13th International IEEE Conference on Intelligent Transportation Systems, Madeira Island, Portugal, 19-22 September 2010

Statistical Modelling and Analysis of Sparse Bus Probe Data in Urban Areas

Andrei Iu. Bejan, Richard J. Gibbens, David Evans, Alastair R. Beresford, Jean Bacon
Computer Laboratory, University of Cambridge
15 JJ Thomson Avenue
Cambridge, UK CB3 0FD
firstname.lastname@cl.cam.ac.uk

Adrian Friday
Computing Department, Lancaster University
InfoLab21, South Drive
Lancaster University
Lancaster, UK LA1 4WA
adrian@comp.lancs.ac.uk

Abstract— **Congestion in urban areas causes financial loss to business and increased use of energy compared with free-flowing traffic. Providing citizens with accurate information on traffic conditions can encourage journeys at times of low congestion and uptake of public transport. Installing the measurement infrastructure in a city to provide this information is expensive and potentially invades privacy. Increasingly, public transport vehicles are equipped with sensors to provide real-time arrival time estimates, but these data are sparse. Our work shows how these data can be used to estimate journey times experienced by road users generally. In this paper we describe (i) what a typical data set from a fleet of over 100 buses looks like; (ii) describe an algorithm to extract bus journeys and estimate their duration along a single route; (iii) show how to visualise journey times and the influence of contextual factors; (iv) validate our approach for recovering speed information from the sparse movement data.**

I. Introduction

Congestion on roads, especially in urban areas, has a large negative social and economic impact on the community and the environment; for example, the cost of congestion to the UK economy was estimated at £12 billion in 2004 [1] and the cost to the US economy in 2007 was estimated at $78 billion [2]. Congestion can be reduced by increasing the capacity of the road network, encouraging drivers to travel on different routes or at different times of the day, or by using public transport. Travellers may often be unaware of alternative means of getting from A to B, regardless of their regular mode of transport, since travelling becomes an automatic and habitual process; fortunately the provision of better information about likely costs and travel times encourages travellers to explore alternative times and modes of transport [3]. Consequently reducing congestion, either through increased capacity or through the provision of better information for travellers, requires good knowledge of the performance of the road network.

The traditional approach to measuring vehicle movement and congestion is to use static vehicle sensors. These can be inductive loops in the road itself [4] or video cameras which detect the presence of vehicles at fixed locations; some newer networks of cameras can measure point-to-point travel times between pairs of cameras using automatic number plate recognition. Unfortunately, these approaches require the installation and on-going maintenance of expensive equipment in a harsh outdoor environment.

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 coordinates recorded over time and contain much more information than is typically available from fixed sensors [5]. Many public transport fleets are now augmented with automated vehicle location (AVL) systems which use GPS to collect probe data [6].

Shalaby and Farhan used AVL data from buses to predict bus arrival and departure times [7]; Uno et al. provided techniques to study variability of travel times and estimated travel-time distributions [8]. Krumm and Horovitz have shown how to predict destination from historic GPS traces [9], and Froehlich and Krumm extended this to predict the route a driver will follow [10]. Liao et al. estimated a person's location and mode of transport together with the individual's goals and trip segments [11]. More recently, Google's "Maps for Mobile" combined location data crowd-sourced from mobile phones with traditional sensor infrastructure to overlay road maps with congestion information on arterial routes [12].

One problem with AVL data is that they are sparse—samples are typically recorded once every 20 or 30 seconds—and therefore techniques developed for probe data having high update rates are not directly applicable. In this paper we analyse sparse probe data collected from a fleet of over 100 buses and we combine these data with descriptions of bus stop locations and the road network as a whole to build a rich vision of traffic patterns and congestion. Specifically, we (i) describe an algorithm to extract bus journeys and estimate their duration along a single route; (ii) show how quantile regression can be used to visualise contextual factors that affect journey times; (iii) show how to recover speed information from sparse probe data using monotonic splines; and (iv) validate this recovery by comparing it to high update-rate probe data.

This research is supported by UK EPSRC grant EP/C547632/1.

Fig. 1: The map on the left shows the city of Cambridge, UK overlaid by the bus coverage. An individual cell on the map is marked if it includes at least one probe in February 2009. The map on the right shows a northern part of Cambridge with Histon Road connecting the town to the villages of Impington and Histon. The map includes bus stop gates (shown by larger symbols), introduced in Section III-A, used to extract the journeys analysed in Section III-B.

II. Data Description

a) Bus probe data: The probe data from buses used in this study were provided by the company supplying real-time information to the largest bus operator in the city of Cambridge, UK. 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. Many of these buses regularly run along main radial roads and multi-lane highways; some connect Cambridge to villages in the surrounding area. The buses cover an area with a radius of approximately 26 km.

