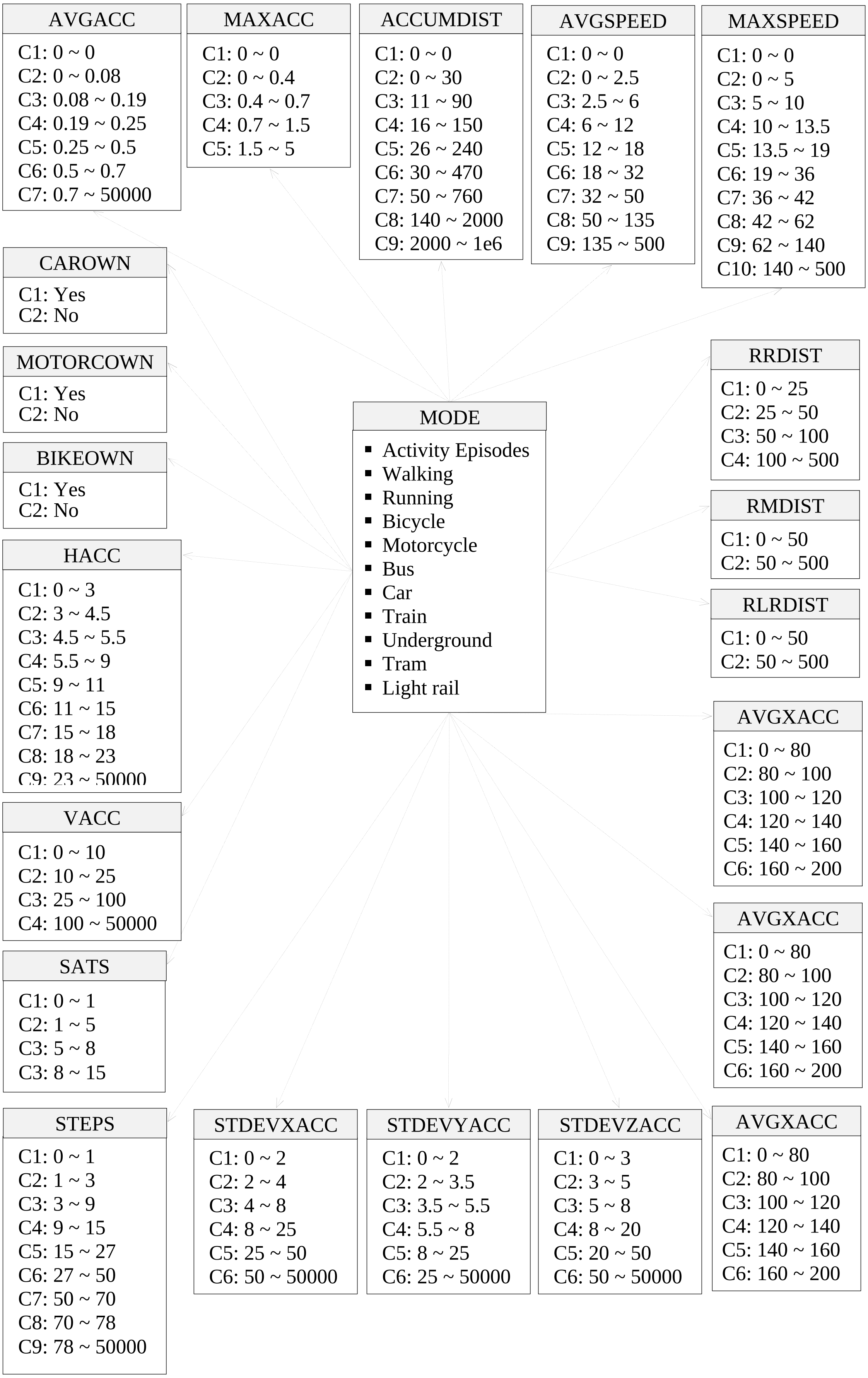
Moreover, the misclassifications between walking, bus and tram in this model also show significant improvements. The percentage of tram epochs misclassified as walking decreased from 10% to six per cent, while the misclassification of bus epochs as walking fell from eight per cent to four per cent.

Table 5.5: Confusion matrix of model GPS-and-Accelerometer for validation data

	Predicted							
Actual	*Walking*	*Bicycle*	*Running*	*Motorcycle*	*Bus*	*Car*	*Metro*	*Tram*
Walking	98%	0%	0%	0%	1%	1%	0%	0%
Bicycle	2%	95%	0%	0%	0%	3%	0%	0%
Running	0%	2%	98%	0%	0%	0%	0%	0%
Motorcycle	1%	0%	0%	99%	0%	0%	0%	0%
Bus	4%	0%	0%	0%	91%	0%	0%	5%
Car	2%	0%	0%	0%	0%	97%	1%	0%
Metro	0%	0%	0%	0%	3%	0%	95%	2%
Tram	6%	0%	0%	0%	6%	0%	0%	88%

5.18 These results pertain only to the application of the BBN to infer transport modes. The ultimate application is however more complex in the sense that the network is also used to detect trip purpose. This is done by detecting activity episodes and differentiating these from travel episodes. Consequently, activity episodes may be misclassified as travel episodes and vice versa and in turn this may affect the inference of transport modes. Hence, to examine the performance of the BBN in this more complex situation, an extended network, including activity type (trip purpose) and train, was developed and assessed. Figure 5.2 shows the structure of this model. Moreover, some unrealistic data were filtered out.

Figure 5.2 Revised model structure and classification settings for the inference of conditional table



AVGACC
C1: 0 ~ 0
C2: 0 ~ 0.08
C3: 0.08 ~ 0.19
C4: 0.19 ~ 0.25
C5: 0.25 ~ 0.5
C6: 0.5 ~ 0.7
C7: 0.7 ~ 50000

MAXACC
C1: 0 ~ 0
C2: 0 ~ 0.4
C3: 0.4 ~ 0.7
C4: 0.7 ~ 1.5
C5: 1.5 ~ 5

ACCUMDIST
C1: 0 ~ 0
C2: 0 ~ 30
C3: 11 ~ 90
C4: 16 ~ 150
C5: 26 ~ 240
C6: 30 ~ 470
C7: 50 ~ 760
C8: 140 ~ 2000
C9: 2000 ~ 1e6

AVGSPEED
C1: 0 ~ 0
C2: 0 ~ 2.5
C3: 2.5 ~ 6
C4: 6 ~ 12
C5: 12 ~ 18
C6: 18 ~ 32
C7: 32 ~ 50
C8: 50 ~ 135
C9: 135 ~ 500

MAXSPEED
C1: 0 ~ 0
C2: 0 ~ 5
C3: 5 ~ 10
C4: 10 ~ 13.5
C5: 13.5 ~ 19
C6: 19 ~ 36
C7: 36 ~ 42
C8: 42 ~ 62
C9: 62 ~ 140
C10: 140 ~ 500

CAROWN
C1: Yes
C2: No

MOTORCOWN
C1: Yes
C2: No

BIKEOWN
C1: Yes
C2: No

MODE
- Activity Episodes
- Walking
- Running
- Bicycle
- Motorcycle
- Bus
- Car
- Train
- Underground
- Tram
- Light rail

RRDIST
C1: 0 ~ 25
C2: 25 ~ 50
C3: 50 ~ 100
C4: 100 ~ 500

RMDIST
C1: 0 ~ 50
C2: 50 ~ 500

RLRDIST
C1: 0 ~ 50
C2: 50 ~ 500

HACC
C1: 0 ~ 3
C2: 3 ~ 4.5
C3: 4.5 ~ 5.5
C4: 5.5 ~ 9
C5: 9 ~ 11
C6: 11 ~ 15
C7: 15 ~ 18
C8: 18 ~ 23
C9: 23 ~ 50000

VACC
C1: 0 ~ 10
C2: 10 ~ 25
C3: 25 ~ 100
C4: 100 ~ 50000

SATS
C1: 0 ~ 1
C2: 1 ~ 5
C3: 5 ~ 8
C3: 8 ~ 15

AVGXACC
C1: 0 ~ 80
C2: 80 ~ 100
C3: 100 ~ 120
C4: 120 ~ 140
C5: 140 ~ 160
C6: 160 ~ 200

AVGXACC
C1: 0 ~ 80
C2: 80 ~ 100
C3: 100 ~ 120
C4: 120 ~ 140
C5: 140 ~ 160
C6: 160 ~ 200

STEPS
C1: 0 ~ 1
C2: 1 ~ 3
C3: 3 ~ 9
C4: 9 ~ 15
C5: 15 ~ 27
C6: 27 ~ 50
C7: 50 ~ 70
C8: 70 ~ 78
C9: 78 ~ 50000

STDEVXACC
C1: 0 ~ 2
C2: 2 ~ 4
C3: 4 ~ 8
C4: 8 ~ 25
C5: 25 ~ 50
C6: 50 ~ 50000

STDEVYACC
C1: 0 ~ 2
C2: 2 ~ 3.5
C3: 3.5 ~ 5.5
C4: 5.5 ~ 8
C5: 8 ~ 25
C6: 25 ~ 50000

STDEVZACC
C1: 0 ~ 3
C2: 3 ~ 5
C3: 5 ~ 8
C4: 8 ~ 20
C5: 20 ~ 50
C6: 50 ~ 50000

AVGXACC
C1: 0 ~ 80
C2: 80 ~ 100
C3: 100 ~ 120
C4: 120 ~ 140
C5: 140 ~ 160
C6: 160 ~ 200

5.19 Table 5.6 suggests that the performance of all three models for the calibration data has been slightly reduced (this compares to Table 5.3). On the other hand, the performance of the models has been improved to the extent that there is now little difference between the models using calibration data and validation data; as one might expect. The accelerometer-only model correctly identifies transport modes for 89% of the calibration data and 89% of the validation data. Table 5.6 shows that over all data, the GPS-only model has a lower level of prediction accuracy than the accelerometer-only model.

