WORKING PAPER

ITLS-WP-13-21

Evaluating and improving software for identifying trips, occupancy, mode and purpose from GPS traces.

By
Peter Stopher, Jun Zhang and Christine Prasad

November 2013

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-13-21

TITLE: **Evaluating and improving software for identifying trips, occupancy, mode and purpose from GPS traces.**

ABSTRACT: Over the past several years, the Institute of Transport and Logistics Studies at the University of Sydney has been developing software to process GPS traces and impute the trip ends, modes of travel, occupancy, and trip purpose. Conducting the GPS-only household travel survey for the Greater Cincinnati Area Household Travel Survey provided an opportunity to assess the accuracy of the existing software and to identify weaknesses and inaccuracies and propose changes to the software to overcome them. Initially, the accuracy of the software was assessed in the pilot survey. This analysis was then extended midway through the main survey and again at the conclusion of the main survey. The results of these assessments are documented in this paper. The paper concludes with a discussion of the accuracy achieved at this point in imputing trip ends, occupancy, mode, and purpose from GPS traces and the suitability of such data for travel demand modelling. The paper also suggests further research in this area that may provide even greater accuracy in the imputation of these attributes that are not measured directly by passive GPS devices.

KEY WORDS: ***GPS, Processing Software; Prompted Recall; Trip Identification; Mode Identification; Purpose Identification.***

AUTHORS: **Stopher, Zhang and Prasad**

CONTACT: INSTITUTE of TRANSPORT and LOGISTICS STUDIES (C13)
The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1824
Facsimile: +612 9114 1722
E-mail: business.itlsinfo@sydney.edu.au
Internet: http://sydney.edu.au/business/itls

DATE: November 2013

1. Introduction

Over the past several years, the Institute of Transport and Logistics Studies (ITLS) at the University of Sydney has been developing software to process GPS traces and impute the trip ends, modes of travel, occupancy, and trip purpose (Stopher et al., 2005; Stopher et al., 2007; Clifford et al., 2007; and Stopher et al., 2008). As a result of the GPS-only household travel survey for the Greater Cincinnati Area Household Travel Survey (GCAHTS), the opportunity arose to assess the accuracy of the existing software. This was principally made possible by conducting a Prompted Recall (PR) survey on a subsample of GPS households. From this survey, possible issues in the software were identified and changes made to the software to overcome them. Initially, the accuracy of the software was assessed in the pilot survey (Stopher et al., 2010), which was repeated midway through the main survey and again at the conclusion of the main survey.