The data represent a set of bus locations recorded over a year (2009). Bus location sampling points are given in a local "easting/northing" (EN) coordinate system (similar to the British Ordnance Survey national geographic grid reference system [13]) which can be regarded as a Cartesian system within scales of a medium size city. The grid squares of the EN coordinate system are of size $10\,\mathrm{m} \times 10\,\mathrm{m}$ and, given the area covered by the city, there is a known transformation between this geographic coordinate system and conventional latitude/longitude coordinates. We estimate the accuracy of each location sample as $\pm 30\,\mathrm{m}$, including the error due to the accuracy of the bus GPS units. The location of each bus is recorded once every δ seconds ($\delta = 20$ or 30) and this reading is transmitted from the bus. The time of the logging of each sample is known up to an offset τ that is fixed for each bus and does not exceed δ. There are missing observations, meaning that there are times of silence when all or some buses are not transmitting their positions due to communication failure, temporary hardware malfunction, weather, and so on. Finally, these data do *not* describe which bus route each bus was taking.

As is shown in the left plot of Figure 1, the bus probe data give good coverage of the main city roads including arteries. Furthermore, these data were available throughout the day and evening, including peak hours when traffic conditions are of most interest.

b) Road network description and features: We use OpenStreetMap (OSM) data for information about the road network, including locations of junctions, traffic lights, and traffic bollards. OpenStreetMap is a collaborative project with the aim of creating a free, public map of the world [14]. The content is contributed voluntarily, using geographic data from portable GPS devices, aerial photographs, and other free sources [15]. The quality of the data has become increasingly high and is comparable to commercially produced maps [16].

c) Bus stops and bus route information: Although some information about locations of bus stops can be obtained using OSM data, we instead use the National Public Transport Access Node (NaPTAN) database. This collates all bus stop locations in the UK and, for each stop, includes the latitude, longitude, and direction of bus travel. For map-matching we use knowledge of the sequence of roads which buses follow while assigned to a certain route. This is derived using timetable information from the bus operator, supplemented by the bus probe data described above.

d) High resolution probe data: In order to validate our approach for restoring bus' behaviour between sampled locations (Section IV), we collected high resolution ("HighRes") traces using a portable GPS device carried on board the buses. The traces record the bus location once per second with an accuracy of $5\,\mathrm{m}$ to $15\,\mathrm{m}$.

III. Journey Times

In this section we describe how we extract bus journeys corresponding to a route of interest and evaluate their durations using our bus data and taking into account the sparseness of the data in time. We also show the effect that the day of the week and time of year have on the journey times and their variability.

A. Extracting Journeys and Estimating their Durations

Let us first formally describe the bus data. We denote the set of buses by $\mathcal{B}$. We also enumerate and refer to the elements of $\mathcal{B}$ using the notation $\mathtt{vid}^{(k)} \in \mathcal{B}$, where $k = 1, 2, \ldots, |\mathcal{B}|$. Each bus $\mathtt{vid}^{(k)}$ updates its location $x_i^{(k)}$ at times $t_1^{(k)} + \tau^{(k)}, t_2^{(k)} + \tau^{(k)}, \ldots$, where each $t_i^{(k)}$ is a regular timestamp such that in the long run the sequence $\delta_i^{(k)} := t_{i+1}^{(k)} - t_i^{(k)}$ is typically a constant sequence of either 20 or 30 seconds. (From time to time, $\delta_i^{(k)}$ is a small multiple of 20 or 30 and in rare instances of silence, it can be larger still.) This means that the movement of bus $\mathtt{vid}^{(k)}$ can be described as a sequence of vectors $\mathbf{y}_i^{(k)} = (t_i^{(k)}, x_i^{(k)})$ whose components are the timestamps $t_i^{(k)}$ and bus positions $x_i^{(k)}$.

Let $r_{AB} \equiv A \rightsquigarrow B$ be a predefined route connecting two points A and B on the road network map. We are interested in extracting bus journeys connecting A and B taking into account the sparseness of the data ($\delta_i^{(k)} \geq 20$ sec). In order to do this, we use information about locations of bus stops and the direction of travel of buses served by these bus stops. The NaPTAN database includes coordinates in latitude/longitude of bus stops and assigns each stop a direction of travel from the set $\mathtt{N}$, $\mathtt{NW}$, $\mathtt{W}$, $\mathtt{SW}$, $\mathtt{S}$, $\mathtt{SE}$, $\mathtt{E}$, and $\mathtt{NE}$.

Let S be a bus stop with bearing $\mathtt{B}_S$. We assume that the Earth's surface is flat within a neighbourhood of S of radius not exceeding two or three hundred metres and define the *bus stop gate* $\mathtt{BG}_S$ for S of length γ to be the line segment perpendicular to B_S with S as its centre and length γ. A journey of bus $\mathtt{vid}^{(k)}$ from A to B is a sequence of points $P_1 P_2 \ldots P_n$ consisting of consecutive location observations of $\mathtt{vid}^{(k)}$ such that $P_1 P_2$ intersects the bus stop gate $\mathtt{BG}_A$, $P_{n-1} P_n$ intersects $\mathtt{BG}_B$, and there are no intersections of the intermediate segments of the trace with either of the two gates. (This rule can easily be generalised to the case when the route of interest is defined by more than two bus stops.)

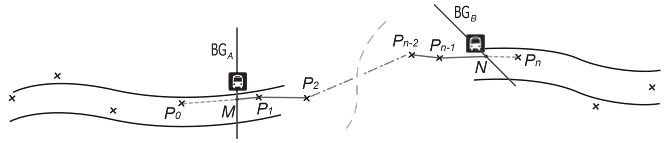

Fig. 2: This figure demonstrates journey time adjustment for intersection with the first and last bus stop gates. In the depicted situation the bearing of the bus stop A is $\mathtt{E}$ (East) and the bearing of the bus stop B is $\mathtt{NE}$ (North-East).

