Travel Time Prediction on Highways

John Davies, Alistair Duke, Sandra Stincic Clarke
BT Group plc
Ipswich, United Kingdom
john.nj.davies@bt.com,
alistair.duke@bt.com,
sandra.stincic@bt.com

Jan Rupnik, Blaž Fortuna
Artificial Intelligence Laboratory
Jozef Stefan Institute
Ljubljana, Slovenia
jan.rupnik@ijs.si,
blaz.fortuna@ijs.si

***Abstract*—We describe the development of a predictive model for vehicle journey time on highways. Accurate travel time prediction is an important problem since it enables planning of cost effective vehicle routes and departure times, with the aim of saving time and fuel while reducing pollution. The main information source used is data from roadside double inductive loop sensors which measure vehicle speed, flow and density at specific locations. We model the spatiotemporal distribution of travel times by using local linear regression. The use of real-time data is very accurate for shorter journeys starting now and less reliable as journey times increase. Local linear regression can be used to optimally balance the use of historical and real time data. The main contribution of the paper is the extension of local linear models with higher order autoregressive travel time variables, namely vehicle flow data, and density data. Using two years of UK Highways Agency (HA) loop sensor data we found that the extended model significantly improves predictive performance while retaining the main benefits of earlier work: interpretability of linear models as well as computationally simple predictions.**

Keywords—travel time prediction, data integration, traffic, machine learning, variable coefficient models

I. Introduction

In this paper, we focus on vehicle journey time prediction on highways. Accurate travel time prediction is an important problem since it enables planning of cost effective vehicle routes (for example, avoiding the congested roads) and departure times (travelling at the most efficient time while still arriving within any required time window), with the aim of saving time and fuel while reducing pollution. The main information source used is data from roadside double inductive loop sensors which measure vehicle speed, flow and density at specific locations. We use a statistical approach of modelling the spatiotemporal distribution of travel times by using local linear regression.

The approach is built upon and extends the results of Rice & van Zwet [1] and Gibbens & Saatci [2] both of whom investigated the use of inductive loop data in journey time prediction. The former tested the use of inductive loop data in the US and the latter then proved its applicability in the UK context. Their key finding was that journey times are more accurately predicted using a dynamic combination of real-time journey time information and historical averages. The use of real-time data is very accurate for shorter journeys starting now and less reliable as journey times increase and/or start times are further in the future. This is because current road conditions at a certain location are less useful as a prediction factor if the vehicle will pass that location in the future. Local linear regression can be used to optimally balance the use of historical and real time data.

The main goal of the paper is to determine the value of using additional sources of information made available by loop sensors - vehicle flow and concentration – for the task of predicting vehicle travel time when using the approach by Rice & van Zwet[1]. Additionally we would like to assess the value of higher order autoregressive features to the model.

The main contribution of the paper is the extension of local linear models with higher order autoregressive travel time variables, namely vehicle flow data, and density data. Using two years of UK Highways Agency (HA) loop sensor data we found that the extended model significantly improves the predictive performance while still retaining the main modelling benefits of the model used by Rice & van Zwet [1]: interpretability of linear models as well as computationally simple predictions.

The paper is structured as follows. We first describe the data, the model and its evaluation. This is followed by an overview of related work, positioning of our contributions against the related work, and concluded by discussion.

II. DATA

The data used in our journey time prediction algorithms is collected by the UK' Highways Agency (HA) as part of its Motorway Incident Detection and Automatic Signalling (MIDAS) system. MIDAS includes a distributed network of traffic sensors, mainly inductive loops, which are designed to facilitate the detection of incidents allowing inter alia variable message signs and advisory speed limits to be set with minimal human intervention.

An inductive loop consists of inductive loop cable which is then connected to detector units housed in a cabinet by means of feeder cables – one cable per loop pair. The inductive-loop system behaves as a tuned electrical circuit in which the loop wire and lead-in cable are the inductive elements. Energy is transmitted from the electronics units into the wire loops (usually at frequencies between 10 kHz to 200 kHz, depending on the model). When a vehicle passes over the loop or is stopped within the loop, the vehicle induces eddy currents in the wire loops, which decrease their inductance. The electronics unit senses this event as a decrease in frequency and sends a pulse to the controller signifying the passage or presence of a vehicle.

