"ANALYSIS OF TFL GPS OBU DATA"

FINAL REPORT FOR THE DUTCH MINISTRY OF TRANSPORT

MARCH 2007

Version	Description	Edits:	Date:
1.0	Interim Report Draft	HA/JG	21st Feb 2007
2.0	Final Report Draft, including feedback from MoT	HA/JG/JS	7th Mar 2007
3.0	Final Report Issued	HA/JG/JS	7th Mar 2007
4.0	Final Report, including feedback from MoT	HA/JG	20th Mar 2007
5.0	Final Report, including feedback from MoT	HA/JG	2nd Apr 2007

TABLE OF CONTENTS

0. Executive Summary

0.1 Introduction

This document presents the summarised results of the '"***Analysis of Transport for London (TfL) GPS OBU Data***" mini-project commissioned by the Dutch Ministry of Transport (The Ministry) and prepared by Mapflow on original data produced during the 'Road User Charging (RUC) Mini-Trials' undertaken by TfL on selected set of vendors between November 2005 and April 2006. These trials are referred to in this report as the Stage 3 TfL trials.

The results presented here have been anonymised to protect the commercial interests of the vendors taking part in the original project. The original report generated by TfL has been published and is available from the TfL website.

0.2 Purpose

The purpose of the Stage 3 TfL trials was to exercise and learn from the GPS-based road pricing systems available from the vendor community. The purpose of this project for the Ministry is to extend that exercise by determining the efficacy of 'simple non-map matching' methods for the calculation of distance travelled.

By using a pre-defined network of road segments with different charge weightings and a number of different tariff schemes, the aim of the original TfL project was to understand the feasibility and resolution of an RUC scheme in London. The goal of this project is to understand whether road usage can be determined sufficiently without the need for detailed mapping, and all of the complexity that implies.

0.3 Project Objectives

The objectives of this study for the Ministry, is to evaluate using the TfL trial data whether non map matching algorithms can achieve performance in line with the results obtained in the "cost monitor" project for which the first phase has been completed and reported on. Three algorithms, ranging from a simple straight line distance calculation, to a simple filter and a more sophisticated algorithm were used in this study to generate distance values without the use of map matching.

0.4 Data used in the analysis

4.2 million GPS points were collected from 17 device types; in some cases a vendor supplied more than one device to the TfL Stage 3 trials. Devices used in the trials ranged from road pricing devices through to telematics units. 1,265 journeys were driven in the central congestion zone in London, totalling 36,617 km. The journeys were typically 24km which unfortunately does not provide a range of journey lengths for the analysis.

0.5 Summary of findings

The three algorithms used in this study demonstrated varying degrees of performance, in terms of road usage distance determined versus actual distance driven:

- The simple “straight line” algorithm provided the best performance for journeys with good GPS data. For journeys with erroneous data the algorithm exhibited very poor performance.
- The simple “Filter” algorithm removed the erroneous data from the calculations improving performance generally. However, this filter also adversely affected performance in cases where GPS was good.
- Reviewing the results of the “Sophisticated” algorithm anecdotally suggests promise. The algorithm is able to interpolate and extrapolate out GPS error and unavailability. However, in an attempt to achieve 0% error, the algorithm was tuned to the point where it became unreliable.

Taking three good vendor devices, the results from the first to methods show that it is possible to achieve 2% average distance deviation and 4.64% standard deviation, for bills of 24km journeys. The analysis shows that these devices would achieve 35% of journeys within 1% error. A margin of 10% would be required for 99% confidence

The following table summarises the performance results for the three individual algorithms for all devices[1]. Note the numbers are quoted as absolute % distance deviation from the actual true distance driven.

Journey Count	Straight Line (Avg) %	Filter (Avg) %	Sophisticated (Avg) %		Straight (StDev)	Filter (StDev)	Sophisticated (StDev)
1145	80.44	6.28	14.03		1516.95	8.47	56.35

These headline figures are very heavily influenced by the poor GPS performance of some vendor system. If we arbitrarily choose '3 good' performing systems the following results are obtained.

Vendor	Count	Straight Line (Avg) %	Filter (Avg) %	Sophisticated (Avg) %	Map Matching (Avg) %
2	51	1.43	1.28	7.32	1.4
11	29	1.59	1.75	6.26	0.2
18	398	2.36	2.55	5.29	3.4

[1] For journeys which were used by all three algorithm methods

We can see that for these three arbitrarily chosen 'good' vendors the weighted average performance is comparable with map matching. If another set of vendors, or another proportion of journeys were chosen for each of these vendors, the results might be very different. Note, for instance with Vendor 11, that map matching is able to achieve 0.2% distance deviation on average.

One of the requirements of this project was to study standard deviations for the 3 algorithms and the 99% confidence interval for three larger billing distances (350km, 1000km and 1,350km). As the journeys in the original trials were generally 24km, to create larger billing periods multiple journeys were aggregated into bills. Unfortunately the TfL trials drove insufficient distances to generate sufficient numbers of bills of this size (>350km) to provide meaningful statistics.

The ARS report uses the Standard Error[2], to scale the Standard Deviation by the square root of the billing distance, as follows

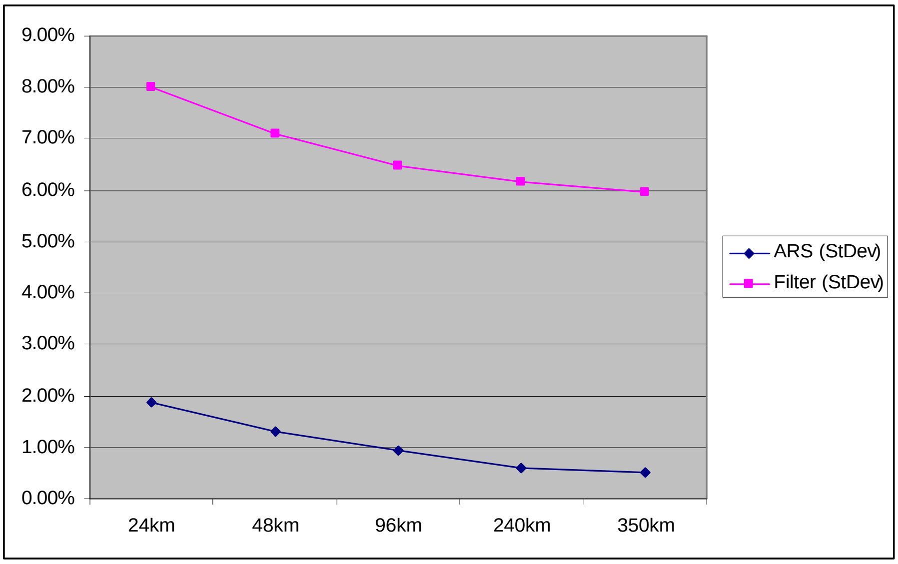
$$S_E = \frac{\widehat{\sigma}}{\sqrt{n}}$$

where

$\widehat{\sigma}$ is an estimate of the standard deviation s of the population, and
n is the size (number of items) of the sample.

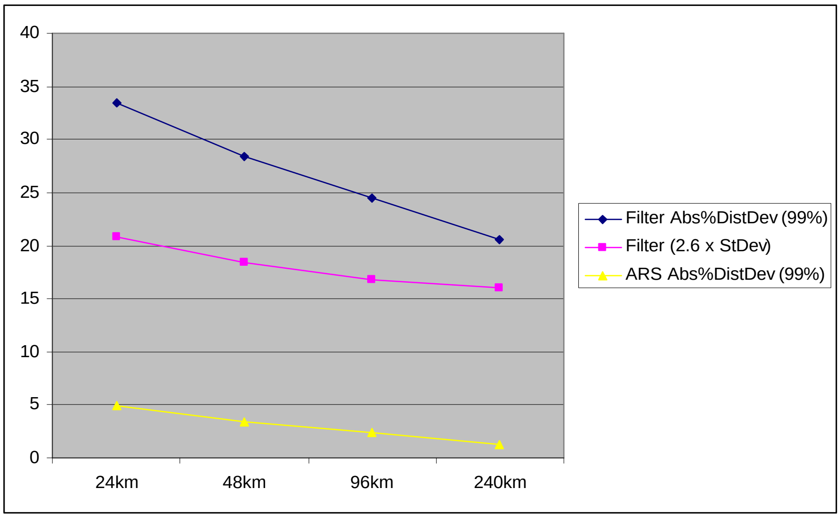
To validate this relationship, journeys were grouped into bill sizes of 48km, 96km, 240km and 350 km. The limited empirical data this produces does seem to follow the relationship described by the Standard Error, although the absolute mean estimator is larger for TfL's data than for ARS' results.

Billing Distance	ARS (StDev) Derived	Straight (StDev) Empirical	Filter (StDev) Empirical
1km	9.20%	N/A	N/A
24km	1.88%	1061.47%	8.01%
48km	1.32%	853.99%	7.11%
96km	0.94%	614.06%	6.47%
240km	0.59%	441.15%	6.16%
350km	0.50%	357.98%	5.95%
1350km	0.25%	N/A	N/A



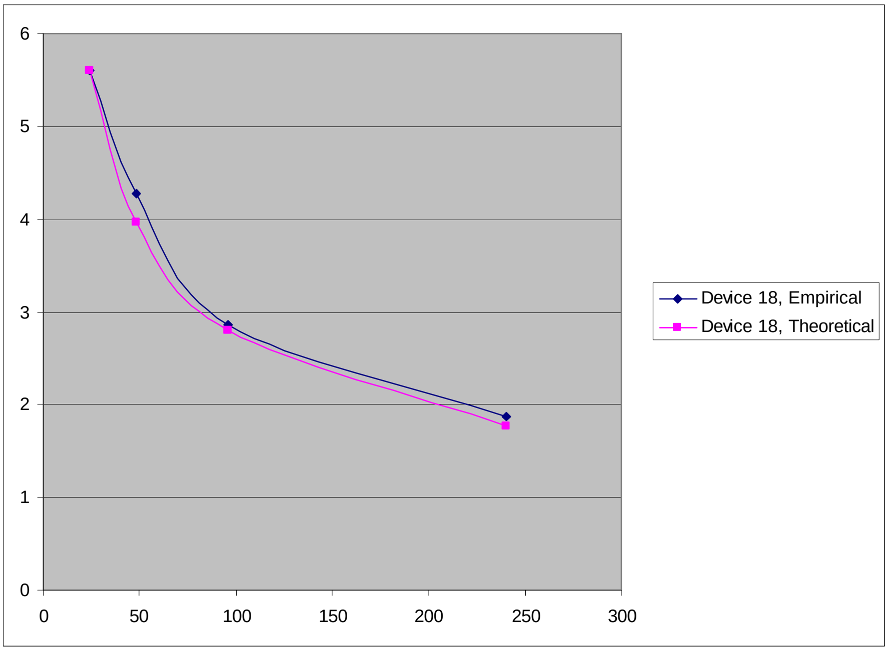
Likewise at a 99% confidence interval for all devices, there is some similarity between the theoretical Standard Error suggested by ARS and the TfL data for these billing periods. The empirical 99% confidence interval results suffer from lack of observed data points for billing periods greater than 24km and therefore the results should be used with caution. Note for instance that only one sample bill of 240km was used to determine the respective 99% confidence interval.

Billing Distance	Samples above 99%	Observed Filter Abs%DistDev (99%)	Filter Abs%DistDev (99%)	ARS Abs%DistDev (99%)
24km	11	33.5	20.83	4.89
48km	6	28.4	18.48	3.43
96km	3	24.5	16.83	2.44
240km	1	20.6	16.00	1.30



[2] See: http://en.wikipedia.org/wiki/Standard_error_%28statistics%29

While empirical bills for 'all devices' show a similar relationship, they do not share an absolute correlation with the theoretical Standard Error. However, when we take an individual device the correlation between empirical and theoretical results is very strong for the billing periods chosen and for a similar environment.



This suggests that the reduction in Standard Error is strongest with a homogeneous system, i.e. with a single device working in a single environment, such as Southwark. When we look at more than one device the reduction with Standard Error is not as strong do the heterogeneous mix of performance. One might assume that if we look at a mix of environments, say dense urban, motorway and rural the reduction due to Standard Error may be similarly diluted. Further investigation is required into the effect of different environments, multiple devices with multiple technologies.

The ARS study also identified and isolated time to first fix (TTFF) as a major source of error. In the original study, TfL did not exclude TTFF from their analysis and their trials were not designed to isolate, identify or control TTFF. To go some way to identify the constraint imposed by TTFF, in this study the performance of the 1st Journey of each day was compared with that for other journeys to see if an effect was noticeable. Our finding is that, for some devices, 1st journeys do appear to perform not as well and that this maybe attributed to TTFF. However, as the TfL journeys were all unusually long i.e. typically 80 minutes in duration, the TTFF may not have been adequately represented. For example if the trials had performed shorter more typical journeys (e.g. 5 minutes) we could have expected to see a greater effect. Similarly, if the ignition off period prior to the journey could be variable and known it would enable a better understanding of TTFF effects.

In conclusion, this study found:

- The highest performing vendor solutions are able to achieve an average 2% absolute percentage distance deviation from the actual distance driven, this compares well with the 1.6% in the ARS study (50 km/h)

- The same subset of vendors provided only 35% of journey within 1% deviation or would require a 10% margin to provide 99% confidence, for 24km journeys. This compares well with the 24% figure in the ARS study (for 1km journeys at 50kph).

- Insufficient data meant that an analysis of billing periods of 350km, 1000km and 1350km was not practical. However an analysis of shorter billing periods was completed to test the relevance of Standard Error suggested by ARS. Using the Filter technique the Standard Error does seem to apply, and suggests that the Standard Deviation of bills will reduce as the size of the bills increase. We find, however, that further work on a larger sample is required to ensure that the findings are correct for the actual billing sizes likely to be generated by a live scheme.

- By comparing the result of grouping bills from all devices versus bills grouped from a single device, we find that the more homogeneous the system the greater the reduction of variance with bill size. That is to say, to rely on billing size to improve consistency, the underlying system should have a single device and operate in a single environment.

- We have found that the 1st Journey of the day does appear to be marginally worse than other journeys, implying a TTFF effect, for some vendors.

The work completed in this study highlights the importance of some further research and study and some next steps recommendations are provided in the findings and conclusions section of this report.

Author's commentary: It is our view that GPS is susceptible to a number of phenomena which lead to erroneous behaviour, in urban areas particularly. By applying simple non-map road usage determination methods, such as these, GPS behaviours are directly translated into road usage distance overestimation or underestimation. From our experience working with GPS data in London and other urban environments, we feel that some GPS behaviours, in some scenarios can lead to overestimation, while other behaviours in other scenarios can lead to underestimation. The Table below lists some of these scenarios.

Behaviour that might lead to apparent underestimation	**Behaviour that might lead to apparent overestimation**
Time to first fix	Multi-path interference (accuracy)
Availability (signal obscuration)	GPS navigation software failure (accuracy)
GPS Reporting interval	Trial operation (driving in addition to the route)
Driving routes with many corners	
Trial operation (actual route was not driven)	

Any one vehicle on any one journey with a given device may experience a combination of GPS behaviours that result in both over and underestimation. In this trial, using the first two simple methods we found that these behaviour tend to combine to underestimate road usage more frequently. The third method tended to overestimate, when presented with the same GPS behaviour.

For instance, we believe the Straight-line algorithm generates more overestimated journeys than the Filter, because it does not exclude GPS navigation software failures which occasionally lead to very erroneous data. Certainly, when the Straight-line method overestimates, it does so to such an extent that if all the journeys were considered the method would generally overcharge even though it tended to undercharge more times than not.

If the Ministry settled on a single robust method (more robust than the filter), the general bias toward under or overestimating could be known and removed, as per the 'baker's dozen'[3]. The second concern then is the variability. By grouping journeys into larger bills the variability can be improved, so long as the underlying system is homogeneous. We believe that there is a good deal of heterogeneity in underlying GPS behaviour, from one device to the next, from one driving scenario to the next and from one environment to the next.

[3] http://en.wikipedia.org/wiki/Baker%27s_dozen

1. Background on the TfL Trials

1.1 Introduction

This project built upon the knowledge gained in three previous sets of RUC trials completed by Transport for London (TfL).

The Stage 1 Trials (February 2004) investigated the use of four different technologies to support congestion charging. One of these was GPS. The performance of 7 different GPS devices were analysed at the location level, using a total of 800,000 samples. The average location error was found to be 9.7 metres.

The Stage 2 Trials (April 2005) focussed on two technologies, one of which was GPS. The average performance of 6 enhanced GPS devices was found to be 9.59 metres, with a fifth device which employed EGNOS technology having an average location error of 8.5 metres. These trials also investigated various map-matching techniques.

The Stage 3 'PIN RUC' Trials (March 2006) focussed on learning from the vendor community, the efficacy of the overall road pricing solutions currently available, when applied to the London environment. This project concluded that a few of the vendors were able to detect driving distance to within 99 percent accuracy and all within 87 percent, despite challenging GPS conditions in London.

The following is a summary of the key elements of the Stage 3 trials:

- Trials totalling more than 14,000 km of journeys in central London were carried out for 17 different vendor systems, plus a Reference device.
- The best-performing system at the location level had an average location error of 5.11 metres.
- The best-performing system at the road segment level correctly identified 98.6% of segments by length, with a further 1.0% of segments incorrectly identified as being in the journey.
- The best-performing system at the overall level had an average magnitude of journey length error of 0.82%.
- The sample interval of the On-Board Unit and the performance of the map-matching algorithm were found to be the most important factors affecting overall system performance.
- The integrity of GPS was found to be too low for it alone to be used to determine a vehicle’s route in central London.

1.2 Study Logic for the Stage 3 trials

All of the RUC systems tested during the original trials operate in the same basic manner:

1. Record a set of GPS locations that correspond to the measured position of a test vehicle at various points in time. This is done using a device fitted to the vehicle called an On Board Unit (OBU).

2. Convert the location information into a road-by-road record of the journey undertaken by the test vehicle. This process is called map-matching, and in the original TfL trials, was a back-office function which took place after each journey was completed.

3. Convert the journey information into a charge by applying the rules applicable to the roads used at the time the journey was undertaken. This process is called billing, and was also a back-office function in these trials.

In reflection of this, the performance of the RUC systems was analysed at three different levels in the Stage 3 trials project:

- **Location Level**

On a point by point basis, is the primary input to the system of a high quality?

This had already been analysed by TfL in the Stage 1 and Stage 2 trials. The benefit of repeating the analysis was to see if any improvements had been made since the previous trials were carried out.

- **Road Segment Level**

On a road by road basis, does the system identify the correct journey?

This is the interim level performance, typically evident in the billing and enforcement processes, where more detail than a journey charge is required. It is important to know the extent to which the system can identify each and every road segment used by the vehicle.

- **Journey Level**

Does the system produce the correct charge?

This is the trial’s 'bottom line'; to what extent does each system produce incorrect charges? Is it too low or high and if so, why? A clear worry here is that whilst the overall journey charge could be generally correct, it may be comprised of a balance of incorrect high and low charges.

1.3 Technologies used in the Stage 3 trials

All of the vendor systems used during the Stage 3 trials were based on GPS satellite positioning technologies.

Some of the systems employed additional technologies in an attempt to improve their accuracy and availability. These included:

- Differential GPS. This uses signals from one or more ground-based systems, whose locations are precisely known, to correct various inaccuracies in the GPS system.

- EGNOS (the European Geostationary Navigation Overlay System). This supplements GPS by reporting on the reliability and accuracy of the satellite signals. In general this leads to improved horizontal accuracy.

- LORAN (LOng RAnge Navigation). This is a terrestrial navigation system based on low frequency radio transmitters. It is primarily used in marine applications, and is significantly less accurate than GPS. However, it can provide location information in situations where the GPS signal is obscured (e.g. underground) or otherwise unavailable.

- Inertial navigation. This detects the acceleration of the vehicle and uses it to provide an independent measurement of its position and velocity.

Although these technologies were used, it was not clear which ones were employed by which vendors. As a result it was not possible to determine what effect (if any) these technologies had upon the performance of the systems. The vendor systems have therefore been treated as simple “black boxes” for the purposes of this trial.

1.4 Stage 3 Trials Design

1.4.1 Methodology

The Stage 3 trials were based upon:

- A geographical area that had been accurately mapped in advance.

- A set of pre-defined routes within that area.

- A charging scheme based on the length of road travelled, the class of road, and the tariff in force at the time of the journey.

The trials were carried out using between 1 and 3 vehicles for each vendor.

Typically, every vehicle carried out 3 pre-defined journeys in the trials area each working day. This was repeated over a 2 week period. The trials for each vendor therefore consisted of between 30 and 90 journeys, depending on the number of vehicles available.

Each vendor's journeys covered a wide range of the possible routes and tariffs.

Each vehicle taking part in the trials was fitted with a vendor OBU and a reference GPS device. This provided an independent record of each vehicle's journeys. This was useful in understanding any discrepancies between the route recorded by the driver and that deduced from the OBU data. It also provided a control from one week of trials to the next, e.g. to allow for variations in GPS performance.

At the end of each day, the positioning data recorded by the vendor OBU and the reference device in each vehicle were downloaded and stored.

A reference device (device 18 in the results tables) is a device which TfL has used in all three stages of their trials. By continuing to use this device in each subsequent trial, TfL can measure improvements over time by the device vendors.

1.4.2 Trial Participants

A total of 12 vendors took part in these trials. Some vendors supplied more than one type of OBU, and so a total of 17 different OBU types are used in the analysis for the Ministry.

The systems fell into two main groups:

- RUC grade solutions (8 OBU types). These were more fully-fledged systems, characterised by high location sample rates, and typically capable of providing information at the road segment and charge level as well as raw location.

- Telematics grade solutions (10 OBU types). These were primarily designed for fleet tracking applications, and typically provide data at the location level only.

Appendix 2 provides some details on the vendor devices used in the Stage 3 trials. Note there isn't a correlation between the order of the devices in the Appendix and the order of the vendors in the results tables.

1.4.3 Trials Area

The original trials were carried out in the area of central London shown in Figure 1. The total length of the all roads in this area was 174 km.

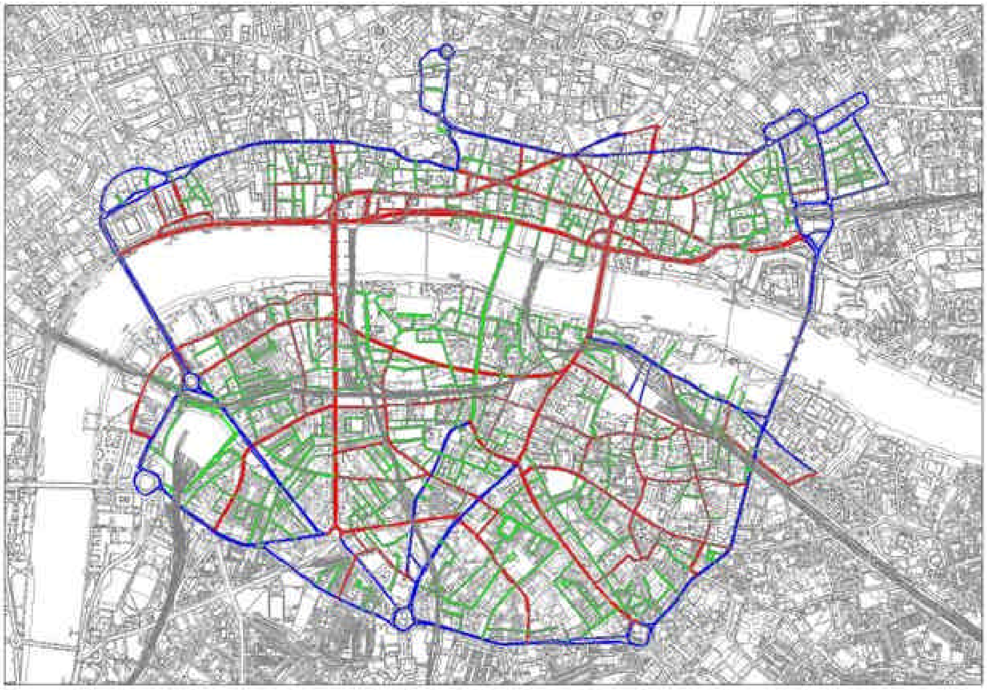


Figure 1 Complete Trials Area

1.4.4 Geographical Truth

Prior to the start of the trials, an extensive Carrier Phase Differential (CPD) survey of the roads in the trials area was carried out.

This was supplemented by road line-work based on OS MasterMap data in areas were the CPD quality was poor.

Together, these two sources of data provided a truth source of location data which underpinned the trials. This was sufficiently accurate to distinguish vehicles travelling in different directions on the same road.

Note a Reference device was used during these trials that utilised pure GPS technology without any enhancements such as EGNOS or LORAN.

1.4.5 Road Segments

The Stage 3 trials area was divided into a more than four thousand road segments. Each of these spanned a section of road between two junctions, such that a vehicle entering the section could only leave it by travelling its entire length.

Lanes of traffic moving in opposing directions on the same road were treated as separate road segments.

Each road segment was assigned the following attributes:

- A unique segment identifier.
- A name.
- A definition of its location and shape. This was expressed using polylines based on the geographical truth derived from the CPD survey.
- A truth length in metres.
- A class (which is not relevant for the current project)

1.4.6 Routes

A total of 30 routes were defined for use during the trials. The average length of each route was 23.3 km, with their lengths ranging from 17.1 to 25.9 km.

Each route consisted of an ordered list of road segments identifiers, known as the truth route. One of the routes used in the trials is illustrated in Figure 2. A close-up of the indicated part of the route, showing the road segment classification, is shown in Figure 3.

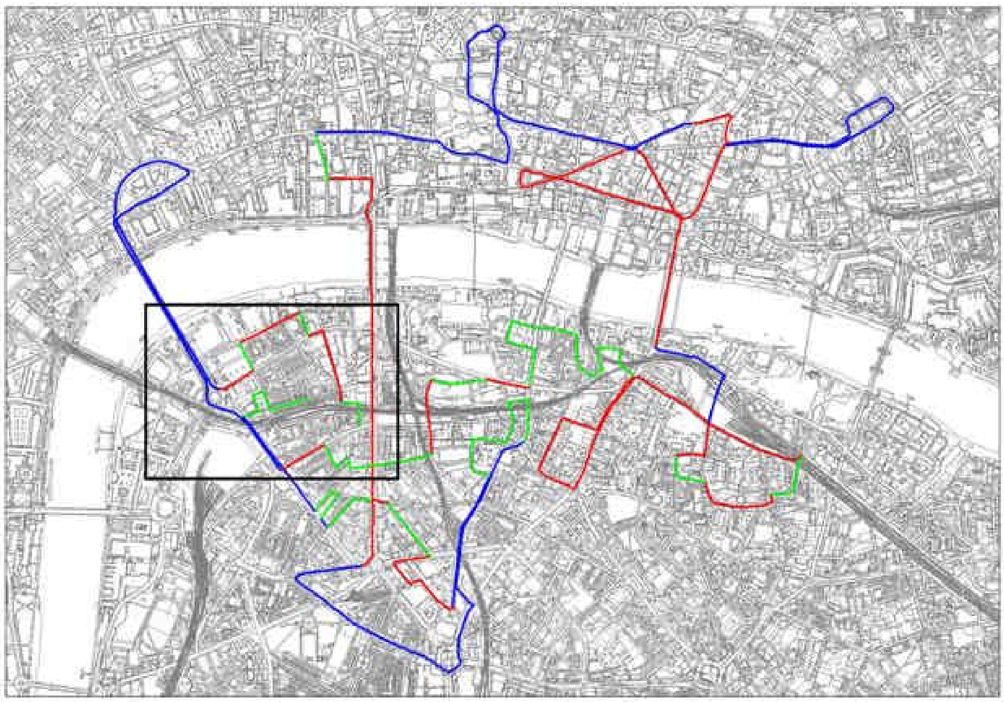


Figure 2 Complete Trials Route

Note: the original road segments were classified into Red, Green or Blue according to the tariff that was applied by TfL in the original study. This has been ignored for this project, since there is no distinction between one road and another.

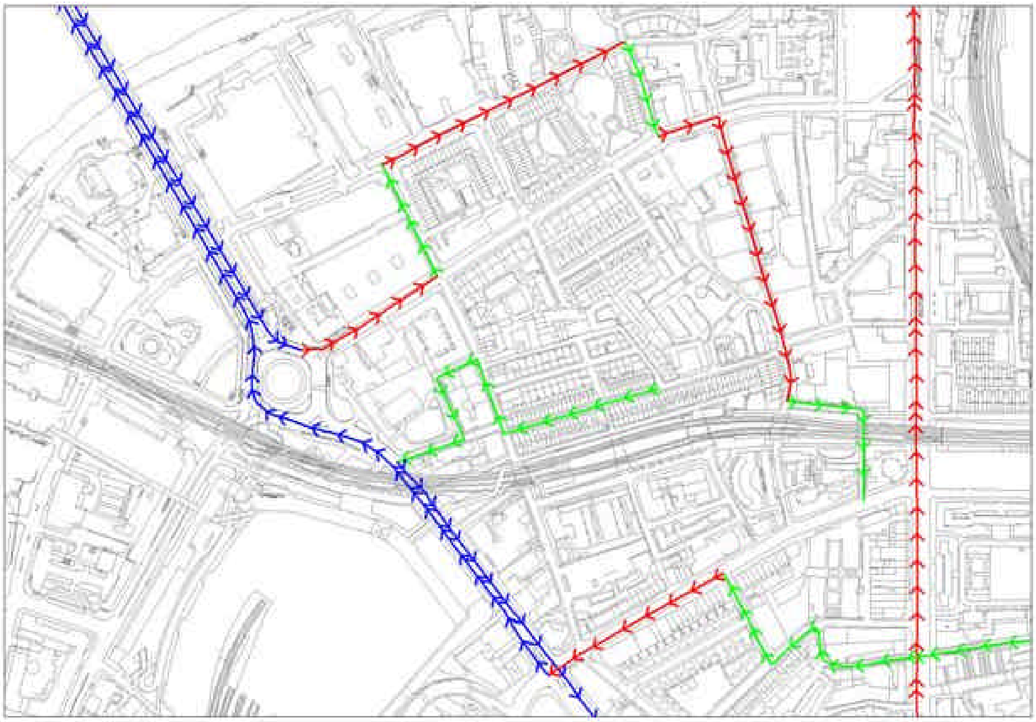


Figure 3 Detail of Route Shown in **Figure 2**

1.4.7 Journeys

Each trials vehicle was scheduled to carry out three journeys per day. Every journey was intended to follow one of the 30 pre-defined routes.