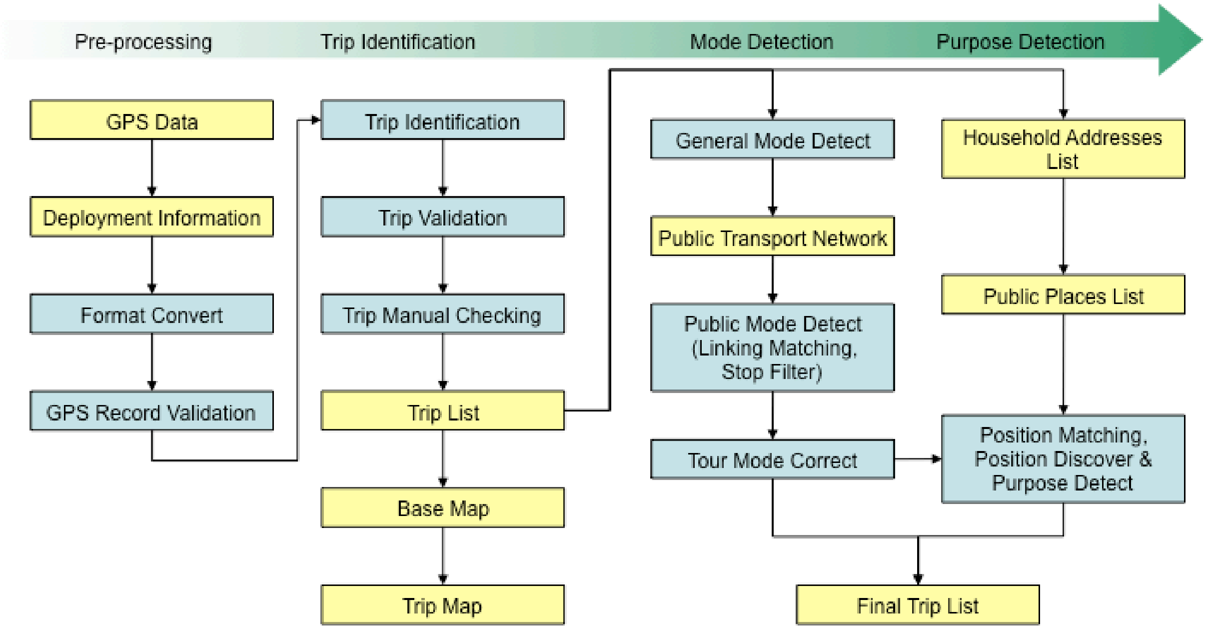
2. The G-TO-MAP software

The ITLS software is known as G-TO-MAP (GPS Trips, Occupancy, Mode and Purpose). It is written using a combination of the Python and GISDK development platforms, and operates on the TransCAD® software platform. TransCAD is used for the purposes of mapping and editing, and as a means to check for the accurate representation of travel. TransCAD manages the GIS layers needed in the processing, as well as displaying the resulting travel records. G-TO-MAP also permits output of maps to GoogleEarth®. G-TO-MAP utilizes the following data from the GPS-PPAL (the GPS Personal Passive Activity Logger):

1. Date
2. Time
3. Position (latitude, longitude)
4. Speed (km/h)
5. Heading
6. Horizontal Dilution of Precision (HDOP)
7. Number of satellites in view

The steps involved in getting GPS data from its raw format to a format suitable for analysis are illustrated in Figure 1, where the yellow boxes are inputs and outputs, and the blue boxes are processes.

Figure 1: Process to analyse GPS data



The procedure combines a number of automated processes with a few manual ones (described in more detail below). The strength of this is that the manual processes are applied early in the processing during the Trip Identification stage, which helps to filter out possible erroneous trips that may have been overlooked by the software, before imputation methods are applied. The weakness is that the manual process requires trained staff to inspect each day of travel.

The imputation method uses a number of complex rules or heuristics combined with a probability matrix and GIS layers, similar to those reported by Bohte and Maat (2009), Chung and Shalaby (2005) and Tsui and Shalaby (2006). Tsui and Shalaby (2006) reported using fuzzy logic for imputation in a test setting and compared GPS/GIS imputation similar to G-TO-MAP with GPS/fuzzy logic imputation and showed little difference in imputation results between the two methods. More recently, Moiseeva, Jessurun and Timmermans (2010) reported on using Artificial Intelligence for mode and purpose imputation but this method is still purely at a research stage.

It must be remembered that the GPS device continues to record for at least 3 minutes after a person stops moving. This is necessary to not lose other valuable information and also to prevent the potential loss of position every time that a traffic stop occurs. For the processing, it is necessary to have GIS layers for the street network, the bus routes, (and train, when this is a potential mode), bus stops (and train stations), and, when possible, land use by parcel. Further, to assist in purpose identification, workplace and school addresses, and addresses of the two most frequently-used grocery stores are collected for each household. These addresses are geocoded at the outset of the process. The three principal processing steps are described in the next sections of the paper.

3. Trip identification

First, the process finds locations where there is no movement recorded for at least 120 seconds and also shorter stops followed by travel in the opposite direction. The latter identifies a stop, such as a pick up or drop off, mailing a letter, etc. The processing software also identifies, where possible, a change in average speed as a trip end, so that the trips are actually identified as trip segments (e.g., a walk to the bus stop is one trip segment, and travel on the bus is another trip segment).

At the end of this process, the identified trips are converted to traces on a TransCAD® map and are then visually examined by trained editors. A text file is produced summarising the information on each trip (start time and location, end time and location, duration of trip, average speed, distance covered). The maps and text files are used to check that the trip identification appears reasonable. Anomalies, such as where a trip appears to have ended in one location and the next trip starts at a different location are flagged by warning messages. The editors then make any necessary corrections in the text file and this information is passed back to the software.

3.1 Improvements made to trip identification

Using information from the PR survey and in-depth analysis of the processing results, a few improvements were made to the Trip Identification process. The first improvement related to the "clouds" of data points that occur when the person carrying the device is more or less stationary. Previously, these clouds (which arise because the GPS continuously solves its position, so that the position keeps moving around) had to be identified and edited manually. Instead, a module was added into the software to locate clouds and remove them, replacing them with a single point that corresponds to the estimated end of one trip and beginning of the next. This resulted in a decrease in the effort required to edit the data and improved the precision and consistency of the trip-end location.

