Evaluation of Velocity Fields via Sparse Bus Probe Data in Urban Areas

Andrei Iu. Bejan and Richard J. Gibbens
Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
firstname.lastname@cl.cam.ac.uk

Abstract—The goal of this paper is to investigate how sparse public transport data such as bus data might be used in evaluating congestion in urban areas and in providing better information for road users and traffic managers. With this aim we use the well known concept of a velocity field and, building on our previous work, we specifically show how velocity fields can be usefully reconstructed from sparse bus data. Buses provide good coverage of cities and, increasingly, are being equipped with satellite navigation devices and monitored in order to display their predicted arrival times to bus passengers. We have had access to a wealth of bus data for the city of Cambridge, England for some years. We describe these bus data and show how its specifics may be used in conjunction with other sources of data, such as OpenStreetMap, in order to reconstruct information on transport traffic flow dynamics. Finally, we consider examples in which we compare our knowledge of the traffic regimes with the outputs of the technique presented.

I. INTRODUCTION

It is widely recognised that congestion on both urban and freeway road networks has a huge negative social and economic impact on the community [1], [2], as well as on the environment [3], [4]. This problem is unlikely to be efficiently approached without significant efforts: the distance driven in the UK by private motor vehicles has been increasing almost linearly from just over 50 billion vehicle kilometres in 1950 to over 500 billion vehicle kilometres in 2008 [5].

Congestion can be tackled either by increasing road network capacity or by decreasing demand. Increasing capacity is very difficult, especially in urban environments, whereas demand can be reduced, at least during peak hours, by encouraging drivers to, for example, switch to using public transport, alter the time of day at which they travel, or perhaps consider avoiding travelling. Fortunately, the provision of better information about likely costs, travel times and traffic speed and its intensity. While such a static infrastructure may be suitable for a freeway environment, it is of limited use in urban settings—traffic flows in urban road networks are highly fluctuating across different network links and over times of the day, and thus area-wide traffic data collection is necessary [10]. In addition, providing comprehensive static sensor-based coverage of a city would be prohibitively expensive for local authorities.

An alternative approach to these conventional data collection methods is to use floating probe data from a sensor, such as a GPS device, attached to a vehicle or person. Probe data consist of a sequence of coordinates recorded over time and may deliver much more information than is typically available from fixed sensors [11], [12]. Recently, Google’s “Maps for Mobile” combined location data crowdsourced from GPS-equipped mobile phones with traditional sensor infrastructure to overlay road maps with congestion information on arterial roads [13] and the Mobile Millenium Project’s Mobile Century field experiment demonstrated the feasibility of traffic monitoring systems based on mobile phone probes [14]. Researchers also explored the feasibility of using commercial transport fleets of vehicles, such as taxis, in estimating traffic conditions on arterial roads—e.g. Herring et al. [15] analysed arterial travel time distributions using a Hidden Markov Model technique applied to sparsely observed taxi vehicles and Braxmeier et al. [16] 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 are now augmented with automated vehicle location (AVL) systems which use GPS to collect probe data [17] and support Real-Time Information (RTI) systems and the potential of using buses as probes has been studied. For example, Bertini and Tantiyanugulchai [18] and Chakroborty and Kikuchi [19] studied transit buses as traffic probes by investigating to what extent travel characteristics of bus vehicles are related to those of general traffic. Pu and Lin [20] and Pu et al. [21] 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 showed that using buses as probes to detect general vehicle traffic conditions could form a real-time traffic monitoring mechanism in an urban advanced
traveller information system. Notably, in all these test 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.