The journeys began at the following set times:

- 09:30 am. These journeys took place within tariff period 1.
- 12:00 am. These journeys straddled tariff periods 1 and 2.
- 14:00 pm. These journeys straddled tariff periods 2 and 3.

Each journey was designed to start and finish with a minimum of 5 minutes “ignition off”. However in practice this was not always possible due to safety considerations.

Key Finding 1: Ignition off time may or may not affect the performance of the RUC systems. However it was impractical in the original trials to enforce ignition-off conditions at the start and end of each journey. Therefore, time to first fix is only likely to appear in the first journey of the day and perhaps the first journey after lunch.

In some cases it was not possible for the driver to follow the pre-defined route exactly, due to road works and other obstructions. These deviations were noted by the trials drivers, but it was noticed in some cases that this manual method did not always provide an accurate record of the journeys.

Key Finding 2: Although great care was taken, the manual intervention required in cases where the driver deviated from the prescribed route may introduce variables unrelated to the RUC systems.

In addition, each journey was halted after 90 minutes if it had not been completed by the point. This was to allow time for the vehicle to be re-positioned ready for the start of the next journey.

At the end of each journey the following information was recorded by the driver:

- Start time
- End time
- Road segment at which the vehicle was located when the tariff change (e.g. from Period 1 to Period 2) took place
- Details of any deviations.

This information was used, where necessary, to create a new truth route specifically for that journey. This took account of any deviations and early finishes.

In some cases a significant difference was noticed between the manually-recorded journey end time and the end time deduced by the vendor from their OBU results. This accounted for up to 5% of the journey duration in the worst cases.

Key Finding 3: The manual intervention required in cases where journeys finished before the prescribed time, may have introduced variability unrelated to the RUC systems.

1.4.8 Vendor GPS Data

Each vendor was requested to provide the following information.

At the Location level, the following data at regular intervals throughout each journey (preferably every second):

- Journey ID
- Timestamp
- Latitude (WGS 84 format)

- Longitude (WGS 84 format)
- Semi-major axis of error ellipse (5 sigma)
- Semi-minor axis of error ellipse (5 sigma)
- Inclination of error ellipse (5 sigma)

1.4.9 Built Environment

The accuracy of any satellite positioning system is heavily influenced by how many satellites are visible. This is a particular issue in urban environments where the receiver can be surrounded by tall buildings which restrict its view of the sky, and hence the number of satellites it can receive signals from.

The effect of the built environment on the accuracy of the vendor systems was investigated by building a model of signal shadowing within the trials area.

3D GIS ‘hill shade’ modelling techniques were used to predict the shadow cast by each building for a given satellite azimuth and elevation. By casting a shadow for satellites at North, South, East and West, at 30 degrees inclination from the horizontal, an overall shadow footprint of each building was developed.

Figure 4 and Figure 5 illustrate the shadows cast from the North and East respectively in part of the trials area.

Figure 4 Buildings Shadow from the North

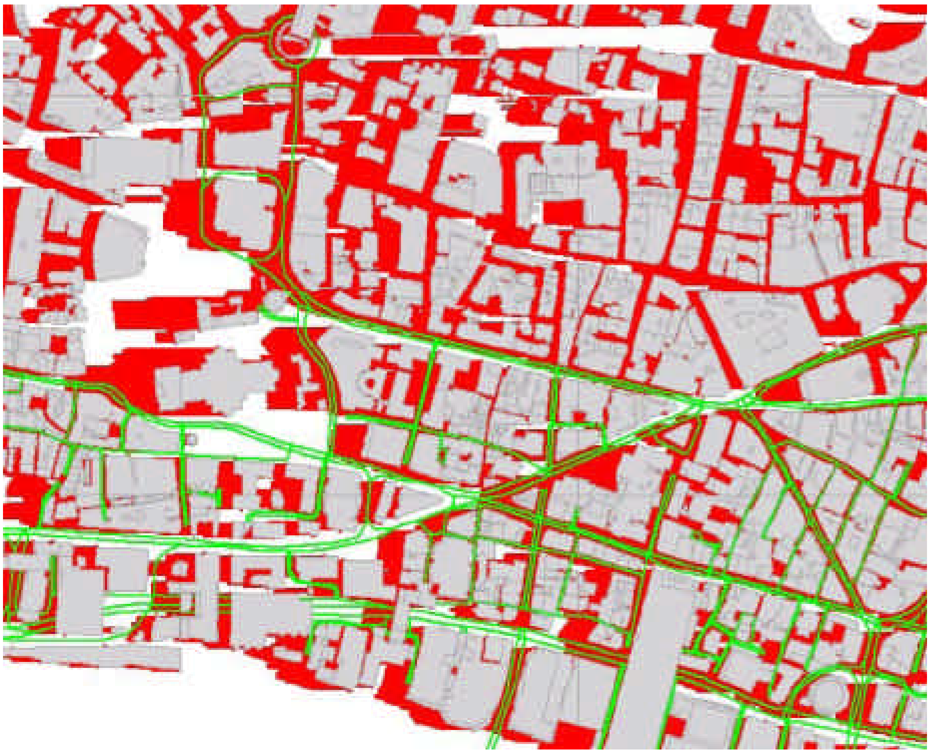


Figure 5 Buildings Shadow from the East

These four overlays were then combined to produce a map of overall signal visibility. Figure 6 illustrates this, with visibility values ranging from 0 (equivalent to a tunnel) to 100% (sky completely unobscured).

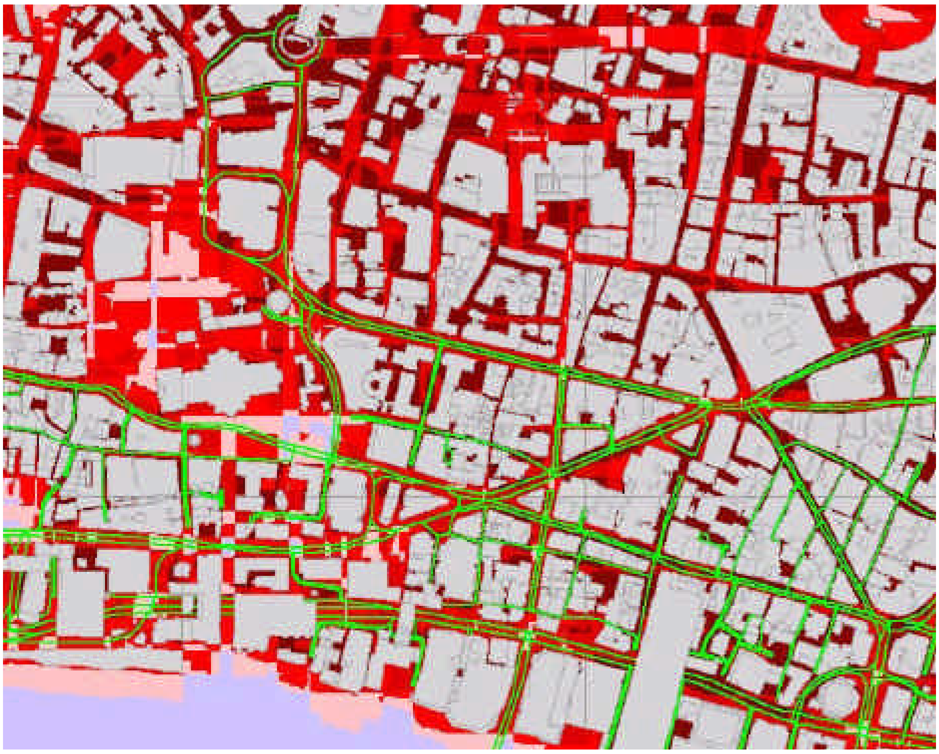


Figure 6 Overall Signal Visibility

Finally, the visibility predictions were overlaid onto the truth road network. This allowed the average visibility for each road segment to be calculated.

This assessment of urban environment was subsequently greatly enhanced by TfL's cooperation with ESA and the use of GNSS simulations tools which took into account the ephemeris nature of the satellite constellation and the signal propagation.

2 Methodology Applied in this Study

2.1 Overview

The objective of the analysis for the Ministry's project is to utilise the data collected as described in Section 2 for the 17 vendor supplied OBUs and the 1 reference OBU to complete a "distance" calculation using simple point to point algorithms. .

2.2 Algorithms

As compared to the Stage 3 TfL project, in this project for the Ministry we are focused on the efficacy of determining distance without the use of a map reference. Mapflow applied a number of simple geometric algorithms to the original GPS data from the vendors, in order to calculate distance. These are summarised below. For reference the results from TfL's map matching analysis are also included in this study.

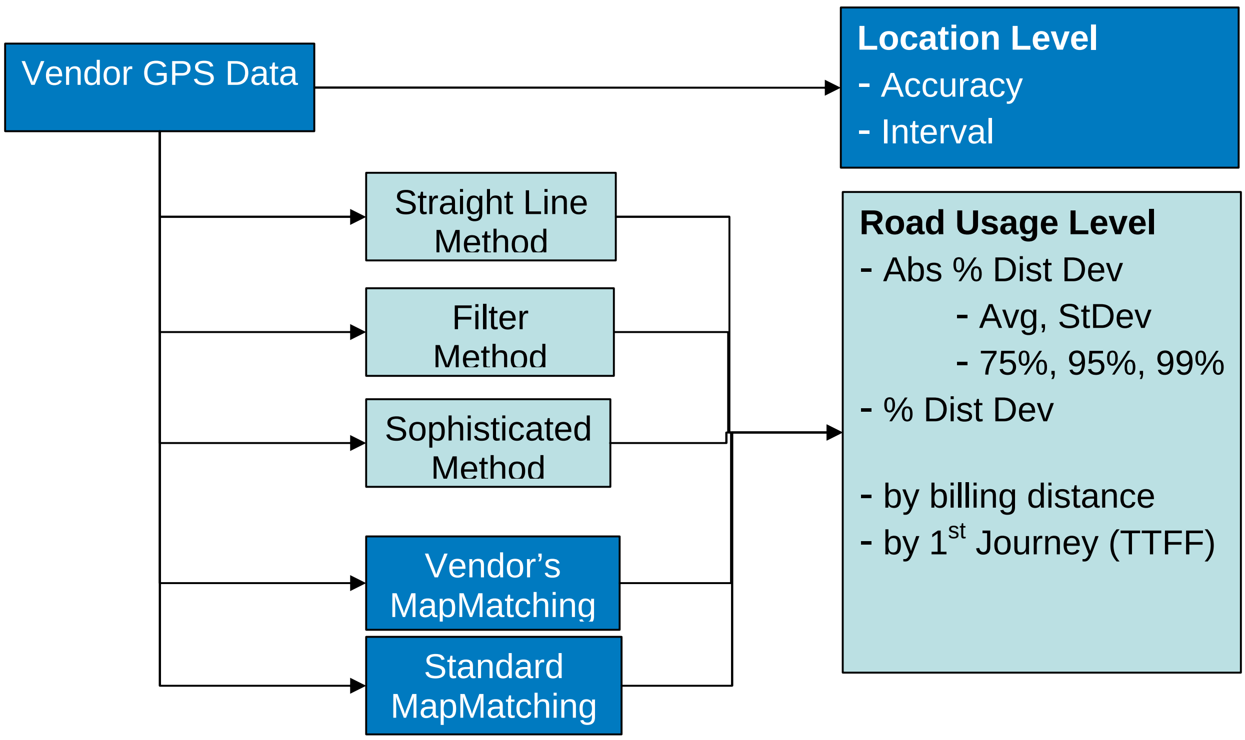


Figure 7 Study Logic

2.2.1 Simple 'Straight-line' Method

In this first, most simple method, we have taken the raw GPS data for each vendor's system on each journey. Based on the time stamp provided, this method calculates the straight-line Euclidean distance between each consecutive point, when projected to a Cartesian coordinate system.

The map below illustrates how, due the normal multi-path interference associated with a built environment like London, GPS scatter (blue) occurs and this approach frequently includes erroneous distances (magenta). Similarly, when GPS data is not available due to signal obscuration this method simply draws a straight-line distance between consecutive points, irrespective of the distance that may actually have been driven (green) in the intervening time

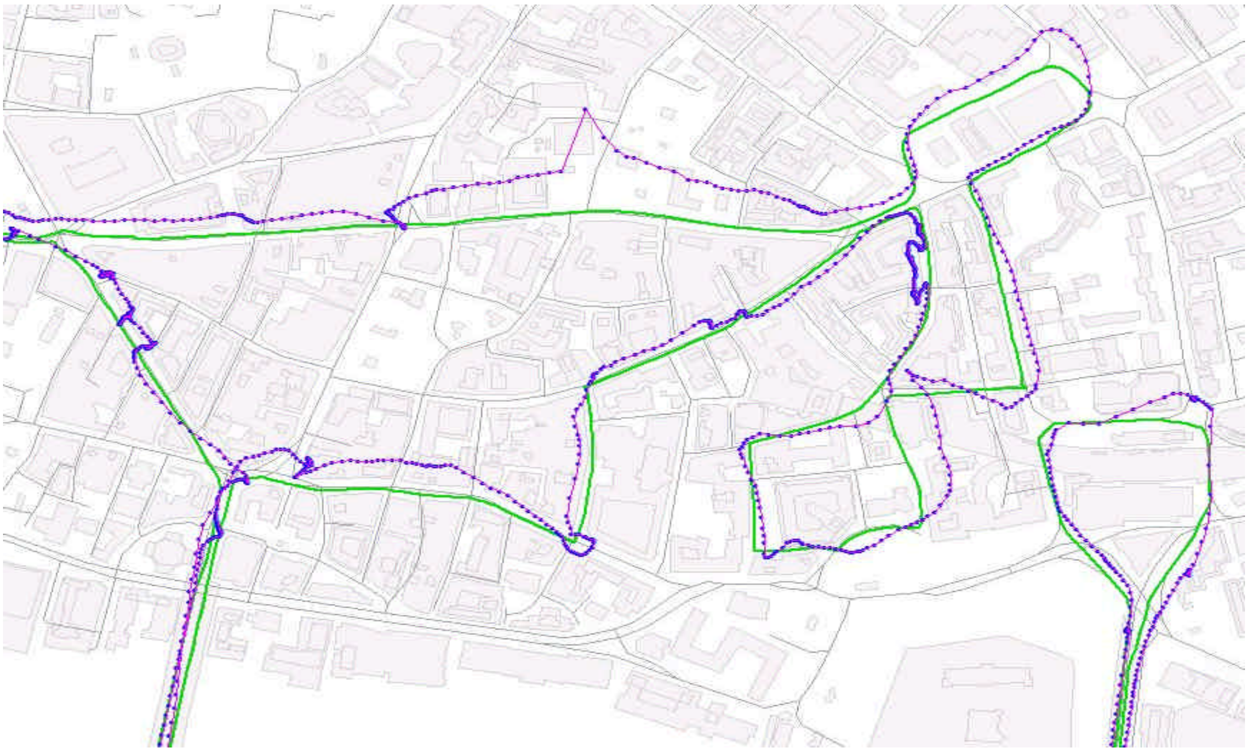


Figure 8 Straight-line distance calculation

2.2.2 Simple 'Filter' Method

In this method we have taken the raw GPS data for each vendor's system on each journey. Based on the time stamp provided, this method calculates the straight-line Euclidean distance between each consecutive point, when projected to a Cartesian coordinate system.

In addition, this method includes a simple filter based on the acceleration of the vehicle. If the two consecutive points show the vehicle travelling faster than $2m/s^2$ this method ignores that point, joining the points before and after the acceleration. The principle suggests that a vehicle, in London, is unlikely to accelerate faster than 7kph in one second, an that the acceleration is due to GPS anomalies.

This simple metric and the absolute number chosen were arrived at purely by observation. The map, below, illustrates how this simple Filter appears to exclude many of the erroneous results without removing much useful data. It should be noted, however, this is a blunt instrument arrived at by experimentation rather than calculation.

Figure 9 Filter distance calculation

2.2.3 More "Sophisticated" Method

Recognising that the two previous methods may provide similar results, we applied a third 'Sophisticated' algorithm. The goal was to illustrate how alternative methods might result in very different road usage detection results.

In this third method, Mapflow applied a less simple strategy than the previous two. This algorithm is capable of extrapolating and interpolating along GPS points. Whereas the simple methods joined one point to the next, this method is influenced by the general movement of the vehicle. In this way, the 'trajectory' is not influenced by each and every erroneous GPS point, instead a curve is formed which approximates the apparent movement of the vehicle over many points. Similarly, the 'trajectory' is not influenced by the momentary absence of points. If points fail, a curve is extrapolated from the trajectory implied by the points immediately prior and immediately after the outage.

The map below illustrates how the output from this method (light blue) appears as a smooth curve approximating the movements of the vehicle. Erroneous GPS points (blue) exert an influence on the output.

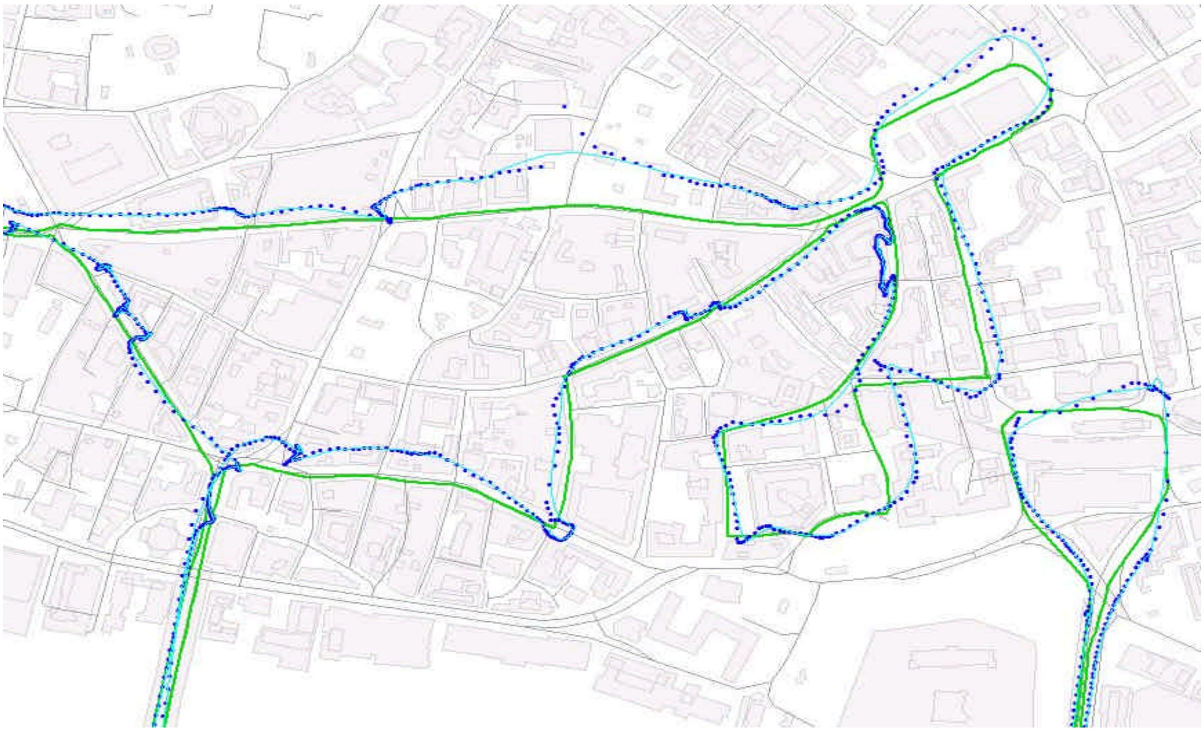


Figure 10 Sophisticated method distance calculation

2.3 Road Usage Metrics

The following summarise the metrics used to judge efficacy in this project for the Ministry.

2.3.1 Abs% Distance Deviation

This metric is calculated by taking the distance determined by the method (Straight-line, Filter, Sophisticated or map matching), subtracting the truth distance actually driven to give the distance deviation. This is then expressed as an absolute % of the actual journey length.

$$Abs\%DistDev = \frac{|(Determined\ Distance - Truth\ Distance)|}{Truth\ Distance}$$

That is to say, this metric provides an indication of the magnitude per km, of the method's deviation from actual distance driven. For instance, if a vehicle actually drove 1km and the method determined 1.1km had been driven, the Abs%DistDev = 10%. Similarly, if a vehicle actually drove 1km and the method determined 900m had been driven the Abs%DistDev = 10%. If these two journeys were combined into single bill, we would judge the average Abs%DistDev = 10%.

2.3.2 Abs% Distance Deviation Quantiles

An empirical cumulative distribution and population histogram have been generated for the absolute % distance deviation (Abs%DistDev) metrics, for each of the methods.

In simple terms, we have counted the number of journeys that showed a given Abs%DistDev, in half % bins. That is to say, we have counted the number of journeys that fall in 0 to 0.5%, 0.5% to 1%, 1% to 1.5% and so forth.

From this simple ECDF (Empirical Cumulative Distribution Function)[4] we can estimate the 75%, 95% and 99% quantiles. For example, we may find that a method may require a margin of 11% Abs%DistDev to include 75% of the journeys.

2.3.3 % Distance Dev

In comparison to Abs%DistDev, this metric is calculated by taking the distance determined by the vendors system (or reference device and HUB), subtracting the truth distance actually driven to give the distance deviation. This is then expressed as a positive or negative % of the actual journey length.

For the avoidance of doubt: when a negative %DistDev occurs the method determines the journey to be shorter than the route actually driven.

$$\%DistDev = \frac{(\text{Determined Distance} - \text{Truth Distance})}{\text{Truth Distance}}$$

That is to say, the second metric provides an indication of the magnitude per mile, of the methods deviation from actual distance driven. For instance, if a vehicle actually drove 1km and the system determined 1.1km had been driven, the %DistDev = +10%. Similarly, if a vehicle actually drove 1km and the system determined 900m had been driven the %DistDev = -10%. A single bill for the two journeys combined, say on a single day, would give %DistDev = 0%. Thus, the second metric does distinguish between whether a charge would be over or under, but combined high and low figures tend toward 0 in the case of averages and standard deviations.

Without specifying the distance or duration over which to average, it would be impossible to provide an average figure. Although it might be argued that 24km of driving is a fair reflection of the normal driving pattern for a single car - in which case each test journey is comparable to a single day of normal driving.

The ECDF (Empirical Cumulative Distribution Function) for %DistDev for each method can be derived. We have chosen to define acceptable bounds within which the Ministry might operate. In the current revision of the document, these bounds are set at -5% and +10%. That is to say the Ministry are willing to allow a method to underestimate road usage by 5% and over estimate it by 10%.

[4] For reference and further information on the ECDF see http://en.wikipedia.org/wiki/Empirical_distribution_function

2.4 Summary of data used in the analysis

Of the journeys carried out during these trials, including the reference device as if it were a vendor, the data supplied for analysis by TfL for this project is follows:

Total Journeys	1,265
Total Duration	100 thousand minutes
Total Distance	36,617 km
Number of GPS	4.2 million

3. Results & Analysis

3.1 Introduction

The figure below summarises the various levels of analysis completed by TfL (Blue) and by Mapflow for this project (Light blue).

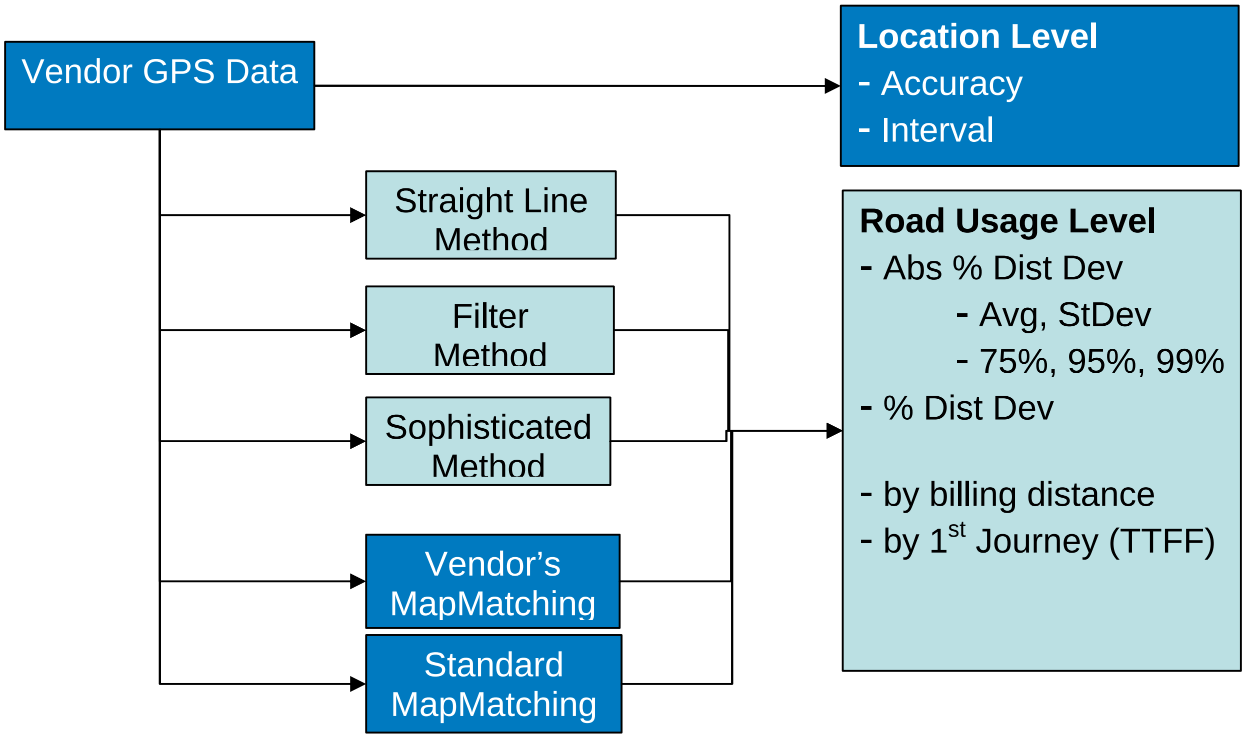


Figure 11 Analysis levels

In Appendix 3 the results and findings from the TfL analysis are provided for reference. In section 3.2 below the main analysis table has been extracted from the TfL analysis as it provides important context to the road usage results section.

3.2 Location Level Analysis

The detailed location error performance of all the OBU types is shown in Table 1, ranked by average location error. 'Vendor Average' is a weighted average which takes into account the number of location samples for each OBU type.

Note that all samples with a location error greater than 500m have been excluded. For the vendor OBUs there were a very small number of these (286 in total). In the case of the Reference device, 8,062 samples were excluded, the vast majority of which occurred during a single day.

OBU Type	Total Location Samples	Average Error (m)	Standard Deviation (m)	75% Quantile (m)	95% Quantile (m)	99% Quantile (m)
10	137,956	5.11	6.48	6.36	14.35	26.17
11	132,129	5.59	6.77	6.68	17.38	32.5
4	52,827	5.92	6.79	7.33	17.85	31.52
8	141,047	5.94	7.55	7.23	18.52	32.33
13	24,179	6.01	7.18	7.2	18.96	34.31
2	237,736	6.02	6.64	7.67	17.54	31.15
3	33,780	6.14	8	7.41	18.56	36.84
12	12,242	6.3	7.31	7.79	19.57	35.6
16	148,652	6.72	14.31	7.46	20.43	43.7
17	181,782	6.75	8.49	8.36	20.26	37.7
14	25,106	6.88	11.71	8.12	21.26	40.31
6	266,622	6.91	8.31	8.67	21.6	39.79
5	108,592	7.02	7.11	9.52	19.05	32.02
15	364,135	7.34	9.02	9.24	22.23	39.7
1	348,478	7.7	10.11	9.47	24.12	44.71
7	17,222	12.1	19.07	14.32	39.96	86.92
Vendor Average	-	**6.67**	**8.55**	**8.26**	**20.22**	**37**
18	3,184,464	8.41	10.56	10.52	27.37	50.17

Table 1 Detailed Location Performance (Source: TfL)

Note device 18 is a reference device which TfL has used in all three stages of their trials. By utilising this device TfL can measure improvements over time by the device vendors.

Key Result 1: The best OBU type had an overall average location error of 5.11 metres.