Table 5.6 Results of error rate of the three models based on filtered data (proportion of epochs for which mode was correctly identified)

Model	Calibration data	Validation data
GPS-only	78.5%	78.4%
Accelerometer-only	88.9%	88.8%
GPS-and-Accelerometer	91.7%	91.7%

5.20 Interestingly, classifying both activity episodes and travel episodes seems to have positively influenced the transport mode-specific hit ratios (Table 5.7) all are higher than they were before (compared to Table 5.4 and Table 5.5).Train is a relatively difficult transport mode to predict: only 83% of epochs were classified correctly.

Table 5.7 Correctly identified hit ratios by activity type based on filtered data

	Accelerometer Only	*GPS Only*	*Combined Accelerometer and GPS*
Activity	33%	84%	83%
Walking	92%	97%	98%
Running	97%	98%	100%
Cycling	88%	100%	100%
Bus	78%	87%	98%
Motorcycle	100%	100%	100%
Car	93%	98%	99%
Train	89%	58%	83%
Metro	86%	98%	99%
Tram	83%	98%	99%
Light rail	98%	98%	99%

5.21 It should be emphasised that these overall error rates (hit ratios) depend strongly on the frequency of observed transport modes in the training data. Because in the present context, these data were collected for specific modes and types of trip, the training data are not a random sample of trips and are therefore not representative. Table 5.7 shows that the hit ratio of the BBN based on the GPS traces is higher than the hit ratio for the accelerometer-only based model for all transport modes, except for train. However, the over-representation of train trips in the training data and the fact that GPS element was missing for many of the train data due to signal issues, the results of the accelerometer-only-model for all data points are better than the GPS-only model (Table 5.6).

5.22 As shown in Table 5.8, there is a considerable overlap between activity episode and train (35%). This indicates that motion patterns inside and outside the train seem similar. It makes a model based on just accelerometer data problematic. The confusion matrix also suggests problems in differentiating between walking and activity episodes and between walking and bus. Most important, however, is the finding that the model based on both types of data improves the results for most transport modes, without simultaneously reducing the hit ratios for other transport modes.

Table 5.8 Confusion matrix of the Accelerometer-only model based on filtered data

	Activity	*Walking*	*Running*	*Cycling*	*Bus*	*Motorcycle*	*Car*	*Train*	*Metro*	*Tram*	*Light rail*
Activity	33%	18%	0%	0%	8%	0%	4%	35%	0%	2%	0%
Walking	4%	92%	0%	1%	1%	0%	2%	1%	0%	0%	0%
Running	0%	2%	97%	0%	0%	0%	0%	1%	0%	0%	0%
Cycling	1%	4%	0%	88%	0%	0%	6%	0%	0%	0%	0%
Bus	7%	15%	0%	0%	78%	0%	1%	1%	0%	0%	0%
Motorcycle	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Car	1%	3%	0%	2%	0%	0%	93%	1%	0%	0%	0%
Train	3%	1%	0%	0%	0%	0%	0%	89%	7%	0%	0%
Metro	0%	0%	0%	0%	0%	0%	0%	14%	86%	0%	0%
Tram	3%	1%	0%	0%	0%	0%	0%	13%	0%	83%	0%
Light rail	1%	0%	0%	0%	0%	0%	0%	2%	0%	0%	98%

Table 5.9 Confusion matrix of the GPS Only model based on filtered data

	Activity	*Walking*	*Running*	*Cycling*	*Bus*	*Motorcycle*	*Car*	*Train*	*Metro*	*Tram*	*Light rail*
Activity	84%	4%	0%	0%	0%	0%	1%	9%	2%	0%	0%
Walking	2%	97%	0%	0%	1%	0%	0%	0%	0%	0%	0%
Running	0%	0%	98%	0%	1%	0%	1%	0%	0%	0%	0%
Cycling	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
Bus	1%	0%	0%	0%	87%	0%	0%	0%	0%	12%	0%
Motorcycle	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Car	0%	0%	0%	0%	1%	0%	98%	0%	0%	0%	1%
Train	0%	0%	0%	0%	0%	0%	5%	58%	36%	0%	0%
Metro	1%	0%	0%	0%	0%	0%	0%	1%	98%	0%	0%
Tram	0%	0%	0%	0%	0%	0%	2%	0%	0%	98%	0%
Light rail	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	98%

Table 5.10 Confusion matrix of the GPS and Accelerometer model based on filtered data

	Activity	*Walking*	*Running*	*Cycling*	*Bus*	*Motorcycle*	*Car*	*Train*	*Metro*	*Tram*	*Light rail*
Activity	83%	5%	0%	0%	0%	0%	1%	3%	8%	0%	0%
Walking	1%	98%	0%	0%	1%	0%	0%	0%	0%	0%	0%
Running	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Cycling	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
Bus	1%	0%	0%	0%	98%	0%	0%	1%	0%	0%	0%
Motorcycle	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%
Car	0%	0%	0%	0%	1%	0%	99%	0%	0%	1%	0%
Train	0%	0%	0%	0%	0%	0%	2%	83%	15%	0%	0%
Metro	1%	0%	0%	0%	0%	0%	0%	1%	99%	0%	0%
Tram	0%	0%	0%	0%	0%	0%	0%	1%	0%	99%	0%
Light rail	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	99%

Conclusions and discussion

5.23 In a previous chapter, we reported that the existing literature reported hit ratios around 80% on average and between 65%- 80% for slower modes of transport. Taking these as a benchmark, it can be concluded that the developed BBN, based on the combination of accelerometer and GPS data, exceeds these benchmarks. This developed model was then used in the application to the NTS data.

6. APPLICATION TO NTS GPS PILOT DATA

6.1 The general purpose system, ‘TraceAnnotator’, which has been developed by the team to process (semi-)automatically multi-day GPS traces was adjusted and extended as described in the previous chapter and subsequently applied to the NTS GPS pilot survey to:

- infer transport modes and activity-travel episodes and stages, using the BBN, combining GPS and accelerometer data, and

- infer activity type/trip purpose, fusing GPS data with GIS land use data and personal data.

6.2 The system was used in two stages of the project: first to develop the BBN for the combined GPS and accelerometer data and test its validity against the training data, and then to process the data from the pilot. Figure 6.1 illustrates how this tool was used to derive the required information after preparing and cleaning the data.

6.3 The application itself involves the following phases:

- Phase 1: Data cleaning and pre-processing
- Phase 2: Annotating the GPS traces
- Phase 3: Deriving patterns from the epoch-level annotated GPS traces

Phase 1: Data cleaning and pre-processing

6.4 GPS data come with different types of noise. In the present study, there were an unexpected high number of occurrences of incorrect data in the GPS files such as wrong dates, impossible coordinates, identical epochs, inaccurate data etc (as discussed in paragraphs 4.11-4.16). Incorrect dates, identical epochs and impossible coordinates were filtered out in a pre-processing stage. Some devices contained data from before/after the travel week (the seven day period starting the day after the interview) because the device was already switched on before the week started or was left on after the week ended. Such surplus data were filtered out before processing the traces. Other noise was annotated in the processing of the traces and therefore was taken into account in inferring transport modes and activities. Such cases were handled in the process of merging epochs into travel and activity episodes.

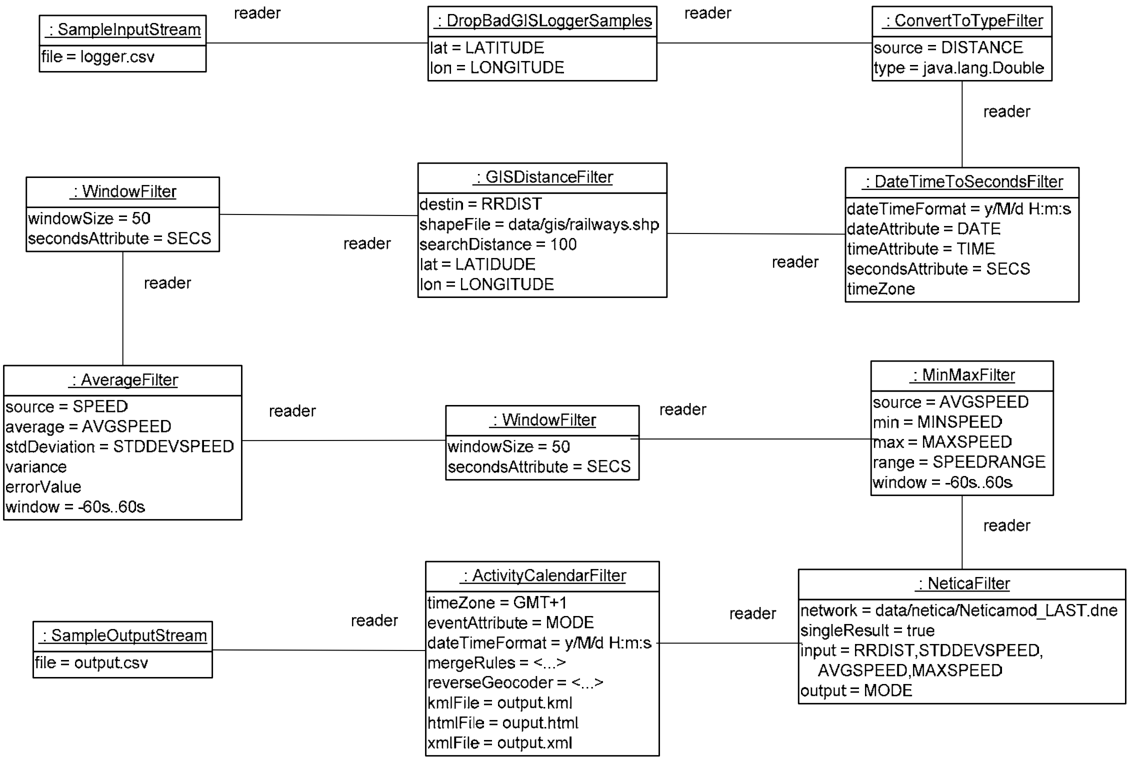
This phase also entails selecting and combining data other than GPS traces that is used in the imputation. Examples include availability of transport modes from the person file.

