Using Real-Time Road Traffic Data
to Evaluate Congestion

Jean Bacon, Andrei Iu. Bejan, Alastair R. Beresford, David Evans,
Richard J. Gibbens, and Ken Moody

Computer Laboratory, University of Cambridge,
15 JJ Thomson Ave., Cambridge CB3 0FD, UK
{firstname.lastname}@cl.cam.ac.uk

Abstract. Providing citizens with accurate information on traffic conditions can encourage journeys at times of low congestion, and uptake of public transport. The TIME project (Transport Information Monitoring Environment) has focussed on urban traffic, using the city of Cambridge as an example. We have investigated sensor and network technology for gathering traffic data, and have designed and built reusable software components to distribute, process and store sensor data in real time. Instrumenting a city to provide this information is expensive and potentially invades privacy. Increasingly, public transport vehicles are equipped with sensors to provide arrival time estimates at bus stop displays in real-time. We have shown that these data can be used for a number of purposes. Firstly, archived data can be analysed statistically to understand the behaviour of traffic under a range of "normal" conditions at different times, for example in and out of school term. Secondly, periods of extreme congestion resulting from known incidents can be analysed to show the behaviour of traffic over time. Thirdly, with such analyses providing background information, real-time data can be interpreted in context to provide more reliable and accurate information to citizens. In this paper we present some of the findings of the TIME project.

Key words: static sensor, mobile sensor, traffic monitoring, middleware, bus probe data, journey times, large scale data analysis, quantile regression, spline interpolation.

1 Introduction

Congestion on roads, especially in urban areas, has a large negative social and economic impact on the community as well as on the environment, with the cost of congestion to the UK economy estimated at £12 billion a year in 2004 [1]. Unfortunately, the problem is unlikely to go away without significant effort: the distance driven in the UK by private motor vehicles has increased almost linearly from just over 50 billion vehicle kilometres in 1950 to over 500 billion vehicle kilometres in 2008 [2].

Congestion can be tackled either by increasing road network capacity or by decreasing demand in congested areas, especially during peak hours. Increasing capacity is very difficult, especially in urban environments, due to the density and importance of the built environment. Therefore some demand must be reduced or displaced, by encouraging drivers to change mode of transport, alter the time of day or day of travel, or not to travel at all.

Congestion is itself a form of demand management, but as it has a severe negative economic and environmental impact there is a strong incentive to find better solutions. One of the major difficulties in reducing congestion on the road network is that it is hard to change the mindset of travellers. Travelling becomes an automatic and habitual process, and travellers are often unaware of alternative means of getting from A to B. Fortunately the provision of better information about likely costs and travel times has the potential to encourage travellers to explore alternative times and modes of transport [3].

Simply providing additional information to travellers may not be sufficient because the private cost of travel predominantly lies up-front, with the result that once a car has been purchased, taxed, and insured, the marginal costs are comparatively low compared with other forms of transport. It remains to be seen whether the recent severe increases in fuel costs will mitigate this effect. Many governments around the world have promoted various forms of road-user pricing, which offers a more sophisticated method of demand management than increasing fuel prices. Charging cars for access to central London is an example of this trend.

Accurate data concerning the use of the road network is essential in order to measure congestion, and measurement in turn is a prerequisite to reducing congestion, whether by increasing capacity or through demand management. In particular, the accurate prediction of journey times is essential in any application intending to encourage travellers to consider alternative modes of transport. Therefore, journey-time measurement is likely to be crucial, regardless of whether the carrot (provision of information) or stick (road-user pricing) approach is used to encourage road users out of their vehicles.

It is expensive to instrument a whole city with static sensors. Increasingly, vehicles are being equipped with sensors and networking facilities to enable them to communicate with each other and with stationary infrastructure. For example, since April 2006 buses used for public transport in and around Cambridge have been equipped with the capacity to detect and transmit in real time their current position. The original reason for this was to provide arrival time estimates at bus stop displays, tackling congestion by encouraging a modal shift to public transport through the provision of better information.

As we shall show, these data from buses can be "re-purposed" to yield more general descriptions of the state of the road network. Archived data can be analysed statistically to understand the behaviour of traffic under a range of "normal" conditions at different times of day; periods of extreme congestion resulting from known incidents can be analysed to show the behaviour of traffic over time. This analysis can provide background information so that real-time

data can be interpreted in context to provide more reliable and accurate information to citizens.

We have found that combining data from disparate sources—and using them in ways that were not foreseen at design time–can yield rich, unique results. For example, we have linked analysis of bus positions with open map and bus stop data to give a detailed picture of traffic patterns and congestion information across the city. Data describing the state of traffic lights can be combined with reports of bus positions to illustrate how lights and buses interact. Variation due to unexceptional context (such as time, day, date, weather, school term etc.) can be determined to answer contextual queries, and the effects of exceptional circumstances, such as scheduled work by utilities, accidents in the city, or closure of the surrounding trunk roads, can be quantified. In general, these analyses incur no extra expense in information gathering because the infrastructure required to collect the data already exists.

The TIME project has focussed on urban traffic measurement and analysis, using the city of Cambridge UK as an example. We have worked with significant players in transport such as Cambridgeshire County Council, who are responsible for transport within and around the city, and Stagecoach, a major local bus operator. We have investigated suitable sensor and network technologies for gathering traffic data and combined these with proprietary systems already in place. Data from these sources are transported, archived, and analysed using a reconfigurable, component-based middleware that we have designed and built [4].

In this paper we give an overview of the TIME project and focus on three applications. We describe the main sensors we have considered in Sect. 2, the data we have made use of is described in Sect. 3, and Sect. 4 presents selected applications. Section 5 summarises and considers the insights the research project has provided for a large, practical problem.

2 Sensors

In this section we describe the sensors we use to determine the state of the road network. Each is *static*, where its location is fixed, or *mobile*, where it moves about.

2.1 Static Sensors

Our static sensors generate streams of events, each representing data about a particular occurrence or point in time. Our two major static data sources are described below.

Inductive Loops at Junctions. The traditional approach to measuring the movement of vehicles in the road network is to use sensors such as inductive loops in the road itself to detect the presence of vehicles at fixed points. A successful example of this is the SCOOT (Split Cycle Offset Optimisation Technique) infrastructure, which is an adaptive traffic control system designed to coordinate

the operation of the traffic signals in an area in order to give good progression to vehicles [5]. Fig. 1 illustrates the idea, showing a junction that includes lanes supporting several vehicle movements (e.g., approaching from the left and turning left towards the top, going straight on, or turning right and heading downwards). The various inductive loops enable the SCOOT algorithms to detect vehicles in each lane and consequently estimate the volume of traffic destined for each movement. The six traffic lights are scheduled appropriately.
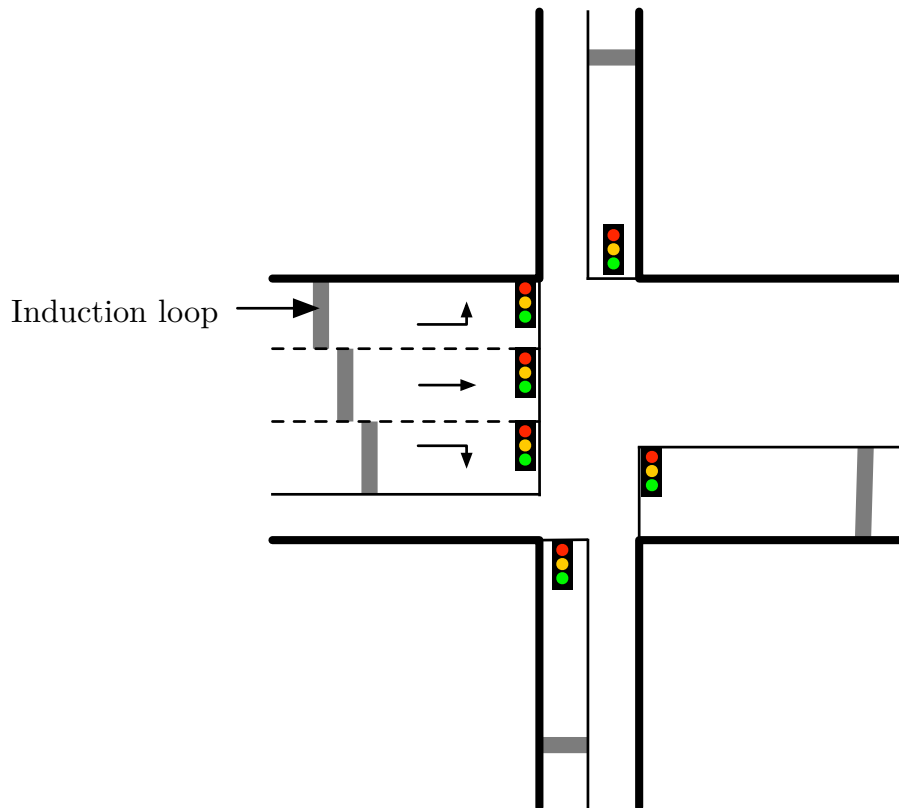