Second, a module was added to map on Google Earth®, instead of TransCAD®. This provides improvement in the identification of trip ends, because Google Earth shows much more information around a potential trip end. It also allows more rigorous removal of "spurious traces", which occasionally occur from a stationary GPS device and usually run in straight lines. When these traces run across buildings, vegetation, etc., it is easier to see and eliminate them. Major intersections as well as potential traffic lights are also easier to visualise on Google Earth, based on the land use and density of an area, such as major shopping centres, highways, train stations, etc., making it easier to determine if a trip end is a traffic stop or a trip end. Third, information about any other household members that appeared to share the same trip or part of a trip were exported into the trip list used for map editing. This prevents inconsistent map editing, where two or more individuals in the same household apparently travelled together, and provides a further check on the map editing process. Fourth, the maps were regenerated from the edited trip list, after map editing, providing a means to double check that the map editing has been effective and is complete.

4. Mode of travel

Using the edited traces, the next step is to identify the mode of travel. This is done from the speeds and acceleration/deceleration and by checking with the underlying GIS layers that are used in the processing. Walk is identified first, because this is most easily identified due to low average speeds, low rates of acceleration and deceleration, and because the trip does not necessarily stay on the road network.

Second, if appropriate, rail trips are identified because they are aligned with the rail lines. If the rail system is partly or entirely underground, trips disappear at station locations and then reappear at station locations, or disappear at tunnel entrances and reappear at the tunnel exits.

Next, trips by bus are identified from the speed of the trips, the number of stops that occur, the location of stops, and the location of the trip itself, the last of which must be entirely along bus routes in the GIS layer. The trip also starts and ends at a bus stop location, and is preceded and followed by a trip by a different mode (usually walk).

The remaining trips are bicycle and car trips. All of the documented software procedures have noted difficulties in reliably separating bicycle and car. First, bicycle is excluded as a mode if the household owns no bicycles. For bicycle-owning households, the average speed, acceleration, and deceleration over the entire trip are used to help isolate bicycle trips. Difficulty in identifying bicycle trips remains one of the weaknesses of the software.

4.1 Improvements made to mode identification

Initially, the software used the fact that a trip began and ended at or close to a bus stop and that the route was predominantly along a bus route as the primary identification for bus. However, it was found that this still resulted in some car trips being identified as bus and too many bus trips being identified as car. Therefore, a modification was made requiring a minimum number of traffic stops that coincide with bus stop locations to be used to define a bus trip. Subsequent analysis showed that the stop search (which has to be circular in a GIS) was picking up bus stops on the wrong side of the road and even on cross streets. Hence, two further changes were made to this process: the search distance was reduced from 45 metres to 15 metres; and the search is restricted, so that only bus stops on the same side of the road as the traveller is moving can be picked up as qualifying bus stops.

In the preliminary results, too many bus trips and too many bicycle trips were being identified, where most such trips were car trips. An in-depth analysis of the GPS travel records showed that many of these bus and bicycle trips were appearing in the middle of a tour, where the rest of the tour took place either by walk or car. As a result, a new module in the processing links trips into tours and then examines the sequence of modes that have been identified for the tour. In the event that a single trip within a tour is identified as bus or bicycle, it is usually replaced by either walk or car, according to the modes of the other trips in the tour. The rules employed are shown in Table 1.

Table 1: Rules for replacing mode in tours

Tour Type	Mode Sequence	Corrected Mode Sequence
2 Trip Tour	Walk-Bicycle	Walk-Walk
	Bicycle-Walk	Walk-Walk
	Bicycle-Car	Car-Car
	Car-Bicycle	Car-Car
3+ Trips	Walk-Bicycle-Walk	Walk-Walk-Walk
	Car-Bicycle-Car	Car-Car-Car
	Bicycle-Car-Car*	Car-Car-Car
	Bicycle-Walk-Walk*	Walk-Walk-Walk

* Bicycle must be first in these sequences.