In this paper we build on our recent work in which we showed how to recover speed information from sparse bus movement data [22]. Our technique uses buses as floating probes to reconstruct dynamics of time-space traffic velocity fields in urban areas. Historical and real-time location data collected from a fleet of moving vehicles, together with an open source map, bus stop and bus routes data, give us a rich picture of traffic patterns and congestion with city-wide coverage at no extra expense. Thus our contribution is two-fold: (i) we show how to infer traffic conditions using bus speeds recovered from sparse bus data and taking into account specifics of buses’ behaviour and road infrastructure usage, and (ii) we demonstrate that with the exception of bus data all other necessary and relevant information can be usefully extracted from an open map, such as OpenStreetMap (OSM) [23]. This paper also extends our earlier work [22] in characterising traffic from a local “per route” description to a global “network-wide” one.

II. USING BUS DATA: PROS AND CONS

Public transport location data recorded over long periods of time as well as “live” feeds containing such data are increasingly becoming available. Over the last decade, the bus fleet in the UK public transport sector has become equipped with Real Time Information (RTI) systems. For example, it was reported in the recent UK Public Transport Technology Survey [17], based on data received from local authorities in England, Wales, Scotland and Northern Ireland in 2008, that 56 Local Authorities (compared with 37 in 2007) had operational RTI systems covering approximately 185 towns and cities each with a population over 10,000.

![A bus in the flow of vehicles, Madingley Road, Cambridge, England.](image)

Local authorities as well as bus operators make use of services delivered by providers of real time location data in order to more effectively manage the fleet of vehicle units in use and deliver to passengers real time information about operating buses. We believe that the collected bus location data can be used successfully for the additional purpose of better understanding traffic and its dynamics on urban and suburban roads.

One advantage of using buses as probes is that buses can give good coverage of the main city roads, including arteries [22]. It is also advantageous that this kind of data is generally available throughout the day and evening, that is when traffic conditions are of most interest.

There are several drawbacks with using location data from buses alone for estimating the dynamics of traffic progression more generally:

- **Buses spend time at bus stops.** A typical concern in this regard therefore is how to efficiently detect and evaluate bus stop dwelling times and how to tell whether an observed slowdown near a bus stop was actually due to stopping there and not due to traffic. To mitigate or remove these effects we analyse the behaviour of buses in vicinities of bus stops and adjust the mapping between bus speeds and the road network accordingly.
- **Roads may be used in a unique way by buses.** For example, roads may have bus lanes and traffic lights may be programmed to favour buses. Furthermore, in places such as central areas of some towns, like Cambridge, buses and taxis may be permitted while other vehicles are barred by barrier or bollard control systems. Although these factors represent a serious limitation in using buses as probes in assessing traffic conditions in certain areas of a road or city, we believe that detailed knowledge about the specifics of road infrastructure and its use by buses and other road users greatly helps in analysing and mitigating associated effects. In particular, the OSM data set has a great potential in extracting such knowledge.

III. DATA DESCRIPTION

A. Description of bus probe data

The bus location data used in this study was provided by the company supplying RTI support to the largest bus operator in the city of Cambridge. In 2009 there were on average 115 buses equipped with GPS units on Cambridge roads on weekdays, 100 buses on Saturdays, and 65 buses on Sundays. In April 2010 the total number of buses equipped with GPS increased by approximately 50 vehicles.

The data represent a set of bus location points recorded over a period of 4 years (2007-2010) for these buses. We estimate the accuracy of these data as ±30m. The location of each bus was logged once every 20 or 30 seconds (more rarely, once every 60 or 120 seconds); furthermore, occasionally, buses were “silent” for much longer periods of time. We did not possess information on the routes which buses were taking at any moment in time.

In Figure 2 we present the positions of all buses running throughout one day (Tuesday, 15 June 2010) along
Madingley Road, Cambridge snapped to this road’s linear representation (the horizontal axis) obtained from OpenStreetMap data [23], and produce a histogram of the snapped sampling points ignoring the time of the day when the observations were made. As would be expected, although buses record their positions at a relatively low rate, the spread and concentration of sampling points identifies places where buses spend relatively more time. Besides bus stops such places are: traffic lights at the M11/Madingley Road junction; the junction at Madingley P&R; near J.J. Thomson Avenue; traffic lights at the junction with Grange Road; traffic lights near Lady Margaret Road; and the Queens Road roundabout.