Induction loop

Fig. 1. A junction with six traffic lights controlled by the SCOOT system. Six inductive loops in the road provide the detection of vehicles used to schedule the lights.

Sensors like inductive loops require specialised equipment to be installed in a harsh environment. Moreover, maintenance requires visits to locations that are often difficult to access safely and repairs disrupt traffic. These difficulties have limited the installation of systems using inductive loops and coverage of the urban road network is rarely dense. Indeed, the number of SCOOT-instrumented junctions in Cambridge has decreased since their installation in the 1980s because of equipment failure and the expense and awkwardness of replacement.

Infra-red Counting. An alternative to induction loops is to detect vehicles optically. We do this using an infra-red detector. This sensor is capable of observing road users (including cars, buses, lorries, cyclists, and pedestrians) and counting how many pass "count lines" placed in its field of view. By aiming the detector at a roadway and using appropriate count lines, it is possible to determine the number of vehicles per time period that pass in each direction. The detector is also capable of measuring a value related to the speed of each road user that crosses a count line.

The output of the sensor can be seen in Fig. 2(a), showing the "blobs" detected; the count lines are shown in green and blue. Fig. 2(b) shows the same scene as seen by a visible-light video camera. The "blobs" correspond clearly to vehicles. Furthermore, the amount of detail that the detector can determine about each vehicle is minimal—image (a) certainly does not reveal number plates or vehicle occupancies—and so the sensing process is inherently privacy-preserving.
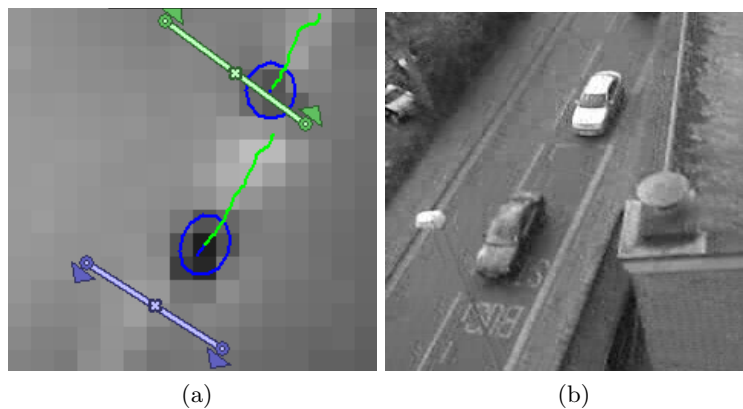

        (a)   (b)

Fig. 2. This figure shows the output of the infra-red vehicle detector (a) and a visible-light view of the same scene (b).

The detector is portable and we have tested it using an erectable mast (see Fig. 3). However, its primary deployment has been at the top of a lamp post, also shown in Fig. 3. Located along one of the radial roads of Cambridge as shown in Fig. 4, it measures the volume of traffic flowing in each direction. To support the detector, the lamp post is equipped with computation facilities in the form of a domestic wireless router, which is inexpensive, has no moving parts, requires little power, and is designed as robust consumer electronics. It is connected to the detector via one of its serial ports and runs OpenWRT.[1] A

---

[1] `http://www.openwrt.org`

line-of-sight WiFi connection to the Computer Laboratory, built using standard high-gain antennae, allows us to retrieve data from the detector.

Fig. 3. Infra-red traffic counting via an erectable mast and a lamp post
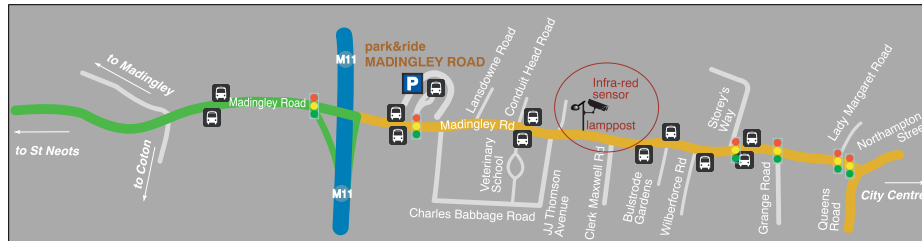

Fig. 4. A radial road into the city, showing the location of the M11 motorway, the Park&Ride terminus, other bus stops, traffic lights and the lamp post with inbound and outbound infra-red monitoring.

2.2 Mobile Sensors

While a static infrastructure is suitable for a freeway environment, it is of limited use in urban settings because traffic flows in urban road networks are highly variable across different network links and different times of day. Therefore, area-wide traffic data collection is necessary [6]. In addition, providing comprehensive static sensor-based coverage of a city would be prohibitively expensive.

An alternative approach is to use *probe* data from a sensor, such as a GPS device, attached to a vehicle or person. Probe data consist of a sequence of co-ordinates recorded over time and may deliver much more information than is typically available from fixed sensors [7, 8]. Use of probe data is commonplace. Google's "Maps for Mobile" combines location data taken from participants' GPS-equipped mobile phones ("crowdsourced" data) with a traditional sensor infrastructure to overlay road maps with congestion information on arterial

roads [9] and the Mobile Millennium Project's Mobile Century field experiment demonstrated the feasibility of traffic monitoring systems based on mobile phone probes [10]. Researchers have explored the feasibility of using commercial transport fleets of vehicles, such as taxis, in estimating traffic conditions on arterial roads. Herring et al. [11] analysed arterial travel time distributions using a Hidden Markov Model technique applied to sparsely observed probe taxi vehicles, and Braxmeier et al. [12] studied positions and velocities obtained from a fleet of 300 GPS-equipped vehicles by regarding the recorded velocities as realizations of a random velocity field sampled at selected points only.

Many public transport fleets have been augmented with automated vehicle location (AVL) systems which use GPS to collect probe data [13] and support Real-Time Information (RTI) systems. This allows buses to be probes and the potential has been studied. For example, Bertini and Tantiyanugulchai [14] and Chakroborty and Kikuchi [15] studied the effectiveness of buses as probes by investigating to what extent their travel characteristics relate to those of general traffic. Pu and Lin [16] and Pu et al. [17] investigated the effect of bus-specific operations and behaviour on bus probe performance as well as the interrelation between bus and car speeds. These studies conclude that using buses as probes to detect general traffic conditions is a viable real-time traffic monitoring mechanism. Note that in all these studies the control characteristic was chosen to be travel time and bus probes' locations were assumed to be recorded at a relatively high sampling rate.