Another improvement was to identify school bus trips, by adding a module to the programming that specifically tests for a school bus, using several characteristics. These include multiple stops, and a destination in the morning and an origin in the afternoon at or close to a school. The nearest stop to the school must also exceed a certain distance, depending on the minimum distance for which school bus service is made available.

The 85^{th} percentile acceleration and deceleration rates are also now used to distinguish between bus, bicycle, and car. The software is provided with maximum rates achievable by each of bus and bicycle. If the 85^{th} percentile acceleration/deceleration rate is higher than the maximum for a bicycle or a bus, then the trip is identified as being by car; if it is above the maximum rate for a bicycle, but below the maximum for a bus, and other characteristics suggest that this may be a bus trip, then it is identified as bus.

Finally, the algorithm that uses the bus route network to detect a bus trip has been fine-tuned to improve detection of bus trips. This was necessary because signal inaccuracy along the route could have the effect of identifying the trace as not being along a bus route for a sufficient proportion of time.

5. Trip purpose

The final processing step is to identify the trip purpose. Purpose is identified mainly from the use of the collected addresses. Given the addresses collected in the survey, home, work, school, and grocery shopping should all be easily identified, and the distinction between home-based and non-home-based trips can be made. All other trip ends were originally classified as "other".

5.1 Improvements made to purpose identification

The capability to identify travellers from the same household, and thus identify a serve passenger purpose when the vehicle occupancy changes, was a capability that was added early in this exercise. The identification of change travel mode was also added as a purpose. These two purposes provided significant improvement to the software. By importing the mode identification results into purpose identification, the change travel mode activity could be identified correctly removing some of the "other" trip purposes. Similarly, by building a "travel companion" database, serve passenger trips could be identified and this purpose also taken out of "other".

In the earlier version of G-TO-MAP, if a person had a geocoded workplace then the software checked to see if any trip made by that respondent originated or terminated within 200 metres of the workplace geocode. If so, the activity at that trip end was designated work. If a respondent did not have a geocoded workplace, but was still recorded as a worker, then the software checked the length of time spent at work and also whether or not the respondent visited the same place on more than one day. In subsequent analysis, it was found that the distance limitation sometimes resulted in no work trip activity being identified for a worker with a geocoded workplace. The latter logic was then extended also to include those with geocoded workplaces.

Further analysis of both the PR data and the 2009 US National Household Travel Survey revealed that only school and work trips normally had an activity duration in excess of 4 hours (a very small number of work-related and school-related trips had a duration in excess of 4 hours, but less than 1 percent of any other purpose exceeded 4 hours). Based on this, a new rule was implemented that, if an activity time for a worker respondent (full-time, part-time, or volunteer) is in excess of 4 hours, then the activity is work; if an activity time for a student respondent (full-time or part-time) is in excess of 4 hours duration, the activity is school. This will misclassify a small number of activities, but was felt to be worthwhile to identify many more work and school activities correctly. It was also noted that a number of respondents to the PR survey had designated a pick up or drop off activity at school as a trip to school. Because of this, it was also defined that any activity at a school that lasted less than 15 minutes before noon, or 30 minutes after noon would be classified as a pick-up/drop-off activity. An analysis of the PR data and comparison to GPS records showed that family members travelling together often did not provide multiple GPS traces for the segment of travel made together with other household members. As a result, it has proven to be quite difficult to find where occupancy changes on multi-occupant trips, to identify a pick-up or drop-off activity.

6. Results from the software processing

In this section of the paper, some comparative results are provided from the software processing. The first set of results is from the pilot survey and represents the results with none of the improvements made. It included a very detailed analysis of several responses, which are quite illuminating. However, space does not permit description of these in this paper. Following this, initial modifications were made to the software and the newer version of the software was used on the full data set. This set of results is presented as the second set. Finally, the latest improvements were implemented in the software, these being principally the introduction of the improved search for bus stops along the route and the checking of the tour mode consistency, and the use of activity durations to identify work and school, and some serve passenger trips.

6.1 Pilot survey analysis

PR and GPS data were obtained from 46 individuals in the pilot for one day. Only trip identification was reviewed in detail from the pilot. From the pilot survey, it was apparent that the G-TO-MAP processing was largely correct and that two principle types of error arose. One of these was respondent error in completing the PR survey, especially where respondents joined together two individual trips into a single round trip, or did not include short stops as something sufficient to define a new trip. The second was map-editing error, which resulted in improved instructions and training to map editors, and also to providing the capability to present the trip data on Google Earth maps for editing. Mode and purpose analysis was not performed on the pilot data, because of the small number of trips available for it and the errors found in the prompted recall data.