Key Result 2: The weighted average of overall location error for all vendor OBUs was 6.67 metres.

Key Result 3: 16 of the 17 vendor OBU types had an overall average location error of between 5 and 8 metres.

It should be noted that in Stage 3, TfL used a simple method to benchmark GPS accuracy, as compared to Stage 2 and 1. In those previous projects a great deal of effort was expended in determining the precise error, including longitudinal components. In this Stage 3 project a simple 'nearest line' approach was used to determine whether GPS performance was comparable with previous trials. The method is likely to be optimistic. For further TfL Location analysis we refer the reader to the Appendices.

3.3 Road Usage Results

The following section breaks out the results for each of the three algorithms used in the study:

- simple Straight line method
- simple Filter method
- Sophisticated method

For each method the following statistics (see section 2.2 for details on the calculations) are provided:

- Absolute % Distance Deviation
- Absolute % Distance Deviation Quantiles (75%, 95%, 99%)
- % Distance Deviation

3.3.1 Straight-line Road Usage Determination

3.3.1.1 Absolute % Distance Deviation

By applying the simple Straight-line method, described above in section 2.2.1, to the raw GPS data provided by the vendors in the original trials, we are able to calculate road usage distance. This can then be compared with the actual 'truth' distance driven to arrive at the Absolute % Distance Deviation metric described above in section 2.3.1.

$$\textit{Abs\%DistDev} = \frac{|(\textit{Determined Distance - Truth Distance})|}{\textit{Truth Distance}}$$

The table below summarises the average and standard deviation of this metric for the 17 vendors and the reference device.

Vendor	Journey Count	Avg Abs % Dist Dev	StDev Abs % Dist Dev	Average Location Error (m)	Average Sample Interval (s)
1	72	1143.60	5980.52	7.70	0.99
2	51	1.43	1.12	6.02	1.01
3	6	1.19	1.10	6.14	4.00
4	58	1.74	1.97	5.92	4.87
5	60	2.62	2.32	7.02	2.50
6	60	5.20	3.86	6.91	1.02
7	87	8.96	4.93	12.10	21.25
8	30	14.69	14.41	5.94	1.04
10	31	3.06	1.17	5.11	1.00
11	29	1.59	1.40	5.59	1.00
12	28	5.67	2.31	6.30	10.45
13	90	8.49	3.36	6.01	18.21
14	60	5.43	8.82	6.88	11.19
15	78	8.07	7.51	7.34	1.04
16	31	163.26	425.49	6.72	1.01
17	36	5.52	3.32	6.75	1.00
Vendor Avg		**113.37**	**553.75**		
18	*458*	*2.36*	*5.60*	8.41	1.00
Overall Avg		**97.51**	**355.16**		
Best Three		**2.17**	**4.64**		

Table 2 Overall Road Usage Performance – Straight-line method

The location error calculations in the Stage 3 TfL project excluded points with 500m error, on the grounds that they would be excluded by the operation of any map matching algorithm. The simple Straight line method has not excluded these points, on the grounds that it is the simplest possible approach.

This explains the massive difference for vendor 1, where a handful of points that were wildly erroneous led to massive errors in the Straight-line method.

Clearly some vendors are affected to a greater extent than others. For a selection of three good performing devices, the average performance is in the order of 2%.

Key Result 4: Average Abs%DistDev is 113% for vendor solutions.

Key Result 5: Average Abs%DistDev is 97.51% for all solutions in the original trial, including the reference device.

Key Result 6: Average Abs%DistDev is 2.17% and a standard deviation of 4.64% for a selection of three good performing devices

Key Finding 4: The simple Straight-line method works reasonably well for good GPS data. If the data is poor the performance of the method makes the algorithm unusable.

Note: throughout this study, we have selected 'good' or 'best' three devices. These have been chosen arbitrarily, based on the performance exhibited by the Straight-line method. By selecting '3 good' devices we hope to illustrate what 'good' performance may look like. By selecting other good devices, perhaps based on results from the Filter or map matching method, very different performance could be achieved. The best performance can be achieved by tuning a given method to the characteristics of a given set of good GPS data. One method applied to another GPS device might provide very poor results, because of a particular susceptibility in the method to a phenomenon that is frequently produced by the device. The reverse is also true. Throughout this study our goal is to provide 'typical' and 'indicative' performance indications.

3.3.1.2 Abs % Distance Deviation Quantiles

Recognising the high level goal of achieving 99% of bills with 1% or less deviation, we are not only interested in the average performance. By counting the number of journeys that exhibited a given level of Abs%DistDev we can generate the histogram and empirical cumulative distribution function, as illustrated by the graphs below.

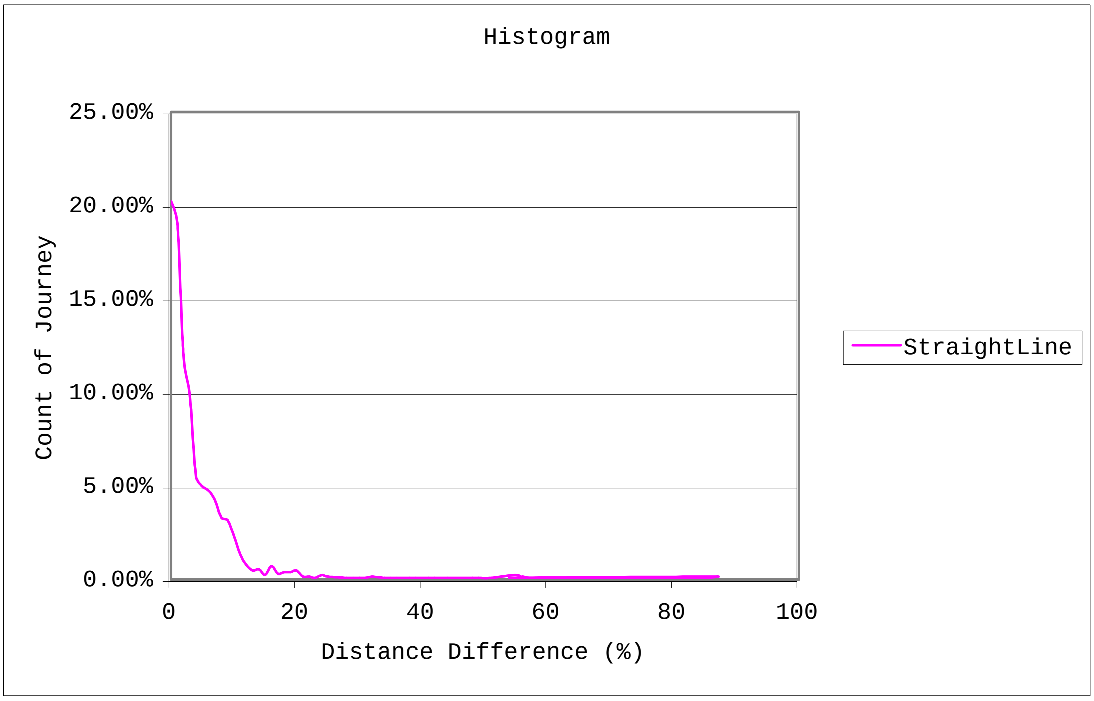


Fig 12 population distribution of Abs%DistDev for the Straight-line method

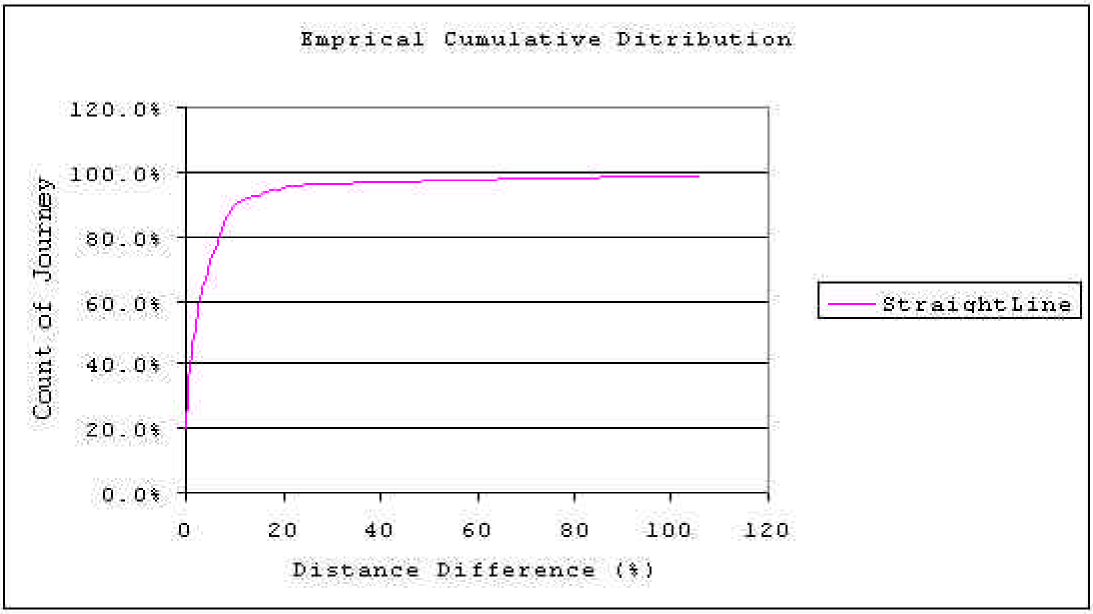


Fig 13 empirical cumulative distribution of Abs%DistDev for the Straight-line method

Key Finding 5: From this we can conclude that with the Straight-line method, GPS is unlikely to provide the proportion of billing events (99%) with the requisite deviation level (<1%), if 24km were the typical billing length.

Key Result 7: In an attempt to achieve the "99% of bills within 1% error" requirement, if all vendors are included;
- The proportion achieving < 1% error is 20.24%

- The margin required to achieve 99% confidence is 106%

Key Result 8: For the three selected good devices (2,11,18):
- The proportion achieving < 1% error is 35.1%
- The margin required to achieve 99% confidence is 11%

3.3.1.3 % Distance Deviation

As noted earlier in this document, compared to the absolute distance deviation, this metric provides an indication as to whether the deviation is over estimated or under estimated. A journey distance that is underestimated by the method is represented as a negative % figure.

By counting the number of journeys that exhibit a given level of %DistDev we can plot the population distribution.

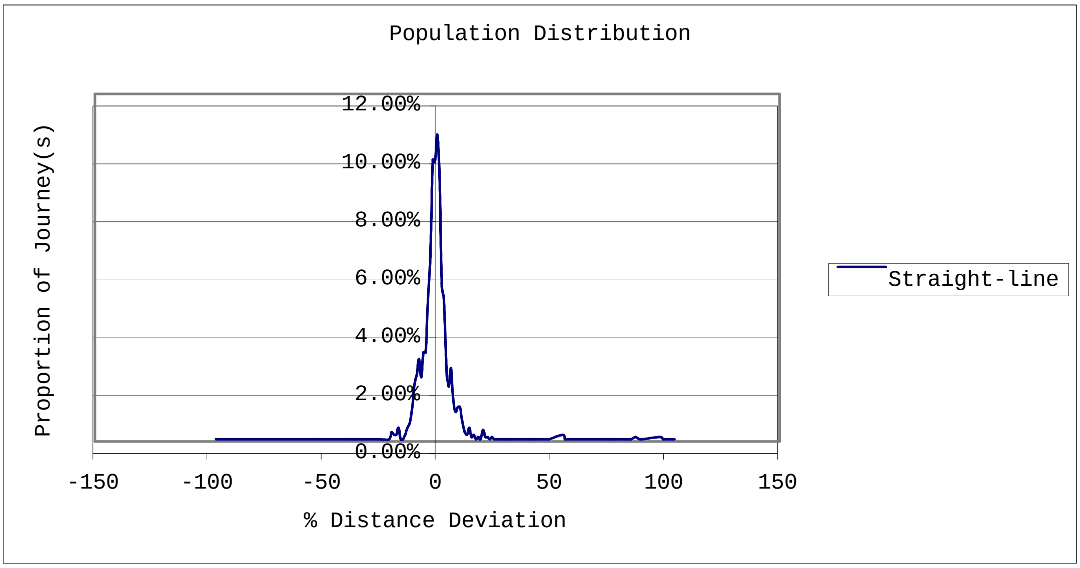


Fig 14: population distribution of %DistDev for Straight-line method[5]

Key Finding 6: From this analysis we find that 38.42% of journeys would be overestimated, while 50.99% of journeys would be underestimated

Notice that even the Straight-line method tends to underestimate more frequently than overestimate. In fact, over all journeys this method overestimates, since in the instances where the GPS error caused an overestimate the deviation was very large, overwhelming underestimates in other journeys.

[5] The charts have been clipped at 100% such that the relative information is visible

3.3.2 Filter Road Usage Determination

3.3.2.1 Absolute % Distance Deviation

By applying the simple Filter method, described in section 2.2.2, to the raw GPS data provided by the vendors in the original trials, we are able to calculate road usage distance. This can then be compared with the actual 'truth' distance driven to arrive at the Absolute % Distance Deviation metric described above.

The table below summarises the average and standard deviation of this metric for the 17 vendors and the reference device.

Vendor	Journey Count	Abs % Dist Dev	StDev % Dist Dev
1	72	9.67	16.55
2	51	1.28	0.95
3	6	4.42	1.70
4	58	4.53	2.55
5	60	3.25	2.63
6	60	3.62	2.12
7	87	19.05	5.93
8	30	4.79	5.53
10	31	3.20	1.09
11	29	1.75	1.55
12	28	12.09	4.75
13	90	17.31	6.15
14	60	8.19	6.85
15	78	5.24	4.16
16	31	10.43	9.93
17	36	3.56	2.30
Vendor Avg		**8.26**	**5.27**
18	*458*	2.55	5.56
Overall Avg		**8.22**	**5.26**
Best Three		**2.33**	**4.62**

Average Location Error (m)	Average Sample Interval (s)
7.70	0.99
6.02	1.01
6.14	4.00
5.92	4.87
7.02	2.50
6.91	1.02
12.10	21.25
5.94	1.04
5.11	1.00
5.59	1.00
6.30	10.45
6.01	18.21
6.88	11.19
7.34	1.04
6.72	1.01
6.75	1.00
8.41	1.00

Table 3: Overall Road Usage Performance – Filter method

As compared to the Straight-line method, the Filter is more capable of excluding these wildly erroneous points (see vendor 1 and 16), hence the very large deviations found in the Straight-line are not as apparent here.

Since the Filter approach is more capable of dealing with outlier points, the spread of vendors is less than in Straight-line. When the good vendors have no outliers, the Filter acts more to eliminate good data then to exclude bad data. So, when we take three good vendors, the average performance is in the order of 2.3%, which is marginally higher than the Straight-line method.

There also seems to be a correlation between the sample interval and poor GPS performance (note vendors 7,12 and 13.

Key Result 9: Average Abs%DistDev is 8.26% for vendor solutions.

Key Result 10: Average Abs%DistDev is 8.22% for all solutions in the original trial, including the reference device.

Key Result 11: Average Abs%DistDev is 2.33% and a standard deviation of 4.62% for the selection of three good performing devices (2,11 and 18)

Key Finding 7: The simple Filter method works well to remove the effect of erroneous GPS data. However, the Filter can impact the performance where it excludes good data from the calculations. There is some correlation between sample interval and poor performance

3.3.2.2 Abs % Distance Deviation Quantiles

Recognising the high level goal of achieving 99% of bills with 1% or less deviation, we are not only interested in the average performance. By counting the number of journeys that exhibited a given level of Abs%DistDev we can generate the histogram and empirical cumulative distribution function, as illustrated by the graphs below.

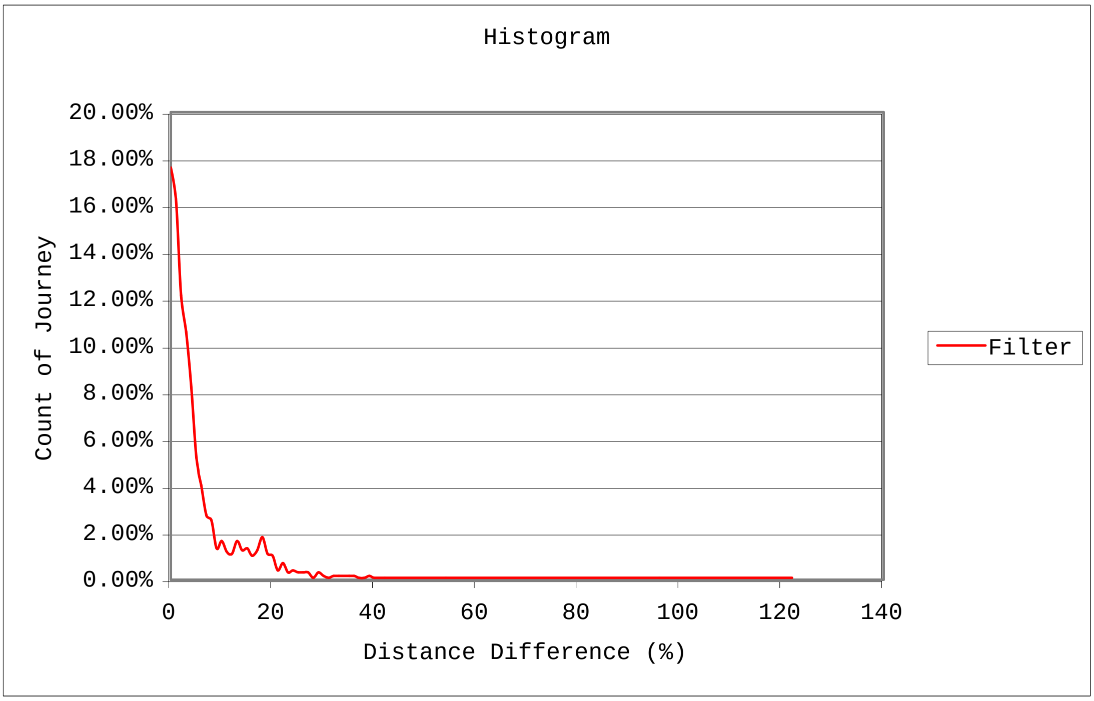


Fig 15 population distribution of Abs%DistDev for the Filter method

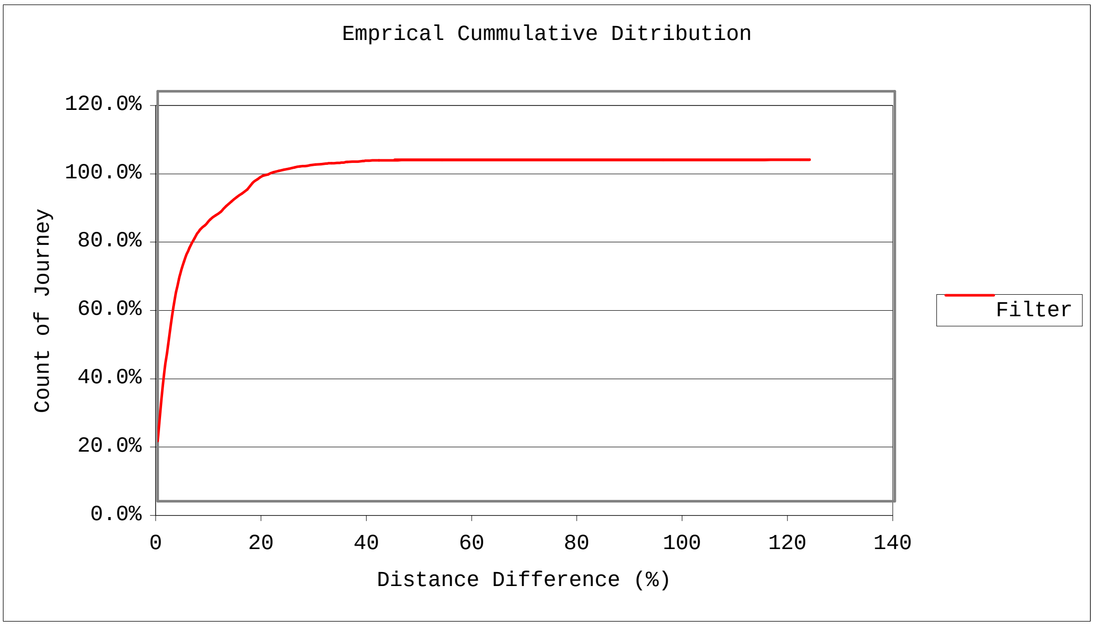


Fig 16 empirical cumulative distribution of Abs%DistDev for the Filter method

Key Finding 8 : From this we can conclude that with the Filter method, GPS is unlikely to provide the proportion of billing events (99%) with the requisite deviation level (<1%), if 24km were the typical billing length.

Key Result 12: In an attempt to achieve the “99% of bills within 1% error” requirement, if all vendors are included;
- The proportion achieving < 1% error is 17.6%
- The margin required to achieve 99% confidence is 33%

Note: this suggests that the Filter method is less able to achieve high performance (I.e. <1%) than the Straight-line method. Suggesting that the Filter is removing good GPS data even when performance of the raw data is good – since in the same scenario the Straight-line method takes the data at face value

Key Result 13: For the three selected good devices (2,11,18):
- The proportion achieving < 1% error is 32.5%
- The margin required to achieve 99% confidence is 10%

3.3.2.3 % Distance Deviation

As compared to the absolute distance deviation, this metric provides an indication as to whether the deviation is over estimated or under estimated. A journey distance that is underestimated by the method is represented as a negative % figure.

By counting the number of journeys that exhibit a given level of %DistDev we can plot the population distribution.

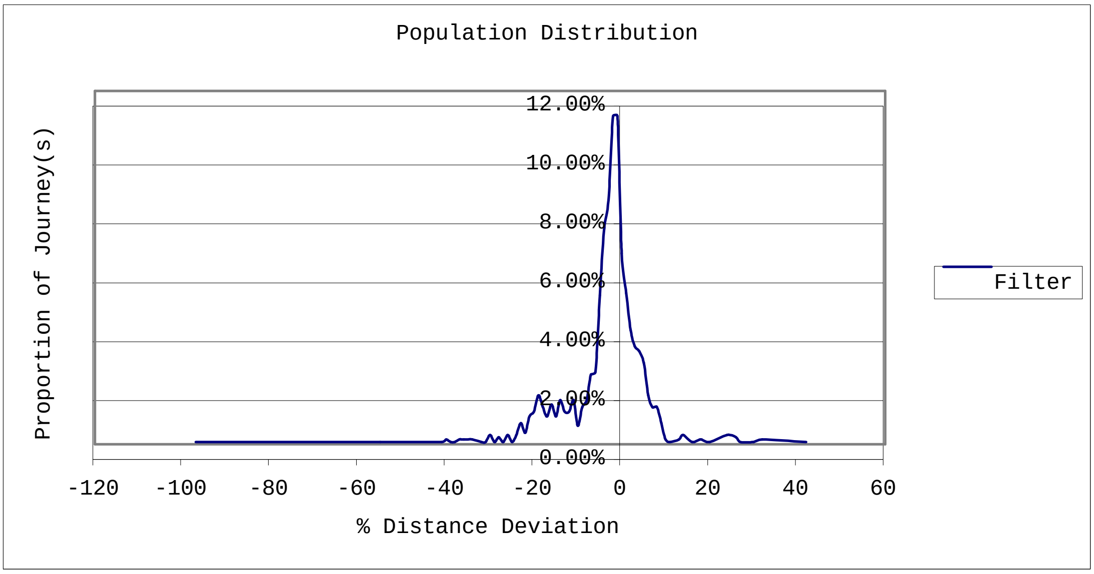


Fig 17: population distribution of %DistDev for Filter method

From this it appears that, as compared to the Straight-line method, which is very much centred on 0%, the median for the Filter is negative (-1.5%). Despite this, we find that, just as in the Straight-line method, 25.22% of journeys would be overestimated, while 68.3% of journeys would be underestimated (ignoring the dead zone at <1% deviation).

Key Finding 9: From this analysis we find that 25.22% of journeys would be overestimated, while 68.3% of journeys would be underestimated

3.3.3 Sophisticated Road Usage Determination

3.3.3.1 Absolute % Distance Deviation

By applying the first iteration of the more Sophisticated method, described in section 2.2.3, to the raw GPS data provided by the vendors in the original trials, we are able to calculate road usage distance. This can then be compared with the actual 'truth' distance driven to arrive at the Absolute % Distance Deviation metric described above.

The table below summarises the average and standard deviation of this metric for the 17 vendors and the reference device.

Vendor	Journey Count	Abs % Dist Dev	StDev % Dist Dev	Average Location Error (m)	Average Sample Interval (s)
1	72	57.08	217.38	7.7	0.99
2	*51*	7.32	1.94	6.02	1.01
3	6	4.71	2.61	6.14	4
4	58	5.28	2.41	5.92	4.87
5	60	4.85	2.54	7.02	2.5
6	60	9.83	3.06	6.91	1.02
7	87	40.04	7.45	12.1	21.25
8	30	10.73	4.57	5.94	1.04
10	31	4	1.28	5.11	1
11	*29*	6.26	2.2	5.59	1
12	28	5.42	3.88	6.3	10.45
13	90	18.7	6.24	6.01	18.21
14	60			6.88	11.19
15	78	11.59	4.95	7.34	1.04
16	31	26.5	11.78	6.72	1.01
17	36	9.36	2.52	6.75	1
Vendor Avg		**16.98**	**23.1**		
18	*458*	5.92	5.84	8.41	1
Overall Avg		**17.30**	**16.85**		
Best Three		**6.07**	**5.27**		

Table 4 Overall Road Usage Performance – Sophisticated method

As compared to the Straight-line and Filter methods, the Sophisticated method is more capable of fitting a curve to a set of points from a sample GPS. However, it is more prone to losing control for periods of time. This is due to the nature of the algorithm itself. Notice the very large average deviations for 1, 7, 16 has re-appeared (as with the Straight-Line method). We feel this algorithm is more subtle and does not filter out (as robustly as the Filter method) the erroneous GPS data. The results are not as poor as in the Straight-line method, but the effect is still evident. Also, in the case of 7, the interval of 21.25sec may have a bigger impact on the subtle algorithm.

Key Result 14: Average Abs%DistDev is 16.98% for vendor solutions.

Key Result 15: Average Abs%DistDev is 17.30% for all solutions in the original trial, including the reference device.

Key Result 16: Average Abs%DistDev is 6.07% and a standard deviation of 5.27% for the selection of three good performing devices (2,11,18).

Key Finding 10: The Sophisticated algorithm has not performed as well as the Straight-line or Filter method. Further tuning of the algorithm may lead to improved results.

3.3.3.2 Abs % Distance Deviation Quantiles

Recognising the high level goal of achieving 99% of bills with 1% or less deviation, we are not only interested in the average performance. By counting the number of journeys that exhibited a given level of Abs%DistDev we can generate the histogram and empirical cumulative distribution function, as illustrated by the graphs below.

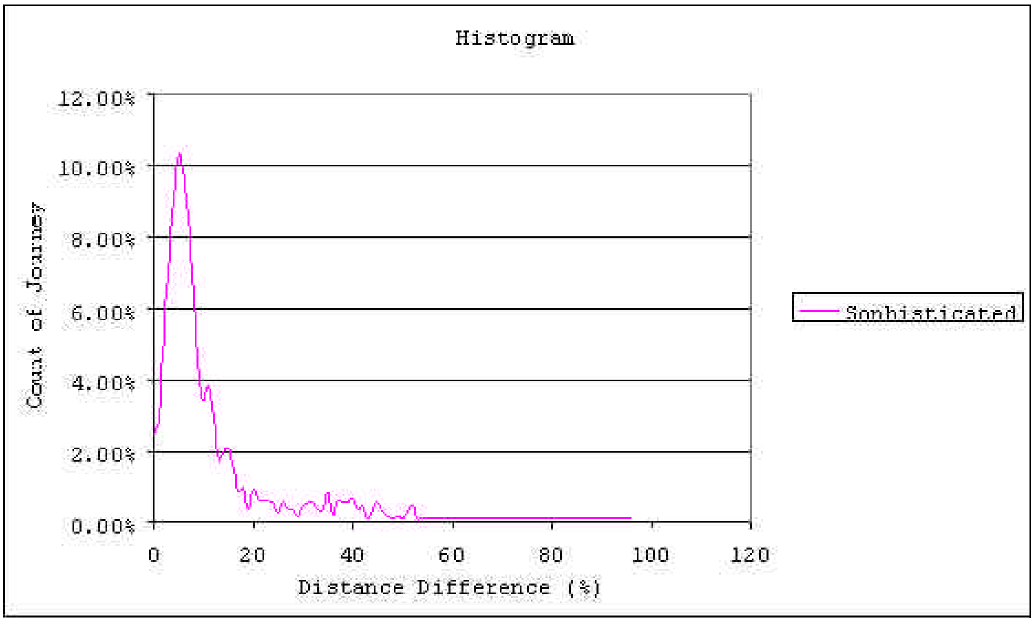


Fig 18 population distribution of Abs%DistDev for the Sophisticated method

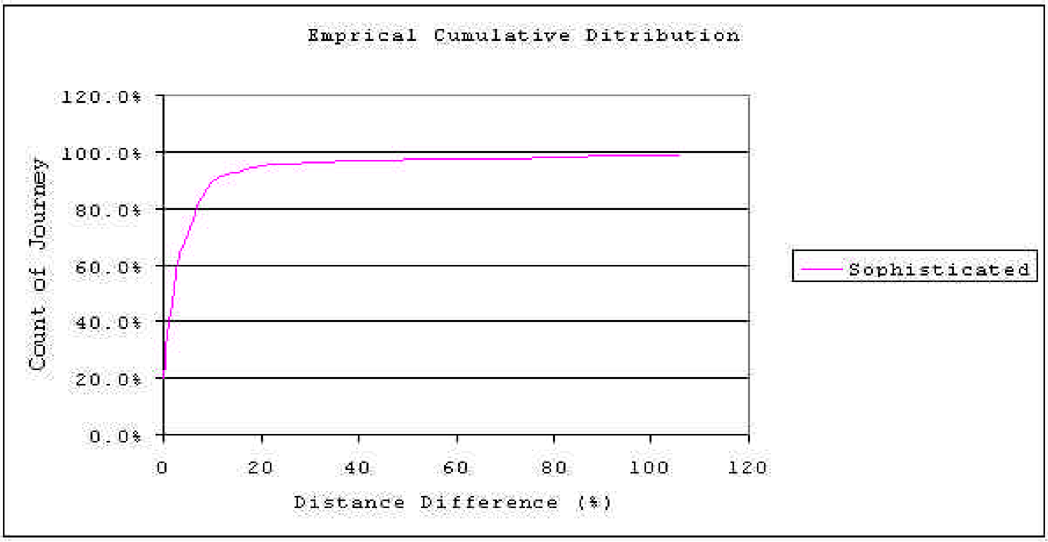


Fig 19 Empirical cumulative distribution of Abs%DistDev for the Sophisticated method

From this we can conclude that with the current configuration of the Sophisticated method, GPS is unlikely to provide the proportion of billing events (99%) with the requisite deviation level (<1%), if 24km were the typical billing length. It is also much more susceptible to going out of control than the Straight-line or Filter methods.