Non-experimental AVL data tend to be sparse—typically samples are recorded once every 20 or 30 seconds—and therefore techniques developed for probe data having high update rates are not directly applicable. In our work within the TIME project we have analysed sparse probe data collected from a fleet of over 100 buses. By combining these data with descriptions of bus stop locations and the road network (as will be outlined in Sect. 3.2) we can build a rich vision of traffic patterns and congestion. We have

- developed an algorithm to extract bus journeys and estimate their duration;
- determined that quantile regression can be used to visualise contextual factors that affect journey times;
- recovered speed information from sparse probe data using monotonic splines; and
- validated this recovery by comparing it to probe data having a high update rate.

Bejan et al. provide detail [18]. Furthermore, we built on our technique to recover bus speed information to reconstruct the dynamics of time-space traffic velocity fields in urban areas [19].

2.3   General Vehicular Mobile Sensing

Current trends in Intelligent Transportation Systems suggest that an increasing number of vehicles will be equipped with wireless transceivers that will enable

them to communicate with each other and form a special class of wireless networks known as vehicular ad hoc networks or VANETs. Researchers and the automotive industry are envisioning the deployment of a large spectrum of applications running on VANETs, including notification services for accidents and traffic congestion. Our TIME collaborators at the University of Oxford have investigated VANETs in the context of probe data [20–22]. They leverage connectivity between travelling vehicles in an urban area to propagate traffic information generated by vehicles to stationary gateways spread across the city.

Due to intermittent connections and an absence of contemporaneous end-to-end paths, the primary concern is to establish a reliable and efficient routing layer. When traffic density is low or when only few vehicles participate, the vehicular network often becomes disconnected. Carry-and-forward protocols are proposed for the reliable delivery of messages between vehicles in dynamically changing network partitions [20]. These data forwarding protocols exploit knowledge of traffic statistics in an urban setting to enable timely delivery of messages from vehicles to stationary gateways, whilst minimizing message transmissions and optimizing bandwidth utilization. To do so, they proactively alternate between two forwarding strategies. Multihop Forwarding refers to the aggressive forwarding of messages to vehicles that are better positioned to deliver them to a gateway; Data Muling describes buffering messages in local memory and carrying them along with the vehicle, in effect doing data transmission at the vehicle's speed. The problem of data forwarding is then explored together with the problem of data acquisition (deciding the rate at which vehicles acquire sensor data), and a joint optimization of these two intertwined aspects of traffic monitoring is proposed [21].

3   Data

Sensors are one source of data to be used by applications. In this section we provide more detail of bus probes and describe data that are not produced by sensors.

3.1   Bus Probe Data

The bus location data used in our work were provided by the company providing real-time information facilities to the largest bus operator in Cambridge. The data represent a set of bus location points recorded over a period of 4 years (2007–2010). Many of the buses in question travel along main radial roads and multi-lane highways; some connect Cambridge to villages in the surrounding area.

Fig. 5 shows the number of buses contributing probes throughout most of 2009 and most of 2010. Values for the following three areas are given: the city of Cambridge (dots and the green line), Cambridge and surrounding villages (crosses and the blue line), and the encompassing area of Cambridgeshire and Bedfordshire (stars and the red line). In 2009 there were on average 115 buses

equipped with GPS units within Cambridge on weekdays, 100 buses on Saturdays, and 65 buses on Sundays. In April 2010 the required GPS units and transmission hardware were installed in approximately 50 more buses.
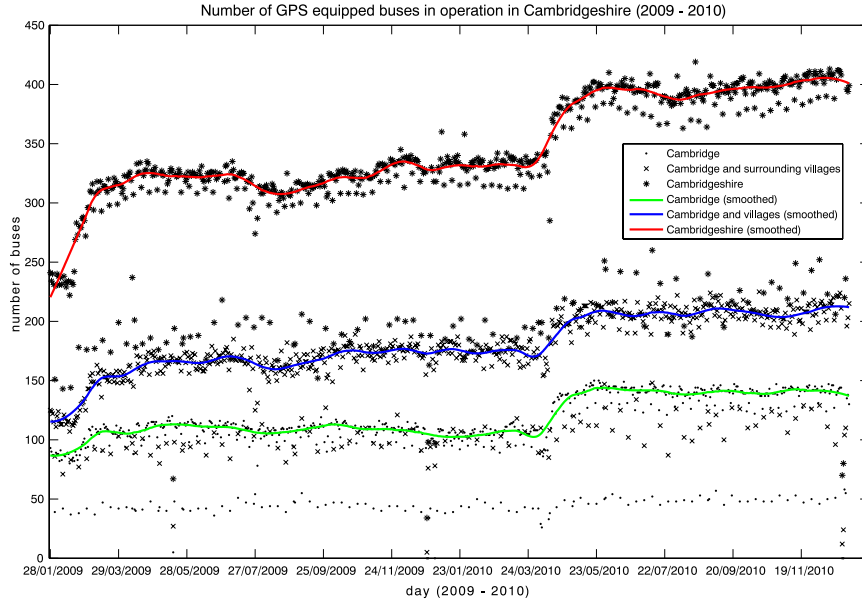

Fig. 5. This figure shows how many buses, equipped with GPS devices, were in operation in Cambridgeshire and Bedfordshire in 2009-2010 and contributed to the data set used in our studies.

Bus location sampling points are given in a local "easting/northing" (EN) coordinate system (similar to the British Ordnance Survey national geographic reference system [23]) which can be regarded as a Cartesian system within the scale of a medium-sized city. The grid squares of this coordinate system are of size 10m×10m and, given the area covered by Cambridge, there is a simple linear transformation to conventional latitude/longitude coordinates. We estimate the accuracy of each location report as ±30m. Each bus reported its location once every 20 or 30 seconds, though rarely every 60 or 120 seconds. There were missing observations, meaning that there were times of silence when some or all buses were not transmitting their positions due to communication failure, temporary hardware malfunction, weather, and so on. Finally, the location reports did not describe which bus route each bus was taking.

As is shown in Fig. 6, the bus probe data give a good coverage of the main city roads including arteries. The temporal coverage is similarly good: the data are available throughout the day, including peak morning and afternoon hours when traffic conditions are of most interest.
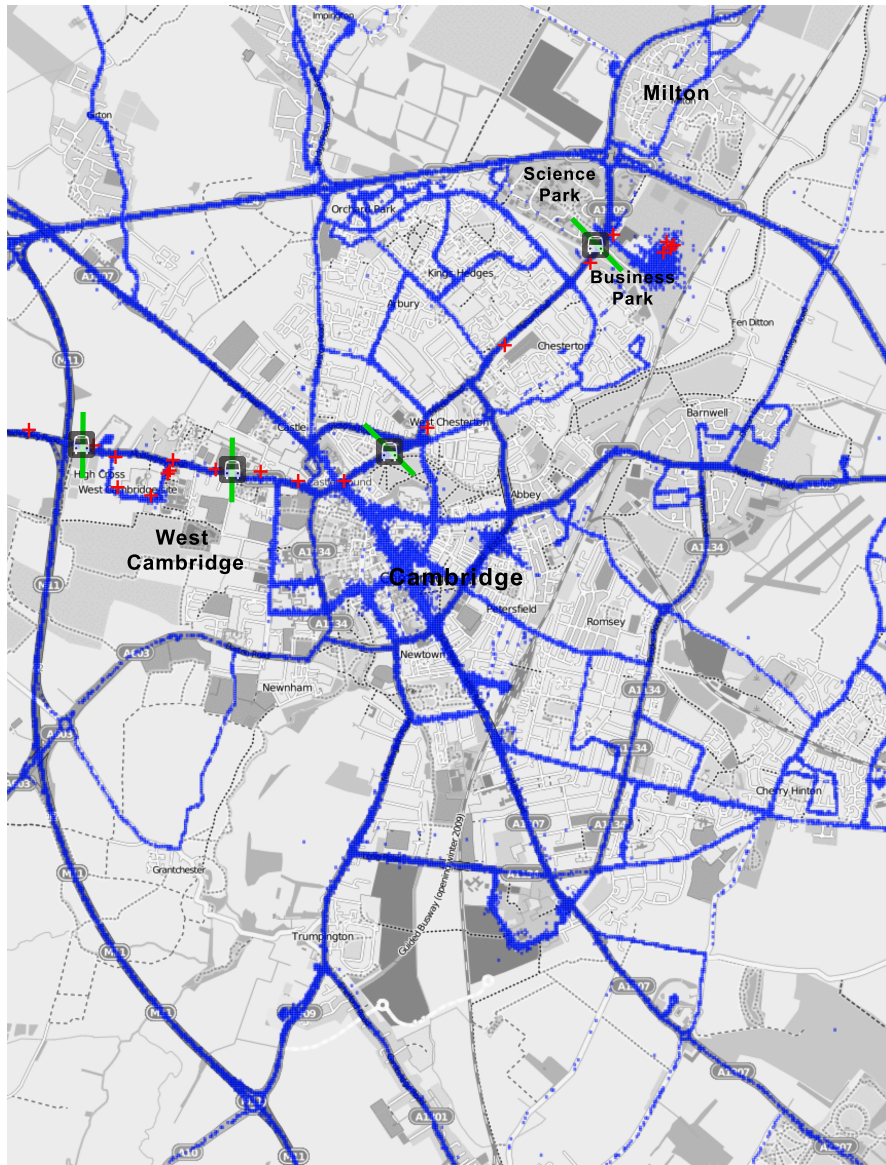
Fig. 6. The map in this figure shows the city of Cambridge, UK overlaid by bus coverage. A pixel in this rasterised map is marked blue if it includes at least one probe from February 2009. The map includes an example of a bus journey (denoted by red crosses) and bus stop gates (shown as green segments), used to extract journeys from the bus data as explained in Sect. 4.2. © OpenStreetMap (`http://www.openstreetmap.org/`) and contributors, CC-BY-SA (`http://www.creativecommons.org`).

