Emerald Journal of Facilities Management

Heterogeneous urban traffic data and their integration through kernel-based interpolation

Journal:	*Journal of Facilities Management*
Manuscript ID:	Draft
Manuscript Type:	Research Paper
Keywords:	Data fusion, Big Data, Transport network , Automatic Number Plate Recognition , GPS , urban loop detectors

Heterogeneous urban traffic data and their integration through kernel-based interpolation

August 2015

ABSTRACT

This paper presents collection and analysis of heterogeneous urban traffic data, and integration of them through a kernel-based approach. The recent development in sensing and information technology opens up opportunities for researching the use of this vast amount of new urban traffic data. In this paper, the data fusion algorithm is developed by using a kernel-based interpolation approach. Our objective is to reconstruct the underlying urban traffic pattern with fine spatial and temporal granularity through processing and integrating data from different sources. The fusion algorithm can work with data collected in different space-time resolution, with different level of accuracy, and from different kinds of sensors. The properties and performance of the fusion algorithm is evaluated by using a virtual test-bed produced by VISSIM microscopic simulation. The methodology is demonstrated through a real-world application in Central London. This paper contributes to analysis and management of urban transport facilities.

Keywords: Data fusion, Big Data analytics, transport network, Automatic Number Plate Recognition (ANPR), GPS, urban loop detectors

1. Introduction

A detailed and reliable picture of spatio-temporal variations of traffic is essential for understanding and managing congestion (Tsapakis et al., 2012; Chow et al., 2014). Much previous research on traffic data has been focusing on freeways where Kwon and Varaiya (2005) provide a review on relevant studies. Compared with freeways or motorways, we see relatively less research work done on urban networks. It is due to the lack of required data and the complexity of problem. Recently, the increasing availability of urban traffic data provides new research opportunities.

Urban traffic data varies greatly in terms of spatio-temporal granularity, latency and accuracy. Typical sources of traffic flow data in urban environment include:

- Fixed sensors – such as loop detectors and automatic traffic counters - provide information of traffic volume, composition of traffic (e.g. proportion of buses, heavy good vehicles, etc), concentration and speed.

- Global Positioning System (GPS) devices – such as smart phones, personal navigators, etc - are attached to vehicles or persons. The GPS devices report location and speed of the attached objects regularly (typically every second). Various information such as trip lengths and journey times can also be derived from GPS data.

- Automatic Vehicle Identification (AVI) – With vehicle (re)-identification techniques, on-road or roadside sensors (e.g. cameras) can provide information including journey times and trip lengths.

Integrating heterogeneous traffic data in a consistent way is always a challenging problem, where Ou (2011) provides a review of different kinds of data fusion approaches. One the most popular approach for integrating traffic data is through the model-based Kalman Filter (KF, Kalman, 1960) and its variants such as Extended KF, Unscented KF, and Particle Filter (PF) (see examples: Wang and Papageorgiou, 2005; Mihaylova et al., 2007; Herrera and Bayen, 2010; Ngoduy, 2011). Under the KF and PF (and their variants) framework, traffic flow estimation is produced by comparing and combining the model estimates and actual measurements. The filtering framework consists of two parts: a state equation to model and predict the traffic state evolution over time, and an observation equation to relate the measurements to the underlying traffic state. Both state and observation equations consider explicitly the associated errors involved. An estimate is then generated through minimizing the associated expected pooled error in state predictions and observations. Such filtering framework is used widely in data fusion algorithms due to a number of its desirable properties. The explicit state and observation models allow data from different kinds of sensors to be incorporated. Moreover, the well-defined statistical measures of uncertainty make it possible to quantify the value of each data source.

Nevertheless, the filtering approach is model based, which mean it has to operate based upon an assumed traffic model. Typical examples of traffic model for this purpose include Cell Transmission Model (Daganzo, 1994), METANET (Papageorgiou et al., 1990), and a recently proposed two-regime traffic model by Balijepalli et al. (2013). As noted in Ou (2011), a major difficulty for this model-based approach is on the selection and calibration of the underlying traffic model. Choosing and calibrating a suitable traffic model is not straightforward, and inappropriate choice will lead to inconsistency and instability of the fusion framework.

This paper presents a kernel-based data fusion algorithm that can integrate heterogeneous urban traffic data with different characteristics. The objective is to reconstruct the urban traffic pattern with fine spatial and temporal granularity through processing and integrating data from different sources. The fusion algorithm does not require assumption of any underlying model and it can work with data collected in different spatio-temporal granularity, with different level of accuracy, and from different kinds of sensors. We need to note the proposed method will be limited to offline application due to the lack of an underlying traffic model as a state estimator. Nevertheless, the fusion algorithm in this paper will be a valuable and cost-effective tool for offline transport planning and policy evaluation through integrating existing sources of data. The performance and properties of the fusion algorithm is evaluated by using a synthetic scenario generated by VISSIM micro-simulation. The algorithm will also be used to integrate actual traffic data collected from a road section in Central London (UK) as an illustration of real-world application.

