Measuring Bus Performance using GPS Technology

By

Philip Bullock and Qingjian Jiang

August 2003

INSTITUTE OF TRANSPORT STUDIES

The Australian Key Centre
in Transport Management

The University of Sydney
and Monash University

NUMBER:	Working Paper ITS-WP-03-15
TITLE:	Measuring Bus Performance Using GPS Technology

ABSTRACT: Assessing the running times of bus services has traditionally been a difficult and expensive task for the majority of bus operators in Australia, and in other parts of the world. Up until recently, travel times have been collected by time keepers positioned at key points along a given route or service corridor who record bus arrival and departure times. These data then need to be manually collated before any kind of analysis can be undertaken. The time consuming nature of this process restricts the ability of operators to collect large and meaningful samples of data. Furthermore, it is difficult, if not impossible, to identify congestion points from such data, and to evaluate the impact that they might have on overall service levels.

Passive Global Positioning System (GPS) technology offers a low-cost means of collecting large amounts of highly accurate data, which can be used in an on-going performance assessment program. Although raw GPS track points can be viewed on most standard GIS packages, on-screen visual analysis is extremely time consuming for even small amounts of data. Programming skills are therefore required to break continuous GPS data into records that are more meaningful to an operator. A number of important tasks need to be undertaken before analysis can take place. Firstly, periods of in-service or out-of-service running need to be defined, and routes need to be identified. This can be a complicated task because operators often design shifts so that buses may switch between different areas and routes, from run to run, to maximise vehicle utilisation. Once routes are identified, records must then be separated into individual runs and matched with a timetable to compare scheduled and actual running times. This paper provides an overview of a number of software applications developed for processing and analysing large GPS data records collected by a bus operator in Sydney in late 2002/early 2003. The data collection process is described, and some examples are presented of output produced by the main trip processing and timetable query program. It is concluded that passive GPS is a highly attractive method of collecting data on performance, even for very small operators.

KEY WORDS:	Global Positioning Systems (GPS), bus performance, on-time running, Geographic Information Systems (GIS)

AUTHORS:	Philip Bullock and Qingjian Jiang
CONTACT:	Institute of Transport Studies (Sydney & Monash) The Australian Key Centre in Transport Management, C37 The University of Sydney NSW 2006, Australia Telephone: +61 9351 0071 Facsimile: +61 9351 0088 Email: itsinfo@its.usyd.edu.au Internet: http://www.its.usyd.edu.au

DATE:	August 2003

1. Introduction

This paper provides an overview of a pilot Global Positioning System (GPS) research project undertaken in late 2002/early 2003 by The Institute of Transport Studies (ITS), The University of Sydney and a bus operator in Sydney. The aim of the project was to develop a cost effective Geographic Information System (GIS) based program to process and analyse Global Positioning System (GPS) data collected on buses operating on a specific route. A secondary aim of the project was to develop a program which could automatically geocode bus stops using GPS data collected from buses, with little or no other input data.

The paper presents an overview of the steps taken to collect the input data used in the project, and details the trip processing and timetable query program developed for processing and analysing the GPS data. An overview is provided of a GIS data visualisation tool and data analysis spreadsheet developed as part of the project. Some examples are presented of output produced by the main trip processing and timetable query program, as well as some of the ways it can be used by schedulers. The bus stop detection program is not reported on in any detail here, but will be the subject of a forthcoming paper. It is concluded that passive GPS is a highly attractive method of collecting data on performance, for operators of all sizes.

1.1 Background: Difficulties of measuring on time running

Assessing the running times of bus services has traditionally been a difficult and expensive task for the majority of bus operators in Australia, and in other parts of the world (Kharola *et al*., 2002). Up until recently, travel times have been collected by time keepers positioned at key points along a given route or service corridor who record bus arrival and departure times. These data then need to be manually collated before any kind of analysis can be undertaken. The time consuming nature of this process restricts the ability of operators to collect large and meaningful samples of data, which could be used to improve timetables and overall service levels. Using this traditional approach, it is difficult if not impossible to identify congestion points from such data, and to evaluate the impact that they might have on overall service levels.

Global Positioning System (GPS) technology offers a low-cost means of collecting large samples of highly accurate data, which can be used as part of an on-going performance assessment program. Although GPS is clearly the best method for collecting travel time data, the rapid pace of change in GPS technology makes investment decisions difficult for many bus operators. In the past few years, a number of sophisticated 'on-line' systems have been developed for tracking vehicle movements in real time, providing information to customers about bus arrival times and for allocating priority at traffic lights (GPS Online 2000, Morehead 2001, Infodev, 2003, NextBus, 2003). Such applications are not cheap to develop however, and may cost in the vicinity of hundreds of thousands, or even millions of dollars. In Auckland, New Zealand , a large scale real time passenger information/bus priority system is being developed that will involve fitting more than 700 buses with GPS equipment, providing on-street variable passenger information displays and modifying traffic lights. The estimated cost of this project is $NZ7 million (Auckland City Council, 2003).

'Passive' GPS technology operating independently of other systems represents a practical, low cost method for collecting travel time data. Over the past few years, GPS technology has improved markedly and accurate GPS data loggers have become very affordable. These devices are well suited to collecting travel time data from buses, irrespective of fleet size. Unless the objective is to manage the movements of an entire fleet, there would be very few benefits in having every bus collecting GPS data. Because a single GPS device can collect tens of thousands of records in the space of just a few days, a situation where several thousand buses might be simultaneously collecting data would create considerable difficulties with data management. An added attraction of using data loggers is that they are portable, and can therefore be moved between buses operating on different routes, and in different regions. Other than a major study undertaken by Kharola *et al.*, (2002) in Bangalore, India, it appears that little work has been done to date using 'off-line' systems to collect GPS data on buses.

Despite the advantages of passive GPS, there are still some practical difficulties that need to be overcome when using portable data loggers. These practical difficulties stem from the fact that the devices collect GPS data independently of other systems within the bus, such as the on-board ticketing system. Output files from data loggers therefore provide continuous streams of spatial and temporal data (i.e., geographic co-ordinates, time and date), but no other meaningful reference information (such as route the bus was operating on, trip start and end times, shift changes, etc.). Although GPS track points can be viewed on most standard GIS packages, on-screen visual analysis is extremely time consuming for even small amounts of data.