Phase 2: Annotating GPs traces

Step 1: preparing input data

6.5 TraceAnnotator, as shown in Figure 6.1 processes each data recording by applying a series of filters. First, it reads the epoch data from the logger.csv file. These epoch data refer to the second-by-second recordings of the GPD device. . Then the ConvertToType filter was used to remove the unit from the DISTANCE attribute and convert it into a Double data type, making further processing faster. The next filter uses the DATE and TIME attributes to construct a SECS attribute. This attribute refers to every second of the traces. Then, the GISDistanceFilter takes the latitude and longitude and search in the given shape file for the closest object within the search distance. A threshold distance of 200 meter was used. The resulting distance is then stored in the destination attribute. After that, a series of filters is used to calculate additional variables that appear in the BBN. The WindowFilter assigns a window to the epoch that references 50 epochs in the future and 50 epochs in the past. Average speed and deviation for the average speed over the window of -60 seconds to 60 seconds are calculated, creating values for the AVGSPEED and STDDEVSPEED attributes. Different values were tried to produce smooth input data. The MinMaxFilter was used to calculate the minimum and maximum values of the AVGSPEED attribute.

Figure 6.1: Process model



: SampleInputStream
file = logger.csv

: DropBadGISLoggerSamples
lat = LATITUDE
lon = LONGITUDE

: ConvertToTypeFilter
source = DISTANCE
type = java.lang.Double

: DateTimeToSecondsFilter
dateTimeFormat = y/M/d H:m:s
dateAttribute = DATE
timeAttribute = TIME
secondsAttribute = SECS
timeZone

: GISDistanceFilter
destin = RRDIST
shapeFile = data/gis/railways.shp
searchDistance = 100
lat = LATIDUDE
lon = LONGITUDE

: WindowFilter
windowSize = 50
secondsAttribute = SECS

: AverageFilter
source = SPEED
average = AVGSPEED
stdDeviation = STDDEVSPEED
variance
errorValue
window = -60s..60s

: WindowFilter
windowSize = 50
secondsAttribute = SECS

: MinMaxFilter
source = AVGSPEED
min = MINSPEED
max = MAXSPEED
range = SPEEDRANGE
window = -60s..60s

: NeticaFilter
network = data/netica/Neticamod_LAST.dne
singleResult = true
input = RRDIST,STDDEVSPEED,
AVGSPEED,MAXSPEED
output = MODE

: ActivityCalendarFilter
timeZone = GMT+1
eventAttribute = MODE
dateTimeFormat = y/M/d H:m:s
mergeRules = <...>
reverseGeocoder = <...>
kmlFile = output.kml
htmlFile = ouput.html
xmlFile = output.xml

: SampleOutputStream
file = output.csv

Step 2: Bayesian inference of activity episodes and transport modes

6.6 In the next step, these variables were used as inputs to the BBN to derive trips and trip stages and classify these into transport modes. The network calculates expected probabilities of different transport modes, given the structure of the BBN and associated conditional probabilities as input. Technically, the network used is a naïve Bayesian classifier.

6.7 The structure of the network is given in Figure 6.2. It shows that child nodes in this network concern distance to tram and light rail, distance to road, average and maximum acceleration, average speed, max speed, deviation from average speed, accumulated distance during every three minutes, the number of satellites that the GPS device used, the estimated measurement error in horizontal and vertical dimension (HACC and VACC), the average number of non-movements of accelerometer data, the average and deviations of acceleration change if the device is moving in three axis.

6.8 In addition to these variables, provided by the GPS tracers, variables that related to personal information such as availability of car, bicycle and motorcycle were used. The availability of this information allows one to improve imputation accuracy of used transport modes.

Figure 6.2: Structure of the Bayesian belief network

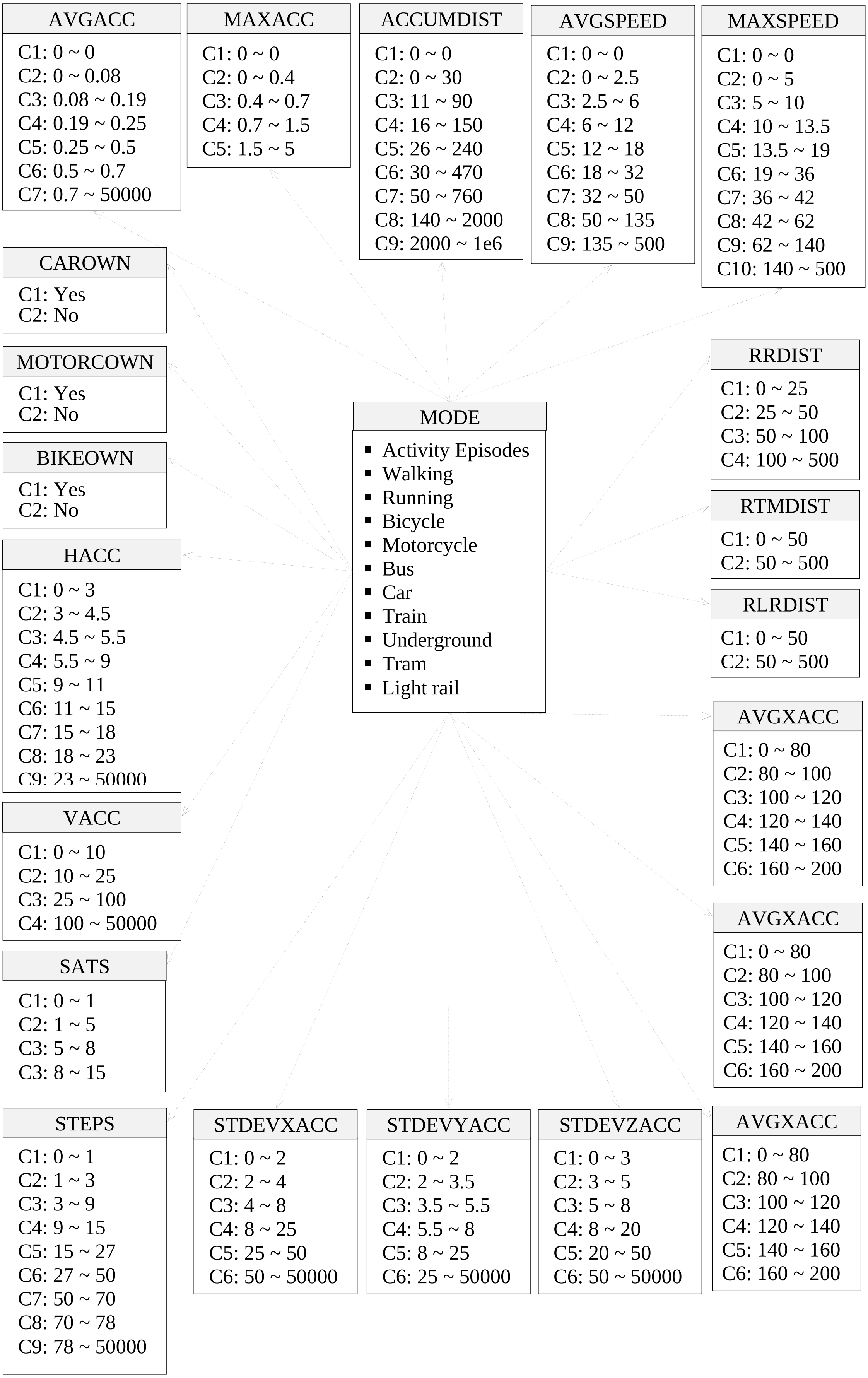
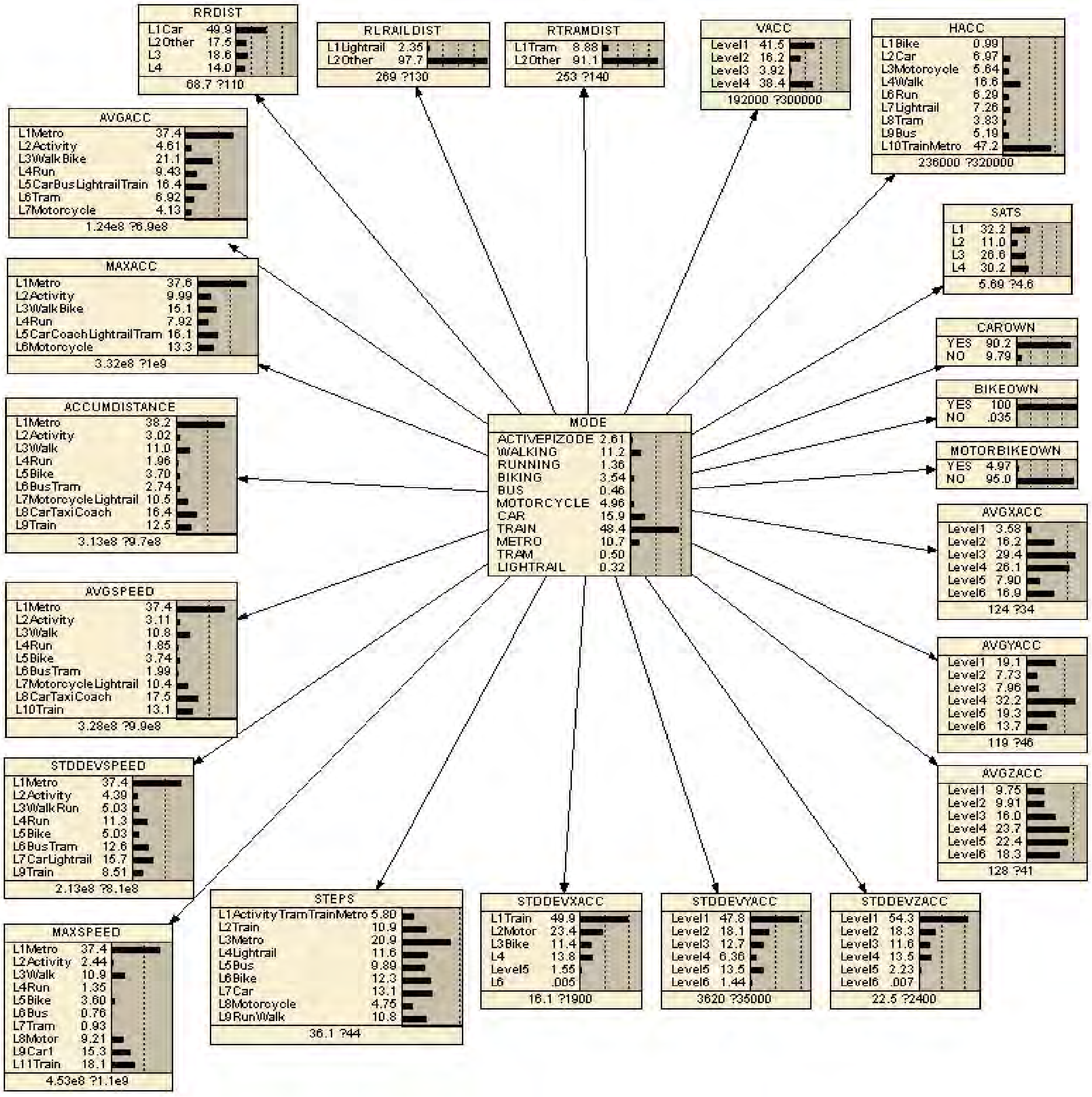


