Using Tracked Mobile Sensors to Make Maps of Environmental Effects

Anthony Steed and Richard Milton

Department of Computer Science

University College London

Gower St

London

WC1E 6BT

Corresponding Author:

A.Steed@cs.ucl.ac.uk

Tel: +442076794435

Fax: +442073871397

Abstract

We present a study the results of a study of environmental carbon monoxide pollution that has uses a set of tracked, mobile pollution sensors. The motivating concept is that we will be able to map pollution and other properties of the real world a fine scale if we can deploy a large set of sensors with members of the general public who would carry them as they go about their normal everyday activities. To prove the viability of this concept we have to demonstrate that data gathered in an ad-hoc manner is reliable enough in order to allow us to build interesting geo-temporal maps.

We present a trial using a small number of global positioning system-tracked CO sensors. From analysis of raw GPS logs we find some well-known spatial and temporal properties of CO. Further, by processing the GPS logs we can find fine-grained variations in pollution readings such as when crossing roads. We then discuss the space of possibilities that may be enabled by tracking sensors around the urban environment – both in getting at personal experience of properties of the environment and in making summative maps to predict future conditions. Although we present a study of CO, the techniques will be applicable to other environmental properties such as radio signal strength, noise, weather and so on.

Keywords

Mobile computing, GPS, pollution monitoring, sensor networks.

1 Introduction

Increasingly the general public are carrying electronic devices whilst they are mobile. Whilst these devices and being miniaturised and integrated in wearable artefacts and even clothing, they are also increasing in power and broadening in capability. They are becoming ubiquitous platforms upon which a wide range of applications might be built. One way in which the range of applications might be expanded is to incorporate sensors that monitor the user and their environment.

  In this paper we want to explore the area of personal monitoring of environmental effects. Whilst there is a lot of interest in static sensor networks, such as the GlacsWeb system [1], a system of small communicating sensors for monitoring glaciers, we believe there is potentially a large area of potential applications of monitoring that could be carried out by people carrying low cost devices. From an environmental monitoring point of, the potential of sensor networks in general is that by using large numbers of cheap sensors, phenomena that are active over a long time and a large geographical scale can be observed at much reduced cost [2]. One of the main challenges in deploying such systems is in creating a network amongst these sensors that communicates results back to the scientists. We can bypass this problem by making the sensors themselves mobile, storing the data locally and thus uploading data reliably when a suitable service becomes available. If we also track our sensors, we can make geographic maps of an environmental effect at a much more detailed level that would otherwise be possi-

ble. Mobility also gives a broader range of sensing conditions and those conditions can better represent an individual's personal experience of the environmental effect in question. We suggest that tracked mobile sensor networks are an interesting complement to fixed sensor networks because they open up a new space of applications.

We want to demonstrate these ideas with an urban pollution monitoring application. Pollution is an important issue in many cities and there are various guidelines and laws regulating acceptable levels (see Section 2 for examples). In some cities there is archival data about various pollutants that goes back 30 years or more. Archival data provides us with a type of ground truth reading with which to compare our own readings. Such archival data will only be for a small number of fixed points, perhaps no more than one per km^2 even in the most densely sampled regions. However, because one of the sources of pollution is traffic and because pollutants are dispersed by wind and rain, the actual pollution level can vary at a very fine scale. We thus propose to deploy a set of mobile sensors carried by commuters that can fill in the gaps between fixed sensor readings. We claim it is not necessary to put thousands of fixed sensors out in to the environment; rather we can find out the nature of spatial and temporal variation by using a smaller number of tracked mobile sensors. We would hope to be able to answer questions such as how much variation is there in an area, and over what scale do changes occur; questions that are impossible to answer with a fixed sensor network without deploying many thousands of sensors. Such questions apply to a host of properties that might be sensed by cheap mobile sensors.

We present a study of carbon-monoxide (CO) pollution that uses a small set of mobile CO sensors. The sensors are tracked using the Global Positioning System (GPS). Our task in this study was two-fold. First we wanted to demonstrate that we could find small scale variations in CO that would be expected based on knowledge about how CO behaves, such as variation over the day, and rises in CO near idling traffic. Finding such effects will give us confidence that maps we generate at a larger scale have some validity. We choose to study pollution rather than, say, noise, because of the availability of archival data and the presence of existing fixed sensors. Second we wanted to investigate the issues of using GPS traces in a situation where error in location reading might be critical.

We find that we can identify expected features of pollution. Without any processing on the GPS logs, we show that we can identify regional variations on the scale of 50m or so. For example we find that on average a relatively traffic free square has a lower CO value than a busier road. If we process the GPS logs using other information then we can isolate hot spots such as bus stops and road crossings. This demonstrates that we can map environmental effects reliably with our approach.

In the following section we will present some related work on environmental sensing and correction of GPS. Then in Section 3 we will outline the sensor package and trials we have undertaken. In Section 4 we give an overview of the type and quality of data that we can capture. Then in Section 5 we discuss more detailed effects that can found in the data by post-processing the GPS logs. In Section 6 we will then generalise and discuss tools that should be more applicable to processing tracked data logs. We also discuss how we have achieved the original objectives of proving we could find known properties of CO in the urban environment and how the sensor packages we have might enable new modes of pollution survey. Finally we conclude and present plans for future work both with CO monitoring and in general with tracked sensing.