Key Finding 11: From this we can conclude that with the Sophisticated method, GPS is unlikely to provide the proportion of billing events (99%) with the requisite deviation level (<1%), if 24km were the typical billing length.

Key Result 17: In an attempt to achieve the "99% of bills within 1% error" requirement, if all vendors are included;
- The proportion achieving < 1% error is 2.45%
- The margin required to achieve 99% confidence is 52%

Key Result 18: For the three selection good devices (2,11,18):
- The proportion achieving < 1% error is 3.1%
- The margin required to achieve 99% confidence is 20%

Of the three methods the Sophisticated algorithm has performed the worst.

Figure 20 below shows in a visual manner how the Sophisticated algorithm is performing.

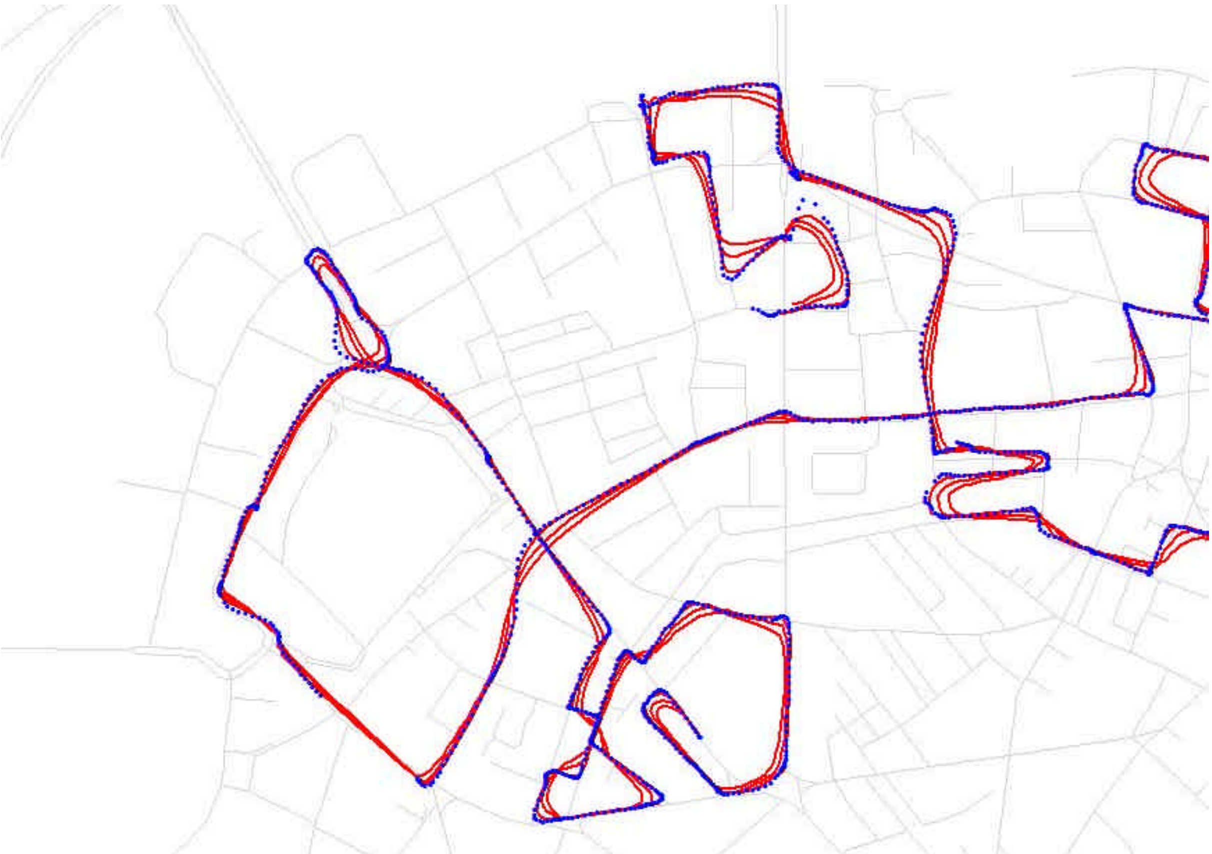


Fig 20: Comparison of Sophisticated method performance (red) for various tuning parameters, against GPS (blue).

3.3.3.3 % Distance Deviation

As compared to the absolute distance deviation, this metric provides an indication as to whether the deviation is over estimated or under estimated. A journey distance that is underestimated by the method is represented as a negative % figure.

By counting the number of journeys that exhibit a given level of %DistDev we can plot the population distribution.

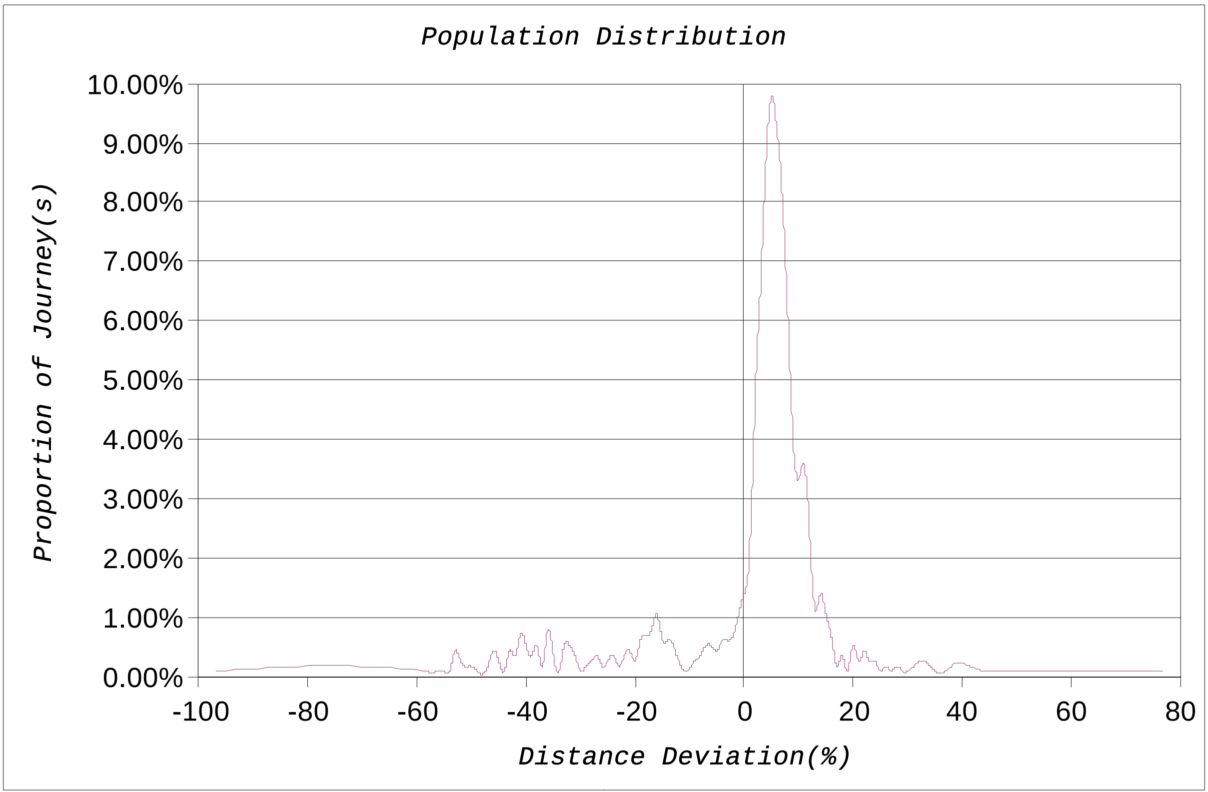


Fig 21 population distribution of %DistDev for Sophisticated method

From the above analysis it appears that the Sophisticated method is positive (5.5%).

Unlike for the other methods, 70.4% of journeys would be overestimated, while 19.2% of journeys would be underestimated.

Key Finding 12: From this analysis we find that 70.4% of journeys would be overestimated, while 19.2% of journeys would be underestimated

3.3.4 Road Usage using Map Matching

3.3.4.1 Using Vendor map matching

By way of comparison, road usage performance from TfL's trials for each of the vendor's own systems are given in Table 5 below. Note that only a subset of vendors was able to provide road usage data, as compared to just GPS data. For the purpose of this project, we have taken the original figures published by TfL, which were categorised into % of journey Correct, Incorrect or Missing and calculated the Abs%DistDev and %DistDev to enable comparison.

All values are percentages by journey length. 'Vendor Average' is a weighted average which takes into account the number of journeys performed for each OBU type.

Vendor Device	Correct	Missing	Incorrect	Abs%DistDev	%DistDev
6	98.60%	1.50%	1.00%	0.40%	-0.40%
5	98.50%	1.50%	0.90%	0.60%	-0.60%
10	97.60%	2.40%	1.10%	1.30%	-1.30%
17	97.50%	2.50%	1.20%	1.30%	-1.30%
2	97.50%	2.50%	1.80%	0.70%	-0.70%
15	97.30%	2.90%	2.80%	0.10%	0.10%
1	96.50%	3.50%	3.60%	0.10%	0.10%
4	94.80%	5.20%	10.20%	5.00%	5.00%
14	93.40%	6.60%	33.40%	26.80%	26.80%
Vendor Average	**96.80%**	**3.20%**	**6.40%**	**3.20%**	**3.20%**

Table 5 Overall Road Usage Performance – Vendor Map-Matching (Source TfL)

Key Result 19: The best vendor's road usage detection deviation is only 0.1% of the actual distance driven.

Key Result 20: The weighted average vendor's road usage detection deviation is 3.20% from actual route driven.

3.3.4.2 Standard Map-Matching

Similarly, TfL in the Stage 3 trials passed all vendor and reference GPS data through a 'Standard' map matching process, to produce road usage information. For the purpose of this project we have taken those figures published by TfL and calculated the Abs%DistDev and %DistDev to enable comparison with the non-map based methods.

Vendor Device	Correct	Missing	Incorrect	Abs%DistDev	%DistDev
10	97.00%	3.00%	4.20%	1.20%	1.20%
2	96.40%	3.70%	5.00%	1.40%	1.40%
17	94.50%	5.50%	5.80%	0.30%	0.30%
6	94.50%	5.50%	5.00%	0.50%	-0.50%
11	93.80%	6.20%	6.40%	0.20%	0.20%
5	93.40%	6.60%	5.80%	0.80%	-0.80%
3	91.90%	8.10%	4.40%	3.70%	-3.70%
4	91.00%	9.00%	4.80%	4.20%	-4.20%
16	89.90%	10.10%	5.00%	5.10%	-5.10%
8	89.70%	10.30%	4.80%	5.50%	-5.50%
12	89.70%	10.30%	5.00%	5.30%	-5.30%
15	88.10%	11.90%	8.30%	3.60%	-3.60%
9	86.40%	6.20%	14.10%	0.50%	0.50%
14	86.10%	13.90%	5.70%	8.20%	-8.20%
7	82.00%	18.00%	12.90%	5.10%	-5.10%
1	81.00%	19.00%	13.60%	5.40%	-5.40%
13	31.50%	68.50%	12.00%	56.50%	-56.50%
Vendor Average	**83.60%**	**16.40%**	**7.60%**	**8.80%**	**-8.80%**
Reference	88.10%	11.90%	8.50%	3.40%	-3.40%

Table 6 Overall Road Usage Performance (Abs%DistDev) for Standard Map-Matching (source TfL)

Key Result 21: The best road usage determination with standard map-matching method was 0.2% of the actual route driven.

Key Result 22: The weighted average of road usage deviation for all vendor OBUs with standard map-matching method was -8.8%.

Key Finding 13: The performance of the standard map matching algorithm applied to vendor data was worse than the performance of any vendor's own algorithms applied to their own data. This may be because a simple conventional curve-to-curve approach was used for the standard algorithm. Also it was tuned to optimise all vendor's GPS data, rather than just the RUC vendors.

3.3.5 Billing Period Analysis

One major topic for this study is to determine the performance associated with aggregating journey level results into larger 'billing' periods. The suggestion (made by ARS) being that Standard Error implies that a greater confidence (lower Standard Deviation) could be achieved if the bill covered a longer distance driven, as follows:

$$S_E = \frac{\widehat{\sigma}}{\sqrt{n}}$$

where

$\widehat{\sigma}$ is an estimate of the standard deviation s of the population, and
n is the size (number of items) of the sample.

To investigate this, Mapflow summated the journey level results (typically 24km) into three larger bill sizes: 350 km, 1000 km and 1350 km. To do this we simple aggregated consecutive journeys until the requisite bill size was generated. An overestimation from one journey would be combined with an underestimation from another. We also aggregated the actual distance driven for these journeys, such that performance at a billing level could be determined.

However, as the usable journey dataset is c.26,000km from the 36,000km driven, such an approach led to a statistically low number of sample bills (for 1,350km there were <20 bills). In order to test the validity of applying the Standard Error, using the TfL data we have summated the journey level results into shorter billing journeys of 24km, 48km, 96km, 240km and 350km. The results are provided in the table below.

Billing Distance	ARS (StDev)	Count	Straight-Line (StDev)	Filter (StDev)	Sophisticated (StDev)
1km	9.20%	N/A	N/A	N/A	N/A
24km	1.88%	1078	1061.47%	8.01%	43.19%
48km	1.32%	539	853.99%	7.11%	34.24%
96km	0.94%	273	614.06%	6.47%	25.15%
240km	0.59%	112	441.15%	6.16%	19.11%
350km	0.50%	83	357.98%	5.95%	N/A
1350km	0.25%	N/A	N/A	N/A	N/A

Table 7 Comparison of billing period standard deviations

The chart below illustrates similar behaviour between the Standard Error, as calculated by ARS and the results from the Filter method.

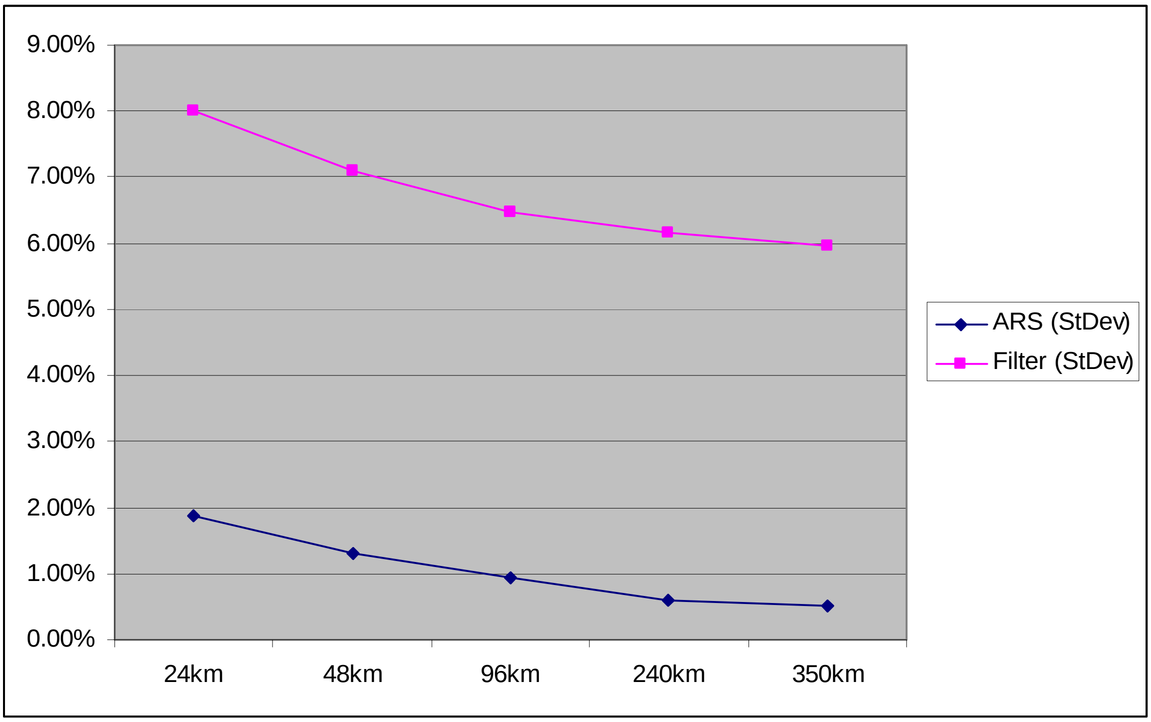


Fig 22 Comparison of billing period standard deviations

Figure 23 below shows how the Standard Deviation changes as the billing size increased, for all devices, the selected three good devices and for the reference device (18).

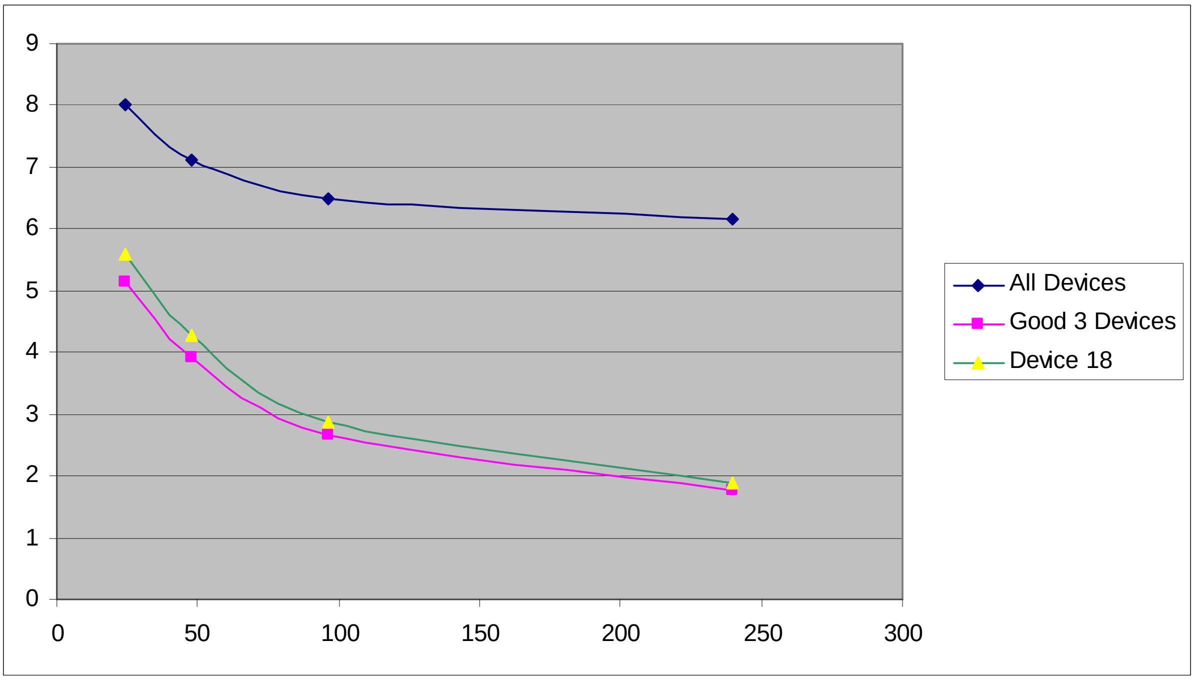


Fig 23 Comparison of billing period standard deviations

While all three curves show a similar behaviour as billing size increases, the single device shows the greatest improvement. Device 18 dominates the “good 3 devices” bill due to the sheer number of journeys driven with this device. This is the reason for the similarity between the best 3 and the single device.

Figure 24 below, illustrates the relationship between the empirical results and the theoretical Standard Error for the single device (18).

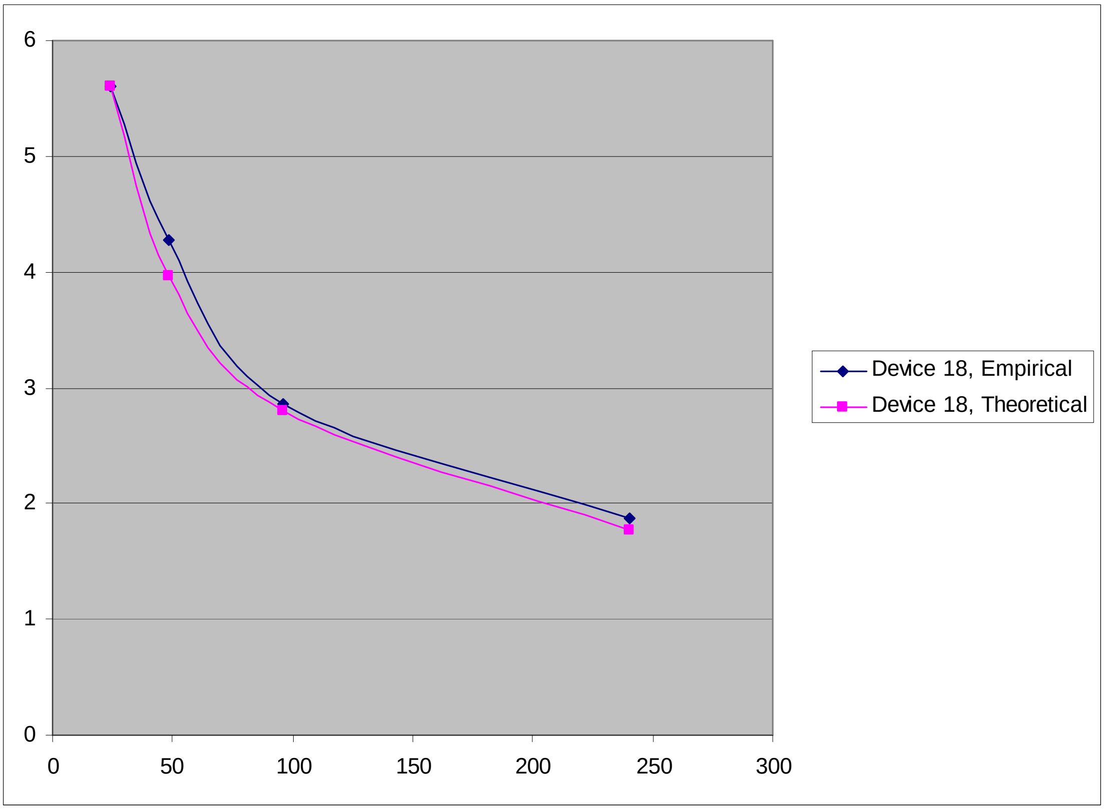


Fig 24 Comparison of empirical and theoretical standard deviation for bills for a single device

Notice the strong correlation between the results. This reinforces the view that with greater homogeneity the behaviour approaches the theoretical Standard Error behaviour.

To compare 99% confidence intervals, the two methods have been compared, firstly an empirical calculation of the 99% interval and secondly by deriving 2.6 times the standard deviation, as might suggested by assuming a Gaussian distribution.

Billing Distance	Bills above 99% confidence interval	Empirical Filter Abs%DistDev (99%)	Derived Filter Abs%DistDev (99%)	ARS Abs%DistDev (99%)
24km	11	33.5	20.83	4.89
48km	6	28.4	18.48	3.43
96km	3	24.5	16.83	2.44
240km	1	20.6	16.00	1.30

Table 8 Comparison of billing period 99% confidence intervals

The chart below though does show similarity between the ARS study results and the Filter results, for all device, using the 2.6 times standard deviation formula for all devices.

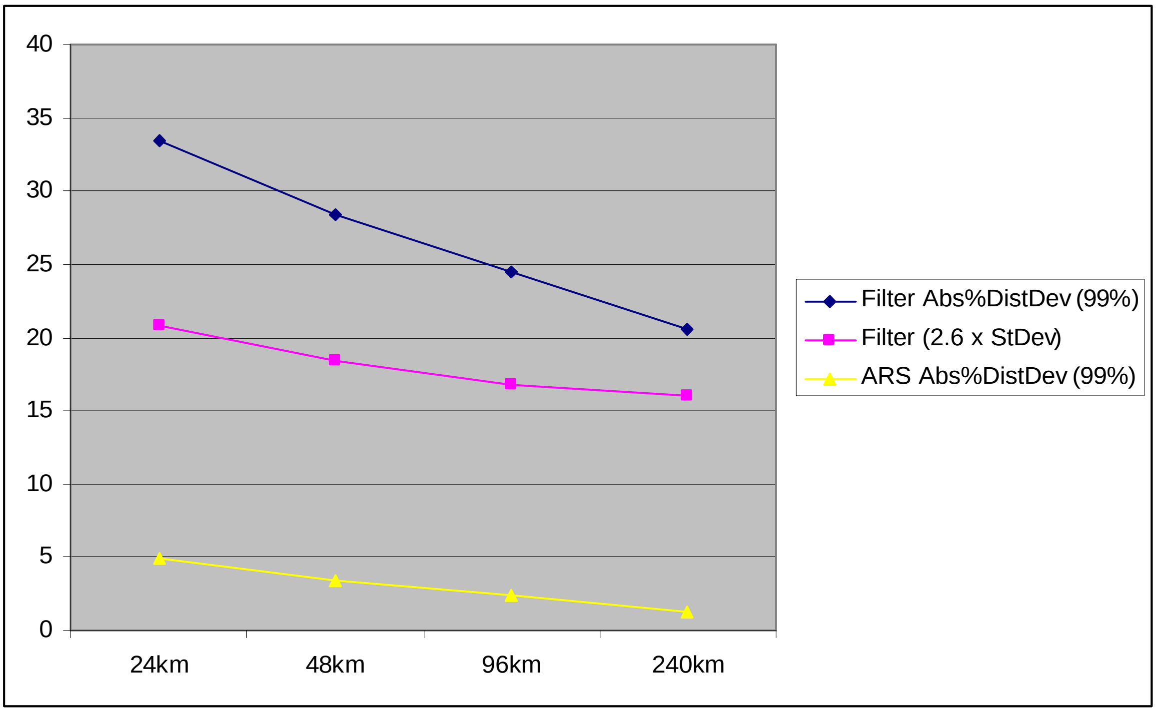


Fig 23 Comparison of billing period 99% confidence intervals

Notice that while the curves of 99% quantiles, for all devices, do tend to follow a similar behaviour, the actual empirical results are very much larger than the theoretical. Firstly, this may be due to the very few samples for large bills - great caution should be taken in considering these results. Secondly, just as in the Standard Deviation, the heterogeneity associated with combining many devices may dominate this comparison. Furthermore, behaviour is unlikely to be Gaussian, a mix of Guassians may be more appropriate.

Note: Insufficient data, limited number of samples to determine 99% quantiles >200km.

Key Finding 14: More analysis is required of a larger sample to achieve the analysis required by the Ministry. That said, the analysis at shorter billing periods demonstrates there is correlation for a single device.

Key Finding 15: When multiple devices are combined the correlation between empirical and theoretical reduces.

Key Finding 16: For a simple method such as Filter, there is a bias (say 6% avg) toward underestimation. Increasing bill size will not change this % bias. Instead the bias could be assumed systematic, and taken out by adjustment (e.g. adding 6%). The billing efficacy is then limited to the variability (StDev) of calculation. This variability reduces with bill size (n).

Key Finding 17: The analysis has been completed on one type of environment. The correlation may be impacted by other local conditions (built environment, TTFF). It was not possible to study such impacts with the data from these particular trials.

3.3.6 First journey of the day analysis (TTFF)

The ARS study highlights the effect that time to first fix (TTFF) could have on the efficacy of determining road usage by GPS. In their original study, TfL did not attempt to isolate or control for TTFF, since they regarded it as part of the normal operation of vendor solutions.

In an attempt to quantify the variation in performance TfL did evaluate the GPS performance during the initial seconds of the journey (chart below). This takes all vendor data, for all journeys, including those where TTFF would not occur.