Figure 6.3 portrays the conditional probability matrices. The probabilities are based on a small pilot study and an assessment of the success of the classifier using that data.

Figure 6.3: Conditional probabilities



6.9 The BBN then used the input data (GPS traces, person information and some GIS data) to calculate the probability of (a) an activity, and (b) a specific transport mode. That is, the application of the BBN results for each epoch in a back-reasoned probability that the epoch profile belongs to each of the classes of the parent node (the various transport modes, activity, etc.). In that sense, each epoch is annotated.

Phase 3: Deriving patterns from the epoch-level annotated GPS traces

6.10 Next, these annotated epoch data need to be aggregated and merged into episode data. This process is completed in a series of steps.

Step 1: Identify activity and travel episodes

6.11 An activity episode is characterized by no or little movement between start and end points. To avoid the situation that a stop at a traffic light would be interpreted as an activity as opposed to being recorded as part of a car trip, episodes of less than 5 minutes of no movement were discarded. Similar fragments may occur in activity episodes. Therefore, if an activity or travel episode lasts less than 5 minutes. Merge rules are activated to merge consecutive fragments. The following rules are applied: (1) simple merging for trips and activities; (2) simple backwards merging for trips; (3) simple forward merging for trips; (4) complex merging. Appendix B outlines these four types of merge in detail.

6.12 Depending on the accuracy of the GPS device, there may still be longer fragments of recorded movement during true activity episodes. For example, even when home or at work, the GPS device can record non-accurate coordinates. In addition, some activities such as shopping may involve movement. In principle, the five minute interval could be extended. However, this would mean that activities and travel of shorter duration than the increased threshold would go undetected. Therefore, in addition to filtered evidence of stops, TraceAnnotator also examines whether the locations of successive stops are the same, i.e. whether start and end points are within 200 meter and whether these are close to public transport. The rules set out in table 6.1 are then applied to differentiate between activity episodes, which are used for identifying travel purpose and travel episodes and to identify round trips.

Table 6.1: determination of trip type

	Start and end points within 200 meter		*Start and end points NOT within 200 meter*
	Location NOT close to public transport	Location close to Public transport	
Movement	Round trip	Activity episode	Travel episode
No movement	Activity episode	Activity episode	Travel episode

6.13 The combination of no movement and Location (except for home location) is the SAME is handled as a travel episode to account for signal loss during a longer duration (metro/train). The result of this step is for the 7 consecutive days a pattern or sequence of activity (A) and travel episodes (T), such as for example A-T-A-A-T-T-A-T-A. The following merge rules are applied here:

A-A-A : A
A-A-T : A-T
A-T-A : A-T-A
A-T-T : A-T
T-A-A : T-A
T-A-T : T-A-T
T-T-A : T-A
T-T-T : T

6.14 As discussed, the A-T-A sequence is used to detect any round trips, while the T-A-T sequence is used to detect any trip stages. Note this principle is applied recursively.

Step 2: Identify trip purpose

6.15 The result of Step 1 is a distinction between travel and activity episodes. For each activity episode, the location where the activity takes place is given. To impute trip purpose, first the personal profile data based on the answers to the address questions in the main interview are checked to see whether there is information on what an individual did/typically does at this location. If not, the GIS data bases are checked. The coordinates of the locations are checked across the coordinates of the geo-coded person and POI data, which are linked to a classification of trip purposes. If there is a match (distance between activity location and data < 300 meters), then the associated trip purpose is assigned to this travel episode.

Step 3: Identify trips

6.16 Trips were identified in a straightforward manner: any travel episode in between two activity episodes makes up a journey/trip.

Step 4: Identify trip stages

6.17 If there is any T-A-T sequence in the data, this (sub) sequence is used to detect any trip stages.

Table 6.2: determination of trip stage

	Location is bus stop/station	*Location is bus stop/station*
Duration A < 60 min	Waiting (W)	T-A-T→ T
Duration A GE 60 min	Activity episode	Activity episode

6.18 This principle means that in the preset application we will not detect any stage that incurs less than 5 minutes and more than 30 minute waiting.

6.19 Location was checked first in the person data. This involves reverse geo-coding and finding the node within 300 meters. If no location is found, distances are calculated based on coordinates. If distance to the bus stop/station is then less than 100 meter, the location is said to be the bus stop/station.

Step 5: Identify transport mode

6.20 The result of step four is a series of trips, which consists of one or more trip stages. Transport modes are assigned to each stage (if there is only one stage to the journey) probabilistically. These probabilities are equal to the average probability of the epochs making up the stage episode for which the transport mode is most likely. The transport mode with the highest probability is assigned to that trip stage.

7. VALIDATION

Context

7.1 As discussed in paragraph 2.8, it was not practicable or feasible to undertake a prompted recall survey to validate the inferences of trip mode and purpose. The only option available is therefore to inspect the imputed activity-travel patterns in terms of a set of indicators that would provide evidence to the plausibility of the inference and detect highly unlikely episodes, and/or (sub) sequences. Such plausibility can also be called face validity.

7.2 A set of rules was formulated to this end. Reasons for doing this include:

i it provides a framework for the valuing of the inference quality

ii it avoid intra- and inter-observer variability in expert judgement

iii different setting allow assessing the impact of more or less stringent thresholds

iv it can be used as a tool in large scale data collections to quickly detect any cases that may need further inspection, cleaning or additional checks and corrections.

7.3 If no framework of reference has been provided and observers have only been informed to check the credibility of the imputed activity-travel patterns using visualised patterns on a network/land use background, they may use different criteria and judgements. Even a single observer may shift the focus of attention due to boredom, previous results, fading concentration and other factors.

Framework and rules

7.4 Rules that were formulated to examine the quality of the inference process. These assessed the plausibility of the generated activity-travel schedules for a set of indicators (plaus=1: highly plausible; plaus=0: possible; plaus=-1: not very plausible. The rules were applied to the activity-travel schedules, imputed from GPS traces to create Excel sheets containing the number of respondents by episodes by set of plausibility indicators

7.5 Every episode of the schedule is defined in terms of date, start and end time, duration, distance, activity type (at), transport mode(tm) and location. At the episode level, the rules formalize threshold values for detecting highly unlikely values on a single variable or highly unlikely combinations of values on multiple variables and tag these as implausible. The threshold values are expert-driven. This approach is also used to select particular (sub-) patterns such as business calls, touring, children playing and off-network travel. Overall inference quality is measured in terms of percentages of episodes satisfying each rule/criterion. The detailed rules used to check the results can be found in Appendix C. In summary the rules check:

i The distance and duration of trips with respect to the mode used;

ii The distance travelled with respect to the trip purpose, and

iii The start time, end time and duration of a trip in relation to the trip purpose.