Given this formulation, extraction of journeys can be expressed as a string matching problem as follows. Let the route of interest r_{AB} be specified by a sequence of bus stop gates $\mathtt{BG}_A \equiv \mathtt{BG}_1, \mathtt{BG}_2, \ldots, \mathtt{BG}_s \equiv \mathtt{BG}_B$ and let $\mathbf{y}_i^{(k)} = (t_i^{(k)}, x_i^{(k)})$ be the i^{th} historical observation of vehicle $\mathtt{vid}^{(k)}$. Let $\mathcal{S} = \{s_i\}$ be a sequence of labels corresponding to intersection points of consecutive bus trace segments $x_{i+1}^{(k)} x_i^{(k)}$ with the bus gates ordered in time as follows. If the segment $x_{i+1}^{(k)} x_i^{(k)}$ intersects the bus gate $\mathtt{BG}_j$, then we set s_i to be '$\mathtt{BG}_j$'; if, however, this segment does not intersect any of the bus gates[1], then $s_i :=$'$\mathtt{O}$'. Treating $\mathcal{S}$ as text and ignoring the presence of the symbol '$\mathtt{O}$' one can then efficiently find positions of the word $w =$'$\mathtt{BG}_1\mathtt{BG}_2 \ldots \mathtt{BG}_s$' in $\mathcal{S}$. These positions fully identify journeys along r_{AB}.

We now describe how we estimate journey times along the route r_{AB}. Let $\rho \equiv P_0, P_1, \ldots, P_n$ be a sequence of observations corresponding to $\mathtt{vid}^{(k)}$ made at times $t_{i_1}^{(k)}, t_{i_2}^{(k)}, \ldots, t_{i_n}^{(k)}$ such that ρ represents a journey along the route r_{AB}. Let $M = P_0 P_1 \cap \mathtt{BG}_A$ and $N = P_{n-1} P_n \cap \mathtt{BG}_B$ as shown in Figure 2. We estimate the corresponding journey time $JT(\rho;\, r_{AB})$ as

$$JT(\rho;\, r_{AB}) = t_n^{(k)} - t_0^{(k)} - \left(\frac{P_0 M}{P_0 P_1} \delta_1^{(k)} + \frac{P_{n-1} N}{P_{n-1} P_n} \delta_n^{(k)} \right).$$

In other words, the time taken on the journey is the time it took for the bus to travel from P_0 to P_n minus the time spent by the bus before it crossed the first bus stop gate and after it crossed the last.

B. Factors Affecting Journey Times

In the previous subsection we described how journey times can be determined from our database of bus probe data. In this subsection we continue with an investigation of the statistical characteristics of these journey times.

We used the bus stop gate technique to extract bus journeys made along Histon Road, Cambridge; four bus stop gates, each of length $\gamma = 300\,\mathrm{m}$, were used as shown on the right-hand map in Figure 1. The data set consists of 13,653 journeys in the direction from north to south (that is, towards the city centre) between Sunday, 2 November 2008 and Saturday, 9 January 2010. The minimum journey time was 2.95

[1]For simplicity we assume that the bus gates are far enough from each other so that a segment $x_{i+1}^{(k)} x_i^{(k)}$ cannot intersect more than one bus gate; however, the described scheme can be generalised when this is not the case.

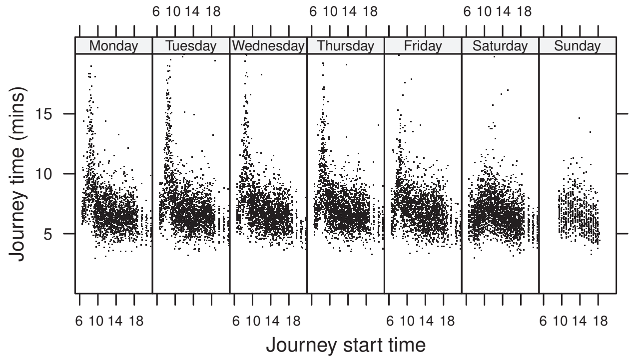

Fig. 3: This figure shows how the time of day that a journey begins affects its estimated duration for each day of the week. All days including bank holidays are used.

minutes and 99.5% of the journeys took less than 20 minutes. The mean journey time was 7.22 minutes with a median value of 6.75 minutes. Associated with each journey is the start time and the date (and thus the day of the week) when the journey took place. We now examine the relationship between journey time, start time, and day of the week.

Figure 3 shows scatter plots of the journey times and start times for journeys that started between 7 am and 10 pm for each day of the week. During weekdays there is a sharp rise in journey times in a morning busy period centred around 8:30 am; during weekends, and especially on Sundays, there is little evidence of a busy period. We shall henceforth consider in detail the daily profile of journey times on weekdays. Within the data set there are 2,431 journeys made on Saturdays and 724 on Sundays, leaving 10,498 taking place on weekdays.