3.2 OpenStreetMap

Road Network Description and Features. We made extensive use of OpenStreetMap (OSM) to obtain information about the road network and some of its infrastructure, including locations of junctions and traffic lights, and for visualisations. OpenStreetMap is a collaborative project with the aim of creating a free, public map of the world. The content is contributed voluntarily, using geographic data from portable GPS devices, aerial photographs, and other free sources [24]. The quality of this volunteered information has become increasingly high and is comparable to commercially produced maps [25].

Bus Stops and Bus Route Information. Our analyses require knowing for each bus stop its location and the direction of bus travel past it (the stop's "bearing"). Although some information about locations of bus stops can be obtained from OSM, our primary source was the UK National Public Transport Access Node (NaPTAN) database, an official list of public transport access points in the UK. The UK Department for Transport has offered to make this database available to the OpenStreetMap project and as of this writing incorporation is in progress.[2]

We use stop location and bearing information to construct a line segment for each stop, perpendicular to buses' travel, called a "bus stop gate." A bus crossing is considered to have gone past the stop. A sequence of these gates defines a bus journey [18]. Fig. 6 includes a trace of a single bus' location samples (shown as red crosses) connecting West Cambridge with the Business Park in the northeast of the city. Green lines show bus stop gates for a series of stops along the way and, with corresponding bearings, define the route taken by the bus. These gates can be used to extract the depicted journey (and any other such journey) from the entire dataset of sparse bus data. Section 4.2 contains an explanation of how this is done.

For map-matching we used knowledge of the sequence of roads which buses follow while assigned to a certain route. This was initially derived using timetable information from the bus operator supplemented by bus probe data. However, at later stages of our study we found that, for Cambridge, this information was sufficiently complete within OSM. (Graphical representation of bus routes that are described in the OSM data can be accessed at `http://www.openbusmap.org/`.)

3.3 High Resolution Probe Data

In order to validate our approach for determining bus behaviour between sampled locations, we collected a set of high resolution ("HighRes") traces by carrying a portable GPS device on board the buses. The traces recorded the bus location once per second with an accuracy of 5m to 15m.

---

[2] `http://wiki.openstreetmap.org/wiki/NaPTAN/Import`

4   Selected Applications

Now that we have set the scene by describing how phenomena are sensed, the data that result, and other sources of useful information, we shall detail three applications. The first combines data from junctions controlled by SCOOT with bus probes to evaluate crudely the efficacy of giving buses priority at lights. This application illustrates the use of the SBUS middleware to combine data from different sources. The second is an examination of the statistical analysis of the bus probe data, archived since 2007, needed to determine journey times. In so doing, we see that contextual factors (in this case whether schools are in or between terms) have a significant effect. Finally, our third application is an illustration of how congestion builds up after incidents (for Cambridge, the closure of the nearby A14 trunk road following an accident is commonplace). We show this by evaluating velocity fields, which describe traffic dynamics in space and time.

We have developed other applications and these are described on the project web site.[3] For example, a service for mobile phones, Minibus, displays bus arrival times at selected stops.

4.1   Buses at Traffic Lights

Suppose that a city is considering giving buses priority at traffic lights by adjusting the lights' timing so that late-running buses are, where possible, presented with green lights. This has the potential to decrease bus journey times but may annoy other users of the road. Evaluating a proposal for such a priority scheme requires quantifying the expected improvement in journey times and evaluating the annoyance. The application we have built provides ingredients to help by estimating how many buses wait at red lights. If few buses wait, then stopping for red lights has little affect on bus journey times and so giving buses priority is unlikely to help.

In order to build the application, we need knowledge of bus positions and of traffic light status. In addition, if the application is to operate in real time and show the current situation, we need continuous feeds of this information. Both are available to us, as described in Sects. 2.1 and 2.2.

Two aspects of the problem in particular make the application interesting.

1. Different organisations are in charge of buses and traffic lights. This means that assessing any interaction between the two requires information from multiple sources; implementing the application would be impossible using any one organisation's data.
2. The application does not need to know how the buses determine their positions or what other purposes these data might have. Similarly, there is no need for knowledge of how traffic lights are scheduled. In fact, the application doesn't need to know that buses are involved at all! It only requires locations for a set of objects and notification of when lights change colour.

---

[3] http://www.cl.cam.ac.uk/research/time

The architecture of our application is shown in Fig. 7. As each bus moves around the city, it creates probes as described in Sect. 3.1. This amounts to a continuous stream of events where each reports the location of a particular bus. Each event includes, alongside a location, the vehicle ID (VID) of the relevant bus and the time of the reading. A component labelled "location storage" in the diagram is responsible for keeping track of the current location of each of a set of objects, each identified by name and a timestamp indicating when the location was recorded. These {`name, time, location`} records are stored in a relational database to be used by other applications. By using a bus's VID as its name and doing some data format conversion, we connect these components to end up with a database (the "location database" in Fig. 7) holding the last-known position of each bus.
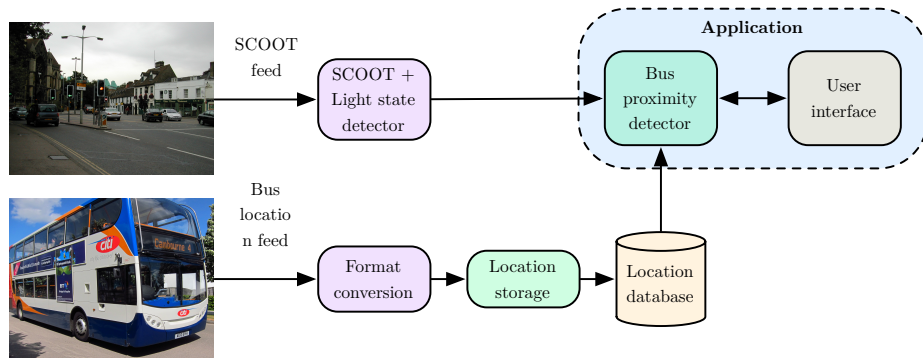

Fig. 7. This figure shows the architecture of the "bus traffic light" application, illustrating how the logic is insulated from the sources of data.