An inductive double loop pair enables not only the detection of a vehicle passing passing the detector but also the measurement of the vehicle's speed.

Loops need to be placed in locations where they are required to detect traffic queues and where signal settings relate as closely as possible to the traffic conditions. Loop pairs are installed in all running lanes of both carriageways, in exit and entry slip road lanes at junctions, in entry slip road lanes to motorway service areas, and throughout the running lanes of motorway-to-motorway link roads. On the HA road network the overall average loop site spacing for a scheme is 500 meters plus or minus 10% and loop joint chambers are provided to house all joints between loop tails and feeders (they are sited off the carriageway).

This enables 4,000 strategic monitoring sites to continuously monitor the state of the UK's main routes (approximately 5000 miles). Data is uploaded every 10 minutes. The data used in our paper consists of flow (vehicles/hour), density (vehicles/km) and speed (km/h). A raw data piece contains the ID of a road segment, a time stamp and the measurements of flow, speed and concentration (density), averaged over the last 10 minutes.

III. MODELS

First, let us introduce some notation. For our purposes, a road network consists of n segments of varying length denoted by $L_1, \dots, L_n$, where their corresponding lengths are denoted by $dist(L_i)$. Several traffic related quantities are measured at time points t_i, where $t_i - t_{i-1} = 600s$. For each link L_k we are given the following time series (where t_i denotes a point in time): flow (vehicles/hour), $Q_{L_k}(t_i)$, density (vehicles/km), $D_{L_k}(t_i)$, and speed (km/h), $V_{L_k}(t_i)$. Average travel time for link L_k over the interval $[t_{i-1} t_i]$ is denoted by $T(L_k, t_i) \coloneqq \frac{dist(L_k)}{V_{L_k}(t_i)}$. Since we model all links independently in this work, we will omit the link indices from now on.

We define δ, the forecast interval, as the time interval between the time t (for which we have real-time data) and the time at which the journey will begin. Variable t is typically the current time. For a given forecast interval δ the aim of the model is to predict the travel time $T(L, t+\delta)$ based on information available up to time t. Let $\tau(t)$ denote the time of day corresponding to t (e.g. number of hours since midnight) and $d(t)$ represent the date corresponding to t, such that

$$t \approx d(t) + \tau(t).$$

We first review the model variable coefficient model proposed by Rice and van Zwet [1]. We start by defining a separate model for each combination of link, current time of day ($\tau(t)$) and the forecast interval (δ). The model is defined as a combination of historical averages and recent information:

$$T(L, t+\delta) = a^L_{\tau(t),\delta} + b^L_{\tau(t),\delta} T(L, t). \qquad (1)$$

Here, $a^L_{\tau(t),\delta}$ and $b^L_{\tau(t),\delta}$ are two model parameters specific to the prediction scenario determined by the link (L), time of making the prediction (t), and the time of starting the journey through link L ($t+\delta$). The final model of link travel time for each link is composed of a large number of (a^L, b^L) pairs which are estimated based on the historical data using weighted least squares regression.

A key feature of the Rice and van Zwet approach is to enforce smoothness of the parameters to avoid overfitting. Each prediction scenario (L, t, δ) corresponds to a specific model $(a^L_{\tau(t),\delta}, b^L_{\tau(t),\delta})$. To achieve local model smoothness for a particular link L we require that $(a^L_{\tau(t),\delta}, b^L_{\tau(t),\delta})$ works well not only for $(\tau(t), \delta)$ but also for similar $(\tau(t'), \delta')$. Gaussian kernels are used to measure prediction scenario similarity:

$$sim\big((t,\delta),(t',\delta')\big) = K\big((t+\delta) - (t'+\delta')\big),$$

where

$$K(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

Given the training set $\{(t_1, T(L, t_1)), \dots, (t_m, T(L, t_m))\}$ for link L, the coefficients $a^L_{\tau(t),\delta}$ and $b^L_{\tau(t),\delta}$ are estimated by minimizing the following objective function (similarity weighted square loss):

$$\sum_{\substack{s \in \{t_1,\dots,t_m\} \\ t+\delta \in \{t_1,\dots,t_m\}}} \Big(\big(a^L_{\tau(t),\delta} + b^L_{\tau(t),\delta} T(L,t)\big) - T(L,s)\Big)^2 sim\big((t,\delta),(s,0)\big).$$