Programming skills are required to break continuous GPS data into records that are more meaningful to an operator. A number of important tasks need to be undertaken before analysis can take place. Firstly, periods of in-service or out-of-service running need to be defined, and routes need to be identified. This can be a complicated task because operators often design shifts so that buses may switch between different areas and routes, from trip to trip, to maximise vehicle utilisation. Once routes are identified, individual trips must then be matched with a timetable to compare scheduled and actual running times.

2.   The Input Data

Three main sources of data were required to develop the programs: bus stop coordinates, timetable information, and in-vehicle GPS data collected from the study route. The following sections describe the methods used to collect and edit the input data. All GPS data used in this project were collected using GeoLogger® passive GPS data loggers, provided by U.S. company, GeoStats. The Geologgers were fitted with Garmin GPS receivers which have an accuracy rating of ±15 metres, although the experience of ITS is that on average it is closer to ±5 metres. All GIS program were developed using the GISDK programming language in Caliper Corporations TransCAD® package.

2.1   Bus Stop Data

Although a database of bus stop locations is held by the NSW Department of Planning, Infrastructure and Natural Resources (DIPNR), these data were not considered to be

sufficiently accurate for this project. Many of stops contained in this database use unprojected latitude and longitude formats, and records have generally not been updated on a regular basis and therefore do not take account of stops that have been moved over time.

Bus stop coordinates were collected by the bus operator in late November 2002, using a data logger and a company vehicle. The driver travelled along the route in both directions and stopped at each bus stop for 1-2 minutes. The data logger was set to poll every second, and the driver of the vehicle recorded the arrival time at major timing points along the route on a clipboard. The GPS data were downloaded by and sent to ITS to be cleaned and put into GIS layers.

The geocoding process itself was relatively straight forward. The bus stops were added to a new point layer by matching the times shown on the GPS points with those recorded by the driver. In most cases it was fairly obvious where the stop was because track points were clustered around areas where the driver stopped. Only a few stops needed to be cross checked with the in-vehicle data collected from the buses. The depot used by buses operating along the study route was also located. The new layers were then loaded on to a base map of Sydney (Sydney Streetworks, Version 6).

2.2 Timetable Data

Timetable data were generated from the scheduling software used by the bus operator and saved in Excel spreadsheets. A number of minor modifications were required to convert the data into a format that could be recognised by the GIS program. Numerical values stored as times were converted to integers, and columns and rows were transposed, so that each row of the table represented a trip, with the columns ST1, ST2, etc. representing the scheduled arrival times at timing point points along the route. The modified inbound and outbound timetables were then saved as dbase files, in a form shown in Table 2.1.

Table 2.1: Extract of Timetable Re-formatted for Program

ID	RUN_ID	GS	ST1	ST2	ST3	ST4	ST5	ST6
1	1	43300		45000	45300	50000	50200	50800
2	2	50300		52000	52300	53000	53200	53800
3	3	51600	54000	55000	55300	60000	60200	60800

2.3 In Vehicle Bus Data

Data were collected from four buses, starting in late December, and finishing in mid March. In terms of program design, it was only necessary to collect two or three weeks worth of data, however it was decided early in the project to extend the survey period so that data could be collected outside the Christmas/school holiday period.

Four buses operating principally on the study route were fitted with data loggers. Because the devices were designed to be plugged into the cigarette lighter of an ordinary motor vehicle, some modifications needed to be made to the power cords so that they could be plugged into the AV accessory outlets of the buses. Other than this,

the devices were relatively easy to install. The GPS antennas could be easily attached to the roofs of the buses because of their magnetic bases. It was not known what polling rate would be most suitable, so two of the devices were set to record data on one second intervals, while the other data loggers were set on five seconds. Since the project was mainly focused on development of methodology, and because only one route was considered, a formal sampling plan was not considered necessary.

The cigarette lighter sockets in most ordinary motor vehicles become inoperable when the ignition it turned off, and so the data loggers do not collect data when the vehicle is stationary. However, the accessory outlets in buses are constantly powered and so the data loggers were in operation 24 hours a day for the entire data collection period of this project. Because of this, the data loggers polling every second needed to be downloaded and cleared every few days.

Because the Data loggers were in operation 24 hours a day, a very large amount of data was collected while the buses were stationary at the depot. This meant that up to two thirds of the data were in fact not relevant to the program developed. This extra data made early prototypes of the program very slow. A stand alone application was therefore created to clean the data before it was processed and analysed by the main program.

3.   The Software

3.1   Trip Processing Algorithm and Timetable Query

3.1.1 Overview

The trip processing algorithm and timetable query was the core program designed to generate travel time output from the GPS data files. There are essentially three main tasks performed by the algorithm within the program: firstly, continuous GPS records are broken into separate blocks of records, or basic trips (trip definition). Secondly the program examines these basic trips, and determines the type of trip made i.e., whether or not the bus was servicing a specific route, or if the bus was involved in some other activity such as driving to or from the depot or the start or end of a route (trip type definition). At this stage, the program examines any trips made on the study route and provides output on running times and travel times between timing points. In the third part of the program, GIS maps, layers, and selection sets are created so that processed data can be viewed and analysed by the user. The following sections provide an overview of each of these main functions.

3.1.2 Trip Definition

There are three criteria used to break continuous data into basic trips. In deciding where to insert a break point, or trip end, the program examines:

1. Whether or not records appear in one of three areas: the Depot (Depot), and the two end points of the route (ST1 and ST17);
2. The number of bus stops travelled through on the study route and;
3. Any reversal in the direction of travel

In the first step taken in the trip definition process coordinates of depot and bus stops are loaded into a temporary array. All GPS records are then matched with inward stops (initially, it does not matter whether inward or outward stops are used) and an additional array is created identifying GPS records that are located within 50 metres of a bus stop, and 120 metres of the depot. When more than one point is located within the radius, the ID of the closest point to the centre is recorded in the array. GPS records are then sorted into a separate group within the GIS layer (selection set) for each day.