2 Related Work

The pollutant we have chosen to study in this proof-of-concept trial is CO. Transport makes the greatest contribution to CO levels and CO affects urban areas more significantly than rural areas. For example, in the United Kingdom overall CO levels have fallen since the 1970s, averaging $1mg/m^3$ [3]. In other areas of the world, especially rapidly growing urban areas, CO is rising.

The Air Quality Site contains archive data from over 1500 UK monitoring stations going back in some instances to 1972 [4]. Such data sources give a good picture of variation from urban to rural areas, the annual cycles and long-term trends. In urban areas some sense of geographic variation is conveyed by the difference in readings between kerbside sensors and sensors placed in green-field areas away from pollutant sources where CO is close to the background level. However fixed sensors don't capture or characterise the detail of per street variation. Indeed these sensors might not be representative, and they almost certainly don't show the highest levels achieved.

CO disperses over a matter of hours, but Croxford et al. have shown that this is affected by local street configuration [5]. Croxford's study used a cluster of sensors in fixed placements in a small area around University College London (UCL), UK. An example set of air guidelines is The Air Quality Strategy for England, Scotland, Wales and Northern Ireland [6], which suggests a standard of 10ppm (11.6mg/m3) running 8-hour mean. In the vicinity of UCL, the Croxford study found a peak CO concentration of 12ppm, but nearby sensors reported much lower values near the background level for CO. Thus, moving pedestrians or vehicles would probably not experience this peak for a long period.

Our approach to bridging this gap between large, sparse and small, dense studies is to use tracked mobile sensors that are carried by pedestrians or mounted on vehicles. There are several other approaches that aim to increase understanding of pollution variation using different sensor deployment strategies. Examples include the dense sensor network of the Discovery Net e-Science project [7], and the combined sensing and air flow modelling approach of DAPPLE [8].

A significant issue in understanding tracked logs is in interpreting and correcting GPS readings. GPS is of course a very significant area of research and there are many existing and upcoming enhancements to GPS that can improve its accuracy [9]. However we wanted to focus on consumer-level GPS because we anticipate sensor package cost being a primary concern when deploying large networks. We also didn't want to weigh down the carriers, nor burden them with any device configuration tasks. Indeed, in our trials the user simply had to power on the devices before going outside. In an urban environment, especially in a densely built city such London where the buildings are typically close to or higher than the road width, we do not expect to get a GPS fix 100% of the time [10]. However we achieve results that are quite similar to other tests of un-enhanced GPS tracked devices in London [11].

Finally, we note that there are parallels to the technologically motivated work on mapping of availability for various types of communication service. A good exemplar is systems that monitor Wi-Fi signal strength in order to indicate areas of reliable coverage such as [12]. While similar in technology because this is a mobile application that involves environmental sensing, radio waves in this case, they have a completely different aim. They are active systems, in that the user model is to interactively use the maps to get a better service. Ours are passive systems, in that the user is not directly involved in the sensing, and isn't expected to use the data interactively. In particular, if GPS isn't giving a fix, or the sensor isn't working, the user is not disappointed or hindered in their normal activity. We might say that our type of mobile monitoring is parasitic in nature.

3 Trial Description

3.1 Overview

The device we will describe is small and fits comfortably in an outside pocket of a rucksack. It is light enough to not be a burden, and during preliminary testing one of the authors carried the device to and from work every day for a couple of months. In future scenarios, we envisage this being a typical mode of use; maps will be built by

people carrying such devices about their everyday travel. They would be motivated to do this by being involved in a large-scale scientific data gathering exercise where their data contribution would lead to a model and understanding that would be of benefit to the whole community. They might also learn about their own exposure and change their behaviour to avoid particular areas that are badly polluted.

To get a dense set of data in order to look at the issues of understanding tracked sensor data we started with a focused trial that looks at a small area of London. We collaborated with members of the Dispersion of Air Pollution and Penetration into the Local Environment (DAPPLE) project [8] which studied a site on Marylebone Road in central London. This site is in a street canyon intersection with a lot of heavy traffic, thus it should provide high and variable readings that our sensors should detect. The DAPPLE project placed a number of sensitive fixed sensors in the environment that covered CO and particulates as well as weather stations and anemometers. They also had people carrying un-tracked sensors around.

We participated in two weeks of data gathering. On both weeks we collected data with pedestrians. On the second week we additionally collected data from cyclists.