This paper starts with Section 2 which discusses the characteristics of different traffic data on urban road networks. We use Central London in the UK as an illustration. Section 2 will also highlight the difference between different data in terms of their spatial and temporal granularity. Section 3 presents the data fusion algorithm that we adopt. Section 4 presents an application of the fusion algorithm to Central London road network. The fusion algorithm is also analysed by using a virtual test-bed generated by VISSIM micro-simulation. Finally, Section 5 gives some concluding remarks.

2. Urban traffic data – London, UK

Journey times are one of the most important performance indicators for urban road networks. As an illustration, journey times in London are measured by using the Automatic Number Plate Recognition (ANPR) technique. In London, there are about 500 cameras for enforcing various policies such as congestion charging and low emission zones. When a vehicle passes by a camera, its plate number will be recognized and recorded along with the associated time. The journey time of the vehicle between two camera sites can then be derived by matching the plate number. The derived journey times are further processed and stored in 5-min averages in the database.

It is noted that various errors may arise in matching the license plate numbers due to various reasons such as misreading of license plates, vehicles stopping en-route, and vehicles taking unusual long route between the two camera locations, data loss due to road closure, and failure of hardware system. Consequently, a set of data filtering and processing rules is adopted to improve the journey time estimation (Robinson and Polak, 2006). Some patching or imputation algorithms may be used for imputing missing data. The associated ANPR journey time data is flagged with a code referring to the type of patching mechanism which is applied. This code is ranging from 0 (best: no patching applied) to 3 (worst: patched by typical profile).

Figure 1 shows a 1-km stretch of Waterloo Road (A301) in Central London, and Figure 2 shows the associated speeds (i.e. reciprocals of journey times) measured along the road in

April 2010. In the figure, the position and width of the bar at each 5-min interval reflects the average and dispersion of the journey times in the month at that particular interval.

Figure 1 Waterloo Road, London (UK)

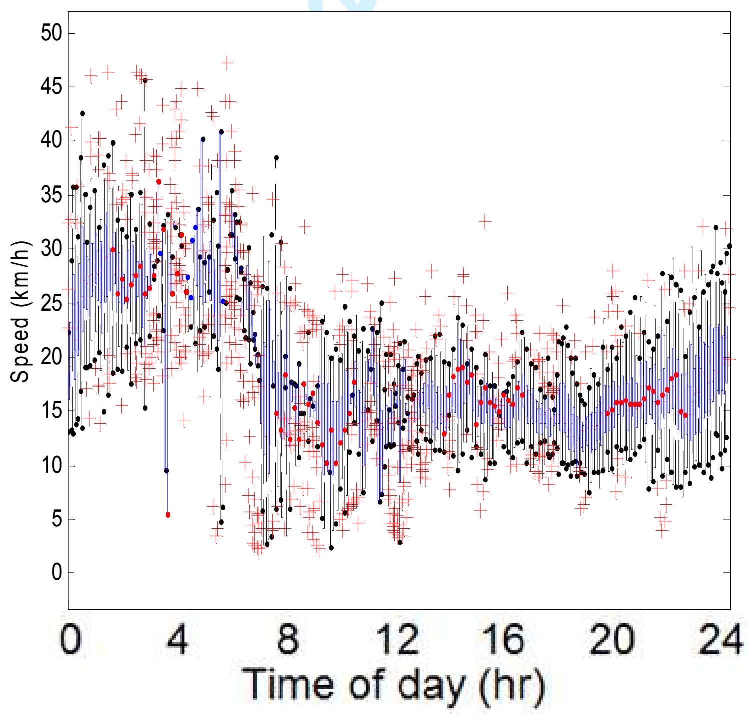

Figure 2 Variations of traffic speeds in April 2010 along Waterloo Road, London (UK)

Figure 3 shows the *average speed field* over time and space produced by the ANPR journey times. The colour scale on the space-time grid represents the average speed of traffic. As can be seen from Figure 3, a major weakness with the ANPR journey times is that they do not capture much spatial feature of traffic. It is because the distance between a pair of ANPR camera sites is typically far apart (in the range of kilometres), which implies a lot of spatial variations are missed along the route.