B. Other sources of data

The road network description and its features, information about bus stops and local bus routes were extracted from OSM data [23]. We also used vehicle count data from our infra-red camera detector installed on one of Cambridge’s radial roads [24]. (The position of the camera is depicted in Figure 2.) Using this camera detector we can estimate traffic intensity, but not speeds of individual vehicles.

IV. EVALUATING VELOCITY FIELDS USING BUS DATA

A. Buses as particles in the traffic flow

First we use the bus location data to characterise the behaviour of traffic on the road network. Because each bus travels through traffic, if the traffic moves quickly then the bus will move quickly. If the traffic’s progress is impeded, the bus will move more slowly. Clearly such slowdowns can have many causes, including road works, accidents, emergency service activities, and congestion.

Figure 3 shows the positions of individual buses over time as the buses travel eastwards along Madingley Road, Cambridge, in the early morning hours. The direction of travel is towards the city centre (a detailed schematic plot of Madingley Road is presented in Figure 2). The upper panel shows buses on the morning of Tuesday, 15 June 2010 while the lower panel is for one week later on 22 June 2010. Each line corresponds to the trajectory of a single bus. The buses begin their journeys far from the city centre (at about 2200 m) before then travelling towards the centre. Not all buses traverse the entire length of the road, which is why different lines start and end at different distances. The slope of each line is related to the speed of the corresponding bus. If the line is near to vertical, travel is rapid but as it becomes near to horizontal, travel slows to a stop. Importantly, however, such a piece-wise linear approximation of bus trajectories is only a rough approximation to the true movement process.

The upper panel of Figure 3 shows a typical day with a pattern of rapid travel preceding a period of slower movement as the bus progresses towards the centre of the city. The degree of slowing down and the location in the journey where this phase begins vary with the time of day. Early in the morning, before 07:30, there is minimal slowing at the end of the journey but a little later between 07:30 and 08:30 journeys slow down further from the centre and remain slow until arriving at the centre. Beyond 08:30 journeys revert to the behaviour prior to 07:30 with little impact of congestion.

The lower panel of Figure 3 shows the severe impact of a lorry fire early in the morning on the A14 bypass to the north-west of the city (the explosion and lorry fire took place

Fig. 2. The upper panel of this figure plots a histogram of occurrences of GPS data readings grouped by location along Madingley Road, Cambridge on Tuesday, 15 June 2010, a school day. Since data readings are logged at regular times, the histogram represents the distribution of speed along the road. We can expect high occurrences of readings near to locations where speeds are often low (such as in the vicinity of road junctions or bus stops). Conversely, the occurrences of readings will be low where buses frequently travel at their highest speeds. The lower panel is a map of Madingley Road showing the alignment of various relevant features with the histogram above.
at 5 am, but as a consequence of this accident the eastbound carriageway was shut and not reopened until 3 pm; see [25] for a news report of this accident and other, separate and less serious, road accidents which happened in and around Cambridge on that day and added to the severe disruption on the city roads. Major congestion followed the accident across the city of Cambridge. Here we see that many buses travelling on Madingley Road throughout the morning period are severely delayed. In later sections we will explore the impact, in terms of speeds and travel times, of this incident on roads right across the city and not just on a single road.

The bus data alone do not explain why the progress of eastbound buses is impeded between about 07:45 and 09:00 on Tuesday 15 June 2010 (and this is typical behaviour for weekdays suggested by experience and confirmed by the data), or why there is an uncommon shift of severe slow downs towards late morning time a week later, on Tuesday 22 June 2010. For comparison we used another source of data to measure traffic volume. We used an infra-red sensor capable of counting vehicles installed on a lamp post on Madingley Road [24] (its position can be seen in Figure 2). Figure 4 depicts two plots of vehicle counts versus time of day for 15 June 2010 (left) and 22 June 2010 (right). While the dynamics of the rates of vehicles travelling eastbound on these two days match relatively well until 09:00 and after 16:30, they exhibit highly different patterns between these times with highly volatile and uncommon behaviour for 22 June. The rise in traffic volume accompanied by the reduction in bus speed seen during corresponding periods of time suggests that congestion is the cause of these changes in traffic behaviour. (Notably, the dynamics of the rates of vehicles travelling westbound on these two days match relatively well throughout the entire day time period.)