3.2 Equipment

The equipment used consisted of a standard PocketPC PDA running our own data logging software, an ICOM device to measure CO and a GPS unit. The type of PDA was not important, and the PDA simply functioned as a data logger with a small diagnostic display to assist with starting the software and verifying log files at the finish. The ICOM is designed as a CO data logger, but we used it in real-time mode where it reports a CO reading every second. Two types of GPS units were used: the first a commercial HAICOM unit which plugs straight into a PDA; the second a "White Box" type that connects to the PDA via a compact flash serial card. The "White Box" type was a unit we assembled ourselves using a GARMIN chipset and its own battery. All the equipment was colour coded and divided into separate sets. The sets were colour coded and parts were never swapped between sets. This allows for easy tracking of the equipment used for each sample run.

Fig. 1. The ICOM CO sensor on the left, the PDA in the middle and the GPS aerial on the right. The HAICOM GPS unit can just be seen sticking out of the top of the PDA.

In use the equipment is placed in a rucksack to allow hands free operation. The GPS aerial was attached to the top of the pack with Velcro. All the pollution monitors were placed in a pocket open to the air on the right hand side of the pack, unless otherwise noted. It is possible to place two packs of equipment into one rucksack and have a CO monitor on both the left and right sides.

Immediately after being switched on, the ICOM units give artificially high CO readings. For this reason, any readings taken in the first 20-30 seconds are unreliable.

Data is sampled once every second from the ICOM CO sensor, combined with the last valid GPS fix and written to a log file on the PDA. Every time the software on the PDA is run, both a GPS log file and a pollution log file are created with a filename that includes the current timestamp. After data collection, these files are uploaded to a database server for analysis.

In principle it would not be difficult to extend the software and hardware to support opportunistic upload of data to a server when the PDA came in to range of a wireless network. We chose not to do this because it would have required a different type of PDA that we did not have available in sufficient numbers at the time of the trial. It would also have added some unreliability to the system.

3.3 System Overview

The log files collected on the PDAs were uploaded to another computer at the end of the day. These log files were then uploaded to a database server via a web server. A java program converted the logs into XML and posted these to a web server using an HTTP-POST message.

The combined web server and database server was a Linux machine running Apache Tomcat version 5.0.18 and Postgres version 7.4.5. The ingestion software was written as a Java servlet that takes the XML information and ingests it into the database.

Two pieces of bespoke software were used to analyse the data collected. Single log files were examined using a Java program that displays a top down view of the recorded GPS position and colour-coded CO level on top of an aerial map. The second piece of software generates 3D visualisations of the area showing various data including the paths traversed and the mean CO values [13]. This software was written as a CGI program in C++. Parameters in the query string passed to the CGI program control the area displayed, the time range of the data and how the data is displayed. The program generates a VRML file for display on the user's browser.

3.3.1 Database Structure

The Postgres database contains two main data tables; the 'carbonmonoxide' table contains all of the sensor data, GPS data and time information while the 'carbonmonoxidepath' table has one entry for each log file uploaded. This allows the original data file to be recreated if necessary. By joining the two tables, the GPS path or the CO data for specific routes can be retrieved. The sensor ID number, start time and finish time from the 'carbonmonoxidepath' table form a unique key into the 'carbonmonoxide' table as a sensor can never be in two places at once.

3.3.2 Visualization

The method of plotting the data was to divide the area into grid squares of between 20 metre and 5 metre spacing and to place data points into the correct square based on GPS position. The mean and variance for each square can then be calculated and plotted. By altering the size of the grid squares, the data can be plotted to within the accuracy limits imposed by the GPS positional error.

3.4 Routes

Two routes along the length of Marylebone Road were used for most of the walking in addition to both walking and cycling around the cycle route (see Fig 2.). Marylebone road is a roughly east-west street and the stretch of road we used has heavy traffic with many buses and bus stops. We thus expect to get high readings and high variation in readings on this stretch. On other occasions, exploratory routes around the area were tried. The use of fixed routes for some trials was to ensure reliability and repeatability of results. The use of more exploratory walks was to investigate variation in the local area to see if there were anomalous highs or lows nearby and to start building a local map of CO. Both walking routes can be walked in either direction and on the north or south sides. Due to the one-way streets on the cycle route, the cyclists must always travel the same way round, while pedestrians can walk the cycle route in reverse. The route taken and direction of walking is noted in the experiment log, allowing the accuracy of the GPS traces to be verified.

Fig. 2. Map showing the two walking routes along Marylebone Road and the figure 8 cycle route.

Between two and four collectors went on up to four excursions each a day at around 10:30, 11:30, 14:00 and 15:00. Each excursion was roughly 45 minutes.