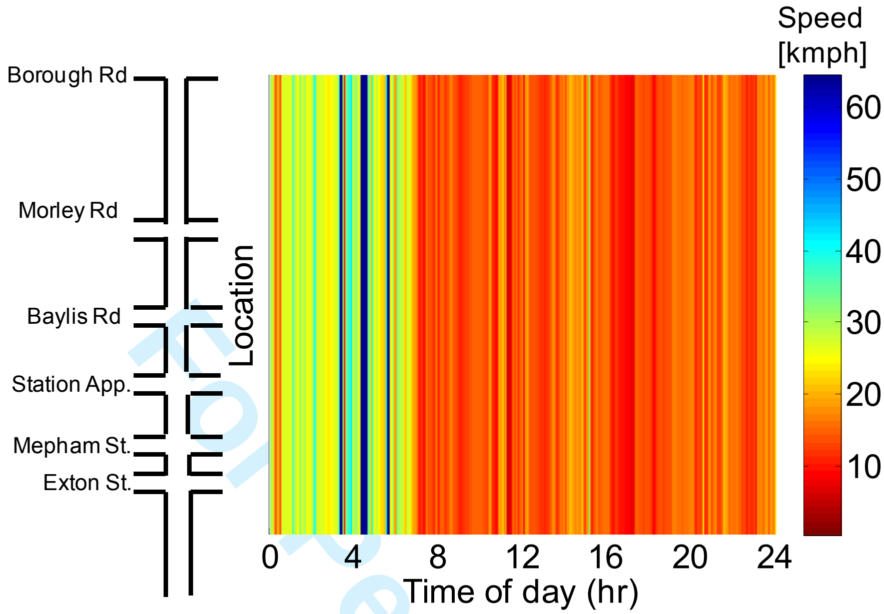

Figure 3 ANPR-Speed field along Waterloo Road, London (UK)

To extract detailed spatial traffic variations, we use floating car data provided by Trafficmaster*. Some vehicles on the road are equipped with Trafficmaster GPS (Global Positioning System) devices. The GPS devices on these vehicles report the locations of the vehicles on a regular basis (~8-10 seconds). Figure 4 shows the corresponding speed field generated by these Trafficmaster data, which can reveal much more spatial feature compared with the ANPR data. Nevertheless, there are only very limited samples of Trafficmaster vehicles on the road (about 1,500 such vehicles in Greater London Area). With such small sample size, Trafficmaster can only reveal limited temporal characteristics of traffic.

* http://www.trafficmaster.co.uk/

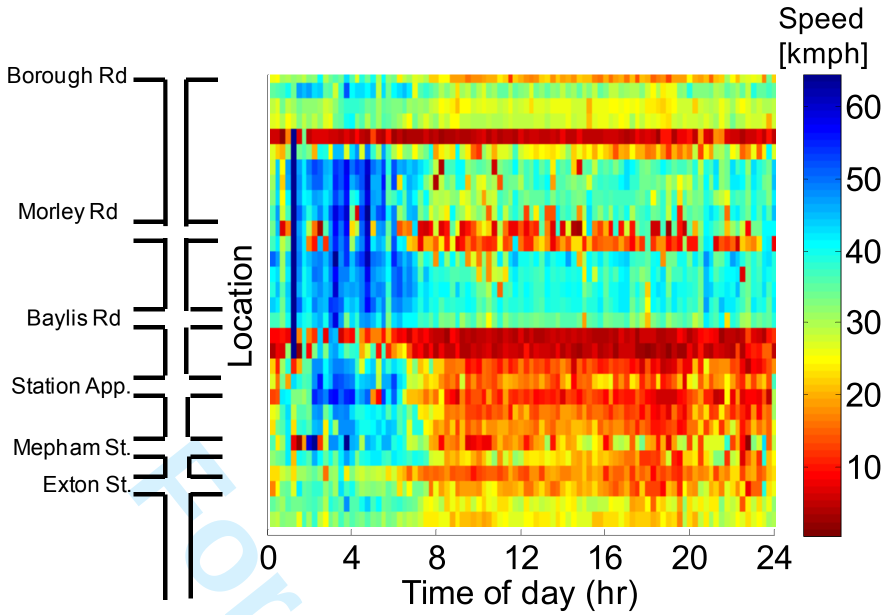

Figure 4 Trafficmaster-Speed contour along Waterloo Road, London (UK)

3. Data fusion

The study presents a kernel-based interpolation method to reconstruct the underlying traffic pattern with fine spatial and temporal details through integrating data from heterogeneous sources. It should be emphasized that the algorithm presented herein is generic and it can be applicable in other scenarios with different data sources from what we present herein. The data fusion algorithm consists of two steps: smoothing and integration.

3.1. Data smoothing

Different traffic data often come in different spatio-temporal granularity. Before integration, it is necessary to first process and reconstruct the data on a common space-time grid. Here we adopt the kernel-based interpolation method which is a popular approach used in data fusion literature (Lanckriet, et al., 2004; Camps-Valls et al., 2006). This kernel-based interpolation can also be regarded as a kind of fuzzy regression approach (Tanaka et al., 1982; Choi and Chung, 2002). Moreover, this kernel-based method is also adopted by Treiber and Helbing (2002 a, b) for fusing freeway traffic data as will be discussed in the next section.