Results

7.6 The following text summarises the proportion of trips which were deemed implausible according to the validation rules described in Appendix C. Tables of these results can be found in Appendix D.

i Activity and distance: 83 per cent of activity episodes had plausible length and five per cent possible, but less plausible length. Purpose could not be identified for 12 per cent of episodes. Although this is substantial, it is less than the benchmarks reported in 2.13;

ii Activity and duration: 78 per cent of activity episodes had plausible duration, five per cent of the activity episodes did not have plausible duration and four per cent were considered possible but less plausible;

iii Activity and start times: 84 per cent of the activity episodes have start times that we considered as plausible and three per cent were considered possible but less plausible;

iv Activity and end times: 84 per cent of the activity episodes have end times that were considered as plausible and four per cent were deemed possible, but less plausible;

v Transport mode and distance: seven per cent of trip stages had distances that were implausible for the mode of transport allocated to the trip and six per cent were considered to be possible but less plausible, and

vi Transport mode and duration: four per cent of trip stages have a duration that was considered implausible for the mode allocated to them and four per cent were deemed possible but less plausible.

8. IDENTIFICATION OF SPECIAL ACTIVITIES

8.1 In addition to the imputation of activity-travel episodes from the GPS traces, the aim of the project was also to flag special cases, which included:

- series of business calls;
- playing in the street;
- round trips, and
- off-network travel.
- in course of work

8.2 DfT will need these flags to ensure that summary results are processed using the same methodology as for the travel diary results and to filter cases such as off-network that are not recorded in the travel diary. These patterns cannot be readily detected from the epoch data, but because information in the (sub)sequence of activity and travel episodes is important in this context, these cases were identified once full schedules were generated.

Series of calls

8.3 According to the NTS manual, respondents will sometimes make a continuous "*series of calls*" for a single purpose. For example, a man might go into several different shops, or a doctor or NatCen interviewer might call at several different addresses. The manual discusses some restrictions, when respondents' travel behaviour can be described as being a series of calls:

- the trip purpose must be either shopping or travel in the course of work;
- the trips included within the series of calls must all have the same purpose – hence a work-related series of calls would be broken if the respondent stopped to do some shopping, and
- the trip must comprise only a single stage – this means that multi-stage trips cannot be treated as a series of calls (that is, they must be on the same mode of transport and – if using public transport – use the same ticket).

8.4 In case of automated detection, these restrictions are problematic. Work trips are based on location information of the job. Business calls can either be at other businesses or at home addresses. Hence, there is no way of detecting travel in the course of work on that basis. In principle, the condition that the series of call must all have the same purpose can be activated, but this likely leads to substantial misclassification. Moreover, ticket information is not available. The same mode of transport could be used but gain may bias the results: for example a mailman may park his car and walk to deliver the mail in adjacent streets; in terms of delivery also logistics often mean the use of different transport modes. Rather than using these principles, would do not seem very appropriate in automatic detection, we decided to flag these series of call in terms of the number of stops made. More specifically, if the number of stops longer than 3 minutes in any subsequence is five or higher, that sequence was flagged as a series of calls. If the purpose of the start activity of a journey is shopping related activity (or work related activity), and the purpose of end activity of a journey is also shopping related (or work related), that was flagged as a series of calls, pl = 1, otherwise, pl = 0.

Playing in the street

8.5 According to the NTS manual, playing in the street is only relevant for the 12-15 year olds. In addition, only walking, running and perhaps cycling would be involved. It seems that two patterns may be relevant: in the case of active behaviour, the traces would show evidence of frequent movement around the home or the location where the children play, in multiple directions, probably with substantial variation in speed and walking/running. In case of passive behaviour, children may be hanging around in a particular area, which if this takes longer than 5 minutes would be reflected in activity episodes.

8.6 Rules:

- If children have a journey with the purpose from and purpose to be one of the listed activities (home, personal other, family&friend, other, unknown), pl = 1, otherwise, pl = 0.

Round trips

According to the manual, roundtrips are those trips where the origin and the destination are the same and where there is no discernible break in their travel. A common example of a round trip is when someone takes their dog for a walk. These trips can be easily detected from the sequence in the sense that the location of any two consecutive activity episodes would be the same. (Sub) sequences obeying to the condition were flagged.

Off-network travel

This type of travel is travel off the public highway. Trips of that kind were flagged on the following rule:: If the average value of distance to the nearest road travelling by car in that journey is longer than 10 meters, pl = 1, otherwise, pl = 0.

In course of work

Some people don’t work at a fixed location or have multiple work locations. This can induce trips in course of work. Such kind of trips was identified in terms of the purposes of start and end of a journey. The following rule was applied:

If both the purpose from and purpose to of a journey are work appointments (or main work), such a journey is in course of work, set pl = 1, otherwise, pl = 0.

References

Auld, J., Williams, C., Mohammadian, A. and Nelson, P. (2009) An automated GPS-based prompted recall survey with learning algorithms. *Transportation Letters*, 1, 58-79.

Axhausen, K.W., Schönfelder, S., Wolf, J., Oliveira, M. and Samaga, U. (2004) Eighty weeks of GPS traces: approaches to enriching trip information. *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington, D.C., USA.

Bao, L. and Intille, S.S. (2004) Activity recognition from user-annotated acceleration data. Pervasive Computing, Proceedings of Pervasive Computing, Second International Conference, PERVASIVE, Vienna, Austria, April 21-23, 2004.

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

Bellemans, T., Kochan, B., Janssens, D., Wets, G. and Timmermans, H.J.P. (2008) In the field evaluation of the impact of a GPS-enabled personal digital assistant on activity-travel diary data quality. *Proceedings 87th Annual Meeting of the Transportation Research Board*, Washington, DC.

Bohte, W. and Maat, K. (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.

Chung, E. and Shalaby, A. (2005) Development of a trip reconstruction tool for GPS-based personal travel surveys. *Journal of Transportation Planning and Technology*, 28, 381-401.

Cooper, A.R., Page, A.S., Wheeler, B.W., Griew, P., Davis, L., Hillsdon, M. and Jago, R. (2010) Mapping the walk to school using accelerometry combined with a global positioning system. *American Journal of Preventive Medicine,* 38(2), 178-183.

Doherty, S.T., Lee-Gosselin, M.E.H. and Papinski, D. (2006) Internet-based prompted recall diary with automated GPS activity-trip detection: system design. *Proceedings of the 85Th Annual Conference of the Transportation Research Board*, Washington, D.C., USA

Draijer, G., Kalfs, N. and Perdok, J. (2000) Global positioning system as data collection method for travel research. *Transportation Research Record*, 1719, 147-153.

Forrest, T.L. and Pearson, D.F. (2005) A comparison of trip determination methods in GPS enhanced household travel surveys. *Proceedings 84th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2005.

Guensler, R. and Wolf, J. (1999) Development of a handheld electronic travel diary for monitoring individual trip making behaviour. TRB, National Research Council, Washington, D.C..

Hato, E. and Asakura, Y. (2001) New approach for collection of activity diary using mobile communication systems. *Proceedings 80th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Li, H., Guensler, R., Ongle, J. and Wang, J. (2005) Using GPS data to understand the day-to-day dynamics of the morning commute behaviour. *Proceedings 84th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Li, Z. and Shalaby, A. (2008) Web-based GIS system for prompted recall of GPS-assisted personal travel surveys: system development and experimental study. *Proceedings 87th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Marca, J., Rindt, C.R. and McNally, M. (2002) Collecting activity data from GPS readings. *Proceedings 81st Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Marca, J., Rindt, C.R. and McNally, M. (2002) The Tracer data collection system: Implementation and operational experience. *Proceedings 81st Annual Meeting of the Transportation Research Board*, Washington D.C., USA

Moiseeva, A., Jessuran, J and Timmermans, H.J.P. (2009) Semi-automatic imputation of activity-travel diaries using GPS traces, prompted recall and context-sensitive learning algorithms, *Transportation Research Record*

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

Ohmori, N., Nakazato, M., Harata, N., Sasaki, K. and Nishii, K. (2006) Activity diary surveys using GPS mobile phones and PDA. *Proceedings 85th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Oliver M., Badland, H., Mavoa, S., Duncan, M.J. and Duncan, S. (2007). Combining GPS, GIS and Accelerometry: Methodological issues in the assessment of location and intensity of travel behaviours. *Journal of Physical Activity and Health*, 7, 102-108.

Ravi, N., Dandekar, N., Mysore, P. and Littman, M.L. (2005) Activity recognition from accelerometer data. Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence, July 9 – 13 2005, Pittsburgh, Pennsylvania.

Schönfelder, S., K.W. Axhausen, N. Antille and M. Bierlaire (2002) Exploring the potentials of automatically collected GPS data for travel behaviour analysis - a Swedish data source. In J. Möltgen and A. Wytzisk (Eds.), *GI-Technologien für Verkehr und Logistik,* Institut für Geoinformatik, Universität Münster, Münster, 220, pp.155-179.

Schuessler, N. and Axhausen, K. (2008) Identifying trips and activities and their characteristics from GPS raw data without further information. Paper presented at the *8th International Conference on Survey Methods in Transport*, Annecy, France.

Stopher, P, Bullock, P.J. and Horst, F.N.H. (2003) Conducting a GPS survey with a time-use diary. *Proceedings 83th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Stopher, P. and Collins, A. (2005) Conducting a GPS prompted recall survey over the internet. TRB Annual Meeting, CD-ROM, Transportation Research Board, National Research Council, Washington, D.C..

Stopher, P.R. and Wargelin, L. (2010) Conducting a household travel survey with GPS: Reports on a pilot study. Proceedings of the 12th WCTRS, July 11-15, 2010, Lisbon, Portugal.