At the same time, a component (the "SCOOT + light state detector") publishes a stream of events describing the changing of the traffic lights controlled by SCOOT. This component will publish an event whenever a traffic light changes state (i.e., from green to red or from red to green—amber is handled by hardware within each traffic light and is not controlled by SCOOT).[4] The "SCOOT + light state detector" also supports a request/response mechanism to retrieve information, such as description and the location, about a given junctions.

The components we have described so far have no semantic connection with finding buses that are near red lights. We have used these source-specific components for applications that are entirely unrelated, such as visualisation of bus movements and queues at junctions. The "bus proximity detector" uses these components to build the application we want and works as follows. It subscribes to the stream of events that indicate changes in the state of traffic lights. When

---

[4] The component publishes other events describing SCOOT's internal state but those are unused by this application.

a light changes from green to red, the location of the light's junction is retrieved using the "SCOOT + light state detector" component's request/response facility. The "location database" is then consulted to find all buses whose last reported position is within a configurable radius of the junction. Events describing these buses are published and a user interface allows viewing these using a standard web browser.

A screen shot can be seen in Fig. 8. The table on the left lists the VIDs of buses that are near lights that have turned red; the column "junction and link" uses SCOOT names to identify the particular light affecting the bus. On the right is a map showing the state of a junction's lights and estimates of the traffic queues at that junction derived from the internal SCOOT state published by the "SCOOT + light state detector" component. This screen shot shows junction 3131, whose lights have affected three buses.

Buses waiting at red traffic lights in Cambridge

Vehicle ID	Junction and link
303	3111G
303	3111I
341	3111C
341	3111G
341	3111I
3045	3111G
3045	3111I
3052	3131E
3052	3131B
3055	3161E
3058	3161E
3060	3131B
3061	3111G
3061	3111I
3081	3131B

Fig. 8. This figure shows a screen shot of the "bus traffic light" application. The table showing buses near red lights and the map illustrating queues of traffic at a selected junction (here junction 3131) are updated in real time and reflect current conditions. Imagery ©2011 DigitalGlobe, Infoterra Ltd & Bluesky, GeoEye, Getmapping plc, The GeoInformation Group. Map data ©2011 Tele Atlas.

4.2 Bus Data Journey Time Estimates

In our data analysis we investigated how to efficiently extract bus journeys corresponding to a route of interest and evaluate their duration taking into account the sparseness of the data both in time and space. We also studied basic statistical characteristics of journey times and showed the effect that the day of the week and time of year have on these journey times and their variability. Below we provide a brief account of this study and some examples, but have published further details [18].

Extracting Journeys and Estimating their Durations. Let r_{AB} be a predefined route connecting two points A and B on the road network. We are interested in extracting bus journeys connecting A and B following the route r_{AB} taking into account that consecutive bus locations were recorded every 20, 30, or more seconds apart. In order to do this, we use information about locations and bearings of bus stops taken from the NaPTAN database, where locations are latitude/longitude coordinates and bearings are members of the set $\{\texttt{N}, \texttt{NW}, \texttt{W}, \texttt{SW}, \texttt{S}, \texttt{SE}, \texttt{E}, \texttt{NE}\}$. We use this information to approximate r_{AB} by a sequence of bus stops and bus stop gates as introduced in Sect. 3.2. We then find journeys corresponding to the route r_{AB} by looking at those piecewise linear path approximations to the bus trajectories which intersect all the bus gates which approximate r_{AB}. For example, we have already explained that Fig. 6 shows a set of bus stops and gates which define a route connecting West Cambridge and the Business Park. Here the orientations of the four bus gates are $\texttt{E}, \texttt{E}, \texttt{NE}$, and $\texttt{NE}$. Consequently, bus trajectories highlighted by red crosses will be picked up as following this route.

The journey time for such a journey is the difference in time between the bus' crossing of the first and last gates. Because it is unlikely that probes were recorded for those positions precisely, we obtain the last probe before and the first probe after the bus follows r_{AB} and compute an adjustment based on the time before the bus crossed the first bus stop gate and after it crossed the last.

Factors Affecting Journey Times. Fig. 9 illustrates journey times for the southbound route towards the city centre along Histon Road during the period 2 November 2008 to 9 January 2010. Each individual black dot refers to a single journey. A quantile regression model was fitted to the data [18]; the red ribbon overlaid on the data points shows the variation in the 10th and 90th percentiles whereas the solid red line indicates the variation in the median journey time (50th percentile). It is clear that the variation in journey time depends on the start time of the journey: there is a morning busy period between 07:30 and about 10:00 where median journey times increase from 7 minutes to over 10 minutes. An even more pronounced increase affects extreme journey times, with the 90th percentile increasing from under 10 minutes to well over 15 minutes.

Fig. 10 describes the same variation using quantile regression lines between the 5th and 95th percentiles (in the left plot) and a grey-scale heatmap of conditional densities (in the right plot). The conditional densities were obtained by
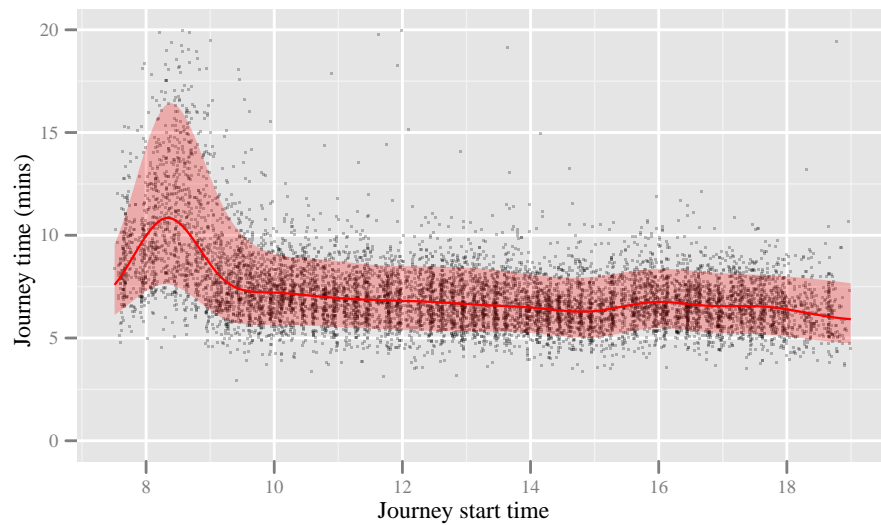
Fig. 9. This figure shows the variation of individual journey times (as black dots) for journeys on week days according to the start time. The bus route is southbound towards the city centre along Histon Road with data between November 2008 and January 2010. Overlaid is a red ribbon showing the fitted quantile regression model for the 10th and 90th percentiles. The red line within the ribbon shows the quantile regression model for the median (50th percentile) journey time. Notice how median journey times increase during the morning busy period coinciding with an even larger increase in the spread of journey times as illustrated by the width of the ribbon.

re-sampling from the corresponding conditional cumulative distribution function for fixed start times combined with kernel density estimation. Examples of the conditional distribution functions and density functions so constructed are given in Fig. 11.

Fig. 10. The left plot in this figure shows many quantile regression lines, from the 5th to the 95th percentiles, fitted to the journey times for weekdays. The right plot shows a grey-scale heatmap plot of conditional densities derived from the set of quantiles from the quantile regression; each such conditional density corresponds to a fixed time at which travel has been made and is obtained by re-sampling from the corresponding conditional cumulative distribution function with subsequent kernel density estimation.

Fig. 12 provides further insight into the variation of journey times by categorising journeys on individual days in 2009 according to the day of the week (including weekends) and whether the journeys start in the morning or in the afternoon. Here we show the proportion of journey times exceeding a fixed threshold (which in this instance is 8 minutes). Many other metrics could be displayed in this intuitive calendar style.