Consider a set of traffic measurements u_i from a source taken as location x_i and time t_i, where $i = 1,2,\ldots, n$ and n is the total number of measurements. With this set of data, Treiber and Helbing's algorithm reconstructs the traffic state on a space-time domain with user-defined spatial interval δx and temporal interval δt.

We define (x,t) be the space-time coordinate on this new space-time domain, the associated traffic state $u(x,t)$ is estimated by

$$u(x,t) = \frac{1}{\Phi(x,t)} \sum_{i=1}^{M(x,t)} \phi_i(x - x_i, t - t_i) u_i(x_i, t_i), \tag{1}$$

where $\Phi(x,t) = \sum_{i=1}^{M(x,t)} \phi_i(x - x_i, t - t_i)$ is a normalizing factor. The notation ϕ_i denotes the value of a kernel smoothing or shape function at $(x - x_i, t - t_i)$, in which $(x - x_i)$ and $(t - t_i)$ are the respectively the spatial and temporal lags between the space-time of interest (x,t), and space-time of the data source (x_i, t_i). The kernel function is added here to capture the 'fuzziness' in the raw data and to smooth out high frequency noise and fluctuations. It is usual the function ϕ_i is symmetric or isotropic in (x, t) in which it depends on the quantities $\left[\frac{(x - x_i)}{\sigma}, \frac{(t - t_i)}{\tau}\right]$, where σ and τ represent respectively the spreads of the spatial and temporal influence regions. Model (1) converges to an ordinary regression model without fuzziness as σ and τ tend to zero, while the model become a simple arithmetic mean of all data points, regardless of their space-time location, as σ and τ tend to infinity.

Treiber and Helbing (2002) and Treiber et al. (2009) adopt the following exponential kernel function:

$$\phi_i(x - x_i, t - t_i) = \exp\left[-\left(\frac{|x - x_i|}{\sigma} + \frac{|t - t_i|}{\tau}\right)\right],$$

as the shape function in their freeway applications. Moreover, Treiber and Helbing (2002) suggests that σ and τ should be taken as halves of the associated spatial and temporal granularities. For example, if the spacing and sampling interval of detection are 500-m and 5-min respectively, then σ should be 250-m and τ should be 2.5 min.

Furthermore, in expression (1), $M(x,t)$ is number of data points that we consider when calculating $u(x,t)$. In extreme case, we can consider all data points in which $M(x,t) = n$. The choice of $M(x,t)$ is a trade-off between computational speed and accuracy: the larger $M(x,t)$, the more accurate the estimates while the heavier computational burden.

3.1.1. Anisotropic filter

It is known in traffic flow theory that the propagation speeds of traffic characteristics are different in free-flow and congested conditions. In free flow, traffic characteristics propagate *along* with the direction of traffic with a 'free-flow' speed v_f; in congestion, traffic characteristics travels *against* the direction of traffic with a speed w. Empirical findings show the propagation speed w is generally less than speed v_f, which suggests the traffic influence is *anisotropic* with respect to direction of influence.

To capture the anisotropic feature, Treiber and Helbing (2002 a,b) propose different formulae of the kernel function for free-flow and congested traffic conditions respectively as

$$u_{free}(x,t) = \frac{1}{\Phi(x,t)} \sum_{i=1}^{M(x,t)} \phi_i \left(x - x_i, t - t_i - \frac{x - x_i}{v_f} \right) u_i$$

and

$$u_{cong}(x,t) = \frac{1}{\Phi(x,t)} \sum_{i=1}^{M(x,t)} \phi_i \left(x - x_i, t - t_i - \frac{x - x_i}{w} \right) u_i , \tag{2}$$

in which $\Phi(x,t)$ is the corresponding normalizing factor for both cases. In the case of urban streets, we take the forward propagation speed v_f as 25 kmph as revealed from the statistics shown in Figure 2, while w is -8 kmph. Note that w is negative which represents that it is travelling against the direction of traffic.

The smoothed speed at each (x, t) is then determined as

$$u(x,t) = \gamma(x,t) u_{cong}(x,t) + [1 - \gamma(x,t)] u_{free}(x,t) , \tag{3}$$

where $\gamma(x,t)$ is weighting factor which manipulates the superposition of the free-flow and congested speeds. The weighting factor is expected to be approximately zero in free-flow (or high speeds), and approximately one at low speeds.