Troped, P.J., Oliveira, M.S., Matthews, C.E., Cromley, E.K., Melly, S.J. and Craig, B.A. (2008) Prediction of activity mode with global positioning system and accelerometer data. *Medicine & Science in Sports & Exercise*, 40(5), 972-8.

Wolf, J. (2004) Applications of new technologies in travel surveys. Proceedings of the International Conference on Transport Survey Quality and Innovation, Costa Rica, August, 2004.

Wolf, J., Guensler, R. and Bachman, W. (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 Ogly, J. (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 Arce, C. (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., Oliviera, M. and Thompson, M. (2003) Impact of underreporting on mileage and travel time estimates: results from Global Positioning System-enhanced household travel survey. *Transportation Research Record*, 1854, 189-198.

Wolf, J., Schönfelder, S., Samaga, U., Oliveira, M. and Axhausen, K.W. (2004) Eighty weeks of GPS traces: approaches to enriching trip information. *Proceedings 83th Annual Meeting of the Transportation Research Board*, Washington D.C., USA.

Appendix A: glossary of terms

Activity	The main business carried out at a location
Activity agenda some time	the set of activities that needs to be completed within period
Activity duration	the amount of contiguous time used to complete an activity
Activity-travel pattern	the set of consecutive activities carried out by an individual in the course of one day, plus the movement involved to perform these activities
Activity profile	a description of an activity in terms of various choice facets, such as type, duration, location
Activity schedule	the sequence of activities, conducted during one day
Bayesian network	a probabilistic graphical model that represents set of random variables and their conditional dependencies via a directed acyclic graph
Diary	a detailed record of activity and/or travel profiles
Episode	the time between the start and end time of an activity or travel
Epoch	the instance of time chosen to record GPS and/or accelerometer data
Imputation	Substitution of some value for missing data
Journey	a tour starting and ending at home
Prompted recall	a test to see how well people remember an object or event in which the respondents are given some help (prompt) which they might associate with the object/event. In the transportation literature, the prompt is the imputed activity-travel diary used to recall travel and activities conducted.

Tour a sequence of trips starting and ending at the same location

Trip movement between an origin and a destination

Trip stage part of a travel episode during which is single transport mode is used

Appendix B: rule for merging fragments of activity or travel episodes

	Example	Description
Simple merging for Trips and Activities	Ta Tb -> Ta Ta	Trip into trip: trip b<3 min between two trips A will be merged in Trip a
	A1 T -> A1 A1	Trip into activity: any short trip T<3 min between two Activity episodes will be merged
	Ta A -> Ta Ta	Activity into trip: any short activity A<3 min between two Trips Ta will be merged into Trip Ta
Simple Backward Merging for Trips	Ta Tb -> Ta Tc	Any short trip<3 min between two different trips will be merged backwards
	Ta Tb -> Ta A	Trips first always merge with trips: a short trip Tb<3 min after trip Ta, followed by activity A will be merged with trip Ta
Simple Forward Merging for Trips	A Ta -> Tb Tb	Trips first always merge with trips: trip Ta after activity followed by trip Tb will be merged with trip Tb
Complex merging	A Ta -> Tc Tb -> Tc Tc	Two different short trips Ta and Tb after activity followed by other trip Tc will be merged into trip Tc
	Ta A -> A Tb -> A A	Subsequent short activity A<3 min and trip Tb<3 min after a trip Ta and followed by activity both will be merged into activity
	A1 Ta -> A1 A2-> A1 A3 A1 Ta -> A1 A2-> A1 Tb	Subsequent short trip Ta<3 min and activity A2<3 min after long activity and followed by trip or activity will be merged with the first activity

Note: these merging rules form an iterative process.

Appendix C: rules for validation of inferred mode and purpose of trip

(1) ***relationship activity type and distance***

IF at = work AND distance LT 150 km pl () = 1
IF at = work AND distance GE 150 LE 250 km pl () = 0
IF at = work AND distance GT 250 km pl () = -1
Note distance here relates to the previous travel episode

IF at = education AND distance LT 150 km pl () = 1
IF at = education AND distance GE 150 LE 250 km pl () = 0
IF at = education AND distance GT 250 km pl () = -1

IF at = grocshop AND distance LT 5 km pl () = 1
IF at = grocshop AND distance GE 5 LE 10 km pl () = 0
IF at = grocshop AND distance GT 10 km pl () = -1

IF at = shop AND distance LT 50 km pl () = 1
IF at = shop AND distance GE 50 LE 150 km pl () = 0
IF at = shop AND distance GT 150 km pl () = -1

IF at = med AND distance LT 50 km pl () = 1
IF at = med AND distance GE 50 LE 150 km pl () = 0
IF at = med AND distance GT 150 km pl () = -1

IF at = eat/drink AND distance LT 30 km pl () = 1
IF at = eat/drink AND distance GE 30 LE 100 km pl () = 0
IF at = eat/drink AND distance GT 100 km pl () = -1

IF at = ent AND distance LT 50 km pl () = 1
IF at = ent AND distance GE 50 LE 150 km pl () = 0
IF at = ent AND distance GT 150 km pl () = -1
IF at = sports AND distance LT 50 km pl () = 1
IF at = sports AND distance GE 50 LE 150 km pl () = 0
IF at = sports AND distance GT 150 km pl () = -1

(2) ***relationship activity type and duration***

IF at = work AND duration GT 540 min pl () = 0
IF at = work AND duration GE 200 min LE 540 min pl () = 1
IF at = work AND duration LT 200 min pl () = -1

IF at = education AND duration GT 540 min pl () = 0
IF at = education AND duration GE 60 min LE 540 min pl () = 1
IF at = education AND duration LT 60 min pl () = -1

IF at = grocshop AND duration GT 90 min pl () = 0
IF at = grocshop AND duration GE 90 min LE 120 min pl () = 1

IF at = grocshop AND duration LT 120 min pl () = -1

IF at = shop AND duration GT 180 min pl () = 0
IF at = shop AND duration GE 180 min LE 300 min pl () = 1
IF at = shop AND duration LT 300 min pl () = -1

IF at = med AND duration GT 540 min pl () = 0
IF at = med AND duration GE 200 min LE 540 min pl () = 1
IF at = med AND duration LT 200 min pl () = -1

IF at = eat/drink AND duration GT 300 min pl () = 0
IF at = eat/drink AND duration GE 60 min LE 300 min pl () = 1
IF at = eat/drink AND duration LT 60 min pl () = -1

IF at = ent AND duration GT 540 min pl () = 0
IF at = ent AND duration GE 200 min LE 540 min pl () = 1
IF at = ent AND duration LT 200 min pl () = -1

IF at = sports AND duration GT 300 min pl () = 0
IF at = sports AND duration GE 240 min LE 300 min pl () = 1
IF at = sports AND duration LT 240 min pl () = -1

(3) ***relationship activity type and start time***

IF at = work AND starttime GE 7:00 LE 16:00 pl () = 1, else 0

IF at = education AND starttime GE 8:00 pl () = 1, else 0
IF at = grocshop AND starttime GE 8:00 LE 22:00 pl () = 1, else 0

IF at = shop AND starttime GE 9:00 LE 22:00 pl () = 1, else 0

IF at = med AND starttime GE 9:00 LE 19:00 pl () = 1, else 0
IF at = eat/drink AND starttime GE 10:00 LE 23:00 pl () = 1, else 0

IF at = ent AND starttime GE 10:00 LE 24:00 pl () = 1, else 0

IF at = sports AND starttime GE 10:00 LE 20:00 pl () = 1, else 0

(4) ***relationship activity type and end time***

IF at = work AND endtime GE 12:00 LE 19:00 pl () = 1, else 0

IF at = education AND endtime GE 10:00 LE 22:00 pl () = 1, else 0

IF at = grocshop AND endtime GE 8:00 LE 22:00 pl () = 1, else 0

IF at = shop AND endtime GE 10:00 LE 22:00 pl () = 1, else 0

IF at = med AND endtime GE 9:00 LE 19:00 pl () = 1, else 0

IF at = eat/drink AND endtime GE 10:00 LE 23:00 pl () = 1, else 0

IF at = ent AND endtime <> pl () = 1, else 0

IF at = sports AND endtime GE 10:00 LE 22:00 pl () = 1, else 0

(5) ***relationship transport mode and distance***

IF tm = walking AND distance LT 5 km pl () = 1
IF tm = walking AND distance GE 5 LE 10 km pl () = 0
IF tm = walking AND distance GT 10 km pl () = -1

IF tm = biking AND distance LT 25 km pl () = 1
IF tm = biking AND distance GE 25 km LE 40 km pl () = 0
IF tm = biking AND distance GT 40 km pl () = -1

IF tm = running AND distance LT 5 km pl () = 1
IF tm = running AND distance GE 5 km LE 10 km pl () = 0
IF tm = running AND distance GT 10 km pl () = -1

IF tm = bus AND distance LT 50 km pl () = 1
IF tm = bus AND distance GE 50 km LE 100 km pl () = 0
IF tm = bus AND distance GT 100 km pl () = -1

IF tm = motorcycle AND distance LT 150 km pl () = 1
IF tm = motorcycle AND distance GE 100 km LE 150 km pl () = 0
IF tm = motorcycle AND distance GT 100 km pl () = -1

IF tm = car AND distance LT 250 km pl () = 1
IF tm = car AND distance GE 250 km LE 400 km pl () = 0
IF tm = car AND distance GT 400 km pl () = -1

IF tm = lightrail AND distance LT 50 km pl () = 1
IF tm = lightrail AND distance GE 50 km LE 70 km pl () = 0
IF tm = lightrail AND distance GT 70 km pl () = -1