For every route covered, the colour of pack used was recorded, along with the route taken. This might be as simple as 'Marylebone Road north walking route, Green Pack', or could include a map showing which roads were walked. These maps were later used to add comment information to the log files and help with identifying the route taken from the GPS data.
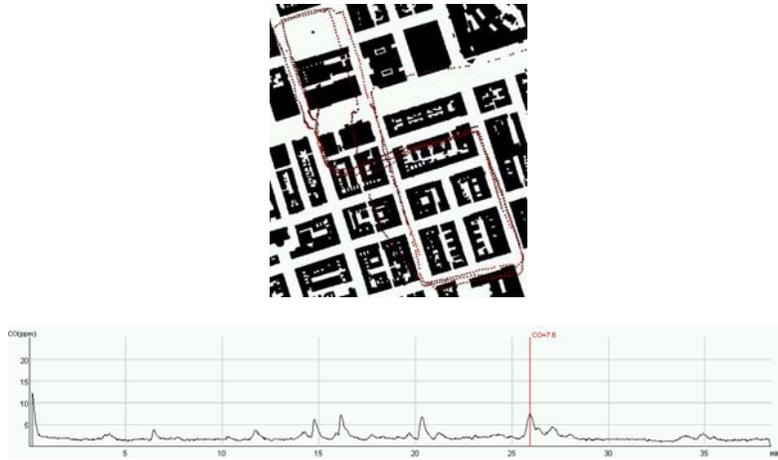
Fig. 3. GPS and CO trace for the cycle route on 19 May 2004 at 11:48

4   Overall Results and Quality of Data

During the data collection period, log files from 87 sample runs were collected, resulting in 227,496 separate data points. During the two-week sampling period, a number of equipment problems were encountered. These resulted in the loss of 3 sample runs due to the GPS not being switched on and 7 sample runs where all or some of the data was lost due to the ICOM sensors not being switched on or the leads coming loose. There were a number of problems due to the stability of the PDAs resulting in the loss of 9 sample runs. With further attention to design of the equipment, physical and software reliability could be improved, but we also note that in longer trials, infrequent unreliability of the majority of device or persistent unreliability of a small number of devices is not a problem because the user is only passively collecting data, and the density of collection will mean that readings from other devices will compensate.

4.1 Reliability of GPS in an Urban Environment

On the first day of data collection, only 61 GPS positions were recorded out of a possible 5806 (61 seconds out of 1.6 hours). This was partly due to using the HAICOM GPS units that provided a higher level of accuracy but with a reduced sensitivity. On all subsequent days the 'white box' GPS units were used, which are less accurate, but more sensitive. Using these units, the percentage of data points with corresponding GPS positions was raised to 42.8%, which still amounts to 34.6 hours of collection time over two weeks. However, a log without GPS is still useful for some forms of analysis (see Section 5.3). By comparing daily GPS data for the pedestrians and cyclists, it can be seen that the cyclists have a marginally higher percentage of valid GPS positions which might be expected because they travel nearer the centre of the urban canyons, see Table 1. The final percentages for all the data collected show that for the pedestrians, 42.8% have a valid GPS position while for cyclists this figure is 45.9%.

Table 1. GPS availability for each day, and for both pedestrians and cyclists. The total number of seconds with valid positions are shown, along with the total length of the trial in seconds and the computed availability as a percentage of the trial.

Day	Pedestrians			Cyclists		
	Valid Positions	Total Seconds	Percentage	Valid Positions	Total Seconds	Percentage
4 May	61	5806	1.1%			
5 May	12821	26504	48.4%			
6 May	16088	36904	44.6%			
7 May	4900	18194	26.9%			
10 May	17141	32598	52.6%			
17 May	4455	12300	36.2%	8217	20193	40.7%
18 May	8202	18280	44.9%	6278	12335	50.9%
19 May	5200	10214	50.9%	5102	10822	47.1%
20 May	1939	4412	43.9%	1447	2492	58.1%
21 May	5342	12776	41.8%			
Total	76149	177988	42.8%	21044	45842	45.9%

4.2 GPS Accuracy

With the GPS data, the situation is complicated by the fact that the positional accuracy is not just contaminated by noise. Due to multi-path reflections and the path to satellites being obscured by close proximity to tall buildings, positions can be consistently inaccurate but still contain a shape that matches a junction layout, even if that shape is displaced. Traces often drift, reaching 20 metres or more from the true location before suddenly jumping back when a junction is reached and alternative satellites are visible for the position fix.

The positional accuracy of the GPS data is impossible to quantify unless the true position is known using another method. One of the traces contains 30 minutes of data from a known static location at a junction where the GPS position can be seen to drift around the centre of the road. By playing back the samples in real-time, obvious discrepancies in the GPS position become apparent at points where pedestrians wait to cross the road.

4.3 ICOM Accuracy

The quoted tolerance of the ICOM CO sensors is +/- 5%, with a small amount of noise that can be seen from the collected traces. The nature of the CO levels being measured is such that it can change very rapidly and returning to the same geographical position after a few minutes can yield significantly different results. In order to test how closely matched the sensors are, a number of tests were carried out where two ICOMs were carried in the same pack. The results show a very good level of coherence:
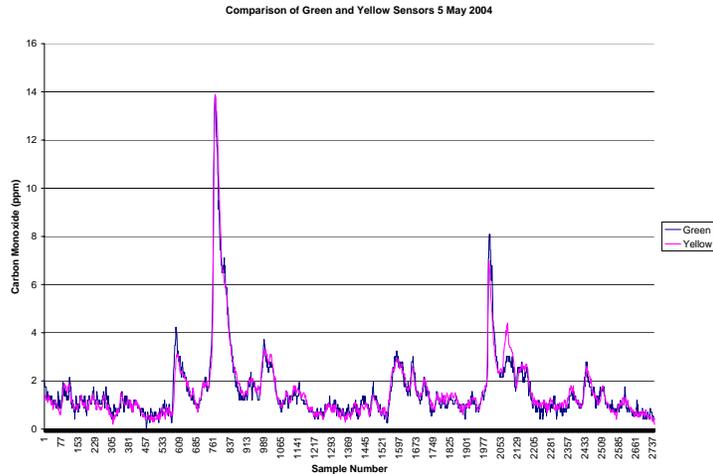
Fig. 4. Comparison of Green and Yellow sensors for 5 May 2004