**B. Velocity Field**

Generally, it is not straightforward to give an exact definition of the velocity field in the context of transport traffic on a road network, especially for the traffic operating in urban environments: analogies with fluids are not always possible for vehicular transport flows, as these may often consist of irregularly moving vehicles with different behavioural properties and varying spacing between them. Furthermore, evaluating velocity fields at any moment in time and space is practically impossible due to the prohibitively high costs of instrumentation of the road infrastructure and tracking of road users. Below we describe an empirical approach to obtaining meaningful approximation as an alternative to this continuous time-space monitoring of traffic using sparse traces of buses spanning the road network.

Our construction of the velocity field makes use of the linear segment representation of the road network \( \mathcal{R} = (N, S) \), where \( N \) is a set of nodes and \( S \) is a set of oriented (one way) and unoriented (both ways) links between them, and is based on the notion of the Average Velocity Table (AVT). This table comprises updates of average speeds corresponding to individual road segments in time and is derived from the raw bus data. The rows of the table are of the form \([\text{seg}_{n}^{(i)}, \text{dir}_{n}^{(i)}, \theta_{n}^{(i)}, \psi_{n}^{(i)}] \), as in Table I, and each such row refers to the \( n \)-th update of the average speed, \( \psi_{n}^{(i)} \), corresponding to the segment \( \text{seg}_{n}^{(i)} \) and made at time \( \theta_{n}^{(i)} \); the direction \( \text{dir}_{n}^{(i)} \) is the direction in which a bus, whose trajectory was used for this update, traversed the road segment \( \text{seg}_{n}^{(i)} \) (index \( i \) here specifies the segment).

<table>
<thead>
<tr>
<th>segment ID</th>
<th>direction</th>
<th>update time</th>
<th>average speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{seg}_{n}^{(i)} )</td>
<td>( \text{dir}_{n}^{(i)} )</td>
<td>( \theta_{n}^{(i)} )</td>
<td>( \psi_{n}^{(i)} )</td>
</tr>
</tbody>
</table>

In order to give further details on the AVT updates we provide a formal description of the bus data. We denote the set of buses by \( \mathcal{B} \) and refer to the elements (vehicle IDs) of \( \mathcal{B} \) using the notation \( \text{vid}^{(k)} \in \mathcal{B} \), where \( k = 1, 2, \ldots, |\mathcal{B}| \). Each bus \( \text{vid}^{(k)} \) updates its location \( x_{i}^{(k)} \) at times \( t_{1}^{(k)} + \tau_{1}^{(k)}, t_{2}^{(k)} + \tau_{2}^{(k)}, \ldots, \) where each \( t_{i}^{(k)} \) is a regular timestamp such that in the long run the sequence \( \tau_{1}^{(k)} := t_{i+1}^{(k)} - t_{i}^{(k)} \) is typically a constant sequence of either 20 or 30 seconds, but can also be a small multiple of 20 or 30 and in instances of silence, it can be a larger quantity. The value \( \tau^{(k)} \) is a small fixed offset unknown to us. Thus, the movement of bus \( \text{vid}^{(k)} \) can be described as a sequence of vectors \( (t_{i}^{(k)}, x_{i}^{(k)}) \) whose components are the timestamps \( t_{i}^{(k)} \) and bus positions \( x_{i}^{(k)} \).