6.2 Interim results

By early April 2010, 214 individuals had provided PR and GPS data for one day covering over 1,200 trips. There are a total of at least 1,200 trips in the records from these respondents. There were 20 cases where the respondent indicated that one or more trips were missing at the end of the day. Because it was not asked if this was one trip or several, it is uncertain about the exact number of trips in the data set.

6.2.1 Trip identification

Of the 214 respondents, 133 (62%) agreed exactly with the number of trips identified by the GPS processing software. After a detailed analysis of the individual trips of the remaining 81 respondents, it was found that most disagreements were a result of respondent error, particularly of joining individual trips into tours. In just 14 cases, it appeared that map editing had resulted in splitting a trip that should not have been split. These last are the only cases where it appears that the GPS data were potentially in error and represents an error rate of just 1.2 percent of the total trips. This represents a significant improvement on the Pilot Survey.

6.2.2 Mode analysis

There are 1,015 trips in the data file where a comparison could be made between the software result and the answers provided by respondents to the PR Survey, shown in Table 2. As can be seen, overall, the processing software is performing well. Summing drivers and passengers in private vehicles shows 885 private vehicle trips in the PR data, compared to 873 in the GPS processing. There are 12 bus trips in the PR data, and 8 were identified in the GPS processing, while there are 15 bicycle trips identified in the GPS processing and none in the PR survey.

Table 3 shows the comparison of the software classification with the PR survey results for those modes that are reported in both sources. The table shows that bus trips were frequently misidentified as car, while the trips identified as bus actually took place by car. A significant number of walk trips were incorrectly identified as car and vice-versa. All of the trips identified as bicycle were either walk or car, apart from one school bus trip. In total, the number of trips that were correctly identified were 916 out of 1068, or about 86 percent.

Table 2: Comparison of mode of travel between PR survey and GPS software

Mode of Travel	Prompted Recall	GPS Processing
Motor Vehicle (GPS)	N/A	873
Driver of Auto/Van/Truck (PR)	820	N/A
Passenger of Auto/Van/Truck/Motorcycle (PR)	63	N/A
Driver of Carpool (PR)	0	N/A
Passenger of Carpool (PR)	1	N/A
Driver of Vanpool (PR)	0	N/A
Passenger of Vanpool (PR)	1	N/A
Motorcycle/Moped (PR)	6	N/A
Bus (GPS)	N/A	8
Bus (Public Transport) (PR)	6	N/A
Demand Response Bus (PR)	0	N/A
School Bus (PR)	6	N/A
Taxi (PR)	0	N/A
Walk	85	105
Bicycle	0	15
Other	15	8
Unknown	13	6
TOTAL	1015	1015

6.2.3 Activity analysis

Again, there were 1,015 trips for which a comparison could be made between what the respondent reported in the PR survey and what the GPS software identified as the activity. The software classified only 435 of the 1015 origins correctly and 431 out of 1014 destinations correctly. This represents correct identification of only 43 percent of activities. Both of these are considered to be unacceptably low, (although comparable to what others have reported – see Bohte and Maat, 2009, among others) and they clearly indicate a need for major improvements to the software. Overall, Table 4 shows the outcome of the activity identification.

Table 3: Detailed comparison of mode identification between GPS and PR

Mode	Car	Bus	Walk	Bicycle	Other	Unknown	TOTAL
Driver of Car/Van/Truck	753	7	37	11	8	4	**820**
Passenger of Car/Van/Truck/Motorcycle	62	0	1	0	0	0	**63**
Motorcycle/Moped	5	1	0	0	0	0	**6**
Passenger of Carpool	0	0	1	0	0	0	**1**
Passenger of Vanpool	0	0	0	0	0	0	**0**
Bus (Public Transport)	5	0	1	0	0	0	**6**
School Bus	4	0	1	1	0	0	**6**
Walk	21	0	59	3	0	2	**85**
Other	10	0	5	0	0	0	**15**
Unknown	13	0	0	0	0	0	**13**
Total	**873**	**8**	**105**	**15**	**8**	**6**	**1015**

Within the comparisons shown in Table 4, there are a number of mismatches, where the Table might seem to imply a match. For example 79 percent of home activities are correctly identified, while 74 percent of work activities, only 15 percent of school activities, and 36 percent of shopping activities are correctly identified. School is most frequently confused with work. The mismatch on shopping is less hard to understand, because the information provided is only of the two most frequently used grocery-shopping locations, whereas there may be many trips to other shopping locations.