After analysing the CO data recorded by the sensors, the range of values of most interest is between 0.3ppm and 5.0ppm. Although much larger values do occur, they are very rare and their location is of more interest than their value. The base level of 0.3ppm corresponds to the global background CO level. There will never be any values below this due to the way that the ICOM devices are calibrated.

5   Detected Variations in CO

In this section we discuss some of variations that we were able to detect with the tracked mobile sensors. We start by showing that we can detect gross regional differences that we might expect between heavy traffic and light traffic areas. We can also find event-related phenomena, in particular peaks when crossing roads. We then analyse the GPS logs to show that, although unreliable in absolute coordinates, they do reveal behaviour at a fine scale and these behaviours can be spotted and used to reason further about location. Finally we find that the mode of transport has an effect. This is significant as it shows that we need to be sensitive to how our passive data collection is done. Although we will be talking about CO, it is important to bear in mind that similar stages would be needed in building confidence and then analysing the reliability of any form of mobile monitoring.

5.1 Gross Geographical Features in the Data

Around the north part of the cycle route is Dorset Square (see Fig. 2 or Fig. 5.). This has a bus stop to the East, a grass area in the middle and a quiet street to the West that has very little traffic. Throughout the collection period, there were noticeably lower levels of CO on the West side of Dorset Square than the East as would be predicted based on knowledge of how traffic produces CO.

Fig. 5. 17 May 2004, mean data plotted on a 5 metre grid. The CO level below 5.0ppm is plotted as a greyscale with values exceeding this in black.

In addition to this feature around Dorset Square, it was also noticed that the Dorset Street (the roughly east – west street at the south of the lower loop of cycling route) often contained high CO peaks. The data for this area shows that CO levels often build up at the crossroads at either end of Dorset Street at the junctions with Baker Street and Gloucester Place (the two roads forming the north-south edges of the lower loop of the cycling route).

One final feature that was highlighted on a composite chart of the area showing both mean and variance is the fact that the section of Baker Street south of Marylebone Road (the east-most, almost north-south part of the lower loop) has consistently low to moderate levels of CO. This might be due to the fact that traffic flows freely along this section with no traffic lights or junctions until it meets with Dorset Street.

These features can be only be detected by repeated visits to the area. On a particular run, these features might not appear; the traffic might be unusually quiet or there might be a parked truck with an idling engine. In fact, by looking solely at the mean CO level, there is little variation and most areas are close to the background level. However with repeated visits, it is clear from the variance, that certain areas have a higher likelihood of having higher levels of CO.

5.2 Recorded CO Levels When Crossing Roads

As all the pedestrian routes involve crossing roads at some point, there is a significant amount of data available in the middle of the road. Road traffic is the main source of CO in the environment, so for our application it would be desirable to be able to identify these occurrences. The problem is how to detect when a sensor crosses the road, from the relatively poor GPS positioning information. One route stands out as a good example of road crossings, which was recorded on 10 May 2004 from 15:39:
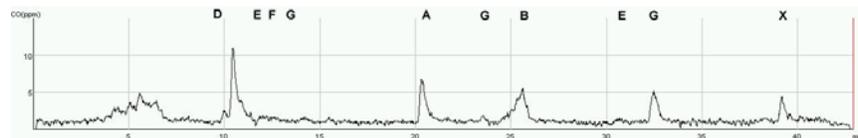
Fig. 7. Map showing positions of CO trace from Figure 6.

The highest peak occurs between points D and E just off of Marylebone Road. The trough is reached at point E on the south side of Marylebone Road. Contrast this with the next point E just after 30 minutes where there is no significant increase in CO when crossing at the same point.

Points A and J are both at traffic lights on major intersections and high levels of CO were seen at both these turnings, although in this trace, only A shows a peak.

Point B is a very slow crossing over six lanes of Marylebone Road with a wait at the island in the middle. Due to the time taken to cross and the large amount of traffic, points near here often show a peak, with points near G also having a similar characteristic. The actual positioning of this point in the trace is determined more by when the GPS is stationary than the actual reported position. Often, the trace will show the shape of the crossing, which displaced from the junction and split in the middle of the road. It is harder to determine for cyclists who cycle straight across the road.

The crossing at point G is by far the worst part of the cycle route, whether cycling or walking. Both walkers and cyclists must either dash across a four lane road or pick their way through stationary traffic at the lights. The final peak is marked as point X as there is no GPS. There is no additional information to identify this point.