In this paper, we extend model by incorporating concentration and flow data, as well as using multiple historical measurements compared to using only the most recent. Extended mode is based on the following equation:

$$\begin{aligned} T(L, t+\delta) = a^L_{\tau(t),\delta} &+ \sum_{i=0}^{k_1} b^L_{i,\tau(t),\delta} T(L, t - i\Delta) \\ &+ \sum_{i=0}^{k_2} c^L_{i,\tau(t),\delta} Q_L(t - i\Delta) \\ &+ \sum_{i=0}^{k_3} d^L_{i,\tau(t),\delta} D_L(t - i\Delta), \end{aligned}$$

where Δ represents the step size (a multiple of $600s$ in our case) and k defines how many historic elements to use.

Although the parameter σ in Gaussian kernel controls the model complexity, we observed that the standard least squares algorithm often encounters numerical issues and overfitting, due to many outliers in the data. The main outliers arise when there are few vehicles on the road such as during the night where the speed measurements typically jump from 0 km/h (no vehicles) to high speeds (free flow conditions). For this reason, we used a bisquare robust regression estimator, with efficiency at the normal set to 95% [3].

A. Remark on the model

The fundamental traffic equation: $Q = D \cdot V$ and the

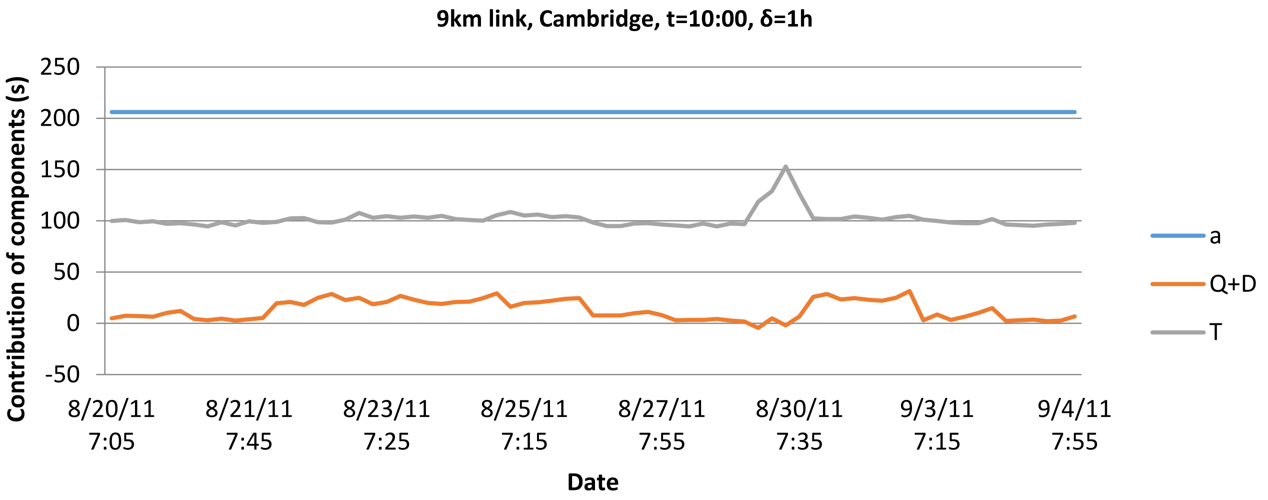

Fig. 1 Flow and concentration (Q+D) are predictive

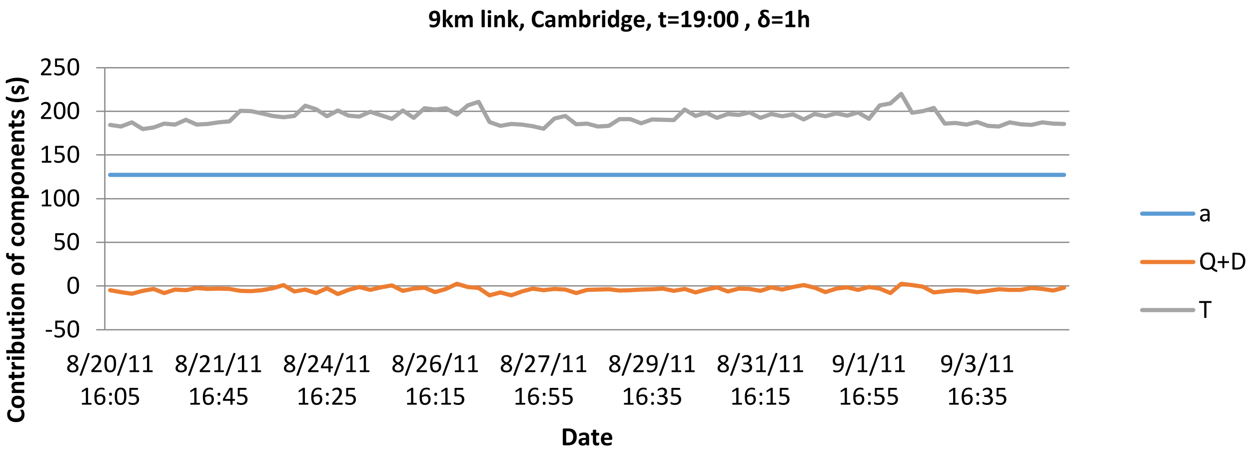

Fig. 2 Flow and concentration (Q+D) are not informative

fundamental traffic diagram [4][5] give relationships between travel time, traffic density and flow, which are not linear. For example, the flow is typically modelled as a quadratic function of the velocity. If for a given $t + \delta$ only a part of the quadratic function is observed, a linear model may be adequate (for example, if the velocities are typically above the critical velocity, which corresponds to the maximal traffic flow). Additionally, if the flow exhibits strong daily periodic behaviour, then the locally linear models for t and δ are appropriate, since $T = const \cdot D$ and Q is incorporated in the $a^L_{\tau(t),\delta}$ term.