In Figure 4 we illustrate several ways in which we can summarise the variation of journey times by day. In the upper panel we show box-and-whisker plots for the journey times binned into 15 minute intervals. The box-and-whisker plot enables us to see how the median journey times vary over the day, at least to a granularity of 15 minutes, and how the upper and lower quartiles behave. It is clear that during the morning busy period median journey times increase but the most dramatic increase is in the dispersion of the data with long tails up to at least 20 minutes. The box-and-whisker plots are particularly suitable when the journey start times fall at a discrete set of times.

The central panel of Figure 4 shows the results from a quantile regression model for 10^{th}, 50^{th} and 90^{th} percentiles which we briefly describe now; Koenker provides further details [17]. Suppose the $N = 10,498$ journey times, z_i, are indexed by $i = 1, \ldots, N$ with corresponding start times t_i. Thus we have bivariate observations $\{(t_i, z_i) : i = 1, \ldots, N\}$ for the random variables (T, Z) and we fit a nonparametric model, $g(t)$, to the r^{th} conditional quantile, $Q_Z(r \,|\, T = t)$ for $r = 0.1, 0.5$, and 0.9. The particular choice of nonparametric model $g(t)$ follows that of [17, Section A.9] and uses B-splines with f degrees of freedom. In Figure 4 the fitted model with $f = 12$ vividly shows how the busy period has the most pronounced effect on the upper percentile of journey

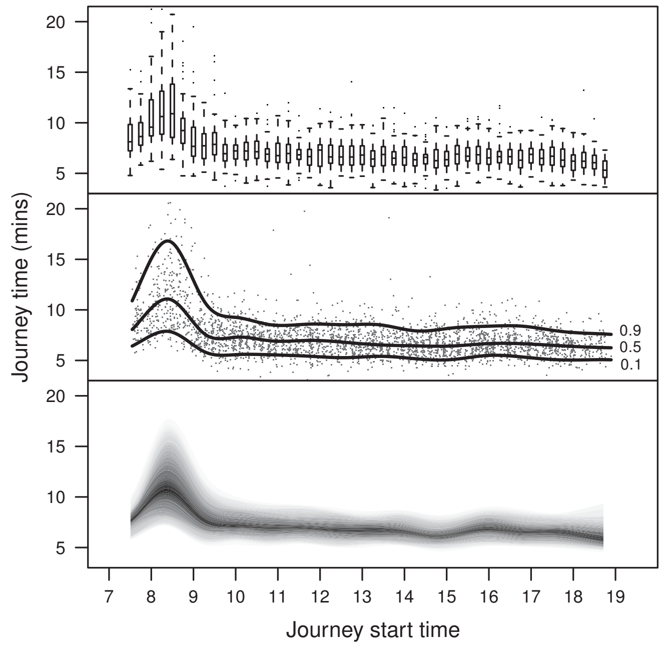

Fig. 4: These figures show journey times for weekdays only. At the top is a box and whisker plot. In the middle plot, overlaid on the data points are three quantile regression lines (10^{th}, 50^{th}, and 90^{th} percentiles) fitted according to the model discussed in the text. The plot at the bottom shows many quantile regression lines fitted, using shading between each as described in the text.

times.

The lower panel of Figure 4 is a variant of the quantile regression plot in the central panel where models are now fitted for a range of r from 0.05 up to 0.95. The figure uses grey colours to fill between successive conditional quantiles. The colour is black at the median quantile and fades to white linearly as the extreme quantiles are approached in both directions. This style of figure captures the variation of journey times by day without making specific reference to any one quantile line. A modification of this technique where the estimated density is shaded is given in [18].

Our graphs and models have so far considered the effects of the time of day and day of week. We now use date information to consider the effect of school terms on journey times. Figure 5 considers weekdays of two types: those within the school term and those outside of term (we have also removed all journeys taken on bank holidays). The two overlaid scatter plots show the striking effect of school terms on journey times. During the busy period in term time journey durations range over a much larger interval. We can see this most clearly with our quantile regression models for the 10^{th}, 50^{th} and 90^{th} percentiles. During school term the quantiles (solid lines) show a strong variation during the morning busy period. However, outside of school term (dashed lines) the variation is much reduced. We shall return to this effect in the next section where we seek to pin-point precisely where, and for how long, journeys are delayed.

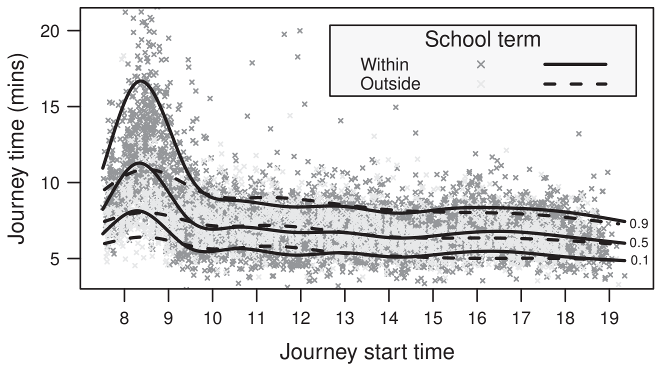

Fig. 5: Dark crosses show the duration for journeys taken on school days (that is, non bank holiday weekdays during the school term). Lighter crosses show journeys for non bank holiday weekdays outside of school term. The solid and dashed lines are quantile regression lines for term and non-term days, respectively.