For each days worth of records, the program searches for points that start or end the depot, stop 1 (ST1) or stop 17 (ST17). In other words, if the first record of the day is found within the depot, the program then looks for the next location that subsequent records appear in. There are three possible locations considered – the Depot, ST1 or ST17 (if more routes were defined within the program structure, more end points would be searched). If a bus drove from the Depot in the morning to one end of the route, stopped briefly to pick up passengers and then made a scheduled trip along the study route, the Depot would be the first location marked, ST1 the second, and ST17 the third.

Within each combination of the three locations (depot – ST1, depot – ST17, ST1 – ST17, etc.), a separate series of sub commands examines the number of stops passed, to determine the likely trip ends. Figure 3.1 provides an illustration of how this process works. If a bus travelled from ST1 to ST10 and passed through 15 to 17 stops, this means that the bus has travelled along study route without deviation and so the trip end is defined as ST17. If the number of stops is less than 15, this means that the bus has travelled only part of the route and so the trip is coded as a 'special trip'. These are actually quite rare and generally only occur when buses run back to the depot via some alternate route to save time. If the number of stops exceeds 17, then this means that it is likely that the bus has made more than one trip, and so the trip end is defined as the point that the bus changes its direction of travel (a change from inbound to outbound). A similar sequence of commands is used to examine records between each combination of stops.

Figure 3.1: Process Used to Define Basic Trips
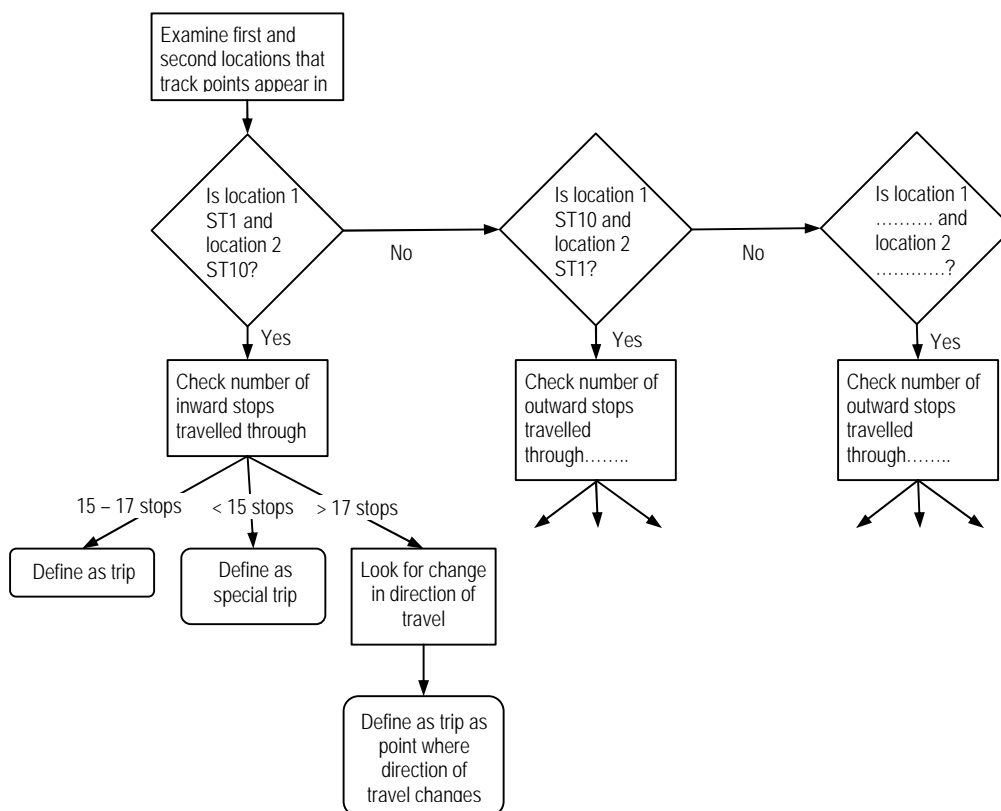
3.1.3 Trip Type Definition and Timetable Query

Once the program has flagged the likely start and end points of trips, the algorithm then defines the type of trip made. Beginning the first trip of the first day, the program examines each set of records and classifies them into one of the following categories: Route A (main study route), route B, route C, trips out from the depot (O_Depot), trips into the depot (I_Depot), trips made out from the depot and straight back to the depot without stopping (D_2_D) and unknown trips (UKNOWN). The direction (inbound or outbound) is also determined for each route. A set of conditions are used to define inbound and outbound trips for each trip type. For instance, Route A inbound trips must start at ST1, finish at ST17, and pass through all major timing points, and match a scheduled trip, an Out_Depot trip must start at the depot and finish at the start point of another route etc.

Whenever the program detects a trip made along the study route, a sub-routine assesses on-time running and measures the time taken to travel between timing points. On-time running is measured by comparing the time the bus arrived at timing point (GPS time) with the time that the bus was scheduled to arrive (timetable time). This requires each GPS trip made along the study route to be correctly matched with trips shown on the timetable. From the data observed as part of the validation procedure, it was noticed that most buses tend to start within just a few minutes of their scheduled start time, and so in most cases, it appeared quite easy to determine which GPS trip belongs to which timetable trip.

Nevertheless some rules were devised to govern this process to control for any cases where there might be a larger discrepancy between actual and scheduled times. In cases where the GPS start time was no more than 5 minutes before the scheduled time, it was assumed that the bus started early, and the trip was therefore matched with the proceeding timetable trip. When GPS start times were more than 5 minutes before the scheduled start time, it was assumed the bus was running late, and the preceding timetable start time was used. For example, if Route A outbound trips were scheduled for 15:00 and 15:30, and a Route A GPS trip started at 15:26, it was assumed that the bus started slightly early, and the 15:30 time was assumed to be the scheduled start. This rule appeared to work quite well and allowed for differences between the time on the ticketing system (or drivers time) and the GPS data time. It appears rare for a bus to start more than 5 minutes before its scheduled start time, and so this rule did not appear to result in any mismatches.