Treiber and Helbing (2002) and Treiber et al. (2009) adopt the "*s*-shaped" hyperbolic tangent function:

$$\gamma(x,t) = \frac{1}{2} \left[1 + \tanh\left(\frac{u_{crit} - u^*}{\Delta u} \right) \right], \tag{4}$$

where $u^* = \min[u_{free}(x,t), u_{cong}(x,t)]$; u_{crit} is a speed threshold distinguishing free-flow and congested traffic; and Δu is the transition window width adopted in the weighting function. Following our observations in Figure 2, we set the u_{crit} be 25 kmph and Δu be 5 kmph for classifying free-flow and congested data.

3.1.2 Processing journey time data

The data smoother above takes fixed data points, while journey times are measurements over a section. The journey time data will first have to be converted into equivalent series of point measurements before they can be used.

Consider a road section with a length L and an estimated journey time ω from a sample of vehicles within a time window Δt (say, Δt is 5 minutes as in the ANPR data). The average speed of this set of vehicles along the section is calculated as $\bar{v} = L / \omega$.

The following two assumptions are made:

- all sampled vehicles travel steadily (i.e. no change in speed) with this speed within the time window;
- the sampled vehicles enter the road section uniformly with a common time headway $\Delta t/(n-1)$, where n is the number of sampled vehicles in Δt.

We can then construct a set of 'virtual trajectories' of the vehicles as shown in Figure 5. The dotted lines in the figure refer to the 'true' trajectories of the sampled vehicles, which are unknown. The solid lines refer to the 'virtual' trajectories constructed. We consider that the vehicles enter the link at times s_0, s_1, …, s_{n-1}, where $s_i = s_{i-1} + \Delta t/(n-1)$ for all i = 1,2,…, (*n-1*). All vehicles travel through the link with a constant speed $\bar{v}$, and exits the link at $s_i + \omega$, where i = 0,1,2,…, (*n-1*). Denote the location '0' and '*L*' be the starting and ending points of the link respectively. For a vehicle enters the link at s_i, data points are sampled every Δs (say, 1 min) at the following space-time coordinates: $(0, s_i)$, $(\bar{v}\Delta s, s_i + \Delta s)$, $(\bar{v}(2\Delta s), s_i + 2\Delta s)$, …, (L, ω). At each of these time and location, it is regarded that the sampled point will report a speed $\bar{v}$, and hence we convert the sectional journey time measurement into a series of point measurements.

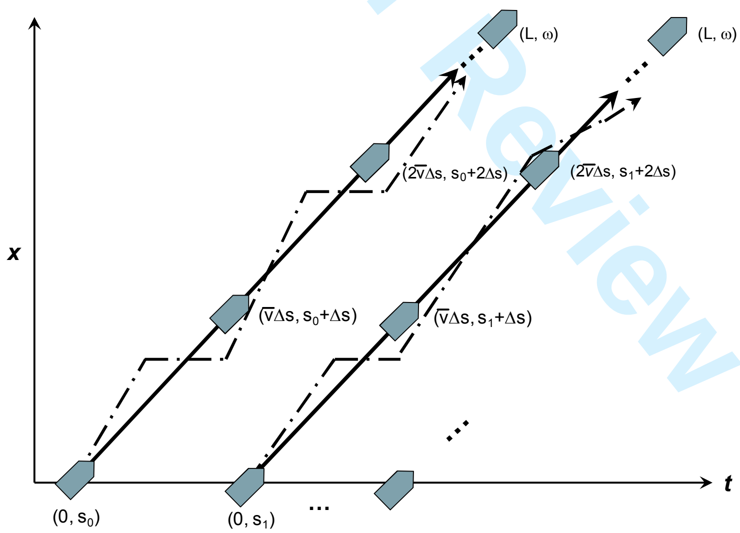

Figure 5 Converting journey times into point measurements

3.2 Data Integration

After smoothing and reconstructing the traffic data onto a common space-time grid, we can then integrate the data with the following formulation:

$$\widetilde{u}(x,t) = \sum_{k=1}^{K(x,t)} \beta_k u_k(x,t), \tag{5}$$

where $u_k(x,t)$ is the smoothed and reconstructed data field from data source k; $\widetilde{u}(x,t)$ is eventual the integrated data field; the weighting factor $\beta_k \in [0,1]$ associated with each source data can be related to various factors such as the accuracy, reliability, number and variance of measurements of data source *k*. This combination method is known as voting technique in data fusion literature (Olkin, 1992; Choi and Chung, 2002), which is essentially a weighted linear combination of information from different sources.

4. Simulation experiment

Before proceeding to the real world application, it is necessary to conduct analysis on the sensitivity and accuracy of the fusion algorithm. Due to the lack of ground truth data, we conduct the analysis on a micro-simulation test-bed. This study chooses VISSIM simulation package after considering the plausibility of the VISSIM model for replicating complex traffic dynamics.