Consider a particular bus \( \text{vid}^{(k)} \in \mathcal{B} \) and one particular path (together with the timestamps) \( (t_{i_{1}}^{(k)}, x_{i_{1}}^{(k)}), \ldots, (t_{i_{n}}^{(k)}, x_{i_{n}}^{(k)}) \). We perform map matching of this path by snapping the bus locations \( x_{i_{1}}^{(k)}, \ldots, x_{i_{n}}^{(k)} \)
to the road network using the fact that the vehicles are buses, and therefore most likely follow bus routes, and accounting for the possibility of ‘false’ snapping to a road at a different altitude. We thus obtain an ordered sequence of snapped points \( w_{i_1}, \ldots, w_{i_m} \) and define numbers \( d_k(t_{i_m}^{(k)}) \), \( m = 1, \ldots, n \), such that \( d_k(t_{i_m}^{(k)}) - d_k(t_{i_m-1}^{(k)}) \) is the length of the shortest path from \( w_{i_{m-1}} \) to \( w_{i_m} \) on \( R \) and \( d_k(t_{i_1}^{(k)}) := 0 \). The discrete values \( d_k(t_{i_1}^{(k)}), \ldots, d_k(t_{i_n}^{(k)}) \) can be seen as evaluations of the function \( d_k(t) \) which measures the cumulative distance travelled by the bus \( v_{\text{id}}^{(k)} \) by time \( t \).

As we showed in our previous work both the cumulative distance \( d_k(t) \) and its derivative can be reasonably well approximated using the interpolation technique applied to sparse bus data [22]. Specifically, we use monotonic cubic spline interpolation to further restore the function \( d_k(t) \) for any \( t \in (t_{i_1}^{(k)}, t_{i_n}^{(k)}) \). This technique provides us with a fast evaluation of the approximation \( \hat{d}_k(t) \) to the continuous derivative \( d_k'(t) \), which reconstructs the speed of bus \( v_{\text{id}}^{(k)} \). (From now on we will drop the index \( k \) specifying a certain bus in our notations.)

After the speed profile of a vehicle has been restored, we map it back to the road network (see Figure 5), updating each of the road segments passed taking into account behaviour of buses near bus stops and usage of road segments which correspond to bus lanes. More specifically, the AVT segment average velocity updates are made in a way which depends on (i) whether there are any (snapped) bus data sampling points between the end nodes of the segment, and (ii) whether the segment contains a bus stop (or, strictly speaking, is the closest segment to a bus stop used by buses in the corresponding direction of travel) in the vicinity of which the velocity was below a certain threshold.

Let us assume that the end points of \( \text{seg}^{(i)} \) are nodes \( n' \) and \( n'' \). The following two cases are possible:

1) There are no sampling points between \( n' \) and \( n'' \); this can be represented as follows:

\[
\begin{array}{cccccc}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
& * & n' & * & n'' & * \\
& x_j & \bullet & x_{j+1} & \bullet & \vdots \\
\end{array}
\]
This scenario results in a single update for the segment $\text{seg}_i(t) = (n', n'')$ made at time $(t_{n'} + t_{n''})/2$ as follows:

$$\bar{v}^{(i)}(t_{n', n''}) = \frac{1}{t_{n''} - t_{n'}} \int_{t_{n'}}^{t_{n''}} d'(t) \, dt,$$  (1)

where $t_{n'}$ and $t_{n''}$ are times of passing the nodes $n'$ and $n''$ (derived using the inverse function of the reconstructed $d(t)$).

2) There is at least one sampling point between $n'$ and $n''$; this can be represented as follows:

$$* \quad n' \quad * \ldots \quad n'' \quad * \quad x_{j-1} \quad x_j \quad x_{j+r} \quad x_{j+r+1},$$

where $r \geq 0$. This scenario results in $r + 2$ AVT updates $\bar{v}^{(i)}(t_{n', x_j}), \bar{v}^{(i)}(t_{x_j, x_j+r}), \ldots, \bar{v}^{(i)}(t_{x_{j+r}, n''})$ for the segment $\text{seg}_i(t)$ at time moments $(t_{n'} + t_{x_j})/2, (t_{x_j} + t_{x_{j+1}})/2, \ldots, (t_{x_{j+r}} + t_{n''})/2$. These average speed updates are of the same kind as in (1).