Once the GPS and timetable start times have been matched, the program then examines the time the bus arrived at each timing point, and calculates the difference between the GPS arrival time, and the scheduled arrival time. Travel times are also calculated between each set of timing points.

3.1.4 Creation of Maps and Output Files

The program opens a base map stored in specified directory, and imports the GPS data in the form of a single GIS point layer. Within this layer, each trip is marked within a selection set. Seven different output files are produced for each data set processed – a trip summary file ([Busno]_[date1]_[date2]_C_SUMMARY), and three different output files for both inward and outward directions:

➢ A timing check file ([Busno]_[date1]_[date2]_C_IN_TRIPS.DBF)
➢ Travel time file ([Busno]_[date1]_[date2]_C _IN_TRIPS_TT.DBF)
➢ Combined timing check and travel time file ([Busno]_[date1]_[date2]_C _IN_TRIPS_ALL.DBF)

All map files and output files generated by the program are saved in a single directory with the same name as the original CSV file. Table 3.1 shows a selection of data contained in the trip summary output file. Start and end times are shown for each trip as well as the time that the bus was stationary between trips (lay-up time). Scheduled travel times are shown for trips that were made along the study route (Route A).

Table 3.1: GPS Data Summary File

Date	Route	Dir.	Start time	End time	Layup time	Travel time	Scheduled travel time
161202	O_Depot	Out	6:39:58	6:54:23		00:14:25	
161202	Route C	In	6:54:32	8:16:23	0:00:09	01:21:51	
161202	Route C	Out	8:40:13	10:05:36	00:23:50	01:25:23	
161202	I_Depot	In	10:06:11	10:15:16	0:00:35	0:09:05	
161202	O_Depot	In	16:26:29	17:18:11	06:11:13	00:51:42	
161202	UNKNOWN	Out	17:19:06	18:33:47		01:14:41	
161202	I_Depot	In	18:34:28	18:43:08	0:00:41	0:08:40	
171202	O_Depot	Out	5:12:08	5:22:08	10:29:00	00:10:00	
171202	Route A	In	5:22:13	6:33:19	0:00:05	01:11:06	01:18:00
171202	Route A	Out	6:57:29	8:30:35	00:24:10	01:33:06	01:43:00
171202	Route A	In	8:59:56	10:50:58	00:29:21	01:51:02	01:38:00
171202	Route A	Out	11:00:23	12:45:10	0:09:25	01:44:47	01:43:00
171202	Route A	In	13:02:11	14:45:53		01:43:42	01:38:00
171202	Route A	Out	14:54:38	16:38:55	0:08:45	01:44:17	01:43:00
171202	Route A	In	17:02:55	18:43:56	00:24:00	01:41:01	01:38:00
171202	Route A	Out	18:57:21	20:30:57	00:13:25	01:33:36	01:33:00
171202	I_Depot	In	20:35:12	20:53:37	0:04:15	00:18:25	
181202	O_Depot	Out	6:47:20	7:05:40	09:53:43	00:18:20	
181202	Route A	In	7:06:05	8:46:41	0:00:25	01:40:36	01:38:00
181202	I_Depot	Out	8:55:41	10:14:51	0:09:00	01:19:10	
181202	O_Depot	In	16:05:46	17:16:52	05:50:55	01:11:06	
181202	UNKNOWN	Out	17:18:32	18:32:32		01:14:00	
181202	I_Depot	In	18:33:07	18:39:17	0:00:35	0:06:10	
191202	O_Depot	Out	5:53:06	6:11:06	11:72:89	00:18:00	
191202	Route A	In	6:11:31	7:45:10	0:00:25	01:33:39	01:28:00
191202	Route A	Out	7:56:41	9:41:17	00:11:31	01:44:36	01:43:00
191202	Route A	In	10:02:22	11:50:48	00:21:05	01:48:26	01:38:00
191202	Route A	Out	11:58:38	13:52:30	0:07:50	01:53:52	01:43:00
191202	Route A	In	14:01:15	15:39:56		01:38:41	01:38:00
191202	Route A	Out	15:54:26	17:46:43	00:14:30	01:52:17	01:43:00
191202	I_Depot	In	17:46:57	18:04:28	0:00:14	00:17:31	
201202	O_Depot	Out	6:38:01	6:54:51	12:92:73	00:16:50	
201202	Route A	In	6:54:56	8:24:46	0:00:05	01:29:50	
201202	Route A	Out	8:44:01	10:19:13	00:19:15	01:35:12	
201202	I_Depot	In	10:20:38	10:29:23	0:01:25	0:08:45	

Table 3.2 shows a sample of output generated from the timetable query. For each trip made by the bus along the study route, the program records some basic descriptive information such as date, day of the week, start and end times of the trip, and the timetable day (i.e., if the day was a weekday, Saturday, Sunday or public holiday). A sample of the travel time information recorded by the program is shown on the right hand of the table. The columns with single timing point names (ST1, ST2, etc.) show the difference between the scheduled arrival time, and the actual GPS arrival time for ach of the timing points along the study route. The columns with multiple timing points (ST1_ST2, ST2_ST3 etc.) show travel times recorded by the GPS between timing points.

Table 3.2: GPS Travel Time Output File

TRP_ ID_S	ROUTE_ NO	DATE	WEEKDAY	TT_ DAY	S_TIME	S_TIME_S	ST1	ST1_ ST2	ST2	ST2_ ST3
5	Route A	281202	Saturday	2	6:33:46	6:30:00	3:46	7:24	1:10	3:06
9	Route A	281202	Saturday	2	12:51:28	12:50:00	1:28	10:09	:37	3:46
11	Route A	281202	Saturday	2	16:51:58	16:50:00	1:58	14:30	5:28	5:00
13	Route A	281202	Saturday	2	20:52:22	20:47:00	5:22	6:15	3:37	2:05
20	Route A	291202	Sunday	3	9:51:12	9:50:00	1:11	8:42	-1:07	5:00
23	Route A	291202	Sunday	3	16:02:30	15:50:00	12:30	14:25	15:55	4:21

3.1.4 Testing the Procedure

Following the completion of the main programming work, tests were run to check for errors and ensure the program was correctly interpreting the GPS records. The programs were validated using eight GPS data files collected from December 2002 to March 2003. Two files were selected from each of the four buses that collected data, to ensure an even spread of dates and a balance between the various polling rates.