In general, it was noticed that CO levels often built up while waiting on traffic islands to cross the road, but this is hard to identify from the GPS data, despite a number of sample runs where people deliberately waited in the middle of the road. Also worth noting is the fact that there are eleven crossing points when walking around the cycle route, not all of which produced peaks in this example and where there was a peak at a junction, it might disappear on the next circuit.

5.3 Effect of Mode of Transport

Due to the problems of obtaining a GPS fix in the centre of London, a large amount of data was recorded without any position information. Data without any valid GPS position was still useful in determining how CO levels varied at different times of the day, and for analysing the difference in mean recorded levels when cycling and walking.

Table 2. Ratios of the Daily CO Recorded for Pedestrians and Cyclists

Monday 17 May 2004		
Cycle	Walk	Cycle/Walk Ratio
1.49 (Red)	1.36 (Green)	1.1
1.41 (Blue)	1.36 (Green)	1.04
1.49 (Red)	1.36 (Yellow)	1.1
1.41 (Blue)	1.36 (Yellow)	1.04
Tuesday 18 May 2004		
Cycle	Walk	Cycle/Walk Ratio
1.11 (Red)	1.65 (Green)	0.67
1.11 (Red)	1.41 (Yellow)	0.79
Wednesday 19 May 2004		
Cycle	Walk	Cycle/Walk Ratio
1.43 (Blue)	1.23 (Green)	1.16
1.43 (Blue)	1.1 (Yellow)	1.3
Thursday 20 May 2004		
Cycle	Walk	Cycle/Walk Ratio
1.29 (Red)	1.47 (Green)	0.88
1.29 (Red)	1.28 (Yellow)	1.01

The mean CO level recorded for a pedestrian and cyclist around the cycle route at the same time of day provides the most directly comparable data, but an alternative method for calculating the cycle/walk ratio is to compare daily averages. The main advantage this has over comparing single routes is that there is significantly more data available for comparison. As the red and blue packs were only given to cyclists and the green and yellow packs to pedestrians, the colour of the pack determines the method of travel.

Firstly, the mean CO level measured for the whole day is calculated for each sensor. Then a ratio is calculated for every combination of cyclist and pedestrian for each day. Table 2 shows the mean CO level for each cyclist or pedestrian for all of the packs used each day.

These results show very closely matched ratios for each day, ranging from 0.67 to 1.3. The results for 18 May are surprising as both ratios are less than 1.0, but as they are so similar, there could be a physical explanation for that day.

6 Dealing with Mobility

6.1 Dealing with Tracking Data

When analysing the GPS traces, it is apparent that the GPS is not reliable enough to place a sensor on the correct side of the road. The quoted accuracy of the raw GPS data is 20m. Occasionally, a log file will contain a distinct trace that sticks to the correct side of the road, but the majority meander along the centre of the road, or drift from side to side randomly. For the Marylebone Road walking routes, without looking at the recorded route notes, it is not usually possible to be certain whether the sensor was on the north or south side of the road, unless additional features are present. For the cycle route this is less of a problem as the cyclists are always in the road and the pedestrians always take the same route, although the problem of detecting junctions accurately remains.
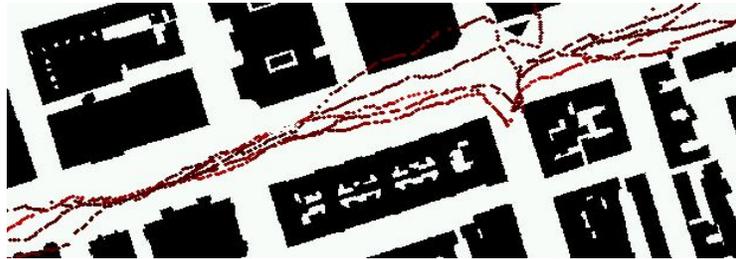
Fig. 8. Example of a GPS trace where you cannot immediately tell which side of the road the pedestrian is on.

On the Marylebone Road walking routes, which are virtually straight lines, there are points where the walkers move off line and down a side road in order to cross at a crossing away from the junction. The most noticeable instance of this is on the south side of Marylebone Road at the junction with Baker Street, where the pedestrian crossing is a short distance south down Baker Street (two examples of such a crossing are shown in Fig. 9.). By recognising certain features present in the data, it is possible to determine which side of the street the samples were taken and also identify when the sensors reach junctions.