Exceptions to these update rules are segments which contain bus stops in whose neighbourhoods the velocity remained below a certain threshold $\gamma$. This can be checked using the interpolation approximation to the function $d(t)$ and function $d'(t)$. For example, it follows from the analysis in the left panel in Figure 5 that the only bus stop in the vicinity (of radius 1 m) of which the bus speed did not exceed $\gamma = 2$ m/sec is the Storey’s Way bus stop (on Madingley Road, see Figure 2). This can be effectively illustrated using the concept of a local time profile function which we introduced and studied in [22]. This function characterises the progression of a bus along the route of interest and measures how long a corresponding bus spent in the small neighbourhood of the point $s$. It is defined as follows:

$$l_s(s) = d^{-1}(s + \varepsilon/2) - d^{-1}(s - \varepsilon/2),$$

where $d^{-1}$ is a generalised inverse function, defined as

$$d^{-1}(s) := \sup \{ t : d(t) < s \}.$$  

This function also takes into account that $d(s)$ is a non-decreasing function and, hence, may not be invertible. It is also robust to the specific choice of the value of $\varepsilon$. When plotted, the spikes of a local time profile indicate locations where the bus spent relatively longer times, whereas long and deep valleys identify parts of the journey which were passed with higher speed. Thus, if we need to detect places $s$ where velocity $v(s) = \varepsilon/l_s(s)$ did not exceed $v = 2$ m/sec, and $\varepsilon = 2$ m, then the criterion for all such points would be the condition $\log l_s(s) \geq 0$. The bottom right plot in Figure 5 depicts the logarithm of the local time profile $l_s(s)$, $\varepsilon = 2$ m, for the journey from the left panel. It can be clearly seen from the plot that the only peak which exceeds the zero threshold and embeds bus stop lines (vertical dotted lines) is the one at Storey’s Way bus stop (it embeds two Storey’s Way bus stop lines, but only one of these corresponds to the bus stop with the right bearing).

After the bus stop segments, with bus speed profiles below the set threshold around the bus stop locations, have been detected, we update average velocities corresponding to those segments as spatial arithmetic means of AVT values $\bar{v}$ of the preceding and subsequent segments (with respect to the bus trajectory). This scheme can be extended to a specific update of a chain of segments containing a bus stop segment and taking into account its length. The empirical distribution of the lengths of road segments and bus stop segments for Cambridge, England, is given in Figure 6. It is perhaps not surprising that the proportion of bus stop segments of any given length is higher than that of any other road segments of the same length, as bus stops tend to be on main urban roads, which tend to be straight and therefore consist of longer segments. However, most of the cumulative distribution mass for each of these two types of segments is concentrated on lengths not exceeding 300 m.

The velocity field can be evaluated using a carefully chosen transformation $f$ applied to the empirical data comprised by the AVT. The following are just two possible examples of such a transformation:

1) $f_{\text{avt}}(t; \text{seg}_i(t)) := \bar{v}^{(i)}(N, t)$, where $N$ is a maximum of all indices $n$, such that $\theta_n(t) \leq t$.

2) $f_{\omega}(t; \text{seg}_i(t)) := \frac{1}{|J|} \sum_{j \in J} \bar{v}^{(i)}(n_j)$, where $J = \{ n : \theta_n(t) \in [t - \omega, t] \}$.

In the former example the transformation $f_{\text{avt}}$ trivially returns the most recent average velocity update corresponding to the segment $\text{seg}_i(t)$. In the latter example the transformation $f_{\omega}$ returns the average of velocity updates for $\text{seg}_i(t)$ made within the time window of length $\omega$, which is a controllable parameter. Generalisations of this transformation function to temporal moving averages or spatial averages are straightforward.

We note that the described approach equally applies to the processing of both historical data and live data feeds with buffering of the most recent streamed data.