GPS summary files and timing check files were compared with fare collection data reports provided by the bus operator. These reports were generated from data downloaded from driver 'smart cards', and represent a record of actual schedules (as opposed to planned schedules developed in the scheduling software).

Overall, the trip detection algorithm worked well. Table 3.3 shows a breakdown of trips detected by the trip processing algorithm for the eight data files. Overall, 251 trips were detected, 96 of which were Route A trips. Route B and Route C trips comprised 11 percent of the trips detected by the program while around 100 trips were made to and from the depot. There were three trips made where the bus left the depot, drove two or three blocks, and then drove straight back to the depot. It is not clear from looking at the data why such trips were be made.

Table 3.3: Trips Detected by Trip Processing Algorithm

Trip Description	No. of trips	Trips as % of total	No. of undetected/ misclassified trips	Undetected/ misclassified trips as % within group
Route A	96	38.2%	0	0.0%
Route B	2	0.8%	0	0.0%
Route C	25	10.0%	0	0.0%
In Depot	50	19.9%	7	12.0%
Out Depot	51	20.3%	5	9.8%
Depot - Depot	3	1.2%	0	0.0%
Unknown	24	9.6%	24	100.0%
Total Trips	251	100.0%	35	13.9%

No errors or inconsistencies were found in any of the 123 trips coded as Route A, B or C, which suggests the program interpreted the data very well. Table 3.5 also shows the number of trips that went undetected or were misclassified by the program. A total of 12

trips were misclassified as either in – depot or out – depot. Of the seven trips within the in – depot group, four were actually Route A trips which appeared to end prematurely. The remaining three trips misclassified as in-depot incorporated travel made on routes not defined within the program structure, and therefore could not be correctly interpreted. Five out – depot trips also incorporated travel on a number of routes which were not defined within the program structure.

A total of 24 trips went undetected by the program, and were coded as unknown. Table 3.4 provides an explanation of what actually took place in the case of each of these trips. Eight of the trips were Route A trips which were missed for a variety of reasons. Seven of these eight trips were missed because of signal loss, principally around the final stop that the bus travelled to in the CBD (ST17). Interestingly, five of these events (and one third of the total unknown trips) occurred in one data file, which suggests there may have been some power problems with the memory storage unit at the time (when the battery of a memory storage unit becomes low, data often become patchy). The other missed Route A trip was caused by the bus driving back to the depot before the route was completed, which was probably attributable to there being no passengers on the bus by the time it reached the depot on its outbound trip. Four Route C trips were also missed because of signal problems. A total of ten trips were coded as unknown because the routes were not defined within the program structure. Two other trips were missed because of a data logger malfunction (duplicate times values), the cause of which was probably low power or a bad signal.

Table 3.4: Explanation of Unknown Trips

Trip Description	No. of occurrences
Route A, with loss of signal	8
Route B and C, with loss of signal	4
Undetectable routes	10
Other Data logger fault	2
Total	24

Urban canyon effects degraded the quality of CBD based travel time output, and unfortunately, these problems could not be fully resolved. Travel times between CBD stops were often coded as missing in output files because no points would be recorded within the buffer areas, despite the fact the bus would have passed the stops. Because the study routes ended just outside the city, urban canyon problems generally caused no problems in the trip definition component of the program. In most cases the signal was generally good around the final stop near the CBD, and hence it was always known where the bus started or finished. If the route finished in some other part of the city, urban canyon effects would have caused significant problems because in many cases, track points may not have appeared in the first or last stops. This would have resulted in a lot more trips being coded unknown.

The algorithm developed to compare GPS times with the timetable also worked well. Start times appeared to be correctly matched against all 96 trips made along the study route. In almost all cases, it was fairly obvious that the correct time was been selected by the program because the GPS start time was no more than a few minutes before or after the scheduled start time and the headway was 30 minutes for most times of the day.

No evidence was found to suggest that a one second polling rate was superior to a five second rate. There was no observable difference in travel time output produced from 1 and 5 second files and a one second polling rate appears to offer no benefits to offset its greater memory storage requirements. If the data loggers used for this project were set to five seconds and only recorded while the bus was driving, the memory storage units could probably have been left in the buses for around two or three weeks before they needed to be changed.

3.2   GIS Data Selection Toolbox

GIS is a very powerful tool for visualising spatial data, however the data query features in most standard GIS packages are relatively simple, and do not allow users to specify multiple attributes or conditions within a single query. A data selection set toolbox was designed as a visualisation tool to allow people not overly familiar with GIS to run advanced queries on large data set.

Figures 3.2 and 3.3 show the selection set toolbox in TransCAD. The toolbox allows the user to filter GPS records using any combination of the following six criteria: speed, day of week, time of day, route, timing points and direction of travel. Records are selected from a layer containing all GPS records, and are placed into a new layer, called 'GPS Data Selection Set'. Each time a query is run, the program saves the selected records as a new selection set. Each group of selected records can be displayed or hidden using standard tools used for normal GIS layers.