Table 4: Comparison of activities between PR and GPS software

Source	Home	Work	School	Social Rec	Shop	Other	Total
PR-Origin	277	163	36	51	142	346	1015
GPS-Origin	244	90	12	174	58	437	1015
PR-Destination	297	155	29	53	136	344	1014
GPS-Destination	243	87	13	176	57	438	1014

A detailed representation of the matching and mismatchng of the PR survey and the GPS software is shown in Tables 5 and 6. It should have been possible to identify correctly 464 trip origins and 456 trip destinations, but which were incorrectly identified. Correction of these alone would raise the correct identification to around 88 percent, which would be considered to be much more acceptable for purpose identification. Further improvements may be possible beyond that by using geographic files of land use, which have not yet been used because the files from the GCAHTS region were not suitable.

7. Findings from the study

The PR data show that a number of respondents are still confused about what constitutes a trip and insist on combining trips in a tour into a single trip. Also, there are a few cases where respondents have attempted to split a trip, but created a second trip of less than 1-minute duration, which may represent an error in trip splitting. Only one trip was added that was missed by the GPS, and the overall combination of map editing and software processing for trip identification has achieved a high level of accuracy.

Table 5: Comparison of PR and GPS software origin activities

PR Origin Activity	GPS Origin Activity						Total
	At Home	Paid Work	School	Social, Recreational, Church	Shop	Other	
Unspecified	0	1	0	0	0	12	13
At Home	191	2	1	31	6	46	277
Paid Work	5	66	1	13	3	75	163
School	4	1	2	12	1	16	36
Volunteer Work	2	0	0	1	0	6	9
Pick up/Drop Off person	4	3	6	8	4	17	42
Social, Recreational, Church	5	1	0	10	6	29	51
Catch a Bus, Train, or Airplane	1	0	0	1	0	2	4
Transfer from One Bus, Train or Airplane to Another	0	2	0	0	1	7	10
Shop	8	3	2	32	21	76	142
Personal Business	5	1	0	19	4	40	69
Eat Meal	5	2	0	18	4	25	54
Go for a Drive	2	0	0	0	1	6	9
Work Related	2	8	0	2	3	35	50
School Related	1	0	0	4	1	7	13
Don't Know/Refused	9	0	0	23	3	38	73
Total	244	90	12	174	58	437	1015

Table 6: Comparison of PR and GPS software destination activities

PR Origin Activity	GPS Origin Activity						Total
	At Home	Paid Work	School	Social, Recreational, Church	Shop	Other	
Unspecified	1	1	0	0	0	12	14
At Home	195	2	1	40	7	52	297
Paid Work	1	63	1	14	2	74	155
School	4	0	1	11	1	12	29
Volunteer Work	1	0	1	2	0	5	9
Pick up/Drop Off person	3	4	7	7	4	16	41
Social, Recreational, Church	6	1	0	10	5	31	53
Catch a Bus, Train, or Airplane	0	0	0	1	0	3	4
Transfer from One Bus, Train or Airplane to Another	0	1	0	0	1	6	8
Shop	5	3	2	30	20	76	136
Personal Business	5	2	0	17	5	40	69
Eat Meal	6	2	0	17	4	24	53
Go for a Drive	3	0	0	0	1	7	11
Work Related	1	8	0	3	3	34	49
School Related	1	0	0	2	1	7	11
Don't Know/Refused	11	0	0	22	3	39	75
Total	243	87	13	176	57	438	1014

Mode of travel is reasonably well identified, although there appear to be some problems with bicycle and bus trips in particular. Further work is needed to distinguish among car, walk, and bicycle, as well as improving the identification of bus trips. Correction of these problems will produce a much better result from mode identification. Discrepancies between walk and car may not be possible to eliminate, if these are predominantly short and slow trips, such as driving from one house to a nearby house in the neighbourhood.

Trip purpose is more troublesome, with a significant number of mismatches from the software that should be possible to eliminate. This was the primary subject of software improvement in the following months.