IF tm = train AND distance LT 200 km pl () = 1
IF tm = train AND distance GE 200 km LE 1000 km pl () = 0
IF tm = train AND distance GT 1000 km pl () = -1

(6) ***relationship transport mode and duration***

IF tm = walking AND duration LT 30 min pl () = 1
IF tm = walking AND distance GE 30 min LE 45 min pl () = 0
IF tm = walking AND distance GT 45 min pl () = -1

IF tm = biking AND duration LT 60 min pl () = 1
IF tm = biking AND distance GE 60 min LE 90 min pl () = 0
IF tm = biking AND distance GT 90 min pl () = -1

```
IF tm = running AND duration LT 20 min pl () = 1
IF tm = running AND duration GE 20 min LE 40 min pl () = 0
IF tm = running AND duration GT 40 min pl () = -1

IF tm = bus AND duration LT 90 min pl () = 1
IF tm = bus AND duration GE 90 min LE 120 min pl () = 0
IF tm = bus AND duration GT 120 min pl () = -1
IF tm = motorcycle AND duration LT 90 min pl () = 1
IF tm = motorcycle AND duration GE 90 min LE 120 min pl () = 0
IF tm = motorcycle AND duration GT 120 min pl () = -1

IF tm = Car AND duration LT 180 min pl () = 1
IF tm = Car AND duration GE 180 min LE 240 min pl () = 0
IF tm = Car AND duration GT 240 min pl () = -1

IF tm = lightrail AND duration LT 45 min pl () = 1
IF tm = lightrail AND duration GE 45 min LE 60 min pl () = 0
IF tm = lightrail AND duration GT 60 min pl () = -1

IF tm = train AND duration LT 120 min pl () = 1
IF tm = train AND duration GE 120 min LE 300 min pl () = 0
IF tm = train AND duration GT 300 min pl () = -1
```

Appendix D: summary of validation results

(1)activity + distance		
	Frequency	Percent
0	831	5.1%
1	13502	82.7%
99	2003	12.3%
Total	16336	100.0%

(2): activity + duration		
	Frequency	Percent
-1	884	5.4%
0	682	4.2%
1	12767	78.2%
99	2003	12.3%
Total	16336	100.0%

(3): activity + timestart		
	Frequency	Percent
0	556	3.4%
1	13777	84.3%
99	2003	12.3%
Total	16336	100.0%

(4): activity + timeend		
	Frequency	Percent
0	593	3.6%
1	13740	84.1%
99	2003	12.3%
Total	16336	100.0%

(5): mode + distance		
	Frequency	Percent
-1	824	6.9%
0	699	5.8%
1	10456	87.3%
Total	11979	100.0%

(6): mode + duration		
	Frequency	Percent
-1	427	3.6%
0	432	3.6%
1	11120	92.8%
Total	11979	100.0%

Appendix E: questions for NTS GPS processing

Placement interview

Household questionnaire

Asked for each child in household aged 0-4
PreSch
Does [name] attend a nursery, pre-school or primary school?
1. Yes
2. No

(IF Aged 5-16 and makes a daily journey to and from school (SchDly = 1))OR (PreSch=1)
SchAdd
What is the name and address of [name's] nursery/school/college?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THEIR SCHOOL/COLLEGE AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

Individual questionnaire

ASK ALL
PrivCar
How frequently [do you/ does name] travel by private car? Do not include taxi.
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO.
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO.
1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL
Ordbus
How frequently do you use local buses?
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO.
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO.
1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL

Coach

(How frequently do you/does name use) an express bus or coach within Great Britain?
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO.
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO

1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL

Train

(How frequently do you/does name use) a train, not including underground, tram or light rail?
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO.
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO.

1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL

TaxiCab

(How frequently do you/ does name use) a taxi/minicab?
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO.

1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL

Plane

(How frequently do you/does name take) an **internal** air flight within Great Britain?
PLEASE COUNT EACH SINGLE TRIP AS ONE JOURNEY AND EACH RETURN TRIP AS TWO.
NOTE: ONLY INCLUDE TRAVEL WITHIN GREAT BRITAIN, OVER THE LAST YEAR OR SO.

1. 3 or more times a week
2. Once or twice a week
3. Less than that but more than twice a month
4. Once or twice a month
5. Less than that but more than twice a year
6. Once or twice a year
7. Less than that or never

ASK ALL

GenCycle

(The next few questions are about cycling.) Excluding exercise bikes, do you... READ OUT...

1. ...own a bicycle yourself,
2. have regular use of a bicycle owned by someone else,
3. or have no regular use of a bicycle?

IF respondent regularly uses bicycle owned by someone else (GenCycle = 2)
CycElse
Is that bicycle owned by someone in your household or someone outside the household?
1. Someone in the household
2. Someone outside the household

ASK ALL
Cycle12
(May I just check,) have you ridden a bicycle during the last 12 months, (that is since [this date last year])?
Helpscreen:
This means independently riding a bicycle. Do not count riding on a child seat or bicycle attached to an adults
1. Yes
2. No
3. Don't know / Can't remember

If aged 16 or over and in full or part-time education (EducN=1 or 2)
EducTr
Do you travel to a school, college or university to undertake this course or is it done through distance learning?
1. Travels to school/college/university
2. Distance learning

If travels to education (EducTr=1)
QEAdd
What is the name and address of your school/college?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THEIR SCHOOL/COLLEGE AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

IF respondent goes to the same workplace each time or at least 2 days a week (WkPlace = 1 or 2)
WkAdd1
NAME
What is the address of your usual place of work?
INTERVIEWER: obtain as full an address as possible, including postcode if respondent can supply this. If the respondent is unsure of exact address/ postcode, please record the Name of their employer/office and as much of the address as they can provide.
Use <ctrl + R> if respondent does not wish to provide the address.
Enter first line of the address.
INTERVIEWER: The journey to work is the most frequently travelled journey for many people. This information will allow the exact distance of this journey to be calculated.

If first line of work address entered (WkAdd1 = Response)
WkPC
NAME
What is the postcode of your usual place of work?
INTERVIWER: Use <CTRL + K> if does not know.

Admin Block

IF Placement Interview completed (StatusQ = 1)
IntroGPS
INTERVIEWER: INTRODUCE THE GPS Devices.
[List who is eligible to carry a device]
Date from: [Day after interview]
Date to: [Day after interview + 7 days]

IF Placement Interview completed (StatusQ = 1)
IntrBGPS
INTERVIEWER: INTRODUCE THE GPS Devices.
Key points to cover:

- How and when to charge the device
- How to switch it on and off
- How and when to carry
- Who to contact if they have any problems with the device
- The incentive
- How the data is processed and used

Asked in turn of each aged 12 or over
WillGPS
INTERVIEWER: IS [name] WILLING TO CARRY A GPS DEVICE

1. Yes: GPS
2. No: not willing to take GPS

Asked of each in turn who is willing to carry a GPS (WillGPS=1)
GPSNo
INTERVIEWER: Record the serial number of the GPS device.

If not willing to carry device (WillGPS=2)
YNoWill
Why are they not willing?
:string

If placement interview complete (StatusQ=1)
PlacTime
INTERVIEWER: HOW LONG DID IT TAKE TO PLACE AND EXPLAIN THE GPS DEVICES? RECORD TO NEAREST MINUTE.

Pick-Up Interview

Household questionnaire

If has vehicle at pick-up interview (StillGot=1)
TWMiles
VEHICLE
How many miles was the [VEHICLE] driven during the travel week?

If response given at TWMiles
KmMile
VEHICLE
INTERVIEWER ASK OR CODE: WAS THE ANSWER TO ' TWmiles' IN MILES OR KILOMETRES?
1. Miles
2. Kilometres

Individual Questionnaire

If accepted a GPS (WillGPS=1)
GPSColl
HAVE YOU COLLECTED A GPS DEVICE FOR [NAME]?
1. Yes
2. No

If device not collected (GPSColl=2)
YNoColl
Why was no device collected?
:string

If returned a GPS (GPSColl=1)
GPSProb
Can I just check did you encounter any problems or difficulties using the GPS device?
1. Yes
2. No

If encountered problems using GPS (GPSProb=1)
YGPSProb
What were the problems/difficulties you encountered?
INTERVIEWER: PROBE FULLY AND ENTER DETAILS.
:Open

If returned a GPS (GPSColl=1)
EasGPS
And can I just check, overall, how easy or difficult did you personally find it to use the GPS device?
Did you find it...READ OUT…
1. Very easy...
2. Fairly easy...
3. Fairly difficult...
4. or Very difficult?