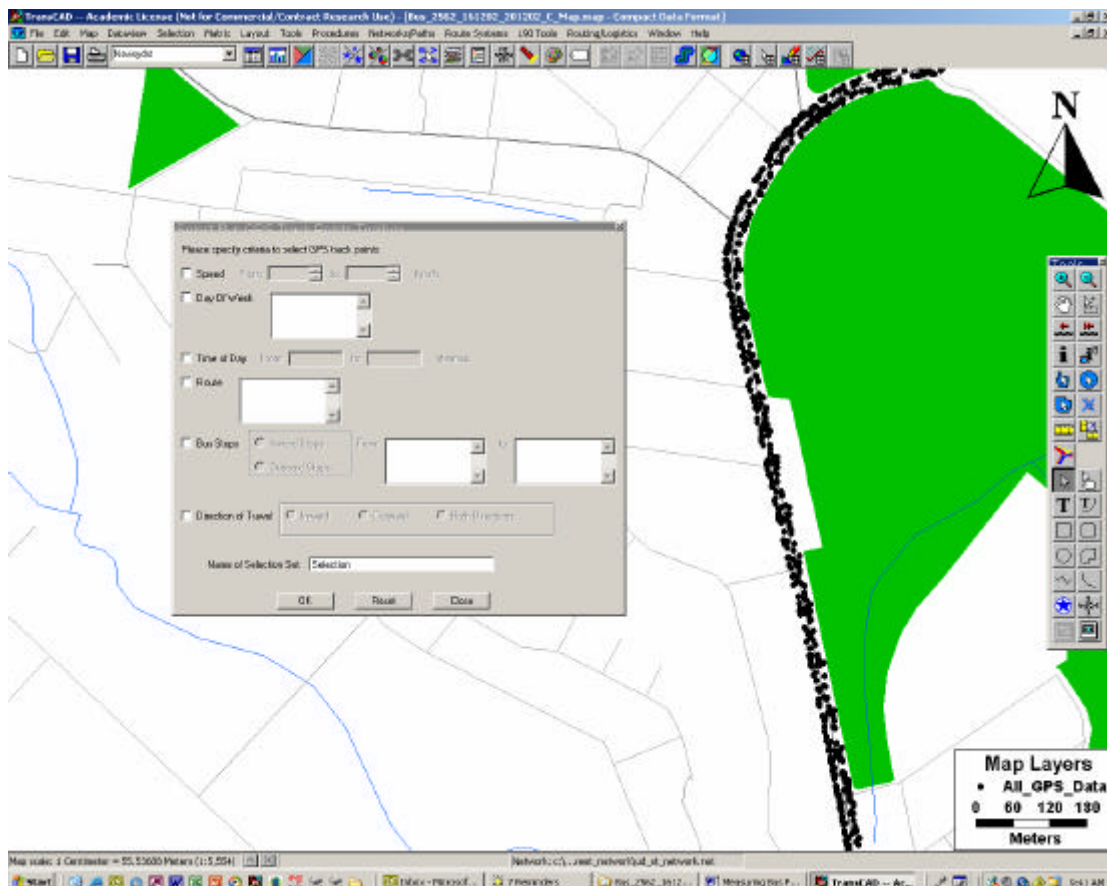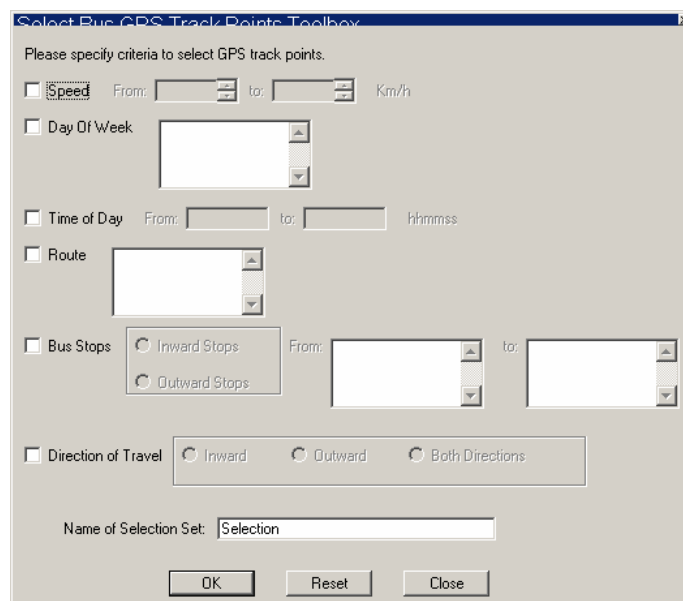*Figure 3.2: Data Selection Set Toolbox in TransCAD*

Figure 3.3: Data Selection Set Toolbox in TransCAD

3.3 Excel Analysis Spreadsheet

GIS has only limited potential for analysing large data set and providing travel time information that is useful to schedule planners. A specialised Excel spreadsheet was developed to allow users to manipulate output files produced by the programs developed in Maptitude, and to generate statistics on travel times and differences between scheduled times and actual running times. The spreadsheet was designed to allow GPS data to be filtered according to date, day of week, time of day, route, bus number and travel times.

The spreadsheet contains a number of relatively simple macros to enable users to filter records according to certain criteria and produce summary tables and graphs from output data. Tables 3.8 and 3.9 and Figures 3.4 and 3.5 show the output produced by the analysis spreadsheet. This output was generated from the eight data files used in the validation process and applies only to inbound trips made along the study route.

Table 3.8 summarises the time differences between GPS travel times and scheduled travel times. Positive numbers represent late running, while negative figures indicate that the bus arrived early. For Route A trips observed in the validation files for, buses arrived an average of 3 minutes and 59 seconds late to the final stop (ST17). Care needs to be taken interpreting this output because the numbers may be influenced by one of two outliers. In this case we can see that the maximum value column shows at least one Route A trip was more than 45 minutes late to ST17. Almost all of the maximum values were attributable to this one Route A trip made on a Sunday, which started 7 minutes late and became increasingly late as the trip went on. Before any meaningful analysis can be done with the spreadsheet, it is obviously necessary to search for outliers like these and flag them, or exclude them from the data set.

The counts shown for each timing point in Table 3.8 vary because for some trips, there were no records located within a 50 metre radius of the stop and so it was not possible to perform a timing check (this means that minimum and maximum values may not

always correspond to the same trip and differ considerably). This occurrence was most pronounced in the CBD because of urban canyon effects.