When crossing Marylebone Road at the junctions to the east or west of the council building, there is a characteristic that often appears in the GPS trace due to the position of the crossings relative to the road.
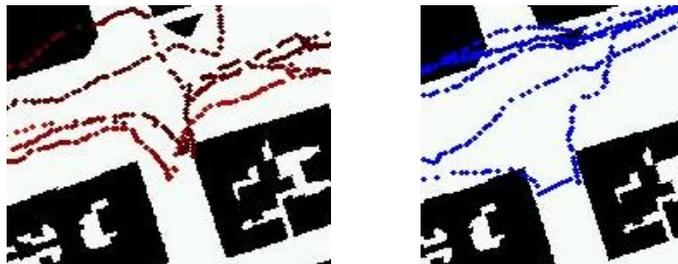

Fig. 9. Identifying a south side route from a pedestrian crossing. The main west-east road is Marylebone Road. The main north-south road is Bake St. Contrast the typical trace on the left with the trace on the right that accurately captures the geometry of the crossing.
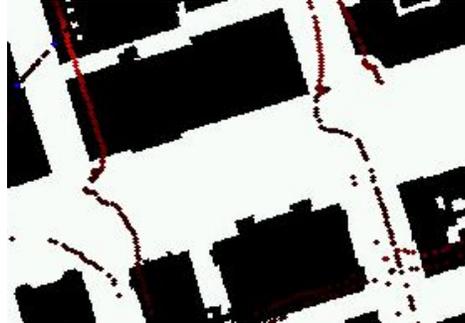
Fig. 10. GPS trace crossing Marylebone Road (the main east-west road).

The significant deviations to the left are a function of where the crossings are placed in relation to Marylebone Road. Both GPS paths are significantly out of position, placing the pedestrian in the middle of the road while he is still waiting to cross. This fact can be seen when the traces are played back in real-time by noting where the GPS position is stationary while the pedestrian is waiting to cross.

The ability to observe these small-scale behaviours from the GPS, despite the gross inaccuracy of the GPS is a great boon to our application: we can use these observations to correct the log files semi-automatically. In the case of the Marylebone Road walking routes, being along the north and south sides of a straight road, it is relatively easy to constrain the path to the correct side of the road, while a similar method can be applied to the figure eight cycle route. This method of correcting the GPS position using knowledge of the route taken has been used to separate the Marylebone Road north and south side data for 5 May 2004. After this correction has been applied, it is possible to attribute the higher CO levels recorded at the east end of the road to the south side, by the bus stops and traffic lights.

With GPS being the limiting factor in the analysis of the tracked sensor data, the ability to use this type of additional knowledge significantly improves the resolution. Using a set of simple GPS corrections [14], physical feature on a scale of 5m can be been extracted from GPS log data that occasionally shows 50m errors. In summary, it is possible to filter the GPS logs using five levels of knowledge:

- Knowledge of the geographic extent of the trial.
- Knowledge of the building footprints
- Knowledge of the route undertaken
- Knowledge of the behaviour of the carrier of the GPS
- Knowledge of expected correlation with other sensors

A set of filters have been incorporated in to a tool that allows us to semi-automatically correct GPS log files. For our application this allows to go from plotting maps of pollution at a 20m scale to 5m scale. The 5m scale allows us to pick out features such as bus stops.

However, with current consumer-grade tracking technologies, there is no ability to go any finer than roughly this 5m scale. Unfortunately this means that certain effects of CO remain will likely remain undetectable with mobile pollution. For example, there will be CO gradient across a pavement as distance from traffic increases. This sort of micro-scale behaviour can be a significant contribution to personal exposure for a traveller.

Other applications may or may not benefit from such corrections. In our case, because the environmental effects is transient and moves with air-flow, it is a very useful correction. With a Wi-Fi signal strength monitoring application, the environmental effect would occasionally be attenuated by people or traffic moving in the direct line between the sensor and the Wi-Fi basestation. However, in general the gradient of signal strength is quite smooth, and fairly even across the width of the road (modulo the distance attenuation of crossing the road). In

this case, then, correcting logs to side of the road is probably not necessary, but other types of correction such as route, geographic area and building footprints are useful. Knowledge of the behaviour of the carrier may be useful to filter based on road direction, velocity etc. depending on the type of vehicle they are using, if any.

It needs to be borne in mind that correcting logs is only semi-automatic and sometimes requires capture of other knowledge, such as the rough route being undertaken. For trials with many devices, this may not be tractable, in which case only crude maps may be drawn. We note though, that in Section 1 we motivated the trial by suggesting that devices would be carried by commuters. One of the reasons we suggested commuters, not visitors, is that they often use the same route to work, so the same set of corrections could be re-used and a model of the behaviour of the carrier developed over time.

6.2 Affordances of Mobile Sensors

With the fixed sensor networks currently used by for monitoring air pollution, there is a vast amount of data available for a few sites around the country. Current fixed sensor sites are designed to provide 8 hour running means in representative parts of the city to ensure that exposure limits are not exceeded Using a small number of tracked sensors, very fine-grained effects can be sampled in a target area, limited only by the number of sensors and accuracy of positioning. This study was able to make direct measurements of CO, both on the pavement using pedestrians and in amongst the traffic using cyclists, something that is not possible with fixed sensors. Data for a number of CO peaks along various roads were captured multiple times during sampling runs, showing identically shaped CO traces that reduced in magnitude over time. By making a composite map of where CO peaks have occurred, it is possible to determine where problem areas exist. With simple software tools, CO data can be correlated with other data sets such at meteorological information and traffic flow rates to investigate the causes.