C. Visualisation of velocity fields

Velocity fields can be visualised either dynamically or statically (by taking a temporal snapshot) via mapping the values of the AVT transformation function $f(t; \text{seg}_i(t))$ into a colour palette. Figure 7 compares static visualisations of
The velocity fields were evaluated using the AVT transformation $f_ω(t; \text{seg}^{(i)})$, where $ω = 30$ min, and the snapshots were taken at $t = 10$ am. This visualisation uses the velocity-to-colour correspondence palette presented in the middle of the figure and shows the city gridlock effect of the lorry fire accident occurred on 22 June 2010 [25] (the place of the accident is denoted by a red cross in the upper left quadrant of the map in the right panel).

D. Commuting Time Maps

Velocity fields represent scalar fields on road networks and change dynamically in time. Hence, velocity fields implicitly determine travel times from point to point on the map. We provide an illustration of using velocity fields based on this observation. Figure 8 depicts a coloured travel-time heat map plot where the colours are used to reflect how much time a road traveller would take to get from a particular point on the road network to central Cambridge arriving at time $T = 10$ am on 22 June 2010: cooler colours are used to denote shorter journeys, whereas warmer colours are used for travel which would have taken longer. We obtained this plot by applying a label-correcting Dijkstra-type algorithm [26] to calculate point-to-point shortest time paths on a road network with time-dependent inter-node travel times and suitably reverting the temporal orientation of the weighted network.

In this example we are greatly motivated by a recent set of diagrams by MySociety.org [27] in which colours and contour lines are used to show how long it takes to travel between one particular place and every other place in the area, using public transport. However, our approach in building such maps uses traffic-dependent velocity fields reflecting the spatial and temporal aspects of road congestion, whereas MySociety’s heat maps are traffic-independent ones, as only transport maps and timetables were used in producing them.

Fig. 8. This figure represents a travel-time heat map for the area of Cambridge and surrounding villages obtained for 22 June 2010 when a major early morning traffic accident on the north-west stretch of A14 paralysed traffic across the city for hours ahead. The colour of each node on the road network corresponds to the time which a road (vehicle) traveller would have needed if he or she were to arrive to central Cambridge (Cambridgeshire County Council) by 10 am.

V. CONCLUSIONS AND FUTURE WORK

We have shown how to effectively make use of sparse location data from a fleet of public transport buses in evaluating traffic behaviour in urban and interurban areas...
using the notion of the velocity field. We have shown in our use of bus location data that a great deal of useful information can be derived from existing data. These data provide good coverage of the urban road network and are already collected and used by transport operators, so the required infrastructure is often already in place.

Velocity fields allow characterisation of road traffic and classification of congestion levels as well as provide a means of estimating journey times along an arbitrary route of interest within the coverage area. This can usefully be combined with other existing sources of data—for instance, real time car park vacancy data can be used together with estimated journey times to inform drivers heading into a city and needing to park their vehicles there.

In our previous studies we have seen that traffic patterns may depend on and regularly vary with such context factors as time of day, day of week, school term time and possibly weather. In this paper we have seen that there are exceptions to any long-established context, however fine-grained, when planned or unexpected incidents occur, for example traffic accidents within and outside the city or roadworks for utilities. By looking at the historical record and using velocity fields derived from public transport data we can see the effects of these incidents. This also helps us to see in real time when an incident may be happening and to an extent predict its consequences. We envisage that developing appropriate tools for quantitative analysis and qualitative display of continuously gathered data will greatly assist planners in optimising the use of the road network and provide road users with better information.

VI. ACKNOWLEDGEMENTS

This research was supported by UK EPSRC grant EP/C547632/1. We wish to thank Stagecoach, ACIS, and Cambridgeshire County Council for the bus location data used in this study. We would like to thank Jean Bacon and David Evans for their careful reading, critical comments and helpful suggestions on the manuscript. We would also like to thank Alastair Beresford and Ken Moody for useful and stimulating discussions.

REFERENCES