The Waterloo Road section (Figure 1) is coded into VISSIM which is used to generate a synthetic scenario. The VISSIM simulation time step is set to be one second and the simulation period is one hour. According to field observations, the demand rates are set to be 900 vph and 100 vph respectively for the mainline (Waterloo Road) and the cross-streets. At the two signal-controlled vehicle intersections, 15% of the mainline traffic will be turning into the cross-streets, while 60% of cross-street traffic turning into the mainline. The signal timings are also set based upon observations from the field. A total of 1,516 vehicles are generated during the simulation period, from which we derive the corresponding speed field (Figure 6) with a space-time resolution at 50-m (Δx) and 1-min (Δt) respectively. At each space-time coordinate (x, t) in the figure, the associated speed $v(x,t)$ is calculated as the *average speed* of all vehicles detected within that (x,t). We regard this speed field $v(x,t)$ as the 'ground truth' for comparison.

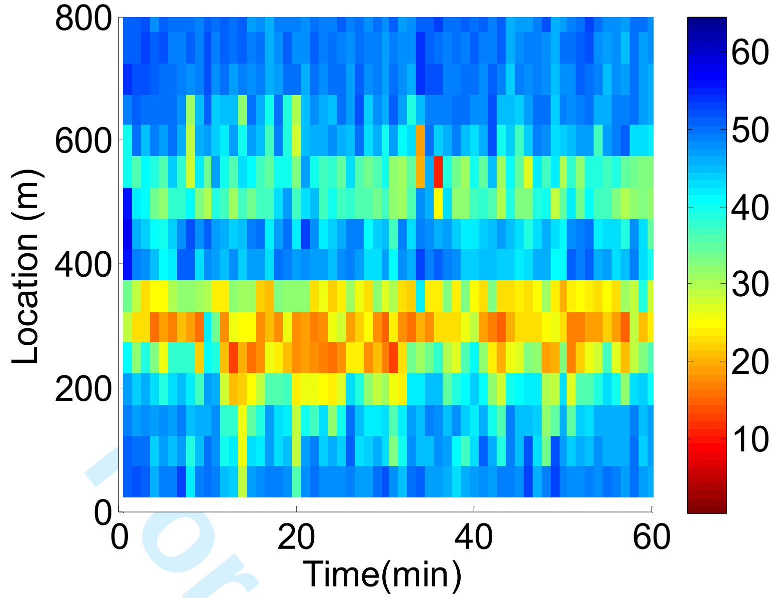

Figure 6 Speed field of VISSIM simulated data
(spatial resolution: 50-m; temporal resolution: 1-min)

To simulate the trafficmaster data, we randomly select 100 out of the 1,516 vehicles (i.e. a penetration rate of 6.6%) and assume that they are equipped with GPS devices that can report the associated positions every second.

To simulate the AVI system, we place two virtual 'cameras' at the entrance and the exit of the road stretch. It is found that 283 out of 1,516 vehicles would travel all way through the stretch and hence their associate journey times will be regarded as the 'ANPR' journey time herein. The journey times of these 283 vehicles are processed into 1-min averages.

The 'raw' trafficmaster and ANPR data are first processed by the ASM specified by formulae (1) and (3) in Section 3.1 and projected onto a common user-defined space-time grid with resolutions at 50-m (Δx) and 1-min (Δt). Following expression (5), the smoothed data are then integrated by using the following linear combination:

$$\tilde{u}(x,t) = \beta[u_{ANPR}(x,t)] + (1-\beta)[u_{GPS}(x,t)], \qquad (6)$$

for all x and t, where the weighting factor β lies between 0 and 1. We adopt a 'data-data consistency' concept (Ou, 2011) to estimate the value of this β. We first regard the overall journey time through the arterial given by ANPR is reliable. The parameter β is determined such that the difference between the corresponding journal times given by the fused data and those given by the ANPR is minimised.

With the 'ground truth' given by the VISSIM simulation, we test this concept and analyse the corresponding the overall RMSNE (Root-Mean-Square-Percentage-Error) with respect to the 'ground truth', where RMSNE is calculated as

$$RSMPE = \sqrt{\frac{1}{N}\sum_{x}\sum_{t}\left[\frac{\hat{v}(x,t)-v(x,t)}{v(x,t)}\right]^2} \, , \tag{7}$$

where $\hat{v}(x,t)$ and $v(x,t)$ are respectively the integrated (estimated) and true speed at space-time coordinate (x, t); The value of β is determined by least square estimation and β is determined to be 0.31 which gives a minimum RMSNE of 7.4%. In addition to the parameter β, we also compare the data fusion results with and without the *anisotropic* formulation proposed by Treiber and Helbing (2002). The isotropic filter can be obtained by assigning a large value to both wave speeds v_f and w in (2). Interestingly, we do not observe significant difference between the isotropic and anisotropic smoothers, where the difference in the RSMPEs produced by the two smoothers is found to be less than 1%. It suggests Treiber and Helbing's anisotropic formulation does not help to improve the traffic estimation on urban streets. An explanation for this is that the spatial and temporal granularity of traffic detection on urban networks is much finer than that in motorway, hence taking the difference in wave speeds into consideration does not have a significant effect in state estimation.