IV. Restoring Behaviour of Buses between Their Consecutive Observations

The scatter plot and quantile regression lines from Figure 5 show that there is a very pronounced effect of school terms on the duration of bus travel in the early morning hours along Histon Road towards the city centre. In order to analyse whether this effect is due to buses spending more time stationary at bus stops serving an increased number of passengers or whether they move considerably more slowly along all or part of the route (or a combination of both), we restore behaviour of buses in the sparse data by approximating their progression using spline interpolation.

As before, we consider a particular bus $\texttt{vid}^{(k)} \in \mathcal{B}$ and consider one of its paths (together with the time-stamps) $(t_{i_1}^{(k)}, x_{i_1}^{(k)}), \ldots, (t_{i_n}^{(k)}, x_{i_n}^{(k)})$ along the route of interest. We perform map matching by *snapping* the bus locations $x_{i_1}^{(k)}, \ldots, x_{i_n}^{(k)}$ to the road network using the fact that our vehicles are buses following bus routes and accounting for the possibility of "false" snapping to a road at a different altitude (as is the case, for instance, with the junction labelles B1049 on the right-hand side of Figure 1 where Bridge Rd/Cambridge Rd cross the A14 at a higher level). We thus obtain an ordered sequence of snapped points $w_{i_1}, \ldots, w_{i_n}$ and define $d_k(t_{i_m}^{(k)})$ to be the length of the shortest path from $w_{i_{m-1}}$ to w_{i_m} setting $d_k(t_{i_1}^{(k)}) := 0$. The obtained values $d(t_{i_1}^{(k)}), \ldots, d(t_{i_n}^{(k)})$ can be seen as evaluations of the continuous function $d_k(t)$ which measures the cumulative distance travelled by bus $\texttt{vid}^{(k)}$ by time t.

We use cubic spline interpolation [19] to further restore the function $d_k(t)$ for any $t \in (t_{i_1}^{(k)}, t_{i_n}^{(k)})$. Specifically, we use monotonic cubic splines [20] in order to comply with the fact that d_k is a non-decreasing function of time. This technique will leave us at worst with the continuous derivative d'_k, which approximates the speed of bus $\texttt{vid}^{(k)}$, and at best with a continuous approximation d''_k to the acceleration of this bus.

Figure 6 presents two traces from bus journeys taken on 23 February 2010. The left hand panels refer to the journey starting just after 8am in the direction towards the city centre whereas the right hand panels refer to a journey in the opposite direction starting just after 8:30 am. In the top panels the trajectory of the distance travelled is marked by the stars. A monotonic cubic spline has been fitted to the observations and this is shown by the solid line. Horizontal dashed lines indicate landmarks such as the locations of traffic lights and bus stops along the journey. The lower panels show bus speeds over time obtained by (numerical) differentiation of the cubic splines. In addition, we show by thin grey lines speeds obtained from HighRes traces and observe that by reducing the resolution we are effectively smoothing the speeds.

In order to characterise the progression of the buses along the route of interest as a *local time profile* function of the distance travelled we use the following function:

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

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

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

The function $l_\varepsilon(s)$ measures how long a corresponding bus spends in the small neighbourhood of the point s. This function also takes into account that $d(s)$ is a non-decreasing function and, hence, may not be invertible.

Figure 7 illustrates local time profiles $l_\varepsilon(s)$ for journeys from Figure 6 by showing $\log l_\varepsilon(s)$, where $\varepsilon = 2\,\mathrm{m}$. High spikes indicate locations where the buses spent relatively longer, whereas long and deep valleys identify parts of the route which were passed with higher speeds. Experimentation suggests that our profiles are robust to the specific choice of the value of the parameter ε.

Finally, Figure 8 shows logarithms of local time profiles $l_\varepsilon(s)$ for seven journeys (from the data set discussed in the previous section) made on weekdays within school terms between 8 am and 9 am along Histon Road, and seven journeys from the same data set and period of the day made on week days outside school terms. These 14 journeys were chosen at random from the total number of 780 journeys (574 journeys within school terms and 206 journeys outside school terms). The profiles suggest that there is a combination of reasons responsible for making journeys made in peak morning hours during school term longer. Notice that additional delay to journeys made in school term occurs between bus stops and traffic light junctions in a manner not observed in journeys made outside of school term. These delay effects are more pronounced at some locations than in others. For example, in Figure 8 portions of journeys before $1000\,\mathrm{m}$ and after $1800\,\mathrm{m}$ show comparatively more delay than occurs in the portion in between.