Some of the potential benefits of using tracked sensors are that it allows us to establish variance bands around the fixed sensors. Because they are fixed, the fixed sensors are probably not at the points of highest or lowest level, so a mobile sensor may be able to locate particular problem areas. Also, a mobile sensor may be able to track more transient phenomena. We believe, though can't establish for certain, that we have managed to track a cloud of CO moving through the environment. In the future we might even speculate that mobile sensors could be sent to investigate such phenomena as they happen.

7 Conclusions

In this paper we have investigated the types of environmental monitoring we are able to do with GPS tracked mobile sensors. We have focussed on CO monitoring in a proof-of-concept trial to demonstrate that such monitoring is reliable and provides useful data. We have been able to study the collected data from three different angles: expected geographic variation at a gross scale; small-scale variation when crossing roads and differences between modes of transport. The features we can find match expectations of how CO should vary based on its being created by road traffic. This certainly shows that the use of tracked mobile sensors has potential in monitoring of CO pollution, but we have argued that similar techniques could be employed in monitoring a variety of other environmental effects.

The key problem with tracked sensors is that the location inaccuracy masks any fine-scale geographical variation. Never the less we have been able to generate these maps at a 50m scale with GPS readings that are accurate to roughly 20m. We have then been able to exploit techniques for correcting GPS readings based on knowledge about the behaviour of participants and knowledge about the environment itself. This has allowed us to start to

observe phenomena at a smaller scale. The correction techniques are quite general, but their utility would depend on whether the environmental effect being studied requires resolution below that provided by the sensors. CO varies at scale smaller than can be accurately resolved with consumer GPS units without semi-automatic correction. With future generations of GPS, it might be the case that we no longer need to correct the logs to monitor CO; however there will always be an environmental effect that we want to monitor that will require higher accuracy than is provided. Currently our tools are offline, but we are working towards making them online. Either offline or online versions of these correction tools will be more generally useful either as an aid to improving the user experience or as a tool to examine behaviour of a mobile application retrospectively.

There is a significant amount of further work to do in determining what types of useful monitoring can be done with large numbers of sensors. We have shown that we can reliably detect certain effects, and these effects are of interest in the environmental science domain. By scaling up the numbers of devices, we could cover a larger area and cover more hours of the working day. Our strategy of using tracked mobile sensors and the techniques we have to developed to deal with imprecise location should be applicable to a potentially large domain of monitoring projects that can help us to reveal hidden aspect of our environment.

Acknowledgements

References

[1] GlacsWeb Overview, http://envisense.org/glacsweb/overview.html (verified 2005-10-14).

[2] Steere DC, Baptista AM, McNamee D, Pu C, Walpole J (2000) Research Challenges in Environmental Observation and Forecasting Systems. In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (Mobicom 2000), Boston, Mass., pp 292-299.

[3] Environment Agency. Air Quality – Carbon Monoxide. http://www.environment-agency.gov.uk/yourenv/eff/air/222825/222913/?lang=_e (verified 2005-10-14).

[4] The Air Quality Archive. http://www.airquality.co.uk/ (verified 2005-10-14).

[5] Croxford B, Penn A, Hillier B (1995) Spatial Distribution of urban pollution: civilizing urban traffic. Fifth Symposium on Highway and Urban Pollution, May 22-24, 1995.

[6] Department for Environment, Food and Rural Affairs (Defra) (1999) The Air Quality Strategy for England, Scotland, Wales and Northern Ireland. Available online at http://www.defra.gov.uk/environment/consult/airquality/pdf/airstrat.pdf (verified 2005-10-14).

[7] Discovery Net. http://ex.doc.ic.ac.uk/new/index.php, (verified 2005-10-14).

[8] DAPPLE (Dispersion of Air Pollutants and their Penetration into the Local Environment). http://www.dapple.org.uk/ (verified 2005-3-06).

[9] O'Donnell M, Watson T, Fisher J, Simpson S, Brodin G, Bryant E, Walsh D (2003) Galileo Performance: GPS Interoperability and Discriminators for Urban and Indoor Environments. GPS World, June, 2003.

[10] Steed A (2004) Supporting Mobile Applications with Real-Time Visualisation of GPS Availability. In S. Brewster and M. Dunlop (eds), MobileHCI, LNCS 3160, pp 373-377.

[11] Ochieng WY, Noland R, Polak JW, Park J-H, Zhao L, Briggs D, Crookell A, Evans R, Walker M, Randolph W (2003) Integration of GPS and Dead Reckoning for Real Time Vehicle Performance and Emissions Monitoring. The GPS Solutions Journal, 6(4), pp 229-241.

[12] Wireless Network Visualisation Project. http://www.ittc.ku.edu/wlan/images.shtml (verified 2005-10-14)

[13] Steed A, Spinello S, Croxford B, Milton R (2004) Data Visualisation within Urban Models. Theory and Practice of Computer Graphics 2004, IEEE Computer Society, pp 9-16.

[14] Milton R, Steed A (2005) Correcting GPS Readings from a Tracked Mobile Sensor. In Strang T, Linnho.-Popien, C (eds) Location and Context Awareness, LNCS 3479, pp. 83-94.