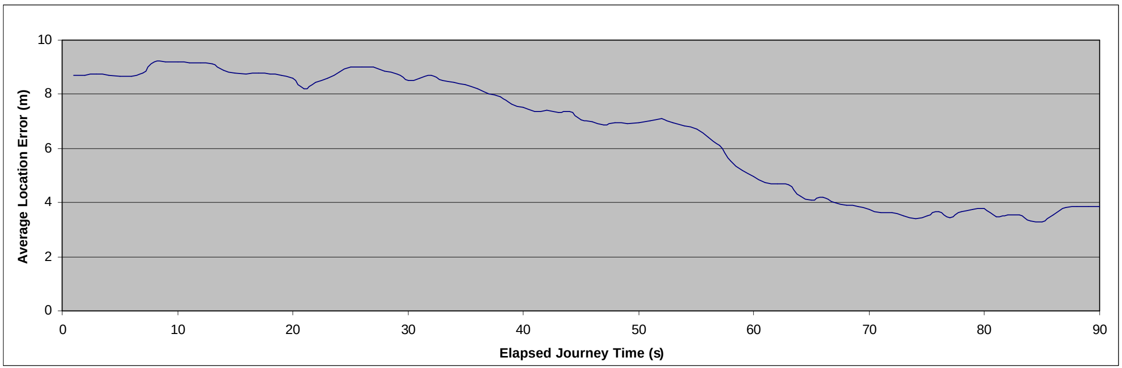


Fig 25 Average location error over lapsed journey time (source TfL)

The following table is the results between the first journey and subsequent journeys for each of the devices used in the trial.

Vendor ID	Count	First Journey	Straight-line Abs%Dist Dev (Avg)	Filter Abs%Dist Dev (Avg)	Soph Abs%Dist Dev (Avg)		Straight-line Abs%Dist Dev (StDev)	Filter Abs%Dist Dev (StDev)	Soph Abs%Dist Dev (StDev)
1	23	TRUE	2,087.64	12.30	83.03		9,285.05	25.49	313.12
1	49	FALSE	700.47	8.44	44.90		3,572.58	10.12	156.47
2	18	TRUE	1.05	1.05	7.25		0.89	0.80	1.85
2	33	FALSE	1.63	1.41	7.36		1.19	1.01	2.02
3	2	TRUE	0.48	3.46	6.62		0.17	0.21	2.49
3	4	FALSE	1.55	4.90	3.76		1.23	1.97	2.37
4	21	TRUE	1.90	4.47	5.74		2.50	2.57	1.79
4	37	FALSE	1.65	4.57	5.02		1.62	2.57	2.69
5	21	TRUE	2.69	3.24	4.93		2.59	3.00	2.38
5	39	FALSE	2.58	3.25	4.80		2.19	2.44	2.66
6	21	TRUE	3.53	2.92	8.46		2.22	1.97	2.67
6	39	FALSE	6.10	3.99	10.57		4.26	2.14	3.04
7	29	TRUE	8.42	19.75	41.18		5.56	6.85	6.63
7	58	FALSE	9.23	18.71	39.47		4.62	5.44	7.82
8	11	TRUE	13.79	6.13	11.39		6.68	8.60	5.50
8	19	FALSE	15.22	4.01	10.35		17.57	2.54	4.05
10	12	TRUE	2.98	3.22	4.06		1.17	0.95	1.31
10	19	FALSE	3.12	3.19	3.96		1.21	1.20	1.29

11	10	TRUE	1.92	2.10	5.92		2.17	2.31	2.53
11	19	FALSE	1.41	1.56	6.44		0.78	0.99	2.06
12	9	TRUE	5.14	10.67	4.28		3.02	4.78	3.96
12	19	FALSE	5.93	12.76	5.96		1.93	4.72	3.83
13	31	TRUE	8.18	16.64	18.06		3.15	5.92	5.92
13	59	FALSE	8.66	17.66	19.03		3.47	6.29	6.42
15	26	TRUE	7.09	4.64	11.18		4.94	2.69	3.87
15	52	FALSE	8.56	5.54	11.79		8.52	4.73	5.43
16	12	TRUE	266.70	11.33	24.30		684.93	11.20	12.62
16	19	FALSE	97.93	9.86	27.89		58.40	9.32	11.35
17	12	TRUE	5.57	3.87	9.46		3.51	2.30	2.38
17	24	FALSE	5.50	3.40	9.31		3.29	2.34	2.64
18	141	TRUE	2.42	2.35	5.93		3.51	3.49	3.83
18	257	FALSE	2.33	2.66	5.91		6.47	6.42	6.70

Table 9 First Journey Analysis by device

In some cases, the results for the first journey are worse than the later journeys whilst for other devices it is vice versa. Likewise there is no correlation between performance between the filters and TTFF. Therefore, it is difficult from the above data to draw a conclusion about the impact of TTFF.

Key Finding 18: There is some indication that TTFF has an impact on performance of devices for first journeys. However, the results are not conclusive and a different approach is required to analyse the impact of TTFF.

4 Findings and Conclusions

4.1 Findings

The key findings from this study are summarised below:

Methods:

Three algorithms were used in the study with each having very different results.

- The simple “Straight line” algorithm provided the best performance for journeys with good GPS data. For journeys with erroneous data the algorithm exhibited very poor performance.

- The simple “Filter” algorithm removed the erroneous data from the calculations improving performance generally. However, this filter also adversely affected performance in cases where GPS was good.

- Reviewing the results of the “Sophisticated” algorithm anecdotally suggests promise. The algorithm is able to interpolate and extrapolate out GPS error and unavailability. However, in an attempt to achieve 0% error, the algorithm was tuned to the point where it became unreliable.

Table 10 and 11 below are a summary of the findings for the three algorithms. Note the numbers are quoted as absolute % distance deviation from the actual true distance driven and are for journeys where results were produced by all 3 methods.

Journey Count	Straight Line (Avg) %	Filter (Avg) %	Sophisticated (Avg) %		Straight (StDev)	Filter (StDev)	Sophisticated (StDev)
1145	80.44	6.28	14.03		1516.95	8.47	56.35

Table 10 Summary analysis results for all three methods

Vendor	Count	Straight Line (Avg) %	Filter (Avg) %	Sophisticated (Avg) %	Map Matching (Avg) %
2	51	1.43	1.28	7.32	1.4
11	29	1.59	1.75	6.26	0.2
18	398	2.36	2.55	5.29	3.4

Table 11 Summary analysis results for the three selected good devices

Taking the three selected good performing devices, it is possible to achieve using the filter algorithm 2.33% and 4.62% distance deviation (average and standard deviation respectively). The map matching performance from TfL demonstrates that it can be as good as 0.2% for certain devices.

Recommendation: the Ministry could consider introducing a 'bakers dozen' bias to adjust for the 2.33% off set. However, this figure is correct for Southwark, for these 'good' devices. The Ministry should seek to understand how this bias would vary with environment, device, driving behaviour and calculation method.

Recommendation: In order to improve the performance of non-map based road usage determination methods, we recommend that more time is spent tuning this and other sophisticated methods. Whereas the current project could only apply a best-guess setup, modest development effort would likely deliver significantly enhanced results, as compared to the very simple filter method.

These vendors would likely deliver 35% of journeys within 1% error. A margin of 10% would be required for 99% confidence.

Recommendation: This analysis of data that was gathered in London raises some concerns and highlights the challenges involved in road usage determination using GPS in that environment. We recommend a short study to identify whether and where environments similar to, or 'worse' (in the context of likely GPS performance) than, London exist in the Netherlands

Recommendation: Secondly to repeat similar fixed route trials with a selection of the best vendor devices in these identified areas. Such trials can help the Ministry in identifying and understanding issues that vendors will need to be able to resolve for a successful KM charging solution.

Billing Size:

Due to the sample size it was not possible to analyse groupings of the TfL 24km journeys into the required 350km, 1000km and 1,350km bills. However, an analysis was completed at billings of 24km, 48km, 96km, 240km and 350km.

For all devices the standard error decreases as expected but at a lesser rate. Analysis for an individual device shows a strong correlation between empirical and theoretical calculation of standard deviation.

Impact of multiple devices, different journey lengths and different environments will impact whether the journey population in a bill follows a Gaussian distribution.

Recommendation: according to the business requirements, the Ministry may wish to introduce a further bias in order to ensure drivers are rarely over or undercharged. The magnitude of such an adjustment would be related to the variability of distance determination. We have found that this reduces with billing size. Therefore the Ministry should consider larger billing

periods where feasible, but should also consider the 'test cases' where smaller billing periods are inevitable. A fixed-fee 'floor' provision in all bills might be useful to mask variability for these shorter bill cases.

Recommendation: Recognising that GPS based road usage determination may not be Gaussian, to quantify how the confidence in performance improves with billing size, we would recommend that a larger study, with different environments which generates bills comprised of realistic journey lengths, be performed. This study should also factor in the fixed effect due to TTFF.

Time to first fix:

TfL took the view that TTFF is a reality of vendor systems and their trials was not designed to isolate, identify or control for it. Even so by comparing the 1st journey of each day for each device, there does seem to be a poorer performance for the 1st journey and this could be attributed to TTFF. It should be noted that TfL's journeys were typically 80 minutes in duration, dwarfing any effect lasting < 1-2 minutes.

Recommendation: the existing TfL trials were not designed to isolate TTFF effect. If the Ministry believes that TTFF can ultimately be solved and can be removed from performance predictions, a specific study should be performed to quantify TTFF.

4.2 Conclusions

The following are key conclusions drawn from this short study on the TfL trials data.

- The best performing vendor solutions can achieve 2% deviation (average), using non-map road usage determination. This compares well with the 1.6% (at 50kph) reported in the ARS study.

- These vendor solutions provided only 35% of journeys with 1% deviation and would require a 10% margin to provide 99% confidence. This also compares well with the ARS study where they report 24% (at 50kph).

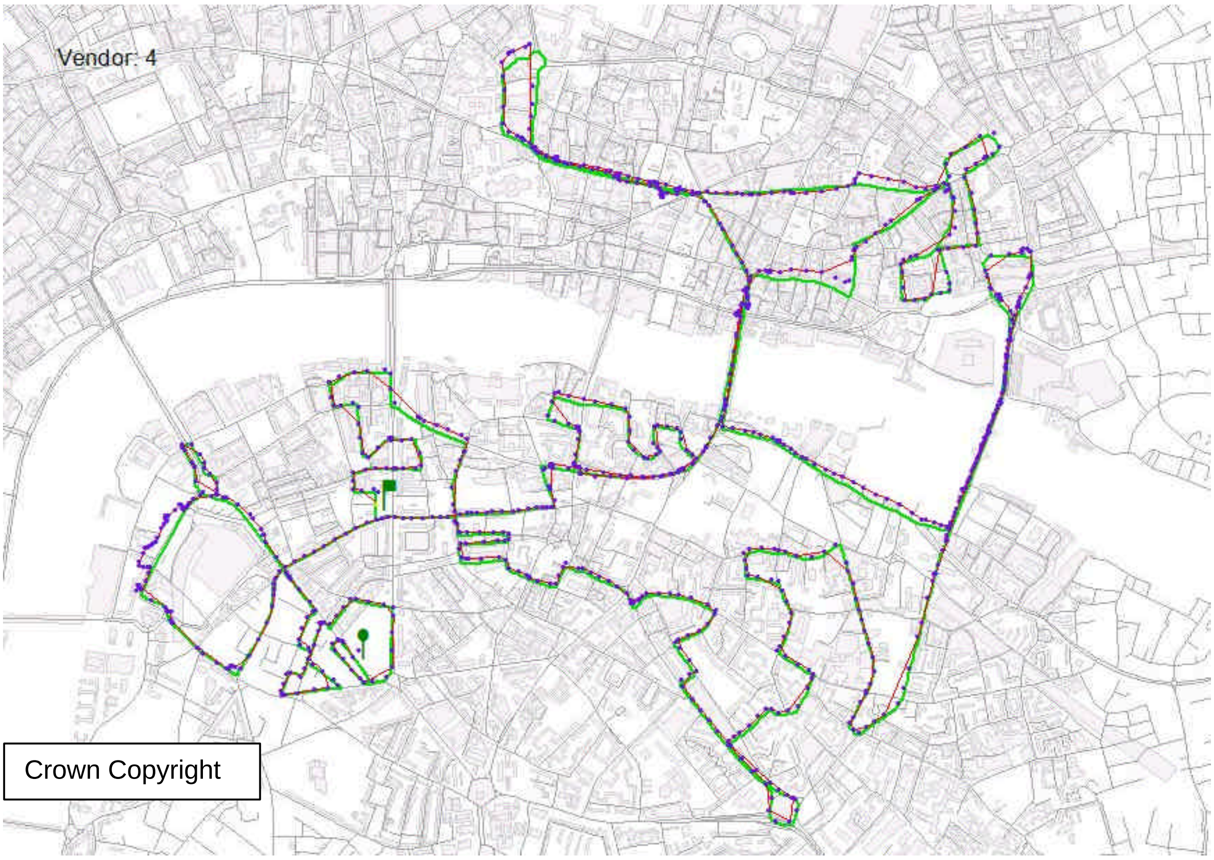
- Due to insufficient and/or inappropriate data an analysis of the 350km, 1000km and 1350km was not possible. However an analysis of shorter billing periods seems to validate the formula approach used by ARS in their study. Note however we recommend further analysis with a larger dataset and different environments to validate this approach.

- We have found that inconclusive evidence that TTFF is affecting performance and we recommend further study.

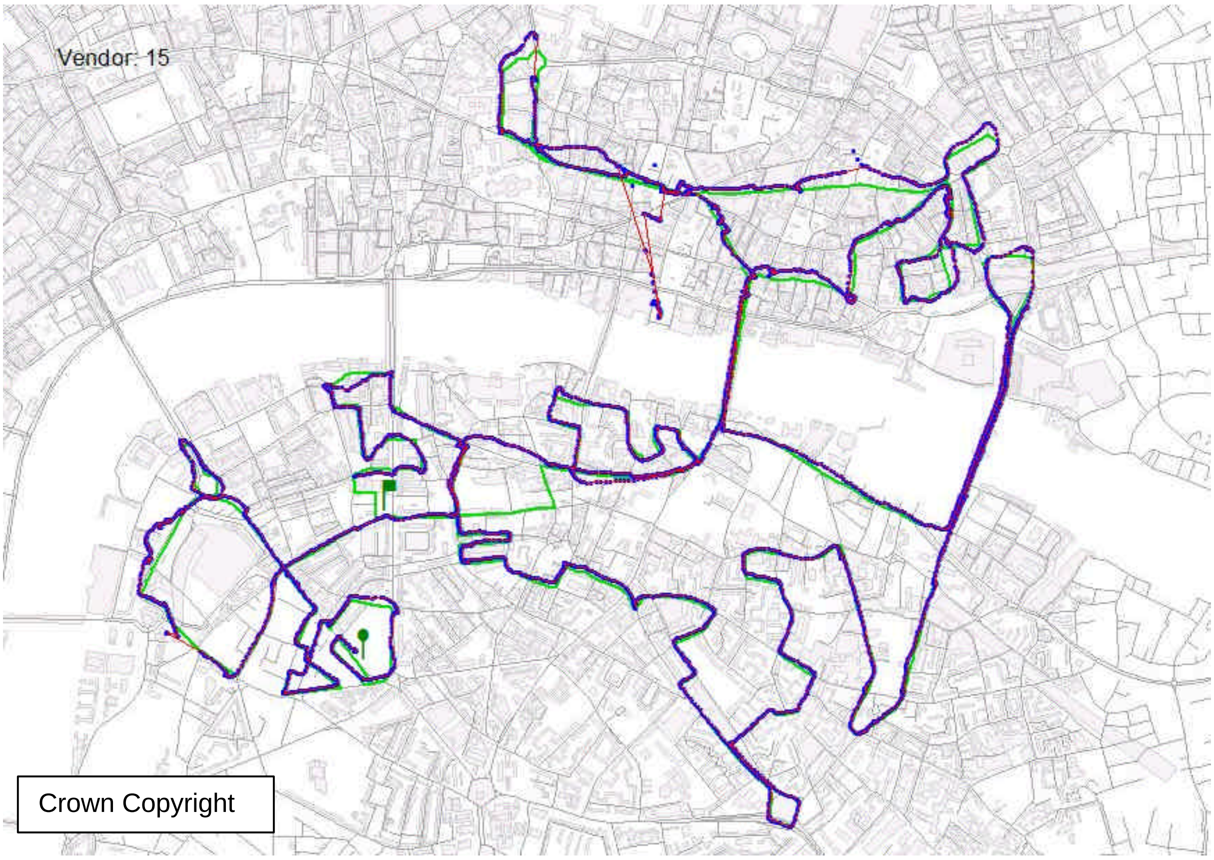
Appendix 1: Example maps

To illustrate the variety and repercussions of some of the issues seen on the ground, we have included a map for a handful of vendors travelling on the same route, where the distance cans been calculated by the Filter method (red). This merely represents a random sample and is intended only to provide the reader with anecdotal evidence. This route (21C) is by no means the most challenging route.

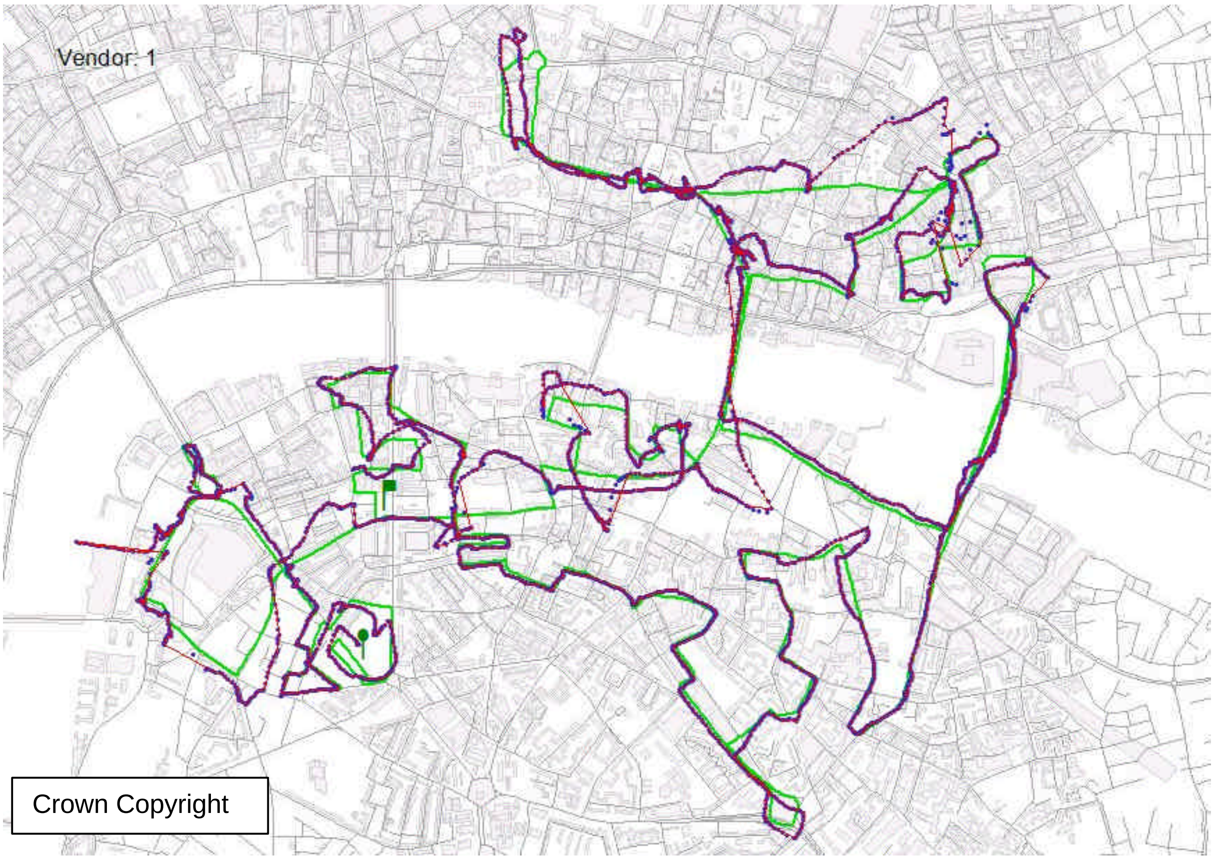
Each of the following maps illustrate a given vendor, indicated in the top left hand corner, and illustrates the result (red) of using the filter method to determine road usage distance from raw GPS data (blue), as compared to the actual route driven (green) from the start (round flag) to the end (square flag).



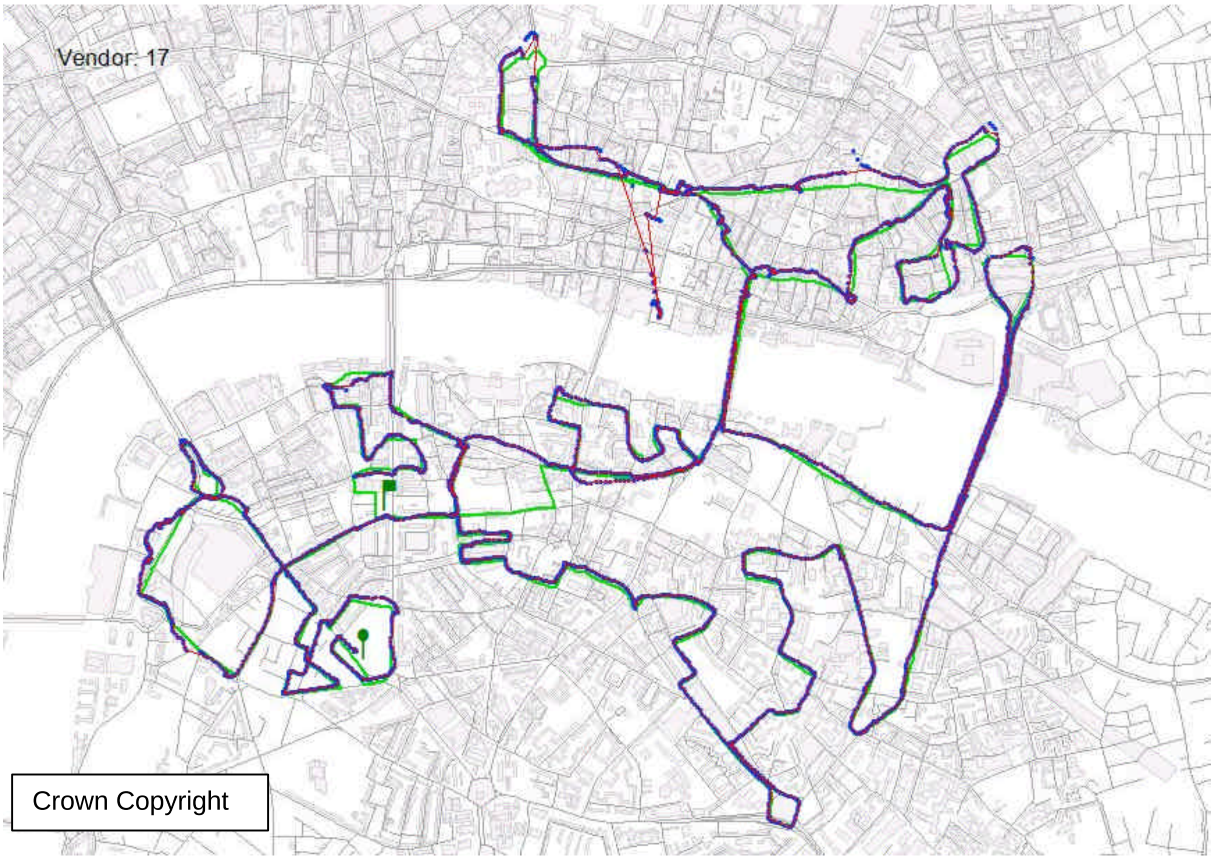
Truth distance: 25,303km. Abs%distance deviation 2.49% using Filter method



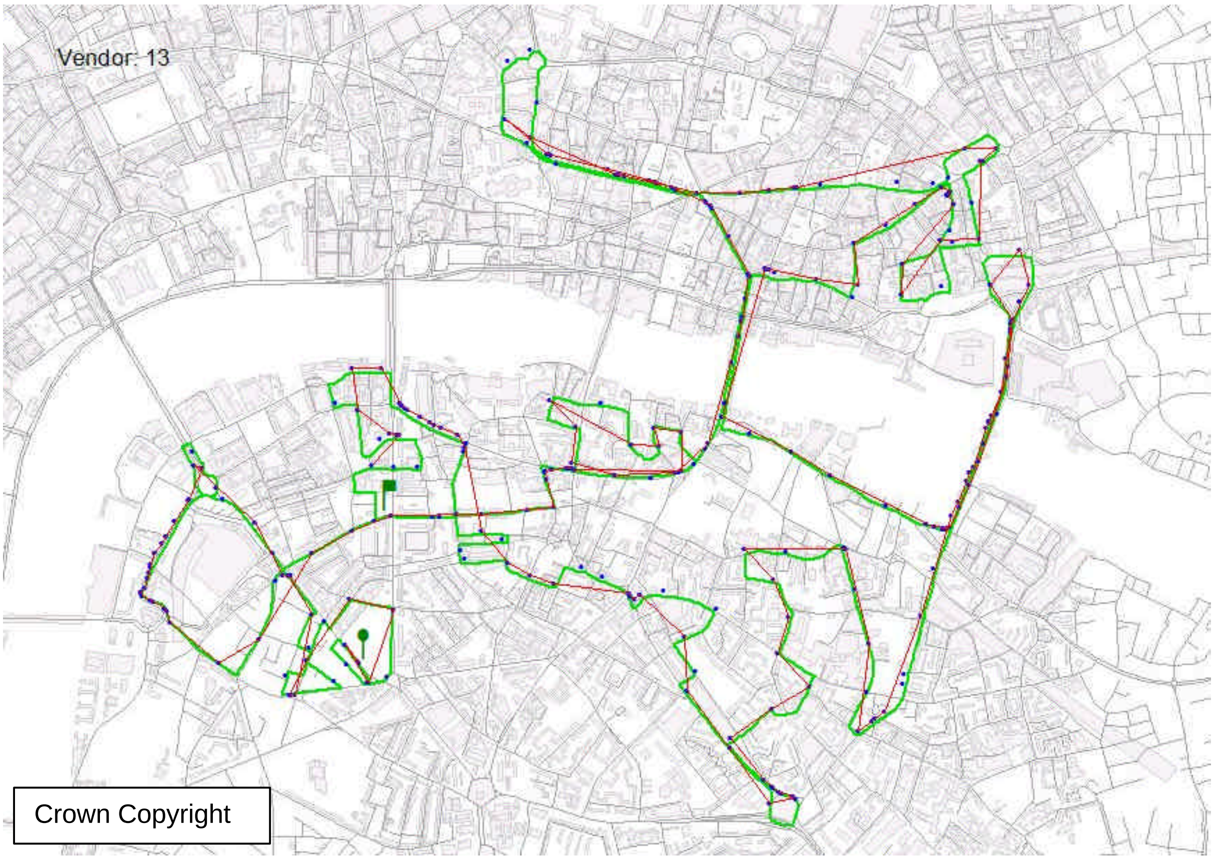
Truth distance: 25,303km. Abs%distance deviation 9.11% using Filter method



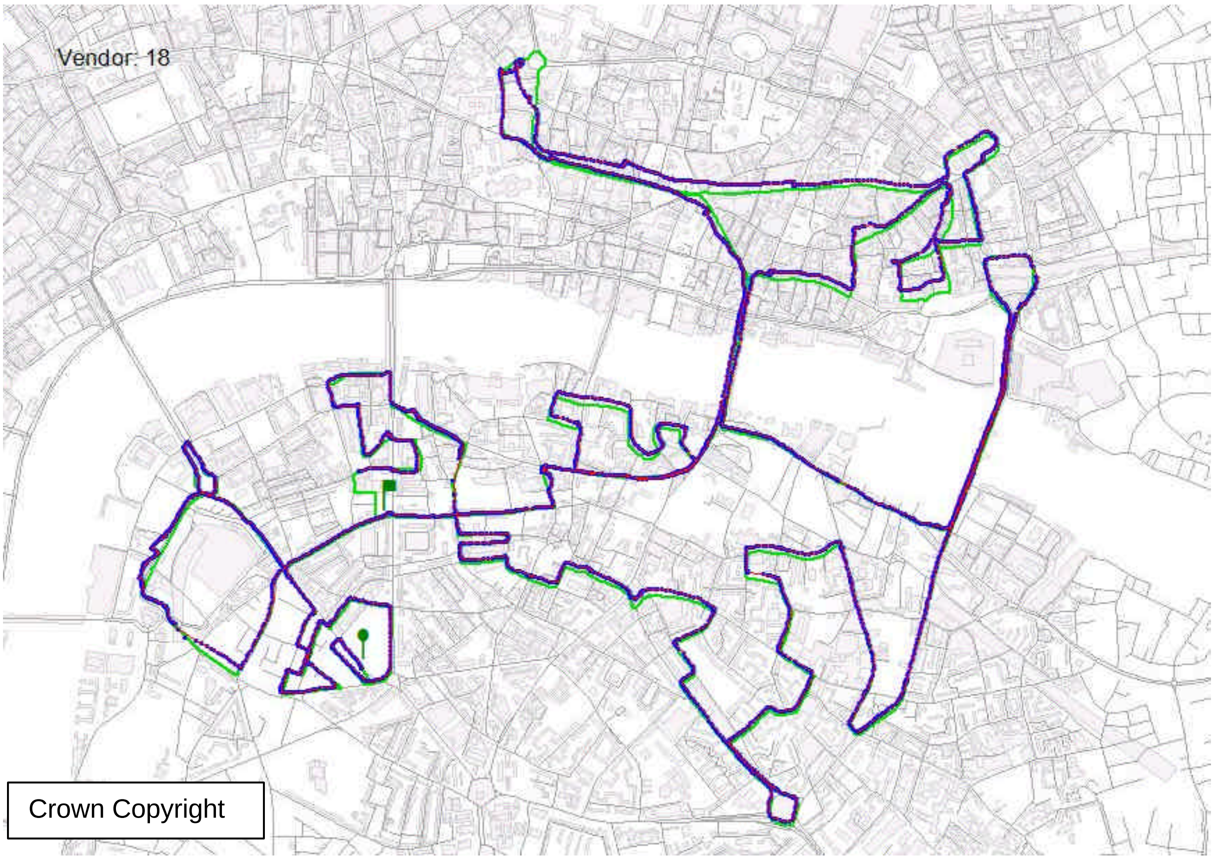
Truth distance: 25,303km. Abs%distance deviation 26.68% using Filter method