7.1 Final results

After completion of data collection and a first round of processing, an in-depth analysis of the results showed several problems: too few motor vehicle trips, too many bus and bicycle trips, too few work and school trips, and too few pick-up and drop-off trips were identified. After undertaking several modifications and improvements to the software, as described earlier in this paper, the following results were found.

7.1.1 Mode identification

Table 7 provides a breakdown of trips by mode from the total completed sample and also provides a household trip rate by mode from the GPS data after processing. As expected, Table 7 shows that the majority of trips recorded were by car. Slightly less than 5 percent of trips could not be identified to a specific mode, usually because this was a trip inserted in map-editing for which there was no trace and therefore mode identification information was missing. (It is planned to reduce this in the case of one-way trips by imputing the mode from the previous or following trip.)

Table 7: Breakdown of GPS trips by mode and daily household trip rate by mode (complete households)

Mode of Travel	Number of Trips	Percent of Trips	Daily Household Trip Rate
Motor Vehicle	53,734	88.2	7.60
Bus	537	0.9	0.08
Walk	3,125	5.1	0.44
Bicycle	585	1.0	0.09
School Bus (GPS and PR)	247	0.4	0.03
Unknown	2,672	4.4	0.38
TOTAL TRIPS	60,900	100.0	8.62

An analysis of the PR data showed that these results from the processing are very close to the percentages in the PR data. The percentages of motor vehicle trips (90.6%) and bus trips (1.3%) are higher in the PR survey results than in the GPS survey, while bicycle trips are higher in the GPS survey results than the PR survey (0.5%). Walk trips show approximately the correct percentage, based on the PR survey (5.0%). The overall trip rate by households in the PR survey is lower than for the GPS survey, partly because of fewer persons per household completing the PR survey, and possibly also because those who did complete it may have been those with fewer trips in the sampled day. While bicycle trips are still overestimated, they are overestimated by a relatively small margin. The lower rate of bus trips shows that the software changes have largely solved this problem. However, the percentage in the GPS data seems quite consistent with other bus ridership statistics for the region.

It should also be noted that an in-depth analysis of a small sample of PR trips (429 trips) showed that about 36 percent of trips had a questionable mode of travel identified, while in 12 percent of cases, respondents failed to identify the mode of travel. These errors in the PR results would suggest that the processed results might be more accurate than the analysis here has indicated. Indeed, looking at detailed path information, speeds, and other trip information, only 3 percent of the cases appeared to be an error in the software processing, suggesting an overall accuracy of mode identification on the order of 97 percent.

7.1.2 Purpose identification

Table 8 shows the distribution of origin and destination activities and rates of these per household per day from the GPS survey. With the most recent version of the software, work, school, and pick up/drop off trips increased as a percentage of total trips. Shopping trips remained approximately the same.

Table 8: Breakdown of trips by origin and destination activity and by household rate from the GPS survey (complete households only)

Purpose	Origin Activity			Destination Activity		
	Number	**Percent**	**Daily Rate**	**Number**	**Percent**	**Daily Rate**
At Home	15,419	25.3	2.18	15,155	24.9	2.15
Paid Work	7,308	12.0	1.03	7,406	12.2	1.05
School	1,871	3.1	0.27	1,905	3.1	0.27
Pick Up/Drop Off	2,243	3.7	0.32	2,224	3.7	0.32
Catch Bus/Train/Plane	1,552	2.5	0.22	1,548	2.5	0.22
Shop	14,452	23.7	2.04	14,457	23.7	2.04
Other	18,055	29.6	2.55	18,205	29.9	2.58
Missing	0	0	0	0	0	0
Total	**60,900**	**100.0**	**8.62**	**60,900**	**100.0**	**8.62**

Work activities are still lower than those shown in the PR data, but this is not surprising, considering the differences in weekend days in the data (almost none in the PR data and about 5 percent in the full GPS data). The only activity that seems substantially higher is that of shopping trips. There are likely to be at least two reasons for this: in the PR survey, respondents often concatenated successive trips into a "tour", which would be likely to remove shopping as an activity from all "tours" that involved travel from one shop or shopping complex to another; and some of the trips identified as shopping trips would also be work trips for sales clerks who work less than a four-hour shift, where the identification rules would not have picked these trips up as work trips, if an accurate geocode for the shopping location had not been provided. Also, where the identified grocery stores are in a shopping complex or shopping centre, these trips may include trips for personal business, eating a meal, and other similar purposes. Overall, therefore, it is considered that the modified purpose identification has improved the accuracy of the purpose identification, without the use of any land-use information, to a quite high degree.