Table 3.8: Output Table Showing Differences Between Scheduled Arrival Times and Actual Times (Inbound Trips)

Timing Point	Count	Average	Median	Standard Deviation	Minimum	Maximum
ST1	49	0:02:31	0:01:31	0:02:56	-0:00:07	0:14:25
ST2	50	0:01:24	0:00:33	0:03:28	-0:02:23	0:15:55
ST3	50	0:00:47	0:00:02	0:03:26	-0:02:31	0:15:16
ST4	50	-0:00:26	-0:01:16	0:03:36	-0:04:34	0:14:16
ST5	50	0:01:17	0:00:43	0:03:46	-0:04:10	0:16:55
ST6	50	-0:00:15	-0:00:53	0:03:44	-0:05:47	0:13:46
ST7	50	0:00:41	-0:00:05	0:03:26	-0:06:04	0:12:55
ST8	50	0:00:28	0:00:40	0:03:23	-0:06:22	0:12:10
ST9	50	0:00:24	0:00:25	0:03:16	-0:05:19	0:11:35
ST10	50	0:01:37	0:01:37	0:03:33	-0:05:19	0:14:35
ST11	50	0:03:45	0:03:10	0:03:51	-0:02:49	0:16:26
ST12	50	0:03:38	0:03:00	0:04:00	-0:03:17	0:16:46
ST13	50	0:04:17	0:03:25	0:07:30	-0:05:58	0:46:47
ST14	50	0:05:15	0:04:46	0:07:47	-0:06:07	0:48:22
ST15	36	0:02:41	0:03:12	0:05:18	-0:09:13	0:11:58
ST16	42	0:02:12	0:02:07	0:05:18	-0:10:07	0:12:37
ST17	50	0:03:59	0:03:05	0:09:10	-0:11:33	0:45:13
TOTAL TRAVEL TIME	50	1:37:10	1:38:46	0:11:49	1:10:30	2:10:43

Figure 3.4 shows the median time differences between scheduled arrival times and actual arrival times. From the limited data observed in the validation process, it can be seen that Route A inbound services experienced their greatest general delays from ST11 to the end of the route.

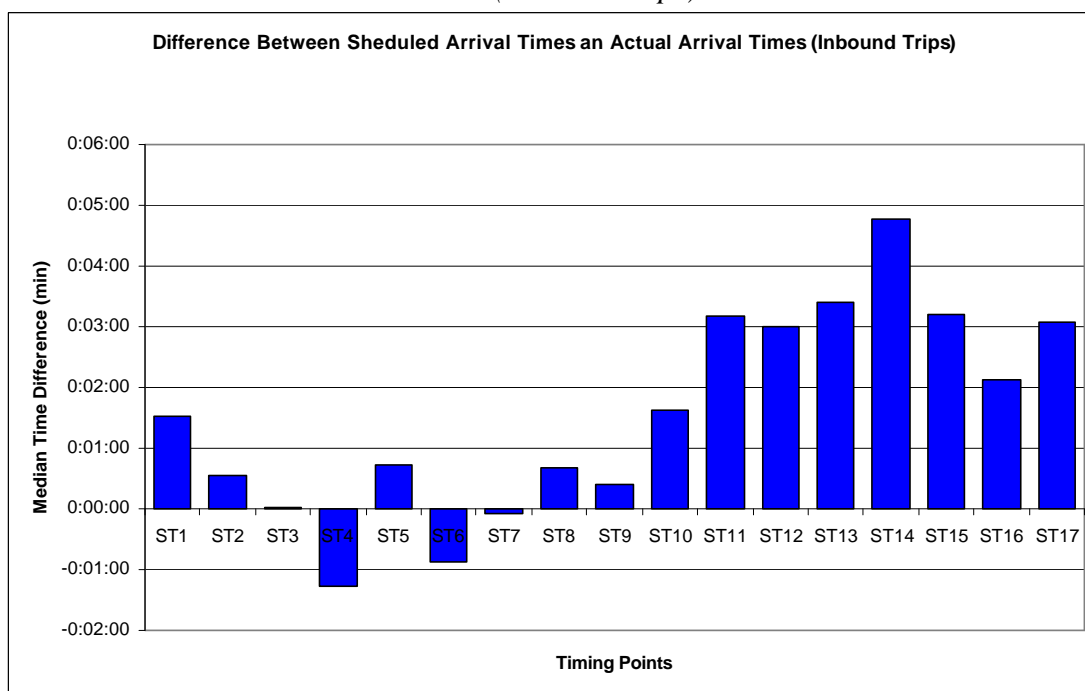*Figure 3.4: Graph Showing Differences Between Scheduled Arrival Times and Actual Times (Inbound Trips)*

Table 3.9 provides descriptive statistics for travel times between timing points. A larger set of descriptive statistics are generated for general travel times because these are more commonly used by schedulers than the time differences shown in Table 3.9. Because good output on travel times has been very difficult to obtain up until now, it is largely unknown which statistic is the best to use for planning schedules. In travel time research literature, median times tend to be favoured over averages because they are less sensitive to outliers (Quiroga, 1997). For the operator of a transport service however, a statistic such as the 85th percentile might be more appropriate. Median travel times and 85th percentile times are therefore displayed in the graphical output generated by the spreadsheet, as shown in Figure 3.5. In general, the 85th percentile times are one to two minutes higher than the median times. It should be noted that the figures discussed here correspond to trips made across all times of the day. For scheduling purposes, it would obviously be necessary to generate output for different time periods and days of the week.

Table 3.9: Output Table Showing Travel Times Between Timing Points (Inbound Trips)

Timing Points	Count	Average	Median	Std. Dev.	Min.	Max.	Cumulative Median Travel Time	Segment Time as % of Total Travel Time	85th Percentile
ST1 – ST2	49	0:09:37	0:09:25	0:01:53	0:06:15	0:14:30	0:09:25	10.01%	0:11:11
ST2 - ST3	50	0:04:01	0:03:58	0:01:03	0:02:05	0:06:25	0:13:23	4.22%	0:05:01
ST3 – ST4	50	0:06:34	0:06:30	0:00:57	0:04:40	0:09:53	0:19:53	6.92%	0:07:23
ST4 – ST5	50	0:04:18	0:04:13	0:01:03	0:02:25	0:07:10	0:24:07	4.48%	0:05:26
ST5 – ST6	50	0:05:29	0:05:33	0:00:52	0:03:31	0:07:34	0:29:39	5.90%	0:06:23
ST6 – ST7	50	0:02:53	0:02:42	0:01:38	0:00:26	0:06:35	0:32:21	2.87%	0:04:38
ST7 – ST8	50	0:09:20	0:09:07	0:01:41	0:06:10	0:12:48	0:41:28	9.69%	0:11:16
ST8 – ST9	50	0:03:56	0:03:48	0:00:55	0:02:39	0:07:05	0:45:17	4.04%	0:04:45
ST9 – ST10	50	0:06:02	0:05:58	0:01:05	0:04:15	0:09:33	0:51:15	6.35%	0:06:55
ST10 – ST11	50	0:05:58	0:05:50	0:01:04	0:04:13	0:08:44	0:57:05	6.20%	0:07:03
ST11 – ST12	50	0:00:53	0:00:31	0:00:36	0:00:21	0:03:15	0:57:36	0.55%	0:01:25
ST12 – ST13	50	0:14:39	0:13:36	0:04:44	0:09:39	0:44:01	1:11:11	14.44%	0:17:03
ST13 – ST14	50	0:04:58	0:05:02	0:01:07	0:02:42	0:09:25	1:16:14	5.36%	0:05:46
ST14 – ST15	36	0:07:46	0:07:44	0:00:59	0:05:49	0:09:39	1:23:57	8.21%	0:08:48
ST15 – ST16	34	0:02:55	0:02:47	0:00:59	0:01:36	0:05:45	1:26:45	2.96%	0:03:37
ST16 – ST17	42	0:07:24	0:07:22	0:01:34	0:04:30	0:10:50	1:34:06	7.83%	0:08:56
TOTAL TRAVEL TIME	50	1:37:10	1:38:46		1:10:30	2:10:43			1:46:27