5. Real world application – Waterloo Road, London

We now present the application of the data fusion algorithm with the real world data. Given the ANPR and trafficmaster data, we find β to be 0.27 using the method described previously, and it is found that this β value is not significantly different from the determined by using simulated data. The corresponding overall RSMPE is found to be 9.6% which is slightly higher than the one we obtained from the VISSIM simulation test-bed. We suggest this is due to the complicated nature (e.g. vehicles may stop unexpectedly) in real world scenario. More sophisticated systems, such as model based fault identification and state prediction, will be developed to improve the estimation errors.

Figure 7 shows the resulting fused speed field along Waterloo Road. It shows that the fusion algorithm is able to retrieve much hidden spatio-temporal feature of traffic with either ANPR or trafficmaster data alone. Moreover, the result also shows that the fusion algorithm is able to smooth out the 'corners' (compared with Figure 4) observed in the Trafficmaster dataset through the underlying kernel function.

We note that there are several areas that require further analysis. In particular, it will be interesting and important to explore whether the fusion algorithm can restore the queue formation and dissipation process – which are important characteristics - in urban networks. Nevertheless, this will require traffic data with finer granularity - say with spatial granularity down to metres and temporal resolution down to seconds. Moreover, it is desirable to validate the performance of the fusion algorithm with real world data rather than simulated ones. We are currently exploring new datasets for further investigation into this. Meanwhile, we are conducting some research into loop detector data collected under the SCOOT urban traffic control system (see Heydecker et al., 2012) which collect flow and concentration information at up to a frequency of 4Hz (i.e. every 0.25 sec). Moreover, we are exploring the possibility of obtaining some other higher quality GPS vehicle data including taxi trajectories from Addison Lee and bus trajectories from London ibus system, which both contain more samples than the trafficmaster dataset.

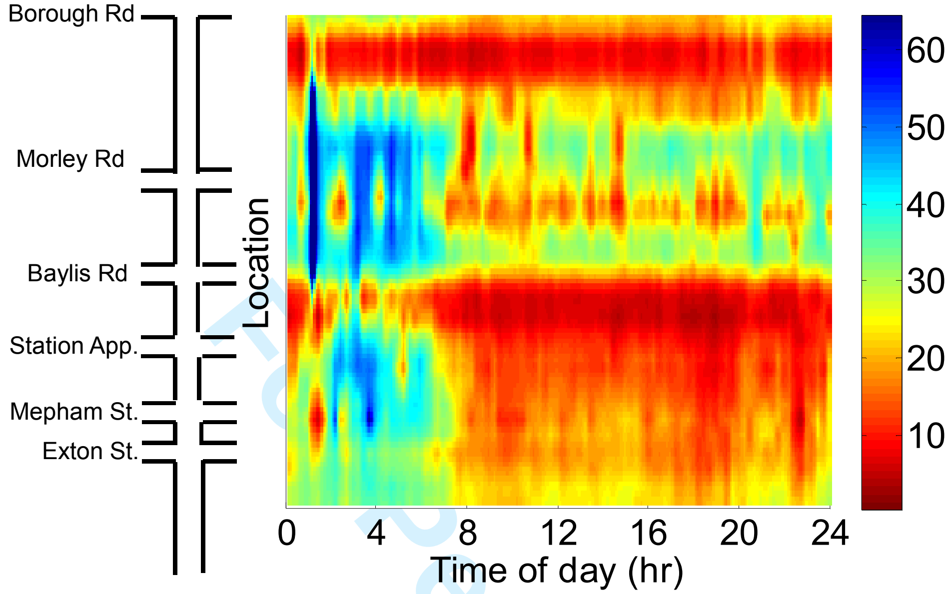

Figure 7 Speed field along Waterloo Road after fusing the ANPR and Trafficmaster data

6. Concluding remarks

This paper presents an application of kernel based interpolation method to urban traffic data fusion. The application is illustration through a case study on Central London road network. Due to the lack of ground truth data, the performance of the fusion algorithm is analysed by using a virtual test-bed generated by VISSIM micro-simulation. It is interesting to note that the anisotropic filter formulation proposed by Treiber and Helbing (2002) does not show significant improvements over the original isotropic one. This is different from the freeway case where Treiber and Helbing's formulation shows significant improvement. An explanation for this is that the spatial and temporal granularity of traffic data required in urban networks is finer than that in motorway (say, minutes in freeway case versus seconds in urban cases), hence taking the difference in wave speeds into consideration does not have a noteworthy effect.