B. Remark on the features

Large forecast intervals are typically less predictable and lead to large $a^L_{\tau(t),\delta}$ parameters, which can be interpreted as historical averages when the other parameters are small. Figures 1-4 illustrate the influence of different sets of features on the predicted travel time. We plotted four prediction settings for a 9km road link near Cambridge over a period of two weeks. We plotted the contributions of the bias term, a, the flow and concentration $(\sum_{i=0}^{k_2} c^L_{i,\tau(t),\delta} Q_L(t - i\Delta) + \sum_{i=0}^{k_3} d^L_{i,\tau(t),\delta} D_L(t - i\Delta))$ and the autoregressive part $(\sum_{i=0}^{k_1} b^L_{i,\tau(t),\delta} T(L, t - i\Delta))$. Figures 1 and 2 illustrate the contributions for a short forecast interval ($\delta = 1\ hour$), as opposed to the longer forecast interval in Figures 3 and 4 ($\delta = 8\ hour$). The flow and concentration features are not always picked up by the model: in Figures 1 and 3 they do contribute to the final prediction, as opposed to Figures 2 and 4 where they do not. In cases where the flow and concentration features are not used by the model (that is, their associated weights are close to zero), the autoregressive part is given a larger influence by the model.

IV. EVALUATION

Evaluation of the extended model was performed on 44 segments of A14 eastbound, a major road in England, selected based on data availability. The length of the selected segments constitutes 40% of A14, against a total length of all the segments of approximately 80 km. The data that we used was collected between 2011-8-19 and 2012-9-5, for a total of 383 days.