V. Conclusions

In this study we have investigated how access to a large repository of bus trajectories combined with other publicly available data can be used to provide a detailed account of

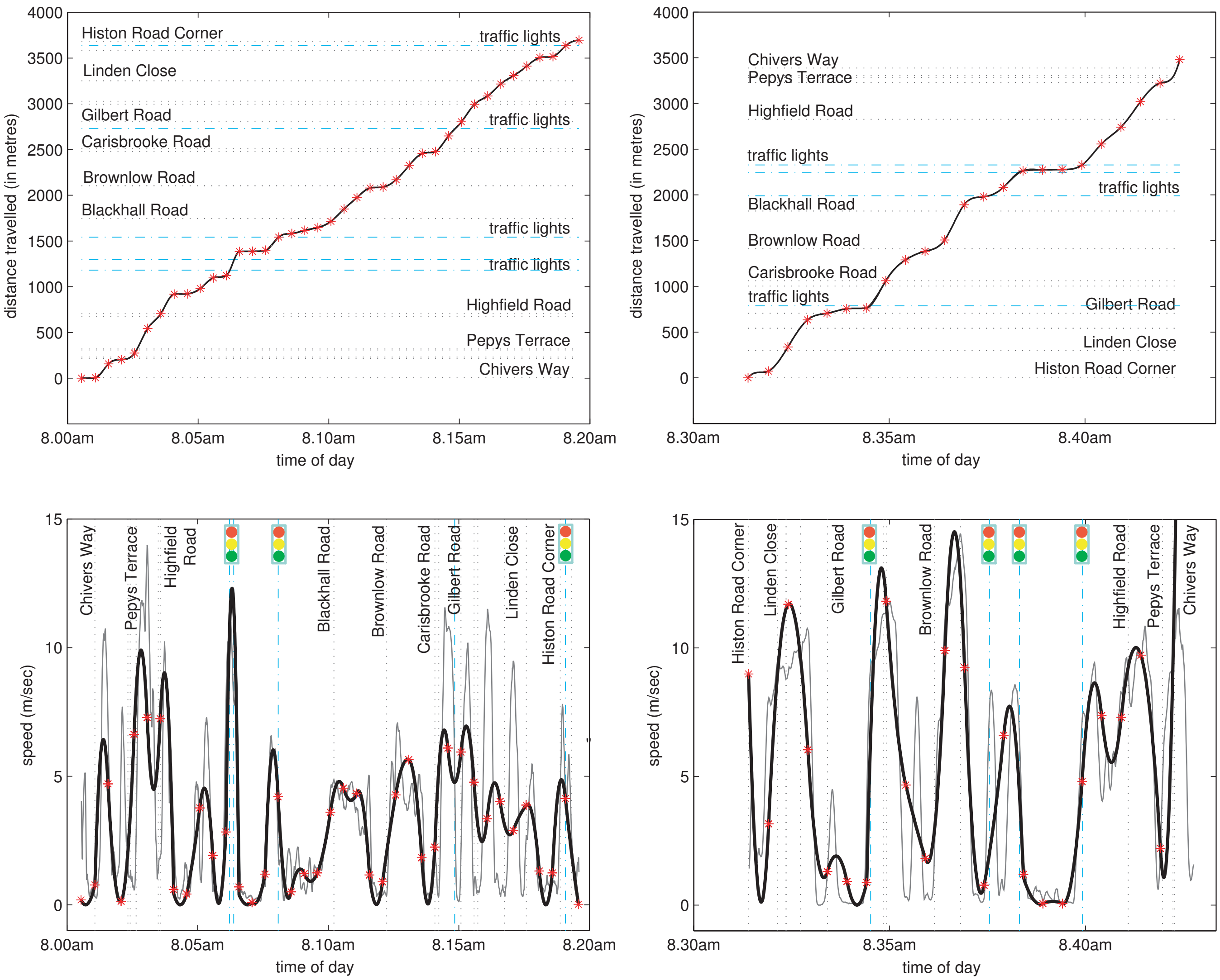

Fig. 6: The upper panel depicts bus observations (stars) and monotonic spline interpolation of the trajectories of two bus journeys made along Histon Road and through part of Impington in the morning hours of 23 February 2010, inbound (left) and outbound (right). The bottom panel depicts corresponding speed profiles derived from fitted splines (thick solid lines) and from HighRes tracks (thin grey lines). Dotted horizontal and vertical lines show the positions of bus stops and dashed lines identify traffic light junctions.

journey times and some of the important factors that affect them. Several types of graphical displays, a nonparametric quantile regression, and monotonic splines technique are used to visualise these effects on journey time and recover speed information from sparse data. We have also been able to study in some detail the sources of delay within individual journeys and have presented our findings using a novel technique based on a notion of local time profile which we define.

Our techniques provide a detailed basis for understanding journey time behaviour in the urban environment which lead to further important research areas such as how to incorporate real-time data to improve prediction of journey times. Similarly, these techniques will be useful to bus operators in helping them to optimally manage their fleets and thus to meet service requirements. Local transport authorities will also be better informed about the nature and extent of congestion and its impacts on travellers within a city.

VI. Acknowledgements

This research is 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. RJG and AIuB wish to thank the Isaac Newton Institute for Mathematical Sciences. We would also like to thank Ken Moody for the useful discussions and support offered.