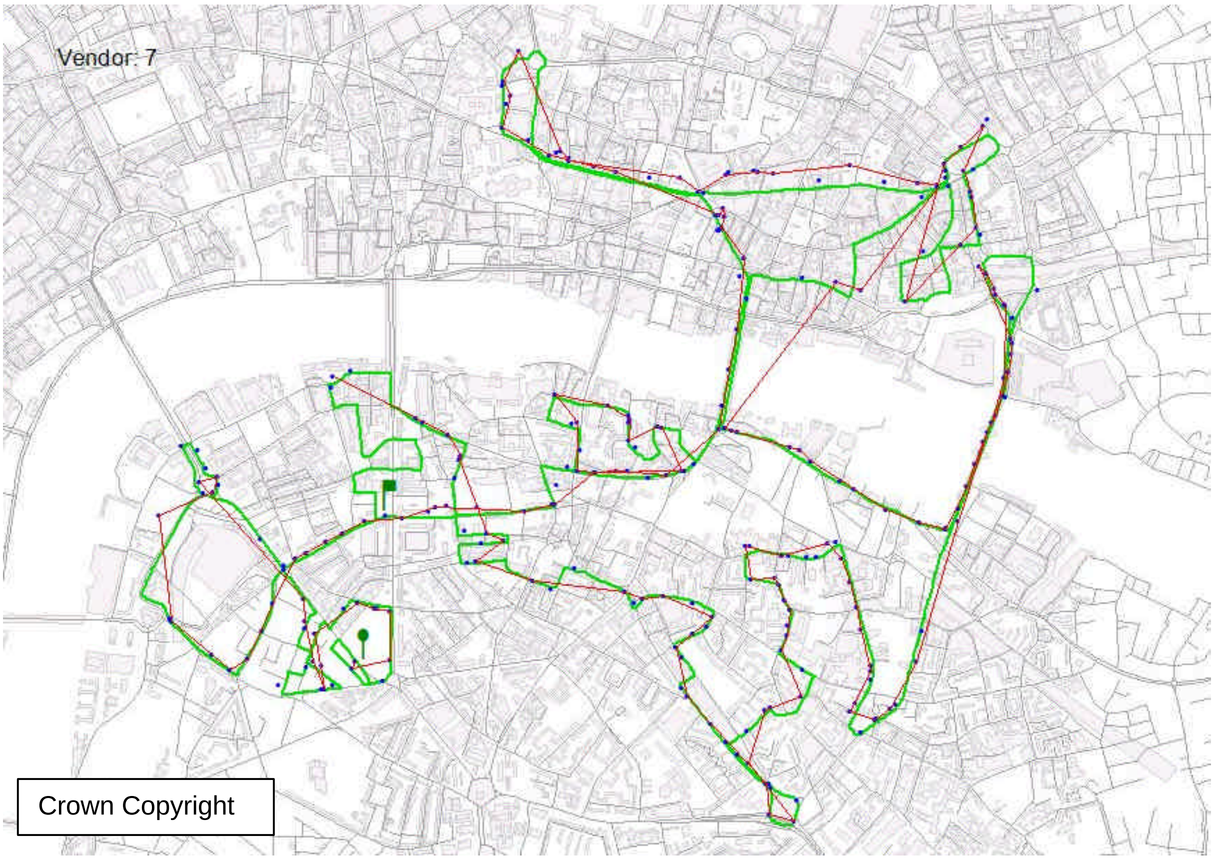
Truth distance: 25,303km. Abs%distance deviation 8.73% using Filter method



Truth distance: 25,303km. Abs%distance deviation 17.34% using Filter method



Truth distance: 25,303km. Abs%distance deviation 1.68% using Filter method



Truth distance: 25,303km. Abs%distance deviation 18.78% using Filter method

Appendix 2: Vendor Technologies

As noted earlier, little information is available on the precise technology within each vendor's device. However, TfL did show case some of the vendor's devices in the 2006 ITS World Congress. In preparation for this they asked the vendors to provide information about their products. That information has been reproduced here, to provide the reader some context.

Note, the actual vendors names are given here, as presented at ITS Congress. For the sake of anonymity no direct translation from the vendors names to the vendor IDs, used throughout the report, has been given.

A2.1 IPL TrakM8 – T4

Highly scalable and easy to integrate, rich vehicle telematics platform from Europe's No. 1 after-market vehicle telematics supplier*.

Low total cost of ownership (TCO) with:

Reduced installation costs, no special tools or laptops required;

- Over The Air (OTA) - configuration for support and upgrades;
- Proven Communications management, including roaming;
- Enhanced GPS receiver for operation in compromised urban locations.

Integrated peripherals:

Driver identification using RFID tags, iButton and Honeywell HiD, enabling:

- reporting driver behaviour;
- meeting duty of care obligations;
- Multiple digital inputs and outputs;
- Built-in battery for security applications;
- Immobiliser and trembler built-in;
- Voice connections for microphone and speaker;
- Pass-through support for third party devices;
- Driver interface button for communications and alarm support.
- Flexible firmware supports multiple applications out of the box.

A2.2 Efkon -TTP OBU

The TTP OBU - is a really high volume product and field proven On-Board Unit platform for traffic telematics applications. The most challenging and successful application is the German TollCollect system, where around 400,000 of these units licensed by EFKON and manufactured by Delphi/Grundig are running in the harsh environment of heavy commercial vehicles.

But, there is much more potential in the box. Comprising GPS localization, GSM communication, CAN interface, Infrared and microwave DSRC, etc., the unit can be used for numerous telematic solutions beyond or combined with tolling including:

- Vehicle and Freight Tracking;
- Fleet Management;
- Driver and Vehicle Surveillance.

Customers will also benefit from a thoroughly proven device and API software and can concentrate on their specific telematic application. EFKON mobility offers a broad range of services and co-operation models around the TTP OBU, such as:

Supply of On-Board Units;

- Manufacturing Licence;
- Customer Specific SW and HW Adaptions;
- Co-Operation in Telematics Projects;
- Training and Technical Support.

A2.3 GMV -Allroad

EGNOS-SISNET compatible

Accuracy (50%)

Open Sky	1.6 m
Typical Urban Environment	6.0 m
Position Availability in Urban Environment	>99.9%
Number of RF channels	12

Sensitivity (dBW)

Tracking	-165 dBW
Acquisition	-173 dBW
Update rate	1 Hz

Specific Integrity Features

Computation of Protection Levels Integrity	99.9999% per epoch
PL size in dynamic conditions (50%)	20m
Open Sky	75m
Typical Urban Environment	

A2.4 Satellic (Member Company of T-Systems) GPS-enabled PDA

Specifications:

- Standard off-the-shelf PDA;
- Road user charging application software based on a combination of T-Systems' standard T-Traffic map matching and T-Navigate positioning software;
- GPS receiver;
- Bluetooth connectivity.

This mass market consumer device enables accurate road user charging while providing the user with access to a range of PDA office and communications applications, as well as a host of other useful telematics services such as satellite navigation and personal location-based applications.

Road user charging application software is loaded onto the PDA and processed "off-board" with map matching software developed by T-Systems specialized T-Traffic subsidiary.

A2.5 Siemens VDO -On-Board Unit 1372

Internal SIM-card reader for : Telephone SIM card:
This card is needed for the GPRS communication with the back office
System ,Internal SIM card reader for the vehicle SIM card: e.g. for recording vehicle data and security keys

Internal GPS module: Integrated Sif Star 2 Module

Internal GPRS module: Dual Band E-GSM 900/1800 MHz
Connection over internal UART interface

Internal DSRC antenna: It contains two DSRC technologies:

- Microwave 5.8 GHz;
- Infrared.

Internal gyro: The gyro supplies data for driving direction identification.

HMI
(Human Machine Interface)

- Two control buttons;
- Display (2 x 16 characters), 5*7 matrix;
- 2 colour LED for Status Display of OBU.

Interfaces:

- CAN-Bus;
- RS 232 interface.

A2.6 Thales Telematics-Telematics Control Unit

The Telematics Control Unit (TCU) is the vehicle-based, mobile component of the Thales Telematics Solution. The TCU’s primary function is to gather a broad range of high value information such as vehicle movements, distance traveled, speed etc. and transmit this back to Thales Telematics Hub.

A2.7 IPL TrakM8- T4

Highly scalable and easy to integrate, rich vehicle telematics platform from Europe's No. 1 after-market vehicle telematics supplier*.

Low total cost of ownership (TCO) with:

- Reduced installation costs, no special tools or laptops required;
- Over The Air (OTA) - configuration for support and upgrades;
- Proven Communications management, including roaming;
- Enhanced GPS receiver for operation in compromised urban
- locations.

Integrated peripherals:

- Driver identification using RFID tags, iButton and Honeywell HiD,
- enabling:
- reporting driver behavior;
- meeting duty of care obligations;
- Multiple digital inputs and outputs;
- Built-in battery for security applications;
- Immobiliser and trembler built-in;
- Voice connections for microphone and speaker;
- Pass-through support for third party devices;
- Driver interface button for communications and alarm support.

Flexible firmware supports multiple applications out of the box.

A2.8 FELA Management AG-Tripon EU

FELA’s TRIPON EU unit was tested among others in the TfL. GRPS mini trial. TRIPON processed GPRS data along with FELA’s optimised map matching algorithm. The trial proved that inner city ,street level tolling is achievable using a mix of optimised hardware and software. TRIPON was the world’s first unit that used GPS(within a tolling scheme) for fraud protection in the Swiss System.

It is now ready for interoperable tolling throughout Europe.

Its characteristics include:

- Trouble free operation in Switzerland for more than six years;
- Full GPS / GSM equipped as well as DSRC and chip card interface;
- Additional gyro sensor for best in its class dead reckoningperformance;
- Additional movement sensors for battery operated fraud protection;
- The first multi-application unit supporting value added services and/or different tolling applications for full system

A2.9 Navicore Personal 2006/1

Navicore Personal is a fully featured portable satellite navigation system for mobile phones, with:

- The fastest and most reliable routing performance on mobile phones: finding a route is as easy as typing an SMS. No other product calculates long routes as fast and reacts to driver deviations from the route as quickly;

- The most convenient map functionality to street number detail with a large Point of Interest directory including phone numbers;

- A fast and reliable map that can be used for browsing and planning even without a GPS receiver;

- The best GPS hardware on the market that can be charged with a Nokia charger and offers the advantage of continued excellent GPS coverage in built-up and wooded areas. Line of sight is no longer a requirement.

A2.10 Satellic (Member Company of T-Systems) In-Vehicle On-Board Unit

- Integrated GNSS (GPS) receiver;
- GSM/GPRS Module;
- CPU processor;
- Memory;
- Removable storage - compact flash (as an alternative form of data capture for trial purposes only);
- Magnetometer;
- Gyro and temperature sensor;
- Built-in power supply.

A2.11 Siemens VDO -MK5 MultiComms Locator

The MK5-MultiComms locator is a fully integrated Automatic Vehicle Location (AVL) unit providing real-time vehicle location data at predefined intervals.

The MK5-MultiComms locator uses a unique combining process of GPS and Datatrak proprietary Low Frequency (LF) Navigation Systems to provide an accurate position under most operating conditions. Data reporting is achieved either via the Datatrak UHF Network or via the GSM network using the GPRS service.

Key features are:

- Automatic calculation of vehicle position, speed, direction &
- status using GPS & LF navigation signals;
- Use of H-Field (magnetic) Antenna to receive LF navigation
- signals for optimum performance in dense urban areas;
- Automatic transmission of AVL data via the Datatrak UHF or
- GSM communications network;
- Use of geo-fences to define areas to trigger user configurable
- events;
- GPS Receiver Fastrax iTrax 100, using Sony 12 channel
- single chip receiver;
- GSM Modem Siemens TC45 Dual band, GSM/GPRS.

A2.12 TRACKER Network -Vehicle Asset Management

The TRACKER Reporter Vehicle Asset Management (VAM) unit is an innovative integrated fleet management unit with superior location and tracking capabilities. GSM GPRS IP communication together with GPS technology ensures inexpensive, yet reliable communications supporting real-time event driven reporting for Fleet management, Vehicle Tracking and Road Charging.

The TRACKER Reporter VAM offers numerous advanced features including:

- **Triple Communication Methods -** IP over GPRS, SMS or CSD.All communication options are fully configurable, with separate controls for home network and roaming scenarios;
- **Log Memory -** When communication networks are unavailable,events are written to memory and transmitted once network coverage returns;
- **GPS Sensor -** 12-channel SiRF GPS with integrated receiver/antenna dramatically increasing sensitivity over conventional coaxial antennae;
- **Over The Air (OTA) Programming -** All the unit's options are fully configurable remotely;
- **OTA Firmware Upgrade -** The firmware of the unit can be upgraded over the air if required allowing new features to be added remotely;
- **NMEA Data Output -** as a source for third party (navigation) systems;
- **Compact Size -** Simplifying installation;
- **Very Low Power Consumption -** Meeting and exceeding vehicle manufacturer requirements;
- **Fully Approved -** EC Directive 95/54, 89/336 EEC, 72/73/EEC, R&TTE, CE.

A2.13 Trafficmaster- Black Box Technology

About Trafficmaster

Trafficmaster Plc is a leading supplier of intelligent vehicle trackingand navigation solutions. Based in Cranfield, Bedfordshire, it also has operations in North America, under the Teletrac brand. Trafficmaster provides advanced satellite navigation, vehicle tracking and diagnostics, fleet management solutions and real-time traffic information.

Trafficmaster's Black Box Technology

Trafficmaster has developed a single platform technology, which provides the driver and companies with a number of motoring services. Using GPS for location data and GSM/GPRS mobile communications, information can be sent to and received from the black box. By incorporating an embedded computer within the box, intelligent services can be downloaded to the car.

Intelligent Driving

Currently Trafficmaster provides the following intelligent driving services:

- Real-time traffic information;
- UK satellite navigation;
- European satellite navigation;
- Stolen vehicle tracking;
- Safety camera alerts;
- Fleet Management solution;
- Emergency and breakdown service;
- Remote vehicle diagnostics;
- Use-based insurance solutions.

Appendix 3: GPS Performance Metrics

The following results and analysis has been provided by TfL for reference. The analysis looks at GPS performance and in particular the various forms of Location Error.

A3.1 Location Error

The error of each raw location point was calculated by measuring the distance from that point to the nearest part of a road segment that constitutes the truth route for any journey (not just the route the vehicle was scheduled to be on).

Although this is not a reliable absolute figure, and is likely to underestimate the true error, it was considered a good relative basis on which to compare the vendors' devices.

The method used to calculate location error in the Phase 1 and 2 trials was more rigorous, and likely to be more accurate, since it was based on knowledge of the vehicle's true position.

All location errors greater than 500m were excluded from the original trial results. This was done on the basis that any working system should be capable of carrying out basic logical checks to exclude such measurements.

The overall location error for each device was characterised by calculating the average and standard deviation of error over all data points recorded by that device.

In addition, the effects of the following variables on the location error were investigated:

- **Elapsed Time**. Do some vendor devices show a delay in acquiring the first accurate location fix following device start-up? This effect is only likely to be present during the first minutes or so each journey, so only that time frame has been considered. Note it may not be possible to see this if the driver could not turn off the ignition prior to the start of the journey, for example due to traffic conditions.

- **Route**. Are particular routes more "difficult" than others, perhaps because they go though certain geographical areas?

- **Built Environment**. What effect does the built environment

The distribution of location error for the following OBU types is shown in Figure 24:

- Best, ranked by average location error.

- Worst, ranked by average location error.

- Average of all OBUs (excluding the Reference device). This average has not been weighted.

- Reference device.

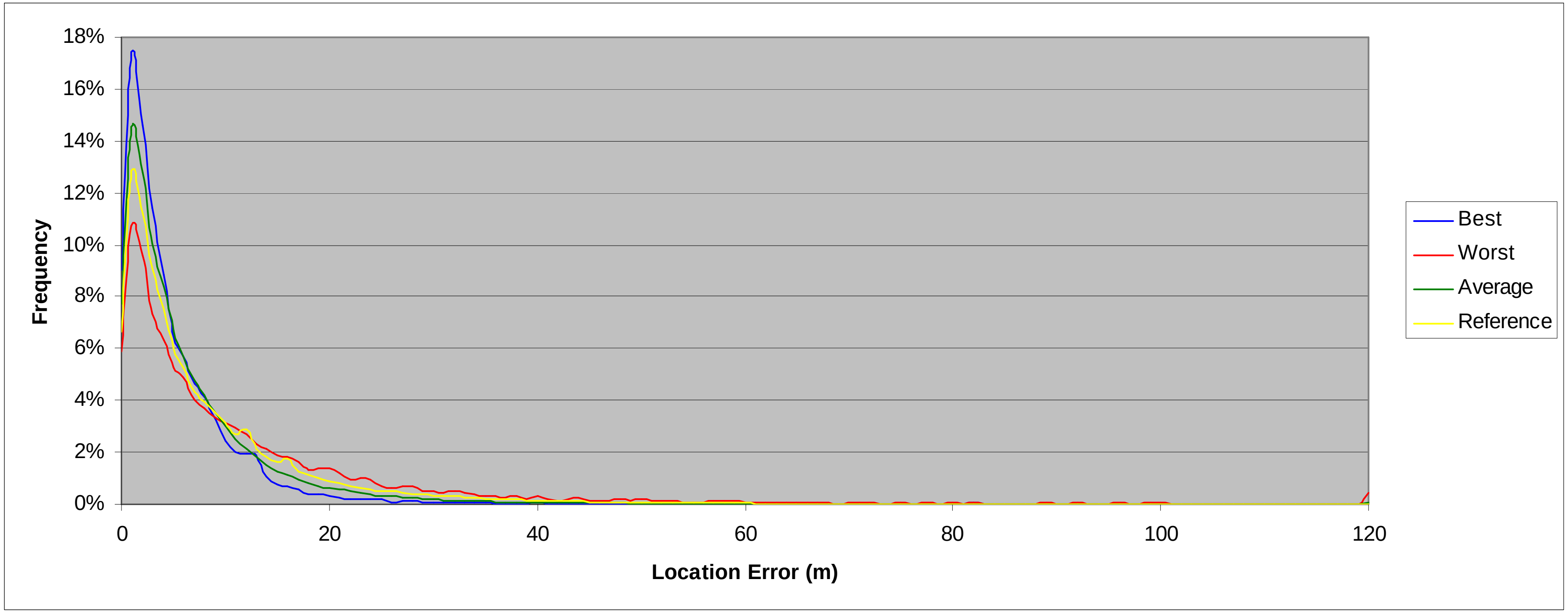


Figure 25 Distribution of Location Error

The cumulative distribution of location error is shown in Figure 4-2.

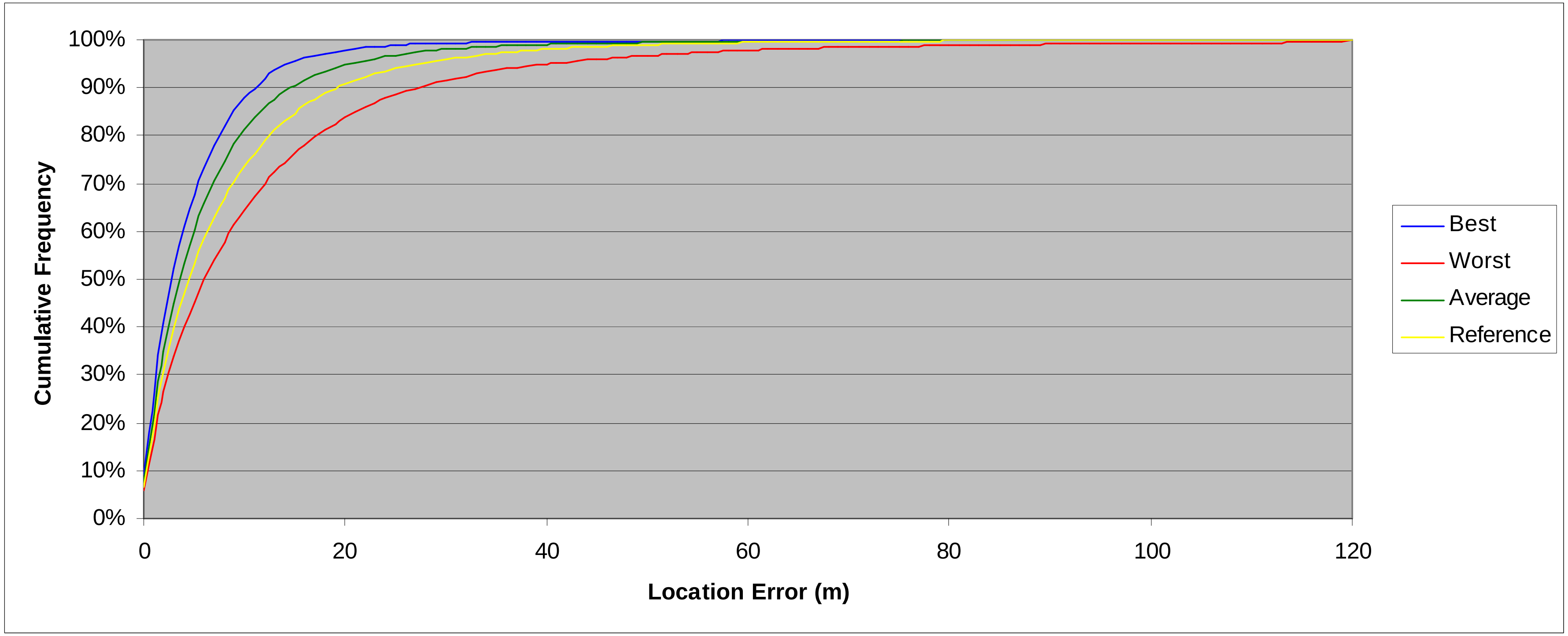


Figure 26 Cumulative Distribution of Location Error

The individual location data points, summarised over all trial journeys, are shown in Figure 26 to Figure 31 for the best and worst OBU types, plus the Reference device. Performance has been ranked by average location error. The same zoomed in area is also shown for each device.