The kernel-based algorithm presented herein does not require assumption and calibration of any traffic model. It is easy to implement and parallelize. For example, the fusion computing task can be parallelized such that each computer can be made be responsible for one specific data source over a specific space-time domain. Such properties make the fusion algorithm suitable for practical large-scale applications. The research contributes to the application of Big-Data analytics to infrastructure management.

References

Balijepalli, N, Ngoduy, D. Watling, D. (2013). Dynamic Network Loading Model using the Two-regime Transmission Model. *Transportmetrica ,* In press.

Camps-Valls, G., Gomez-Chova, L, Munoz-Mari, J, Rojo-Alvarez, J, Martinez-Ramon, M (2008). Kernel-based framework for multi-temporal and multi-source remote sensing data classification and change detection. *IEEE Transactions on Geoscience and Remote Sensing*. 46(6), 1822 – 1835.

Choi, K and Chung, Y (2002) A data fusion algorithm for estimating link travel time. *Journal of Intelligent Transportation Systems*, 7(3), 235 – 260.

Chow, AHF, Santacreu, A., Tsapakis, I., Tanaksaranond, G., Cheng, T. (2014) Empirical analysis of urban congestion. *Journal of Advanced Transportation*, 48(8), 1000-1016.

Herrera, JC and Bayen, A (2010) Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B*, 44(4), 160-481.

Heydecker, BG, Tsapakis, I., Chow, AHF (2012) Exploring Macroscopic Fundamental Diagrams for the Transportation Network in Central London. *Proceedings of the 1st European Symposium on Quantitative Methods in Transportation Systems*, September 4-7. Lausanne, Switzerland.

Kalman, RE (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35 – 48.

Kwon, J, Varaiya, P (2005) The congestion pie: delay from collisions, potential ramp metering gain, and excess demand. *Proceedings of the 84th Annual Meeting Transportation Research Board*, 9-13 January, Washington, D.C.

Lanckriet, GR, Deng, M, Cristianini, N, Jordan, M, Noble, W (2004) Kernel-based data fusion and its application to protein function prediction in yeast. *Proceedings of Pacific Symposium on Biocomputing*.

Mihaylova, L, Boel, R, Hegyi, A (2007) Freeway traffic estimation with particle filtering framework. *Automatica*, 43(2). 290 – 300.

Ngoduy, D. (2011). Low rank unscented Kalman filter for freeway traffic estimation problems. *Transportation Research Record*, 2260, 113-122.

Olkin, I (1992) Meta-analysis: Methods for combining independent studies, *Statistical Science*, 7(2), 226 – 236.

Ou, Q (2011) Fusing Heterogeneous Traffic Data: Parsimonious Approaches using Data-Data Consistency. *PhD thesis, Delft University of Technology, The Netherlands.*

Papageorgiou, M, Blosseville, J, Haj-Salem, H (1990) Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris. *Transportation Research Part A*, 24, 345 – 359.

Robinson, S, Polak, J (2006) Overtaking rule method for the cleaning of matched license plate data. *ASCE Journal of Transportation Engineering*, **132**(8), 609 – 617.

Tanaka, H, Uejima, S, Asai, K (1982) Linear regression analysis with fuzzy model. *IEEE Transaction on Systems Management and Cybernetics*, 12(6), 903 – 907.

Treiber, M, Helbing, D (2002a) Reconstructing the spatio-temporal traffic dynamics from stationary detector data. *Cooper@tive Tr@nsport@tion Dyn@mics*, **1**, 3.1–3.24.

Treiber, M, Helbing, D (2002b) An adaptive smoothing method for traffic state identification from incomplete information. *arXiv:cond-mat/0210050*.

Treiber, M, Kesting, A, Wilson, RE (2009) Reconstructing the traffic state from fusion of heterogeneous data. *Computer-aided Civil and Infrastructure Engineering*, **26**(6), 408–419.

Tsapakis, I., Turner, J, Heydecker, B, Emmonds, A, Cheng, T, Bolbol, A (2012) Effects of Tube Strikes on Journey Times in Transport Network of London. *Transportation Research Record*, 2274, 84 – 92.

Van Lint, JWC (2010) Empirical evaluation of new robust travel time estimation algorithms. *Transportation Research Record*, **2160**, 50-59.

Van Lint, JWC, Hoogendoorn, SP (2009) A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-aided Civil and Infrastructure Engineering*, **25**(8), 596-612.

Wang, Y and Papageorgiou, M (2005) Real-time freeway traffic state estimation based on extended kalman filter: a general approach. *Transportation Research Part B*, 39(2), 141-167.