References

[1] R. Devereux, J. Dawson, M. Dix, G. Hazel, D. Holmes, S. Glaister, S. Joseph, C. T. Macgowan, B. Nimick, M. Roberts, L. Searles, R. Turner, S. Gooding, S. Hickey, and W. Rickett, “Feasibility study of road pricing in the UK,” Department for Transport, Tech. Rep., July 2004. [Online]. Available: http://www.dft.gov.uk/stellent/groups/dft_roads/documents/pdf/dft_roads_pdf_029788.pdf

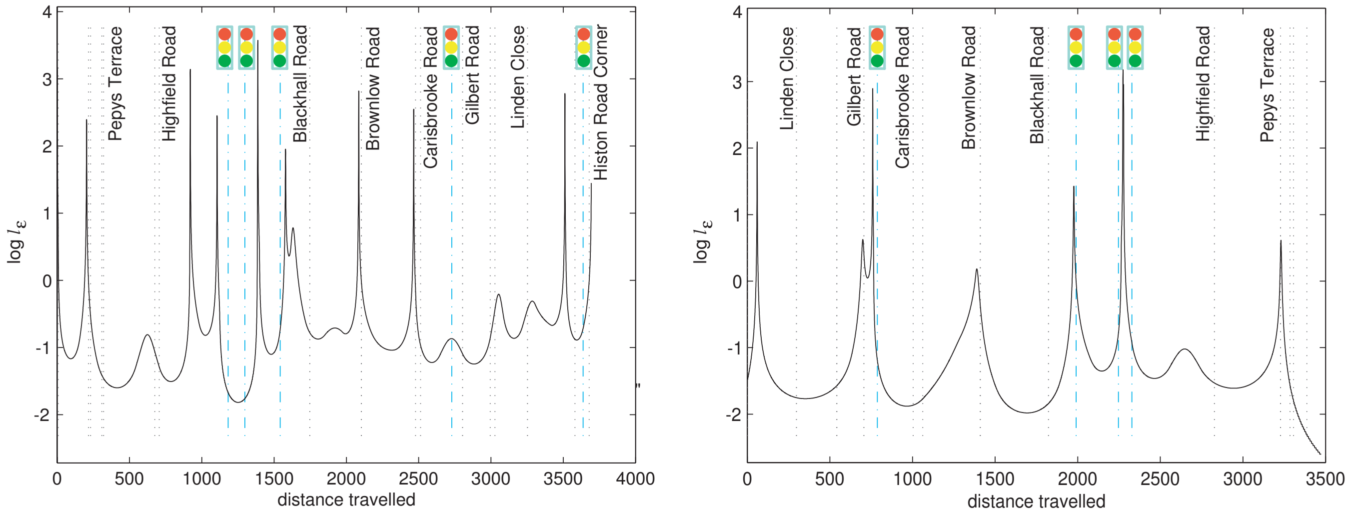

Fig. 7: These plots depict the logarithm of the local time profiles $l_\varepsilon(s)$ derived from spline interpolation of the bus trajectories shown in Figure 6 ($\varepsilon = 2\,\text{m}$).