We now explore how the context of a journey can affect its duration. Fig. 13 shows journey times divided into two categories. One category consists of journeys undertaken on school days, shown on the left, while the second category consists of non-school days (including bank holidays), shown on the right. Journeys made on school days take longer compared with those made on non-school days: median journey times on school days rise to a level only reached by the 90th percentile of journey times on non-school days. The 90th percentile of journey times on school days exceeds that on no-school days by over 5 minutes.

4.3 Velocity Fields and Congestion Spreading After Incidents

Our third sample application shows how bus probe data can be used to visualise the effects of an incident. This is possible because each bus travels through traffic. If the traffic moves quickly then the bus will move quickly and if the traffic's
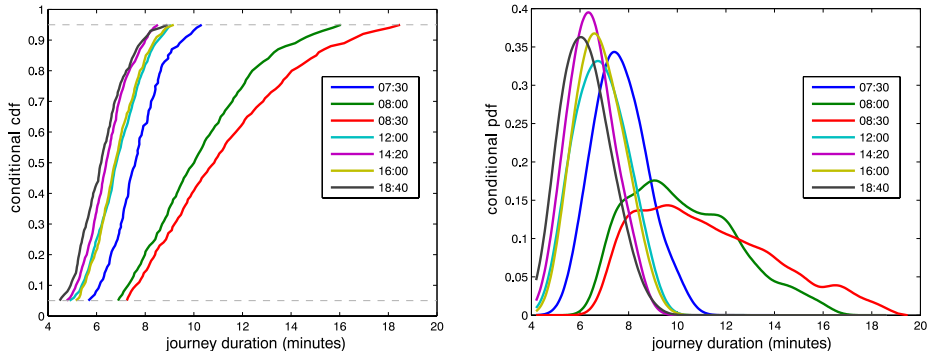
Fig. 11. This figure shows conditional cumulative distribution functions (left) and density distribution functions (right) reconstructed from the quantile regression discussed in the text.

progress is impeded, the bus will move more slowly. This means that bus probe data can be used to characterise the behaviour of traffic on the road network.

Clearly, slowdowns can be consequences of many factors, including road works, accidents, emergency service activities, and, importantly, congestion. In addition, buses spend time at bus stops and may use roads in a unique way (e.g., roads may have bus lanes and traffic lights may be programmed to favour buses). We have investigated how to infer traffic conditions using bus speeds recovered from sparse bus probes, taking into account specific bus behaviour and road infrastructure usage [19]. Our goal is the approximation of the traffic velocity field, as it gives continuous time-space monitoring of traffic dynamics. The field is constructed by mapping recovered bus speed profiles onto the linear segment representation of the road network. The speed profiles are adjusted where necessary, entailing (i) making a specific choice of the transformation used to map the profiles onto the road network (this transformation can be thought of as a way of summarising observed buses' behaviour) and (ii) accounting for the uniqueness of bus behaviour as discussed above.

We recover speed profiles from probe data as follows; Fig. 14 provides an illustration. We first apply map matching to a bus trajectory and perform monotonic spline interpolation to the cumulative distances travelled (step A); red stars in the upper panel correspond to probes and the solid curve is the fitted monotonic spline. Differentiating this spline we derive a continuous approximation to the bus speed (steps B and C). Shown as a black solid curve in the middle panel, this is a good approximation to a smoothed version of the real speed profile, shown as a thin grey line in the same panel, as taken from "HiRes" traces. We then detect behaviour of the bus in the proximity of a bus stop using the concept of a "local time" (step D). The local time profile corresponding to a bus trajectory characterises the progression of the bus along its route and measures how long that bus spent in the small neighbourhood of each point. When plotted, as in the bottom panel of Fig. 14, the spikes indicate locations where the bus
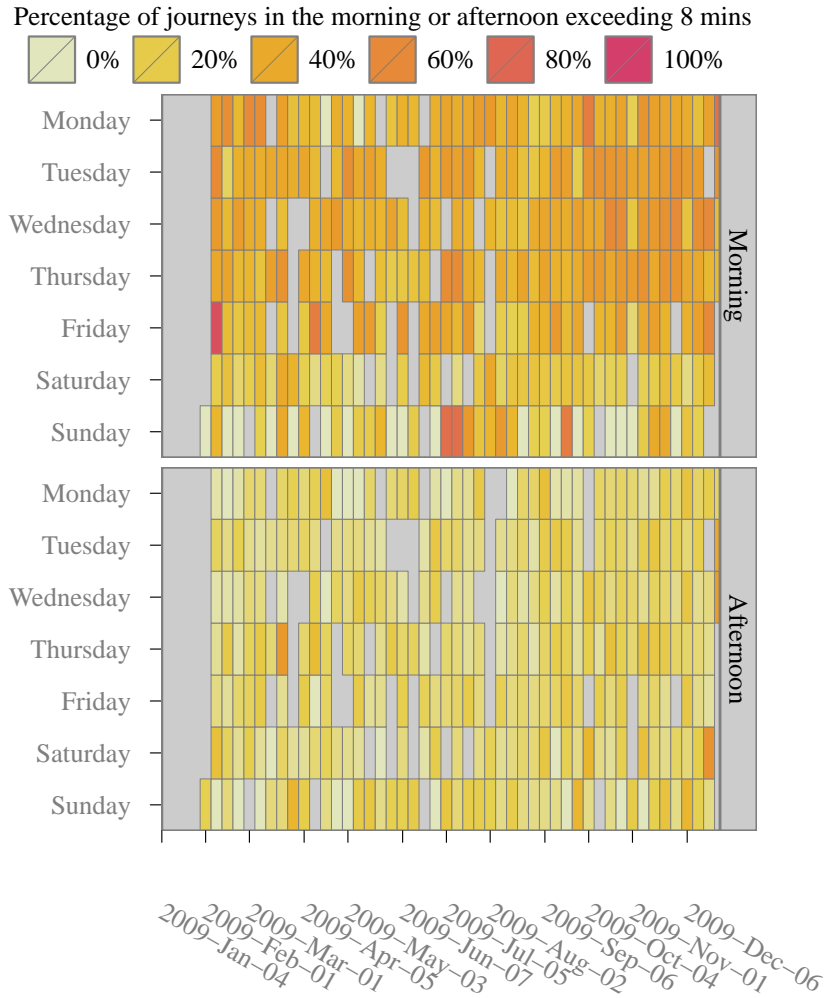
Fig. 12. This figure shows the variation of journey times taken from Fig. 9 according to day of the week, whether morning or afternoon and for individual days throughout the one year period of 2009. The colours indicate the percentage of journeys exceeding 8 minutes with the background grey colour used for missing data.
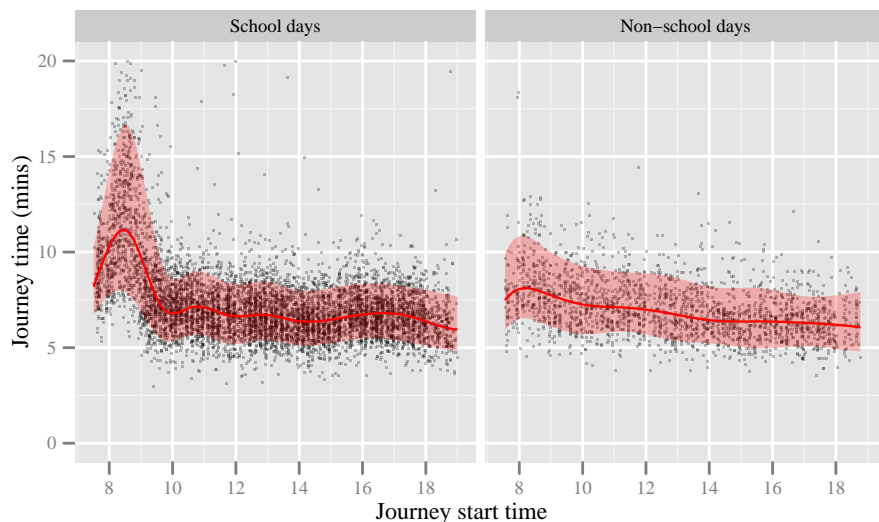
Fig. 13. This figure shows the journey times from Fig. 9 in two separate panels according to whether on the day schools were open or closed. We can see that much of the adverse impact on journey times during the morning busy period occurs on school days. Non-school days show much milder effects in comparison.

spent relatively longer times, whereas long and deep valleys identify parts of the journey that were passed with higher speed.