Figure 3.5: Graph Showing Travel Times Between Timing Points (Inbound Trips)
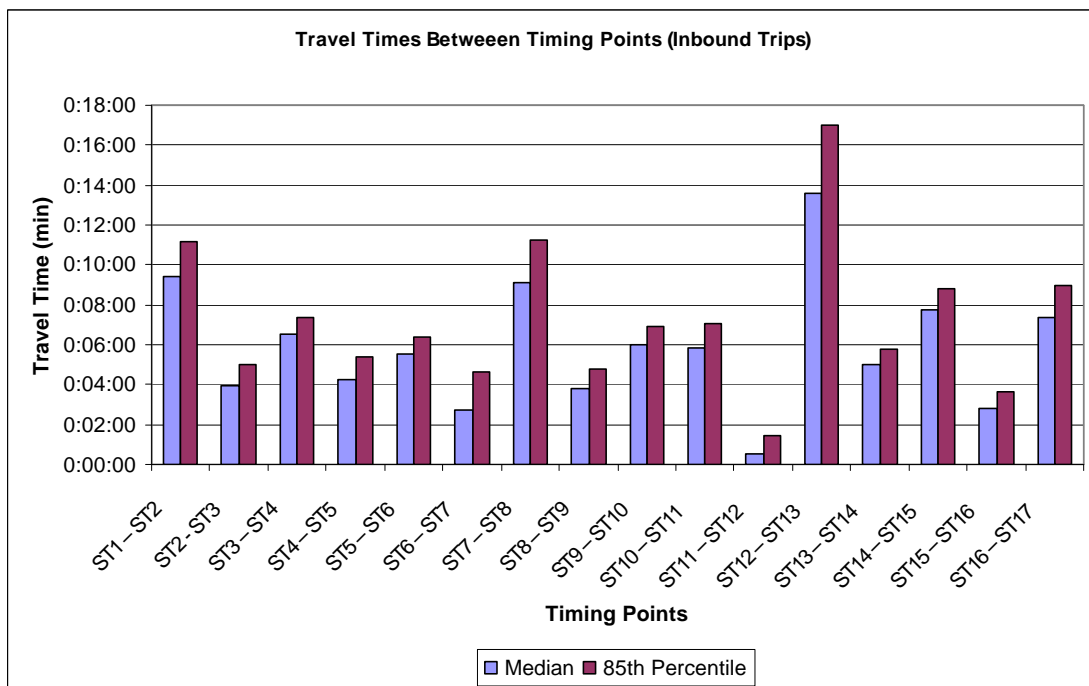
4. Conclusion

This pilot project has shown that it is feasible to collect accurate travel time data using simple, passive GPS devices operating independently of bus drivers and existing on-board computer systems. The approach taken for this project represents a viable, low cost method for collecting accurate travel time data which can be used to measure on-time running and provide data useful for schedulers.

The data collection costs associated with the approach taken in this project are very low. Hundreds of hours worth of data were collected on the study route for little more than the cost of coordinating the movement of data loggers between the depot and head office. The challenge in using GPS to collect travel time data is no longer how accurate data can be collected, but how data can be collected and managed for buses operating in a number of different areas.

In our opinion, portable data loggers are well suited to measuring travel times and on-time running. It is not necessary to have an entire fleet of buses equipped with GPS to provide information useful to schedulers. With a small investment in just 10 to 20 data loggers, it would be possible to implement a continuous survey of many different routes. Data loggers could be rotated through different depots every few weeks and a large travel time database could be built and expanded over time.

References

Auckland City Council (2003) *Real Time Passenger Information* [Online], Available: http://www.aucklandcity.govt.nz/auckland/transport/busesfirst/realtime.asp, [2003 July 01]

GPS Online (2000) *GPS Online Tracks Sydney Buses: 1 September 2000* [Online], Accessed: http://www.gpsonline.com.au/news&events/news_index/news2000.html [2003 July 01]

Infodev (2003) Automatic Vehicle Localization System (AVLS-GPS) [Online] Accessed: http://www.infodev.ca/english/pagebus/avls.shtml [2003 July 01]

Kharola, P.S., Gopalkrishna, B. and Prakash, D.C. (2002) 'Fleet management using GPS and GIS' [Online] Accessed: http://www.gisdevelopment.net/application/Utility/transport/utilitytr0010.htm [2003, July 01]

Morehead (2001) 'GPS Keeps Track of Buses', *Civic.Com*, June 26 [Online] Available: http://www.civic.com/civic/articles/2001/0625/web-gps-06-26-01.asp [2003 July 01]

NextBus (2003) *How NextBus Works* [Online], Available: http://www.nextbus.com/corporate/works/index.htm [2003, July 01].

Quiroga, C. A. (1997), *An Integrated GPS-GIS Methodology for Performing Travel Time Studies* Ph.D. Dissertation, Louisiana State University, Baton Rouge, LA