Best OBU Type

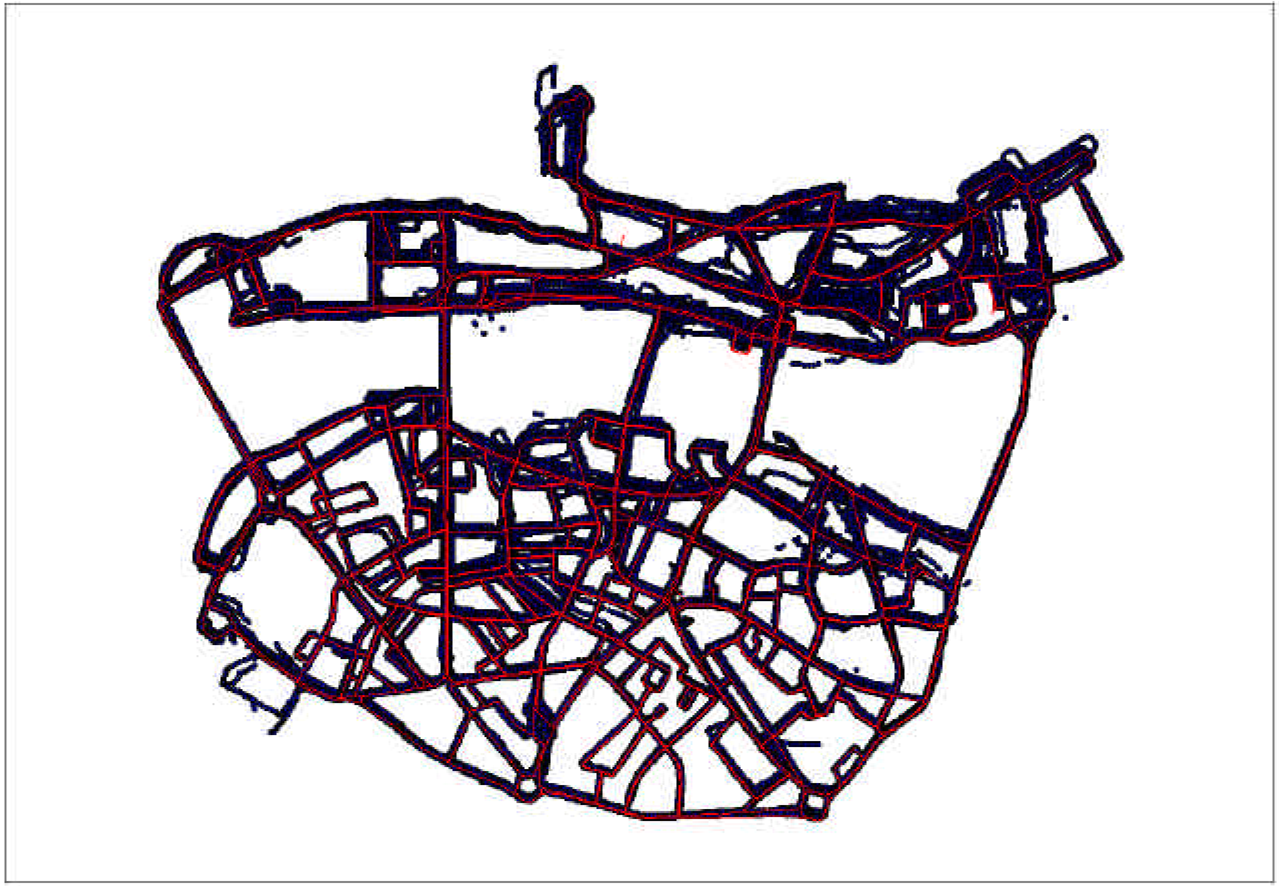


Figure 27 Location Data Points – Best OBU Type

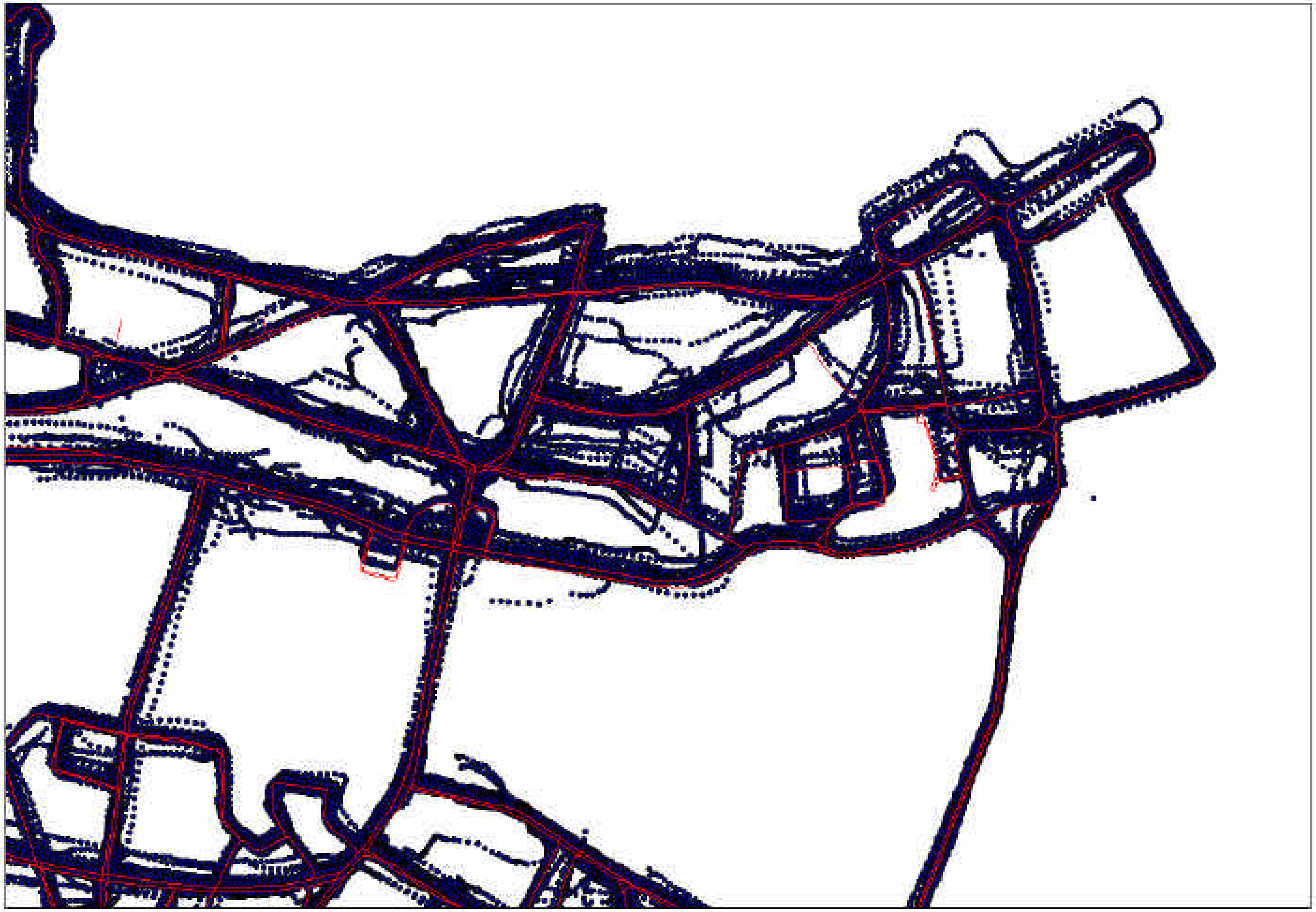


Figure 28 Zoomed Location Data Points – Best OBU Type

Worst OBU Type

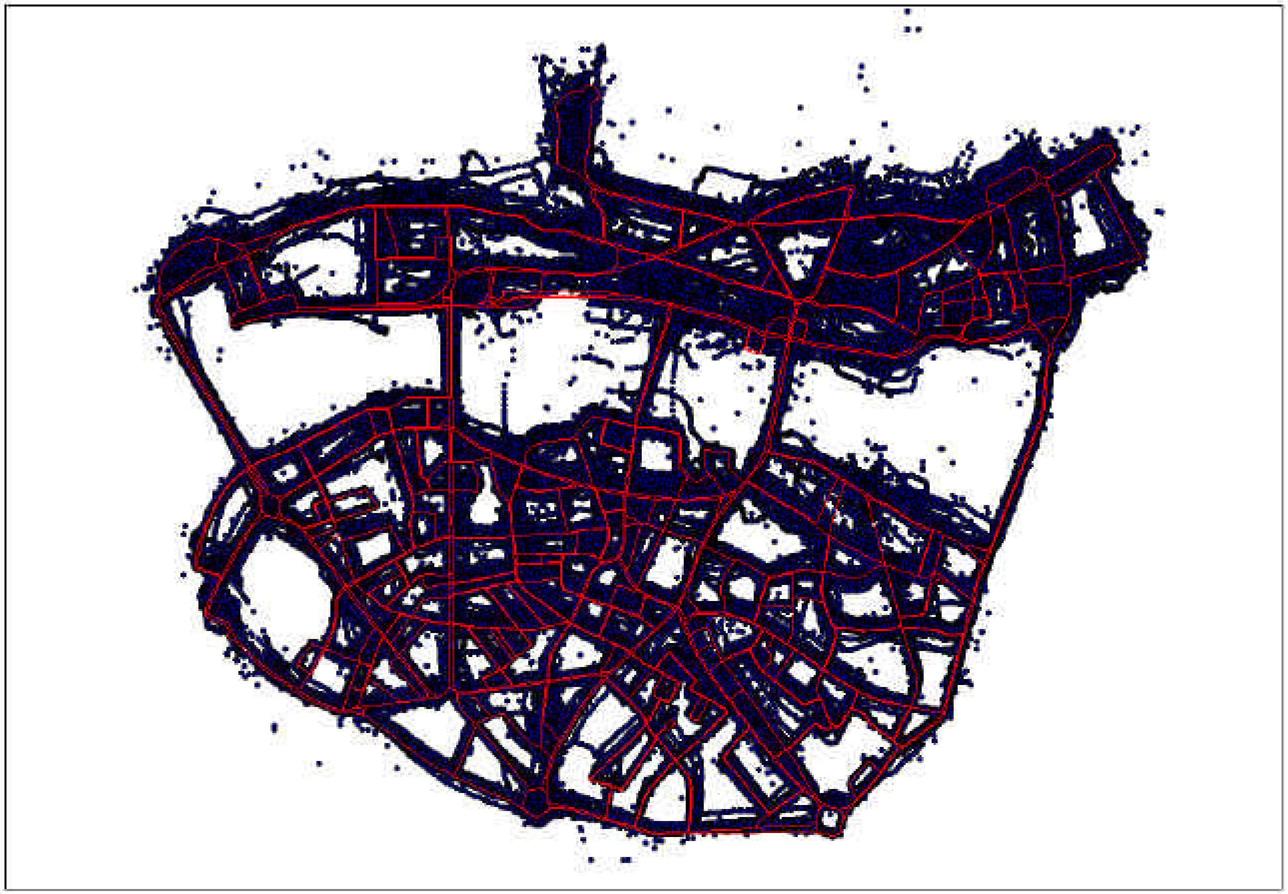


Figure 29 Location Data Points – Worst OBU Type

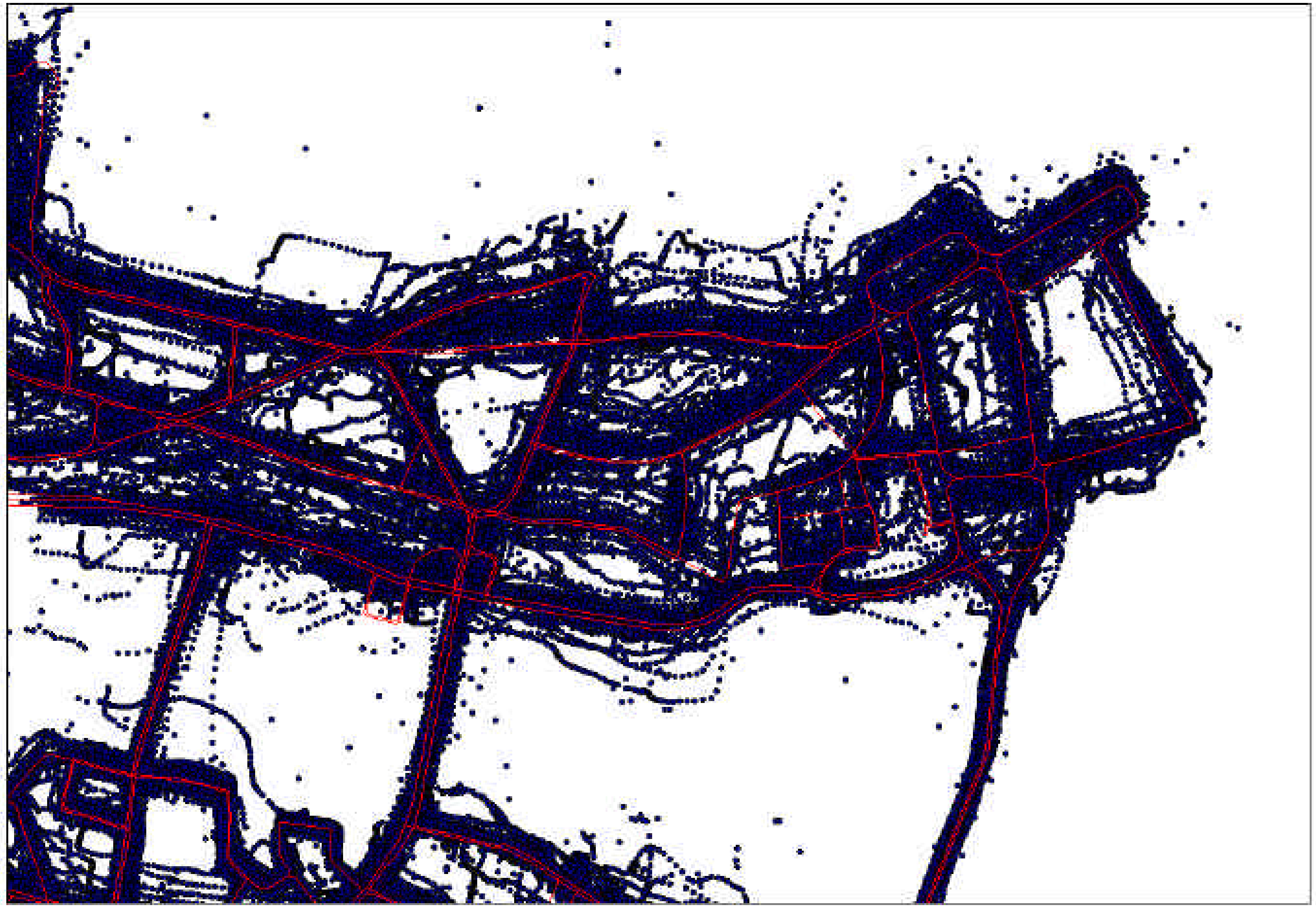


Figure 30 Zoomed Location Data Points – Worst OBU Type

Reference

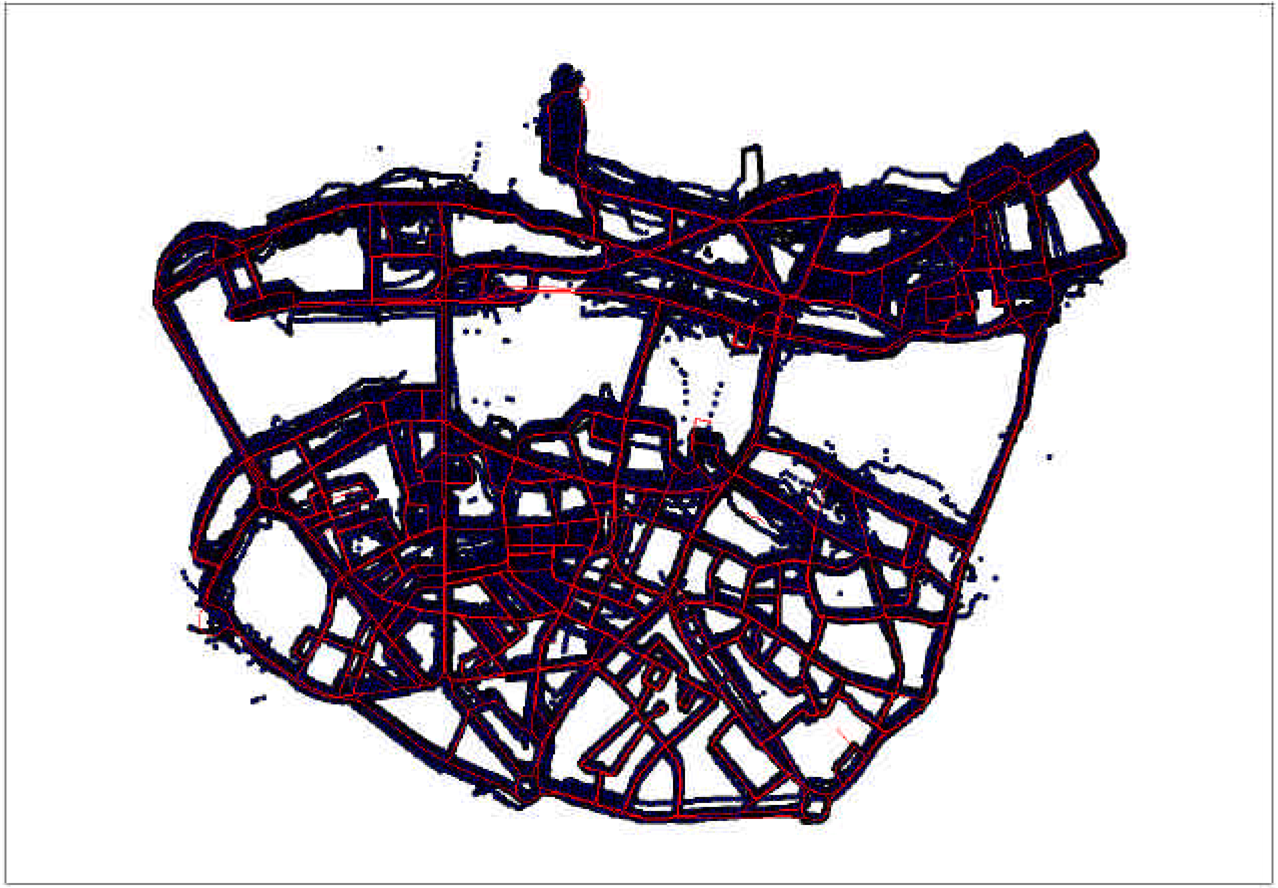


Figure 31 Location Data Points – Reference

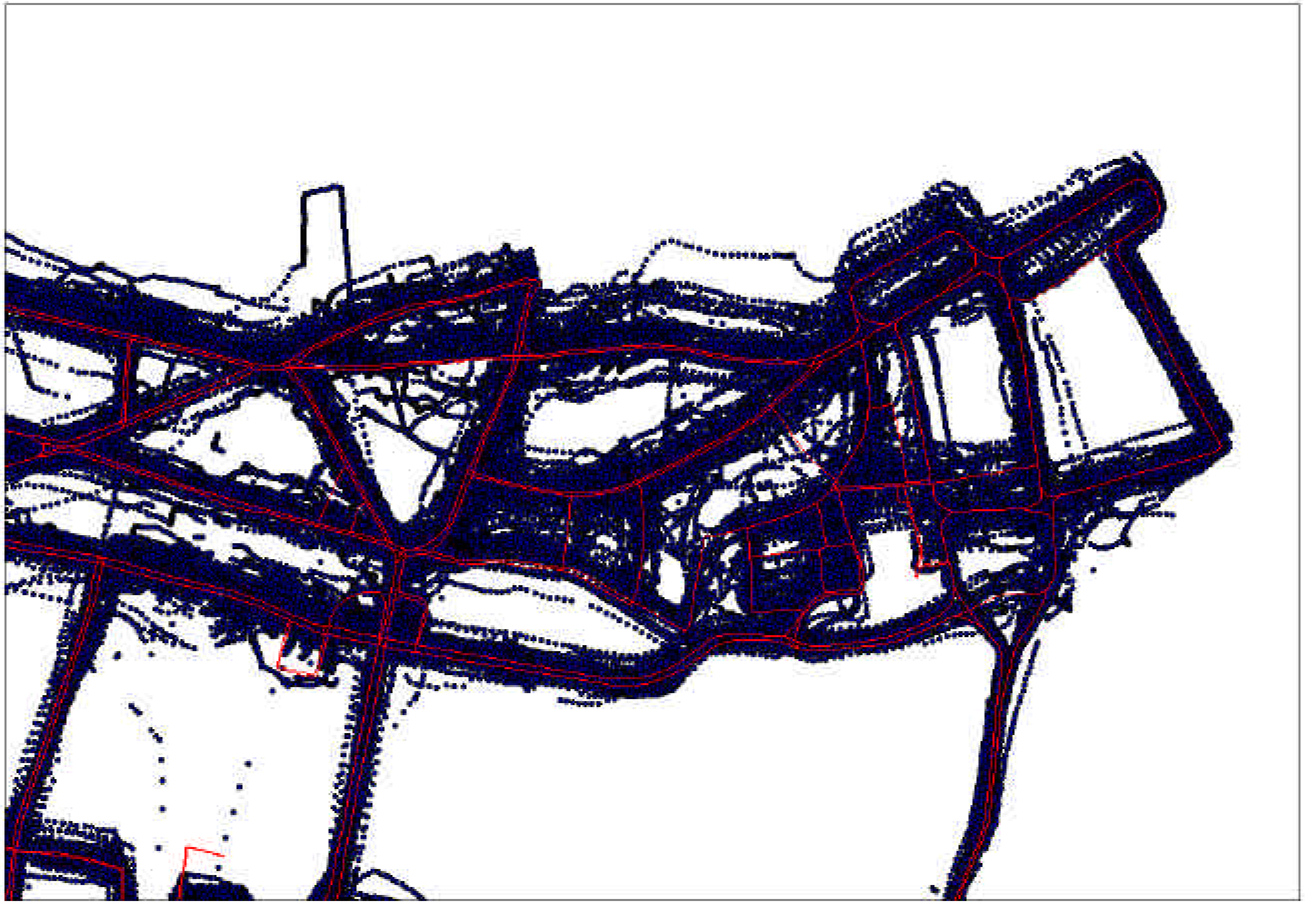


Figure 32 Zoomed Location Data Points – Reference

The detailed location error performance of all the OBU types is shown in Table 4-1, ranked by average location error. ‘Vendor Average’ is a weighted average which takes into account the number of location samples for each OBU type.

Note that all samples with a location error greater than 500m have been excluded. For the vendor OBUs there were a very small number of these (286 in total). In the case of the Reference device, 8,062 samples were excluded, the vast majority of which occurred during a single day.

OBU Type	Total Location Samples	Average Error (m)	Standard Deviation (m)	75% Quantile (m)	95% Quantile (m)	99% Quantile (m)
10	137,956	5.11	6.48	6.36	14.35	26.17
11	132,129	5.59	6.77	6.68	17.38	32.5
4	52,827	5.92	6.79	7.33	17.85	31.52
8	141,047	5.94	7.55	7.23	18.52	32.33
13	24,179	6.01	7.18	7.2	18.96	34.31
2	237,736	6.02	6.64	7.67	17.54	31.15
3	33,780	6.14	8	7.41	18.56	36.84
12	12,242	6.3	7.31	7.79	19.57	35.6
16	148,652	6.72	14.31	7.46	20.43	43.7
17	181,782	6.75	8.49	8.36	20.26	37.7
14	25,106	6.88	11.71	8.12	21.26	40.31
6	266,622	6.91	8.31	8.67	21.6	39.79
5	108,592	7.02	7.11	9.52	19.05	32.02
15	364,135	7.34	9.02	9.24	22.23	39.7
1	348,478	7.7	10.11	9.47	24.12	44.71
7	17,222	12.1	19.07	14.32	39.96	86.92
Vendor Average	**-**	**6.67**	**8.55**	**8.26**	**20.22**	**37**
18	3,184,464	8.41	10.56	10.52	27.37	50.17

Table 12 Detailed Location Performance

The best and vendor average of overall location errors were both significantly better than the values measured in the Stage 1 and 2 trials (see Section 2.1). However this may be due to differences in the way the error values were calculated.

- The best OBU type had an overall average location error of 5.11 metres.
- The weighted average of overall location error for all vendor OBUs was 6.67 metres
- 16 of the 17 vendor OBU types had an overall average location error of between 5 and 8 metres

A3.2 Location Error by Elapsed Time

We are mindful that time to first fix (TTFF) can have a large impact on the performance of RUC. In all cases, TTFF is included in the analysis. As noted above, in these trials TfL was unable to ensure that the ignition was always on or always off prior to the execution of each journey. Instead TfL studied the performance during the initial period of each journey.

The variation in average location error during the first 90 seconds of each journey for three selected OBU types, plus the Reference device, is shown in Figure 32 to Figure 35.

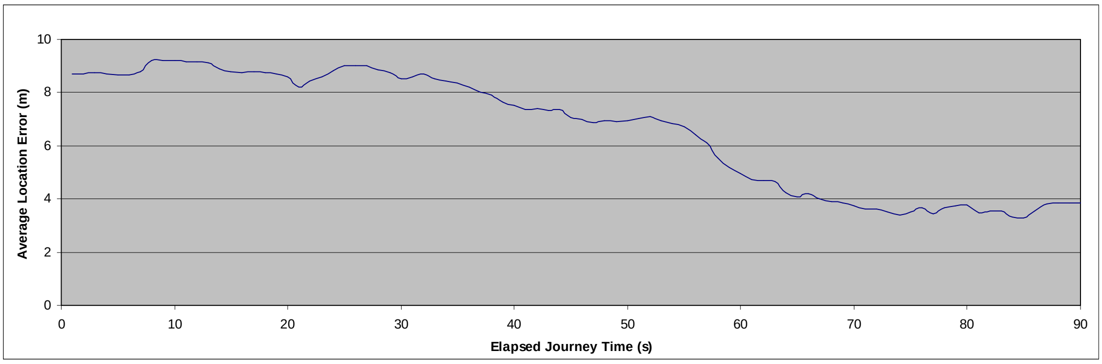


Figure 33 Location Error by Elapsed Time – OBU Type A

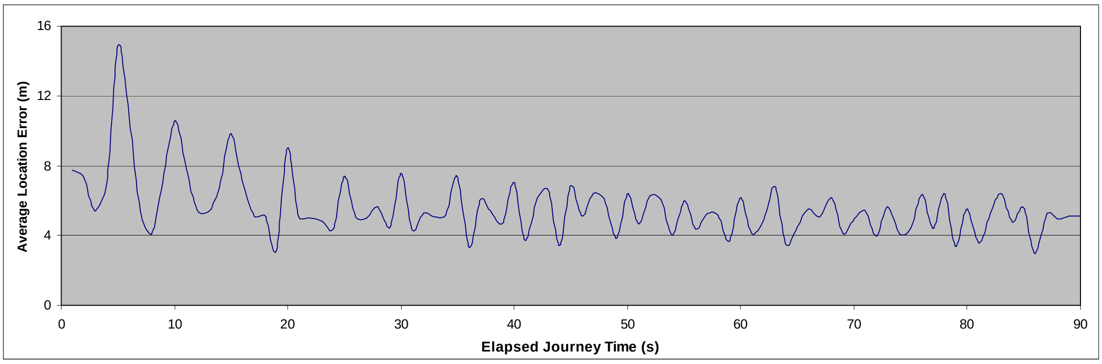


Figure 34 Location Error by Elapsed Time – OBU Type B

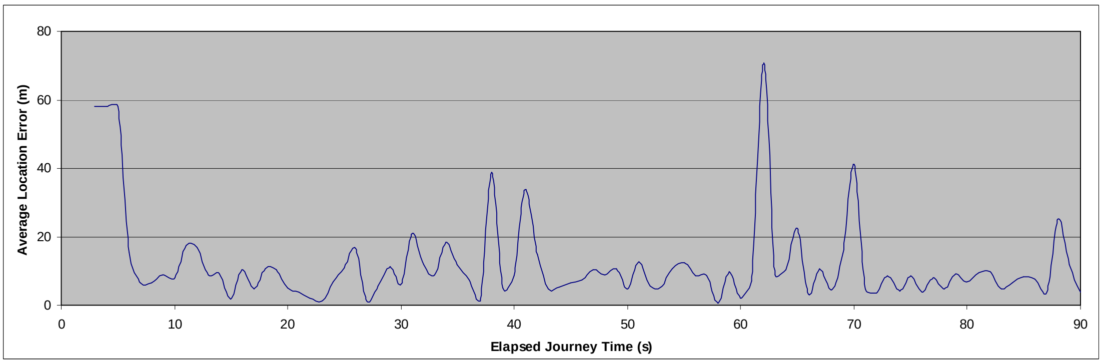


Figure 35 Location Error by Elapsed Time – OBU Type C

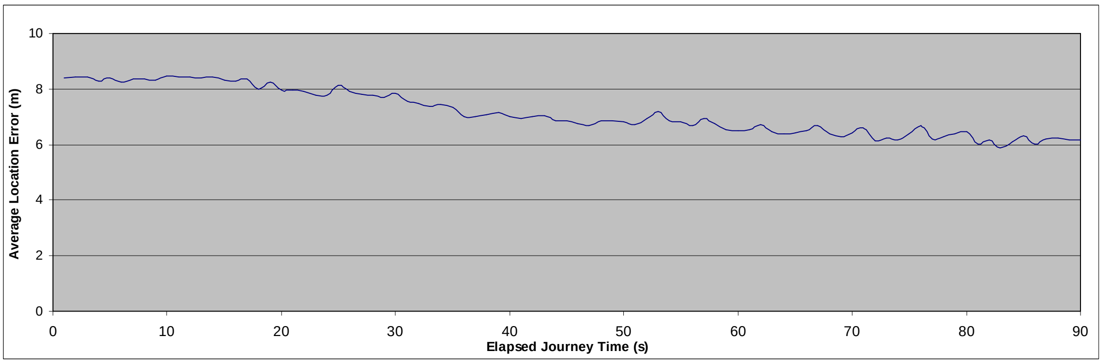


Figure 36 Location Error by Elapsed Time – Reference

OBU Types A, B and C all showed a significant deterioration in location error performance at the start of journeys. However the ‘signature’ was different in each case.

Other OBU types, including the Reference device, did not exhibit such deterioration.

We understand that in the initial seconds the GPS devices attempt to acquire the satellites and during this period their accuracy can be poor. In other cases, the GPS device can store the satellites acquired from the last positive fix such that the time to first fix is reduced. In still further cases the power may not have been entirely cut from the device before the start of the journey, for practical reasons.

Lacking this information, we cannot conclude the reasons for these variations, and recommend that provision be made in future trials to study these boundary conditions.

Some OBU types showed a significant deterioration in location error performance at the start of journeys. Other devices showed no such behaviour. The reasons for this are not fully understood.

OBU start-up behaviour may or may not affect the overall performance of the RUC systems. However it was not possible in these trials to study this in detail. Future trials may make more provision for these boundary conditions to be tested.

A3.3 Location Error by Route

The variation in average location error by route, for the best, worst and average of OBU types, plus the Reference device, is shown in Figure 4-13. Performance has been ranked by the overall average location error.

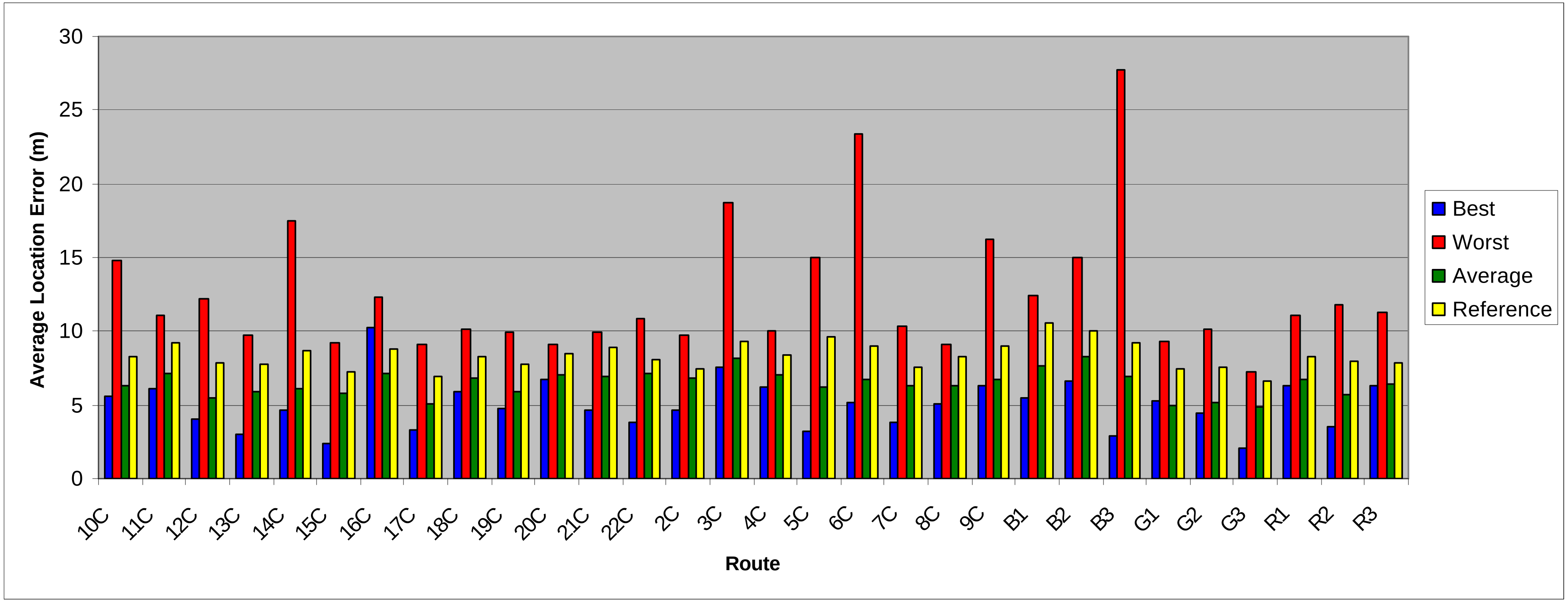


Figure 37 Location Error by Route

A3.4 Location Error by Built Environment

The variation in average location error with signal visibility, for the best, worst and average of OBU types, and the Reference device, is shown in Figure 37.

Performance has been ranked by the overall average location error.

Signal visibility was measured as a percentage over the length of each road segment.

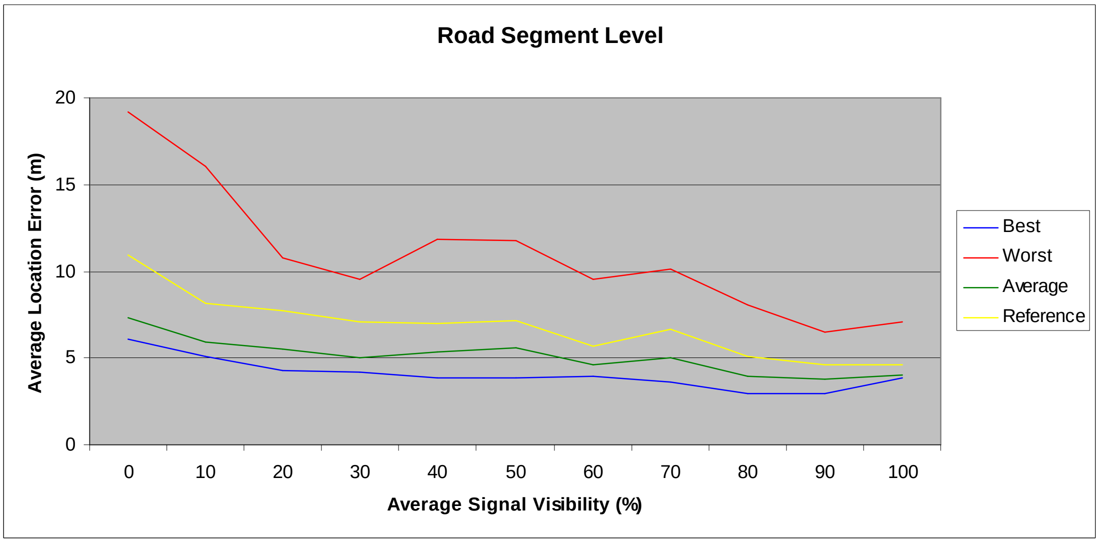


Figure 38 Location Error by Built Environment – Road Segment Level

As expected, a clear relationship between signal visibility and location error is apparent in all four cases.