If returned a GPS (GPSColl=1)
NoJrny
Were there any days where you did not make any journeys?
CODE ALL THAT APPLY
1. Day 1
2. Day 2
3. Day 3
4. Day 4

5. Day 5
6. Day 6
7. Day 7
8. No
9. (Can’t remember)

If returned a GPS (GPSColl=1)
NotChrg
Were there any days where you forgot to charge the device?
CODE ALL THAT APPLY
1. Day 1
2. Day 2
3. Day 3
4. Day 4
5. Day 5
6. Day 6
7. Day 7
8. No
9. (Can’t remember)

If returned a GPS (GPSColl=1)
NoCarry
Were there any days where you made journeys but did not carry the device with you?
CODE ALL THAT APPLY
INTERVIEWER: IF LEFT SOMEWHERE OVERNIGHT (E.G. WORK), RECORD AS NOT CARRIED ON BOTH DAYS.
1. Day 1
2. Day 2
3. Day 3
4. Day 4
5. Day 5
6. Day 6
7. Day 7
8. No
9. (Can’t remember)

For each day not carried at NoCarry
AllSome
On day [Day number] was the device not carried for all journeys or for some journeys?
1. Not carried for all journeys
2. Not carried for some journeys

For each day not carried at NoCarry
YNoGPS
Why was this?
:Open

First parent in household (aged16+), with children aged 12-15, interviewed face to face.
ChldPrb
In this survey we asked everybody in the household aged 12 and above to use a GPS device. Were there any reasons that [Child’s name] did not want to, or could not use the device?
1. Yes
2. No

If ChldPrb=yes
YChldPrb
What was the reason for that?
:Open

If age>=12 and there are 12 to 15s in the household
ChldGPS
Do you think we should give GPS devices to children aged 12-15 as part of this survey?
1. Yes
2. No

If doesn't think children should carry GPS (ChldGPS=2)
YChldGPS
Why is that?
:Open

If returned GPS device (GPSColl=1) and recorded as being in work at placement interview (DVIL03a=1)
ExcJob
SHOWCARD BB
Does your job mainly involve transporting either goods or people around or does it involve transporting specialist equipment or tools necessary to do your job? Some examples of these types of jobs are shown on this card?
1. Yes
2. No

If in job with excluded work trips (ExcJob=1)
JobDay
For people who work in these sorts of jobs such as yours, we need to be able to separate out travel done in the course of your work. It would therefore be helpful if you could tell us at which time you started and finished work on each day of the week that you worked.
Firstly, could you tell me which days of the travel week you worked on?
1. Day 1
2. Day 2
3. Day 3
4. Day 4
5. Day 5
6. Day 6
7. Day 7
8. None

For each day worked at JobDay
ExcStrt
(And) On day [1/2 etc] at what time did you start work?
INTERVIEWER: If respondent travels to a depot or other common place before starting work, please record the times they arrived at and left this place each day. If respondent works from home please record the time the left their home and arrived back there each day.
INTERVIEWER: ENTER TIME USING 24 HOUR CLOCK.

For each day worked at JobDay
ExcFin
And what time did you finish work?

If returned GPS device (GPSColl=1) and travels to different places to work (WkPlace=3) and not in excluded occupation ExcJob=no
WkDay
You mentioned in the placement interview that you work at different places each day. It would therefore be helpful if you could tell us at which time you started and finished work on each day of the week that you worked.
Firstly, could you tell me which days of the travel week you worked on?
1. Day 1
2. Day 2
3. Day 3
4. Day 4
5. Day 5

6. Day 6
7. Day 7
8. None

For each day worked at WkDay
WkStrt
(And) On day [1/2 etc] at what time did you start work?
INTERVIEWER: If respondent travels to a depot or other common place before starting work, please record the times they arrived at and left this place each day. If respondent works from home please record the time the left their home and arrived back there each day.
INTERVIEWER: ENTER TIME USING 24 HOUR CLOCK.

For each day worked at WkDay
WkFin
And what time did you finish work?

If Age>=12
CarTrip
Thinking about the most recent journey you made by car or van: were you the driver or the passenger?
INTERVIEWER: This question refers to private vehicles, do not include taxis.
INTERVIEWER: In most cases this is likely to be a return journey to their home.
1. Driver
2. Passenger
3. Can't remember

If Age>=12
CarVan
And, was this trip made in a car or van?
1. Car
2. Van

If Age>=12
NoAdult
Including you, how many people were in the car or van on that trip?
Enter number of adults (aged 16 or over)

If Age>=12
No1215
Enter number of children aged 12 to 15

If Age>=12
NoChld
Enter number of children aged under 12

If Age>=12
TaxiTrp
How many taxi trips did you make in the travel week?
0. None
1. One
2. Two
3. Three
4. Four
5. Five or more

If made 5 or more taxi trips (TaxiTrp=5)
XTaxiTrp
How many taxi trips did you make?

If made a taxi trip (TaxiTrp>0)
TaxiFare
(Thinking about the most recent of these) how much did you pay for the taxi fare?
THIS SHOULD BE THE AMOUNT THE RESPONDENT CONTRIBUTED TO CAB FARE, NOT WHAT WAS THEIR SHARE OR TOTAL VALUE OF CAB FARE.

If made a taxi trip (TaxiTrp>0)
TaxiNo
And, including you, how many passengers were in the taxi on that trip?

If returned a GPS (GPSColl=1)
TWSma
It would also be helpful to know some of the places you visited during the travel week. Did you visit any supermarkets during the travel week?
1. Yes
2. No

If visited supermarket (TWSma=1)
QSmaAdd
What was the name and address of the (first) supermarket you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.
INTERVIEWER: RESPONDENTS CAN CHECK THE ADDRESS ON A RECIPT FROM THE SHOP.

If visited supermarket (TWSma=1)
SmaOth
Did you visit any other supermarkets during the travel week?
1. Yes
2. No

If SmaOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWGym
Did you visit any gym or fitness centres during the travel week?
1. Yes
2. No

If visited gym or fitness centre (TWGym=1)
QGymAdd
What was the name and address of the (first) gym or fitness centre you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited gym or fitness centre (TWGym=1)
GymOth
Did you visit any other gym or fitness centre during the travel week?
1. Yes
2. No

If GymOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWCin
Did you visit any cinemas or theatres during the travel week?
1. Yes
2. No

If visited cinemas or theatres (TWCin=1)
QCinAdd
What was the name and address of the (first) cinemas or theatres you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited cinemas or theatres (TWCin=1)
CinOth
Did you visit any other cinemas or theatres during the travel week?
1. Yes
2. No

If CinOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWOthL
Did you visit any other leisure facilities during the travel week?
1. Yes
2. No

If visited other leisure facilities (TWOthL=1)
QOthLAdd
What was the name and address of the (first) other leisure facilities you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited other leisure facilities (TWOthL=1)
OthLOth
Did you visit any other leisure facilities during the travel week?
1. Yes
2. No

If OthLOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWDen
Did you visit any dentist / doctor / hospital during the travel week?
1. Yes
2. No

If visited dentist / doctor / hospital (TWDen=1)
QDenAdd
What was the name and address of the (first) dentist / doctor / hospital you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited dentist / doctor / hospital (TWDen=1)
DenOth
Did you visit any other dentist / doctor / hospital during the travel week?
1. Yes
2. No

If DenOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1) and employed (DVIL03a=1) and work in the same place every day or at least 2 days a week (WkPlace=1 or 2)
TWWApp
Did you visit any work appointments during the travel week?
INTERVIEWER: Include any journeys made in the course of work to somewhere other that their usual place of work
1. Yes
2. No

If visited work appointments (TWWapp=1)
QWAppAdd
What was the name and address of the (first) work appointment you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited work appointments (TWWApp=1)
WAppOth
Did you visit any other work appointments during the travel week?
1. Yes
2. No

If WAppOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWCCen
Did you visit any community centres or church/religious centres during the travel week?
1. Yes
2. No

If visited community centres or church/religious centres (TWCCen=1)
QCCenAdd
What was the name and address of the (first) community centres or church/religious centres you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited community centres or church/religious centres (TWCCen=1)
CCenOth
Did you visit any other community centres or church/religious centres during the travel week?
1. Yes
2. No

If CCenOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWECL
Did you visit any evening or weekend classes during the travel week?
1. Yes
2. No

If visited other evening or weekend classes (TWECL=1)
QECLAdd
What was the name and address of the (first) evening or weekend classes you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited other evening or weekend classes (TWECL=1)
ECLOth
Did you visit any other evening or weekend classes during the travel week?
1. Yes
2. No

If ECLOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1) and there is a child under 12 years in the household
TWCCL
Did you visit any children's (under 12) clubs/classes during the travel week?
1. Yes
2. No

If visited children's (under 12) clubs/classes (TWCCL=1)
QCCLAdd
What was the name and address of the (first) children's (under 12) clubs/classes you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited children's (under 12) clubs/classes (TWCCL=1)
CCLOth
Did you visit any other children's (under 12) clubs/classes during the travel week?
1. Yes
2. No

If CCLOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
TWFrF
Did you visit any friends or family during the travel week?
1. Yes
2. No

If visited friends or family (TWFrF=1)
QFrFAdd
What was the address of the (first) friends or family you visited?
INTERVIEWER: OBTAIN AS FULL AN ADDRESS AS POSSIBLE, INCLUDING POSTCODE IF RESPONDENT CAN SUPPLY THIS. IF THE RESPONDENT IS UNSURE OF EXACT ADDRESS/ POSTCODE, PLEASE RECORD THE NAME OF THE ORGANISATION AND AS MUCH OF THE ADDRESS AS THEY CAN PROVIDE.

If visited friends or family (TWFrF=1)
FrFOth
Did you visit any other friends or family during the travel week?
1. Yes
2. No

If FrFOth=Yes ask for address. Collect up to 5 addresses.

If returned a GPS (GPSColl=1)
OverNt
Were there any days during the travel week that you were away from home overnight?
CODE ALL THAT APPLY
1. Day 1
2. Day 2
3. Day 3
4. Day 4
5. Day 5
6. Day 6
7. Day 7
8. At home all nights

Admin Block

If any GPS devices collected (GPSColl=1)
Despatch
Prepare a despatch note to be returned with the GPS devices.
:Enter to continue

IF Pickup Interview completed (StatusQ=2)
PickTime
INTERVIEWER: HOW LONG DID IT TAKE TO PICK UP AND CHECK THE GPS DEVICES?
RECORD TO NEAREST MINUTE