A similar in-depth analysis of over 400 trips showed that about 23 percent of identified purposes in the PR data were questionable, while also identifying only 11 percent as ones where it appeared that the software erred. Based on this analysis, the accuracy of the purpose identification is about 89 percent.

8. Conclusions

From the results documented in this paper, it is apparent that some substantial improvements have been made to the processing software, G-TO-MAP. The results of these improvements suggest that trip identification is between 95 and 99 percent accurate, mode identification is between 90 and 97 percent accurate, and purpose identification is 85 to 90 percent accurate. These levels of accuracy probably exceed those associated with self-report data by a substantial margin, although accuracy levels of such data have never been analysed in detail in reported research. However, given that research shows that there is usually a 10-20 percent under-report of trips from self-report surveys, it is almost certain that the accuracy of G-TO-MAP is far superior.

This suggests that, if more detailed mode and purpose information is not required for modelling, it is now possible to collect GPS data and process the data to produce information that is suitable to support travel-demand modelling activities. This is even more the case, given the higher quality of the time and distance information contained in the GPS data. Further improvements are also possible in the processing software that should be able to improve these results beyond these levels of accuracy, especially on mode if the inserted trips are provided with mode information, and on purpose if detailed land use data are available at the parcel level.

8.1 Future research

Apart from use of GIS parcel land use data, the primary improvement in the software would be to use polygons rather than points in purpose identification. At present, the software uses a search distance of 200 metres around each address location (e.g., home, workplace, school, etc.). Enlarging this diameter is not the solution, because this will lead to losing some short trips, and also will compound the misidentification process. However, replacing the point information with a polygon that aligns with the boundaries of the developed parcel would solve the problem.

In common with most other processing algorithms, the G-TO-MAP software has difficulty distinguishing between car and bicycle and sometimes bus. In the future, it would be useful to include in the web/telephone recruitment survey questions to ascertain frequency of bicycle and bus use by a household, in addition to ownership of bicycles. Some additional questions that would be extremely helpful in improving the mode identification are:

1. Number of bicycles in working order available to members of the household;
2. Respondents who used a bicycle in the past week;
3. Respondents who used a public bus in the past week; and
4. Respondents who used a school bus in the past week.

By asking these questions, it may be possible in many instances to rule out that any trips by a household are by bicycle or by bus or by school bus, and conversely, to identify households where trips by one or more of these modes are likely to be found. This would improve the accuracy of identification significantly.

References

Bohte, W. and K. Maat (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. Transportation Research Part C 17, pp 285–297.

Chung, E.-H. and A. Shalaby (2005). A trip reconstruction tool for GPS-based personal travel surveys. Transportation Planning and Technology 28 (5), 381–401.

Clifford, E., J. Zhang and P. Stopher (2007). Determining Trip Information Using GPS Data, paper presented to the IGNSS 2007 Symposium, Sydney, December.

Moiseeva, A., J. Jessurun and H. Timmermans (2010). Semi-automatic Imputation of Activity-Travel Diaries Using GPS Traces, Prompted Recall and Context-Sensitive Learning Algorithms. Transportation Research Board 89th Annual Meeting.

Stopher, P., Q. Jiang, and C. FitzGerald (2005). Processing GPS Data from Travel Surveys, paper presented to the 28th Australasian Transport Research Forum, Sydney, September.

Stopher, P.R., E. Clifford, J. Zhang, and C. FitzGerald (2007). Deducing Mode and Purpose from GPS Data, paper presented to the 11th TRB National Planning Applications Conference, Daytona Beach, FL, May.

Stopher, P.R., E. Clifford, J. Zhang, and C. FitzGerald (2008). Deducing Mode and Purpose from GPS Data, paper presented to the 87th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2008

Stopher, P. C. Prasad, and J. Zhang (2010). Comparing GPS and Prompted Recall Data Records, paper presented to the WCTR Conference, Lisbon, July.

Tsui, A. and A. Shalaby (2006). Enhanced system for link and mode identification for personal travel surveys based on Global Positioning Systems. Transportation Research Record 1972, 38–45.