Figure 38 shows the variation in average location error with signal visibility by route. Each data point represents a single route, with its location error averaged over all vendor journeys.

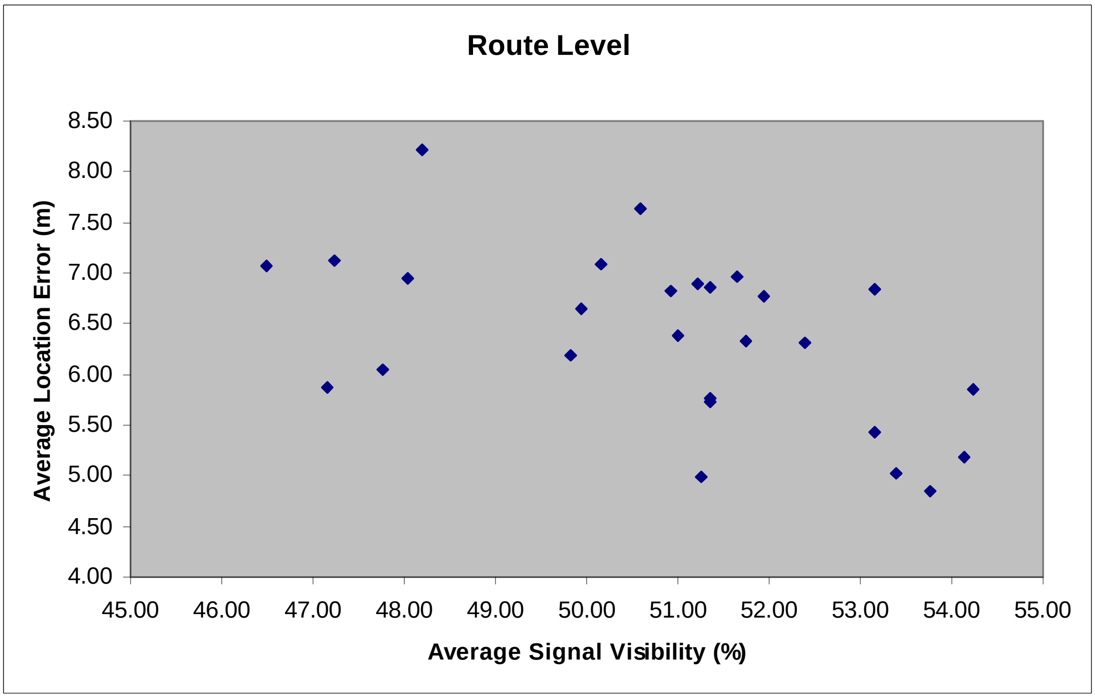


Figure 39 Location Error by Built Environment – Route Level

No clear relationship is apparent at this level. This is probably due to the length of the routes, each of which includes a large number of road segments with different signal visibility characteristics. In fact the worst three routes by average location error (3C, B1 and B2) all had an average signal visibility that was the same or better than the overall average.

No relationship was apparent between signal visibility and location error at the route level. This was probably due to the length of the routes, each of which included areas with widely differing signal visibility characteristics.

A3.5 GPS Sample Interval

A3.5.1 Overall Sample Interval

The overall sample interval of each OBU type is shown in Table A3.1, ranked by average sample interval.

'Vendor Average' is a weighted average which takes into account the number of location samples for each OBU type.

OBU Type	Average Sample Interval (s)	Standard Deviation Sample Interval (s)
Fastest	0.99	3.69
	1.00	0.04
	1.00	0.82
	1.00	3.24
	1.01	0.76
	1.01	0.95
	1.02	1.30
	1.04	1.99
	1.04	0.99
	1.13	11.11
	2.50	1.00
	4.00	0.16
	4.87	1.60
	10.45	9.97
	11.19	11.18
	18.21	13.08
Slowest	21.25	29.73
Vendor Average	**1.70**	**2.56**
Reference	1.01	2.34

Table 13 Overall Sample Interval

The distribution of sample interval for the fastest and slowest OBU types, and the Reference device, is shown in Figure 39 to Figure 41. Performance has been ranked by the overall average sample interval.

Fastest OBU Type

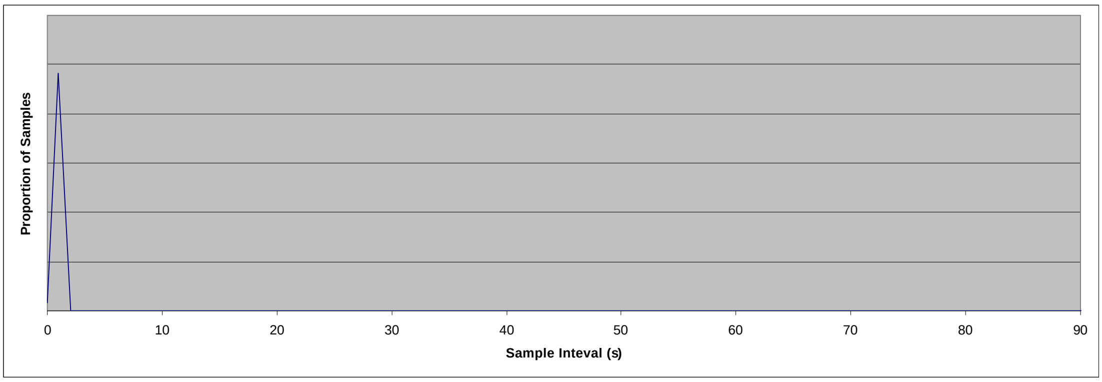


Figure 40 Distribution of Sample Interval – Fastest OBU Type

Slowest OBU Type

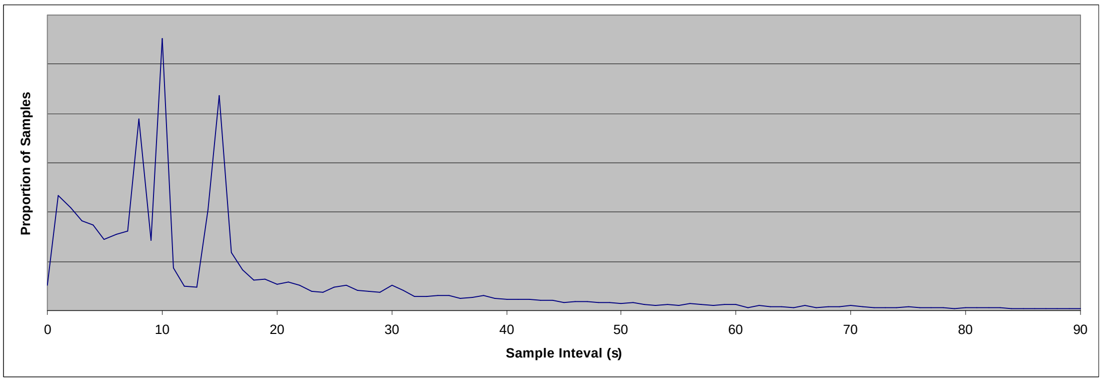


Figure 41 Distribution of Sample Interval – Slowest OBU Type

Reference

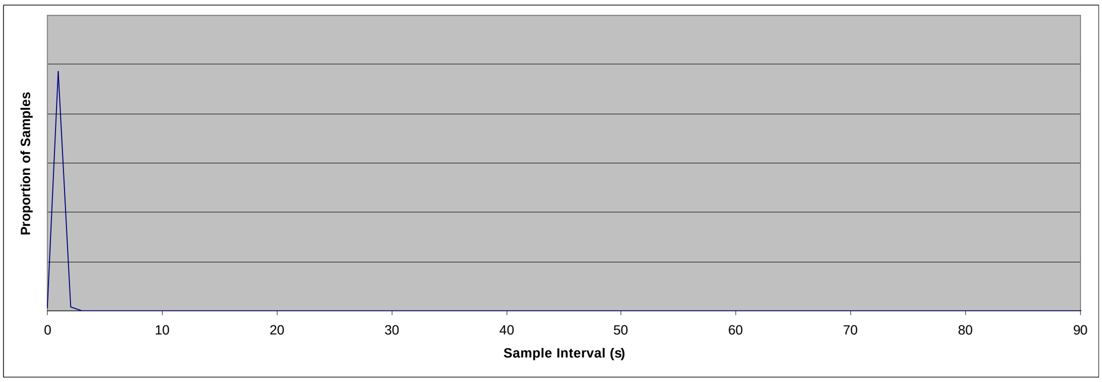


Figure 42 Distribution of Sample Interval – Reference

In the case of sample interval, the OBUs fell into two distinct categories. Nine of the OBU types had a sample interval close to 1 second, with a small standard deviation. This behaviour was typical of the RUC grade solutions. The remainder had longer sample intervals with more variation. This was typical of the Telematics grade systems.

A3.5.2 Sample Interval by Elapsed Time

The original intention of TfL was to report on the availability of each device. Availability is defined as the percentage of time a service is available for use, without outages, whatever the cause.

However, this was not possible because some of the devices report location fixes at irregular intervals. This is often the design intention of the system vendor, to reduce the amount of communication or irrelevant information.

As an alternative TfL reported on the sample interval. This allowed them to characterise the temporal behaviour of each device, and also to investigate its start-up performance.

The variation in average sample interval during the first 90 seconds of each journey for three selected OBU types, plus the Reference device, is shown in Figure 42 to Figure 45.

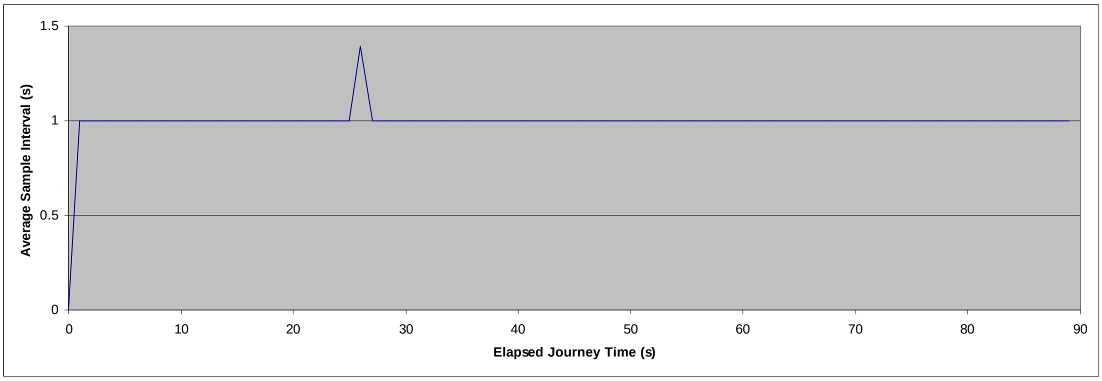


Figure 43 Sample Interval by Elapsed Time – OBU Type A

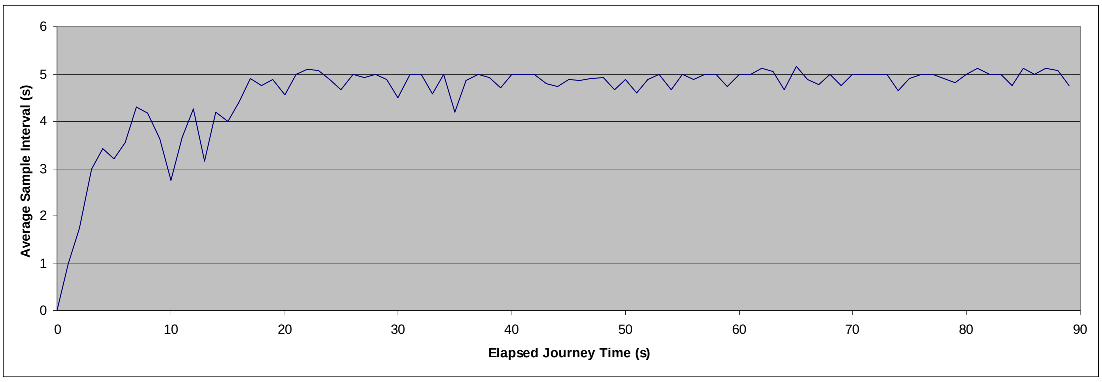


Figure 44. Sample Interval by Elapsed Time – OBU Type B

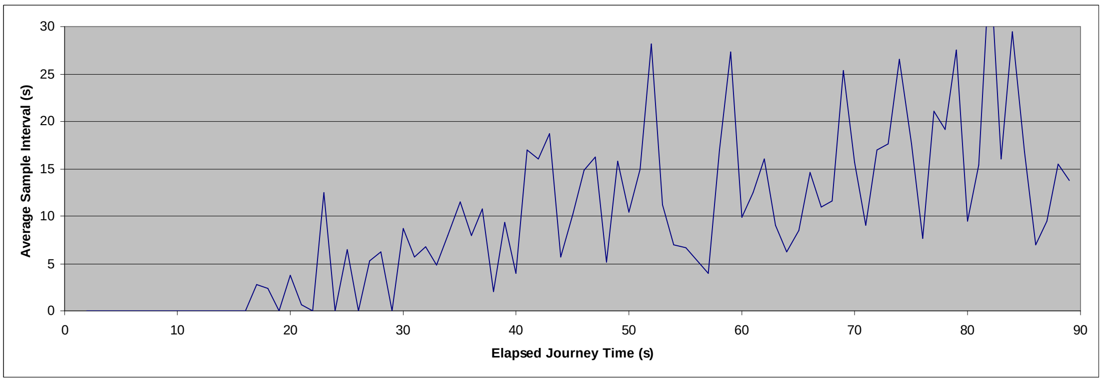


Figure 45. Sample Interval by Elapsed Time – OBU Type C

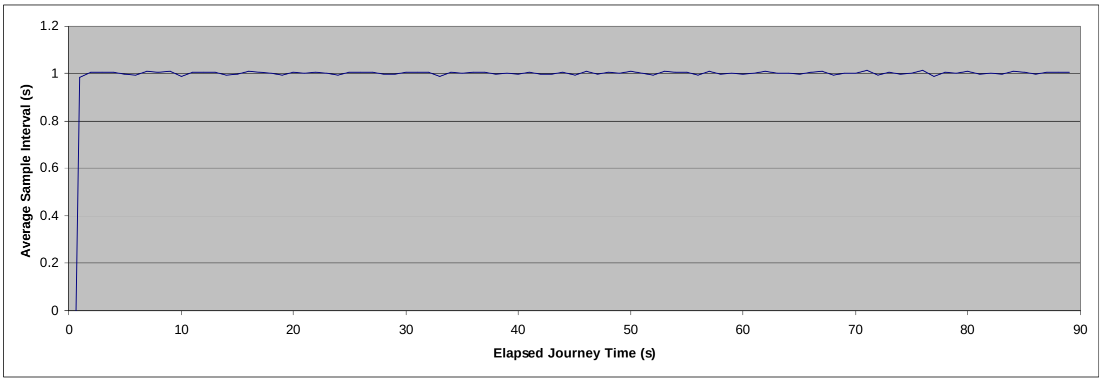


Figure 46. Sample Interval by Elapsed Time – Reference

Some OBU types showed a significant deterioration in sample interval at the start of journeys. Other devices showed no such behaviour. As for location error, the reasons for this are not fully understood.

A3.6. Location Latency

Location latency is a combination of availability and accuracy.

A latency figure represents the average elapsed time that passes before a given location accuracy (in metres) is achieved. For example, one could state “if the required accuracy were 10m, latency would on average be 4 seconds”.

Latency is typically quoted for a specific distance value, known as a ‘required level of accuracy’. The following accuracy levels were used in this study:

- 3 metres
- 5 metres
- 10 metres

The results for the best, worst and average of OBU types, plus the Reference device, are shown in Figure 46. The average has been weighted to take into account the number of location samples for each OBU type. Performance has been ranked by the average latency at the 3 metre accuracy level.

The figures should be interpreted as per the following example:

> *The average time required for the worst OBU type to produce a location accurate to within 3 metres was just over 70 seconds.*

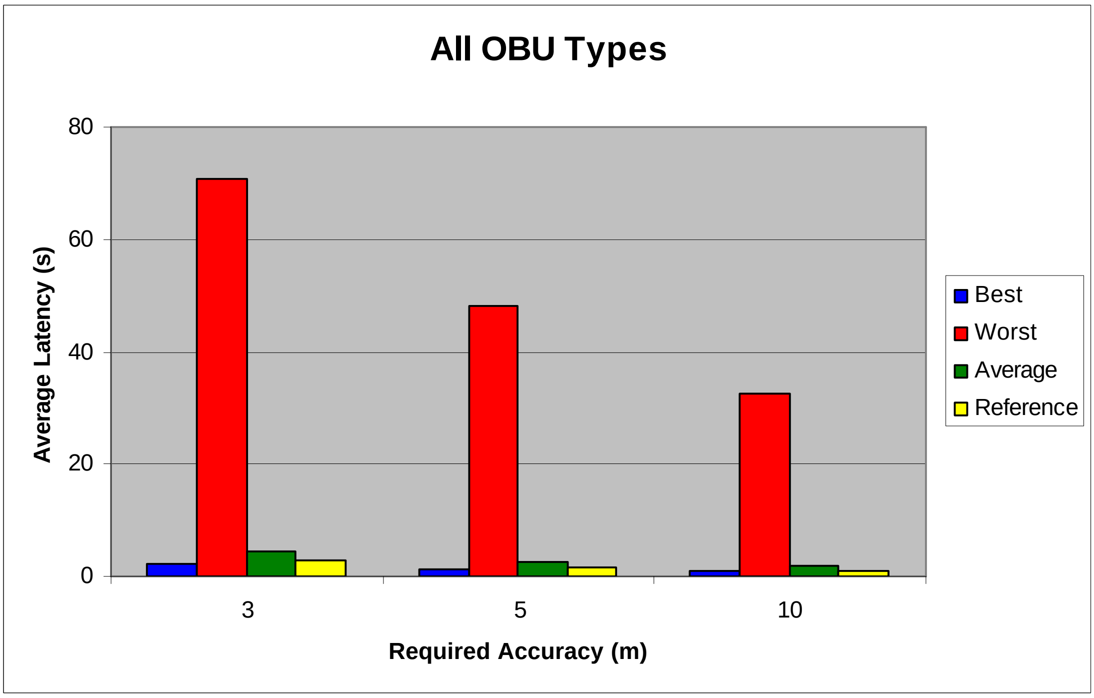


Figure 47. Location Latency – All OBU Types

Figure 47. demonstrates that there is a fundamental relationship between the sample interval of an OBU and its latency performance.

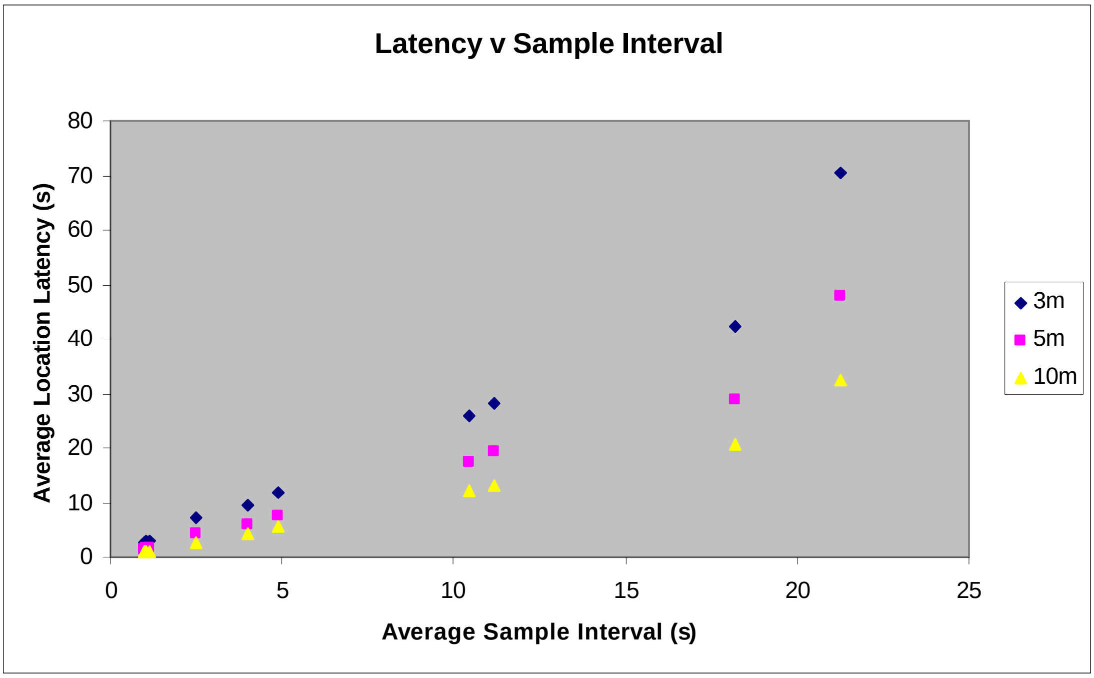


Figure 48. Relationship Between Sample Interval and Latency

It is therefore useful to divide the OBU types into two distinct groups: those that have a sample frequency of approximately 1 Hz, and those that sample more slowly than this. Figure 48 and Figure 49 show the results for the best, worst and average of the OBU types in these two groups, plus the Reference device for the 1 Hz group.

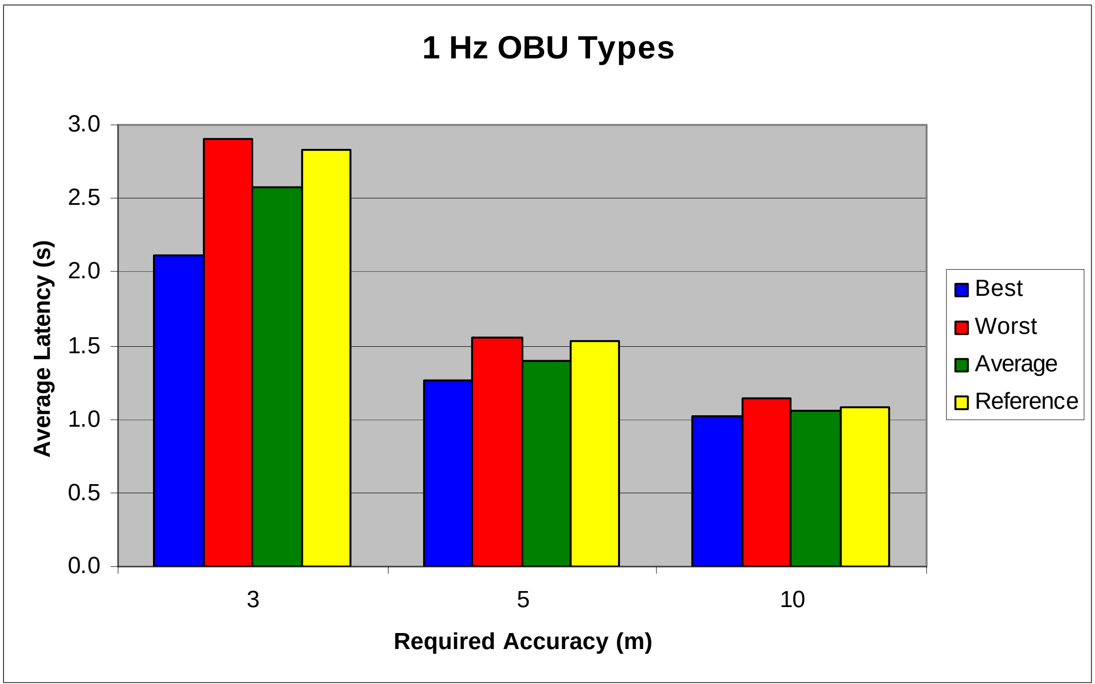


Figure 49. Location Latency – 1 Hz OBU Types

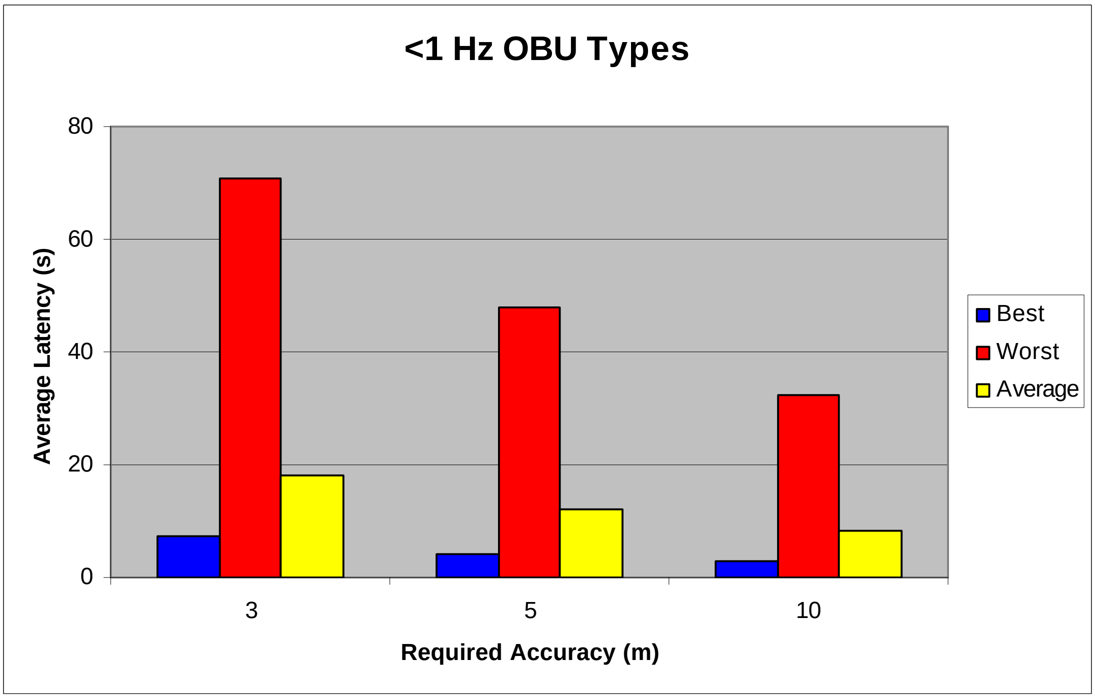


Figure 50. Location Latency – <1 Hz OBU Types

The detailed location latency performance of each OBU type is shown in Table A3.2, ranked by average latency at the 3 metre accuracy level.

OBU Type	3 metres		5 metres		10 metres	
	Average (s)	Std Dev (s)	Average (s)	Std Dev (s)	Average (s)	Std Dev (s)
Best	2.12	10.27	1.26	6.57	1.02	4.73
	2.24	9.63	1.29	4.6	1.02	0.95
	2.45	10.44	1.38	5.77	1.03	1.38
	2.46	9.38	1.33	4.29	1.04	1.18
	2.47	9.6	1.35	4.92	1.04	1.22
	2.54	8.56	1.36	3.77	1.03	1.06
	2.56	8.98	1.37	4.11	1.03	1.15
	2.7	11.25	1.44	5.67	1.06	3.37
	2.81	9.95	1.48	5.16	1.09	2.52
	2.9	21.52	1.56	16.21	1.14	7.71
	7.19	17.01	4.21	9.82	2.78	3.65

OBU Type	3 metres		5 metres		10 metres	
	Average (s)	Std Dev (s)	Average (s)	Std Dev (s)	Average (s)	Std Dev (s)
	9.39	20.3	6.02	11.19	4.4	4.58
	11.93	23.3	7.57	13.54	5.53	7.14
	25.87	42.02	17.38	27.11	12.2	14.17
	28.14	50.41	19.26	37.04	13.15	19.4
	42.41	56.67	28.86	38.88	20.6	20.49
Worst	70.64	103.04	48	70.19	32.5	47.95
Vendor Avg	**4.39**	**13.08**	**2.62**	**7.42**	**1.89**	**3.37**
Reference	2.84	11.96	1.53	6.44	1.08	2.47

Table 14. Detailed Latency Performance

Key TfL findings were:

The best OBU type took 2.12 seconds to produce a location accurate to within 3 metres on average.

The average time to produce a location accurate to within 3 metres, averaged over all vendor OBUs, was 4.39 seconds.

A strong relationship was observed between the sample interval of an OBU and its latency performance. This is a natural consequence of the way in which latency functions.

There is a relationship between the required accuracy and the elapsed period between acceptable data. For 3m accurate data, the systems can, on average, produce results within 4.5 seconds. However if this requirement is loosened to 10m accuracy, acceptable data would be available within 2 seconds. There is a clear trade off between the required accuracy and the concurrent periods where acceptable data may not be available.

END OF DOCUMENT