After processing a set of trajectories, we apply a velocity field transformation to the restored speed profiles (the result of the differentiation in step C) to map speeds back to the road network (step E). The utility of these velocity fields can be seen by an example. In the early morning of Tuesday 22 June 2010 a lorry caught fire on the A14 main road north-west of the city. Major disruption followed the incident with adverse consequences across Cambridge. Fig. 15 shows traffic counts taken using the lamp post sensor described in Sect. 2.1. The volume of eastbound traffic on 22 June 2010 is very different compared with that on the surrounding Tuesdays, 15 and 29 June 2010. Fig. 16 shows the velocity field mapped to the road network based on snapshots taken at 10:00 on 15, 22, and 29 June 2010. On the day of the lorry fire, severe disruption has spread widely across the city.

5 Conclusions and Future Directions

Trunk roads and city ring roads are becoming well-instrumented, the recorded data presented to road users by displays *in situ* and via web-based applications showing estimated journey times and congestion. However, information on urban
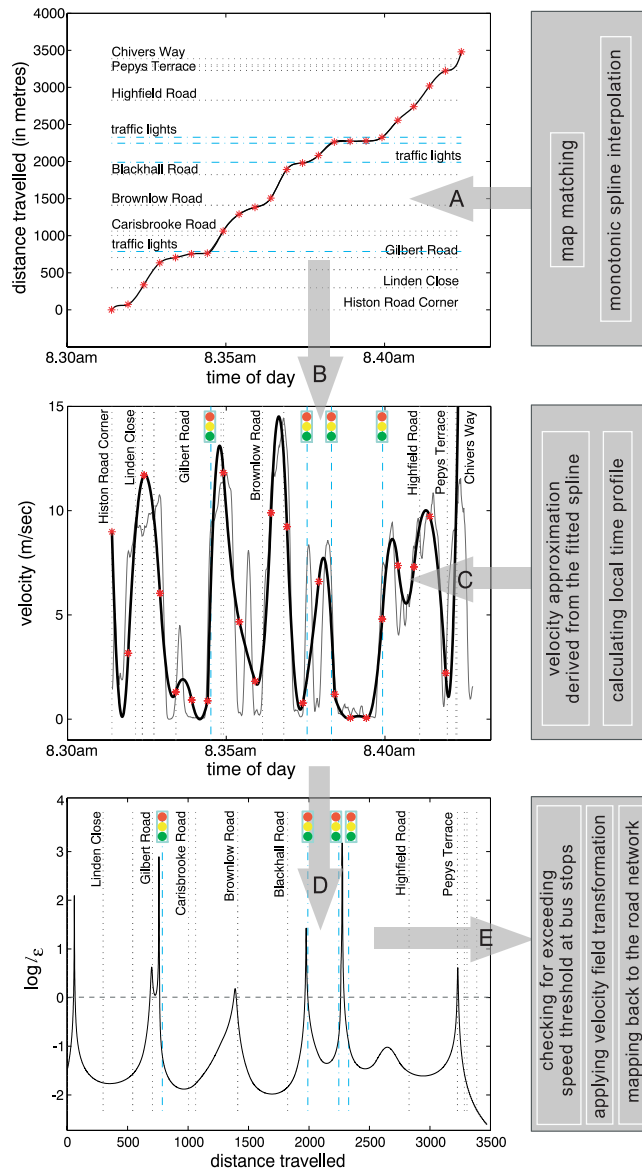
Fig. 14. This figure represents a diagram showing how speed profiles are restored from sparse bus data and mapped back to the road network. This is shown for the trajectory of a bus journey made along Histon Road in the morning hours on 23 February 2010, outbound. Details on each of the steps in this diagram can be found in the text.
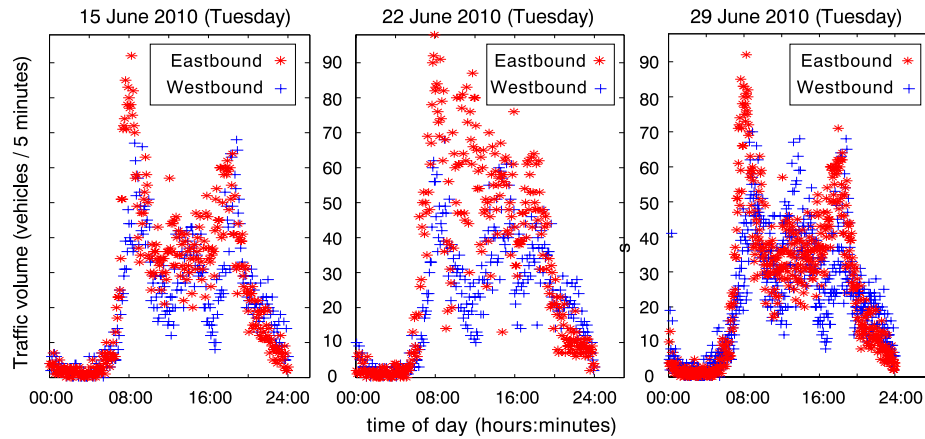
Fig. 15. This figure shows traffic counts recorded by the infra-red camera installed on Madingley Road, Cambridge for the Tuesdays, 15, 22, and 29 June 2010. The vertical axis describes traffic counts measured in vehicles per minute with the horizontal axis giving the time of day. The red and blue markers show the eastbound and westbound traffic, respectively. The figures give an indication of the change in traffic on 22 June 2010.

traffic is more difficult to acquire because of the dense network of roads and junctions that comprise a city. Instrumenting an entire city would be prohibitively expensive and so a limited number of static sensors is typically deployed, for example to assist in scheduling traffic lights. However, this leads to necessarily patchy views of the state of the road network, and data tend to be used in ways that are narrowly defined. For example, in Cambridge car-park occupancy is shown in displays on radial roads, but is not used as an indication of general traffic conditions.