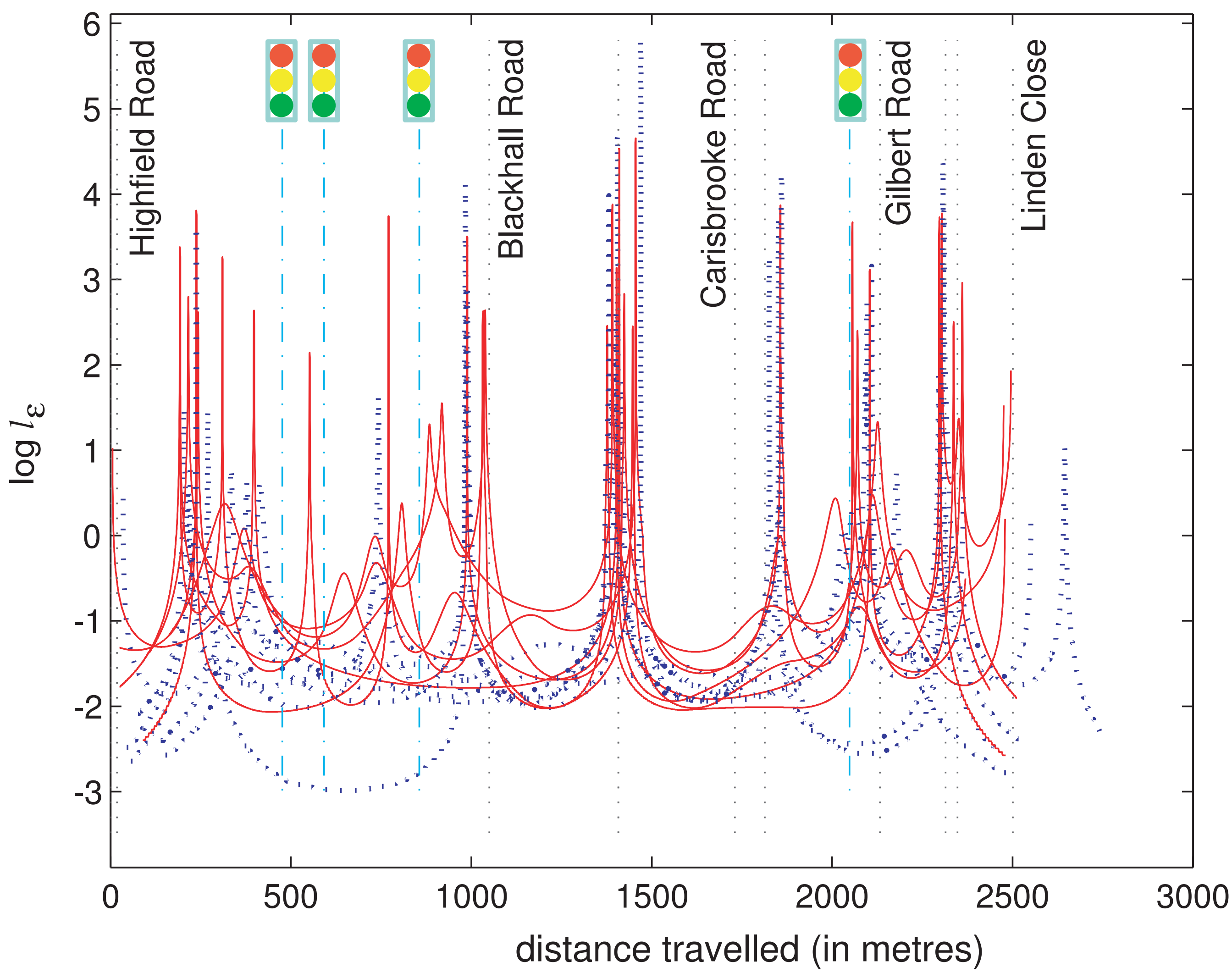

Fig. 8: This figure shows 14 profiles of the function $l_\varepsilon(s)$ (where $\varepsilon = 2\,\text{m}$) for 7 journeys made within school terms (solid lines) and 7 journeys made outside school terms (dashed lines). These 14 journeys were chosen at random from a large number of journeys made between 8 am and 9 am along Histon Rd towards the centre of Cambridge.

[2] D. Schrank and T. Lomax, “The 2007 urban mobility report,” Texas Transportation Institute, Texas A&M University, Tech. Rep., September 2007.

[3] S. Kenyon and G. Lyons, “The value of integrated multimodal traveller information and its potential contribution to modal change,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 6, pp. 1–21, March 2003.

[4] P. B. Hunt, D. I. Robertson, R. D. Bretherton, and R. I. Winton, “SCOOT—a traffic responsive method of coordinating signals,” Transport and Road Research Laboratory, Tech. Rep. LR1014, 1981.

[5] G. Rose, “Mobile phones as traffic probes: Practices, prospects and issues,” *Transport Reviews: A Transnational Transdisciplinary Journal*, vol. 26, no. 3, pp. 275–291, 2006.

[6] L. Knoop and T. Eames, “Public transport technology in the United Kingdom: Annual survey 2008,” RTIG Ltd, Tech. Rep., Feb. 2009.

[7] A. Shalaby and A. Farhan, “Prediction model of bus arrival and departure times using AVL and APC data,” *Journal of Public Transportation*, vol. 7, no. 1, pp. 41–62, 2004. [Online]. Available: http://www.nctr.usf.edu/jpt/pdf/JPT%207-1%20Shalaby.pdf

[8] N. Uno, F. Kurauchi, H. Tamura, and Y. Iida, “Using bus probe data for analysis of travel time variability,” *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, vol. 13, no. 1, pp. 2–15, 2009.

[9] J. Krumm and E. Horovitz, “Predestination: Where do you want to go today?” *IEEE Computer*, vol. 40, no. 4, pp. 25–27, April 2007.

[10] J. Froehlich and J. Krumm, “Route prediction from trip observations,” in *Society of Automotive Engineers (SAE) 2008 World Congress*, 2008.

[11] L. Liao, D. J. Patterson, D. Fox, and H. Kautz, “Learning and inferring transportation routines,” *Artificial Intelligence*, vol. 171, no. 5-6, pp. 311–331, 2007.

[12] Google, “Arterial traffic available on Google,” http://google-latlong.blogspot.com/2009/08/arterial-traffic-available-on-google.html, August 2009. [Online]. Available: http://google-latlong.blogspot.com/2009/08/arterial-traffic-available-on-google.html

[13] 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. [Online]. Available: http://www.ordnancesurvey.co.uk/oswebsite/gps/docs/A_Guide_to_Coordinate_Systems_in_Great_Britain.pdf

[14] Open Street Map, “The free wiki world map,” http://www.openstreetmap.org/, March 2010. [Online]. Available: http://www.openstreetmap.org/

[15] M. Haklay and P. Weber, “OpenStreetMap: User-generated street maps,” *IEEE Pervasive Computing*, vol. 7, no. 4, pp. 12–18, October–December 2008.

[16] M. Haklay, “How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets,” http://www.ucl.ac.uk/~ucfamha/OSM%20data%20analysis%20070808_web.pdf.

[17] R. Koenker, *Quantile Regression*, ser. Econometric Society Monograph Series. Cambridge University Press., 2005.

[18] C. H. Jackson, “Displaying uncertainty with shading,” *The American Statistician*, vol. 62, no. 4, pp. 368–370, November 2008.

[19] P. Dierckx, *Curve and Surface Fitting with Splines*. New York, NY, USA: Oxford University Press, Inc., 1993.

[20] G. Wolberg and I. Alfy, “An energy-minimization framework for monotonic cubic spline interpolation,” *Journal of Computational and Applied Mathematics*, vol. 143, pp. 145–188, 2002.