The TIME project has attempted to address this problem of poor data and lack of integration. We began by investigating the range of sensors that could be used for monitoring urban traffic and the communications networks that would be needed to transmit resulting data in real-time. This paper has summarised that work and presented the sorts of data available from these sensors and other sources such as OpenStreetMap. In parallel, we developed a component-based, reconfigurable middleware (SBUS) for linking sensors, data processors, displays, and archival facilities. Its versatility is demonstrated by the ease of development of the application described in Sect. 4.1, where data from different administrative domains are fused to create real-time visualisations. We have described further applications that are made possible by sensing of urban traffic, focussing on real-time position reports by buses. Because buses must serve the entire city, data describing their locations over time is a rich source of information on city-wide traffic conditions. Using such "probe" data, we have estimated journey times by
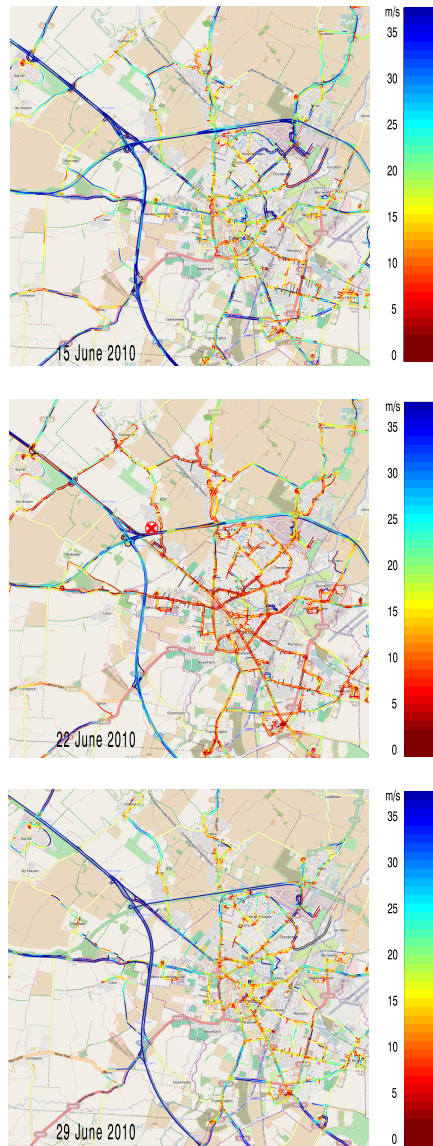
Fig. 16. This figure shows velocity fields corresponding to snapshots taken at 10:00 on the Tuesdays 15, 22, and 29 June 2010. Velocities are measured in metres per second and are indicated by a palette of colours ranging from slow speeds in dark red through to higher speeds in dark blue. The first and last Tuesdays are similar but the middle Tuesday (22 June 2010, the day of a lorry fire on the A14) shows widespread and severe reduction in traffic speeds across the city.

bus, shown how these can be affected by context, and explored the effects of unusual occurrences such as accidents.

Plans for the future include integration of crowd-sourced traffic data, as it is a natural fit for our techniques and is complementary to public transport data. We plan further work on statistical analysis of bus data in response to real-time queries. Our first project is to produce on demand, as annotations to a map of Cambridge, historical journey times for road segments under various conditions. Responding to such a query requires an analysis of context (time of day, wet, dry, summer, winter, school term) and selection of the appropriate historical estimate. Doing this efficiently is challenging given the number of historical probes that are accumulated.

In summary, we have explored a range of traffic monitoring data derived from static and mobile sensors. We believe that combining data types from multiple administrative domains can give as full a picture as is possible; all too often, transport has relied on proprietary applications with a single purpose. For reasons of cost, and coverage in terms of space and time, public transport data from buses has been our richest source of data. We have made a start on analysing these data, and we believe that many future projects could be based on them.

References

1. Devereux, R., Dawson, J., Dix, M., Hazel, G., Holmes, D., Glaister, S., Joseph, S., Macgowan, C.T., Nimick, B., Roberts, M., Searles, L., Turner, R., Gooding, S., Hickey, S., Rickett, W.: Feasibility study of road pricing in the UK. Technical report, Department for Transport (July 2004)
2. Boucher, A., Bowman, J., Noble, B., Hird, D., Lloyd, D., Rose, W., Shah, H., Dennis, S., Lees, A., Grieve, I., Riley, S., Robinson, D., Salathiel, D., Crane, E., Laffan, W., Medhurst, C., Beg, O., Coleman, B.: Transport statistics bulletin road statistics 2008: Traffic speeds and congestion. Technical report, Department for Transport (2009)
3. Kenyon, S., Lyons, G.: The value of integrated multimodal traveller information and its potential contribution to modal change. Transportation Research Part F: Traffic Psychology and Behaviour 6(1) (March 2003) 1–21

4. Ingram, D.: Reconfigurable middleware for high availability sensor systems. In: Proceedings of the Third International Conference on Distributed Event-Based Systems (DEBS'09). (2009)

5. Hunt, P.B., Robertson, D.I., Bretherton, R.D., Winton, R.I.: SCOOT—a traffic responsive method of coordinating signals. Technical Report LR1014, Transport and Road Research Laboratory (1981)

6. Ehmke, J.F., Meisel, S., Mattfeld, D.C.: Floating Car Data Based Analysis of Urban Travel Times for the Provision of Traffic Quality. In: Traffic Data Collection and its Standardization. Volume 144 of International Series in operations Research & Management Science. Springer Science+Business Media (2010) 129–149

7. Rose, G.: Mobile phones as traffic probes: Practices, prospects and issues. Transport Reviews 26(3) (2006) 275–291

8. Naranjo, J.E., F. Jiménez, F.J.S., Zato, J.G.: Comparison between floating car data and infrastructure sensors for traffic speed estimation. In: IEEE ITSC2010 Workshop on Emergent Cooperative Technologies in Intelligent Transportation Systems, Madeira Island, Portugal, IEEE Intelligent Transportation Systems Society (September 2010)

9. Google: Arterial traffic available on Google. `http://google-latlong.blogspot.com/2009/08/arterial-traffic-available-on-google.html` (August 2009)

10. Herrera, J.C., Work, D.B., Herring, R., Ban, X.J., Jacobson, Q., Bayen, A.M.: Evaluation of traffic data obtained via GPS-enabled mobile phones: The *Mobile Century* field experiment. Transportation Research Part C (18) (2010) 568–583

11. Herring, R., Hofleitner, A., Abbeel, P., Bayen, A.: Estimating arterial traffic conditions using sparse probe data. In: 13th International IEEE Conference on Intelligent Transportation Systems, Madeira Island, Portugal, IEEE Intelligent Transportation Systems Society (September 2010) 929—936

12. Braxmeier, H., Schmidt, V., Spodarev, E.: Kriged road-traffic maps. In: Interfacing Geostatistics and GIS. Springer (2009) 105–119

13. Knoop, L., Eames, T.: Public transport technology in the United Kingdom: Annual survey 2008. Technical report, RTIG Ltd (February 2009)

14. R.L.Bertini, S.Tantiyanugulchai: Transit buses as traffic probes: Use of geolocation data for empirical evaluation. Transportation Research Record (1870) (2004) 35–45

15. P.Chakroborty, S.Kikuchi: Using bus travel time data to estimate travel times on urban corridors. Transportation Research Record (1870) (2004) 18–25

16. Pu, W., Lin, J.: Urban travel time estimation using real time bus tracking data. In: Transport Chicago 2008, Chicago, Illinois, USA (2008)

17. Pu, W., Lin, J., Lon, L.: Real-time estimation of urban street segment travel time using buses as speed probe. Transportation Research Record (2129) (2009) 81–89

18. Bejan, A., Gibbens, R., Evans, D., Beresford, A., Bacon, J., Friday, A.: Statistical modelling and analysis of sparse bus probe data in urban areas. In: 13th International IEEE Conference on Intelligent Transportation Systems, Madeira Island, Portugal, IEEE Intelligent Transportation Systems Society (September 2010) 1256–1263

19. Bejan, A., Gibbens, R.: Evaluation of velocity fields via sparse bus probe data in urban areas. Submitted for publication (2011)

20. Antonios Skordylis, N.T.: Delay-bounded routing in vehicular ad-hoc networks. In: ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). (2008)

21. Skordylis, A., Trigoni, N.: Jointly optimizing data acquisition and delivery in traffic monitoring VANETs. In: 24th Annual ACM Symposium on Applied Computing (ACM SAC'09). (March 2009)

22. Skordylis, A., Trigoni, N.: Efficient data propagation in traffic monitoring vehicular networks. IEEE Transactions on Intelligent Transportation Systems (to appear) (2011)
23. Ordnance Survey: A guide to coordinate systems in Great Britain. `http://www.ordnancesurvey.co.uk/oswebsite/gps/docs/A_Guide_to_Coordinate_Systems_in_Great_Britain.pdf`
24. Haklay, M., Weber, P.: OpenStreetMap: user-generated street maps. IEEE Pervasive Computing 7(4) (2008) 12–18
25. Haklay, M.: How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and Planning B: Planning and Design 37(4) (2010) 682–703