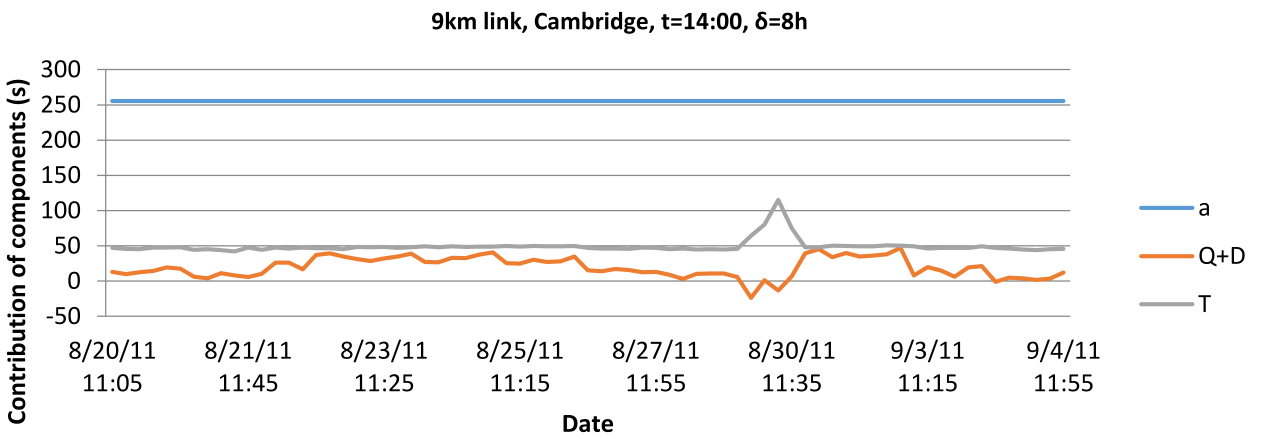

Fig. 3 Flow and concentration (Q+D) are predictive

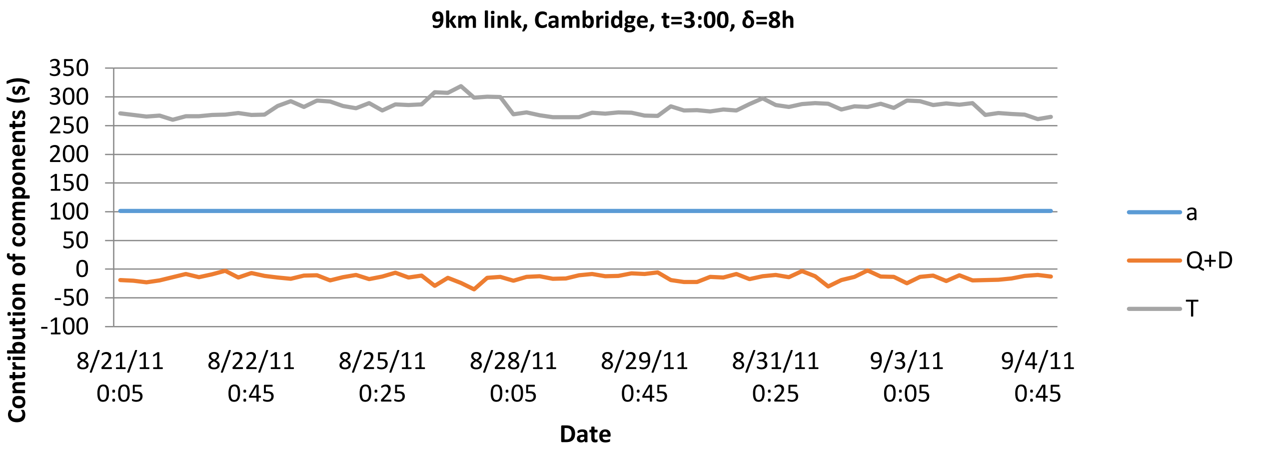

Fig. 4 Flow and concentration (Q+D) are not informative

For each month in the data set we assigned the first twenty days to the training set and the rest to the test set. In cases where several autoregressive features were used, we removed all the test instances whose information overlapped with the training set. In all tests we used time window of size 12 hours for historic data. More specifically, parameters k_1, k_2, k_3 were set to 12 and the step size Δ was set to 1 hour.

We evaluated two extreme regimes for the smoothing parameter: small σ, which correspond to locally linear models for each forecast scenario and large σ which corresponds to a single linear model (coefficients independent of $\tau(t)$ and δ).

We use the following notations in the figures presenting evaluation results:

- w1_single_S: original model with a large σ
- w1_hourly_S: original model with a small σ
- w12_single_SFC: extended model with a large σ
- w12_hourly_SFC: extended model with a small σ

We consider two error metrics: root mean squared error (RMSE) and mean absolute error (MAE). We measured RMSE since it is commonly measured in the related work, as well MAE, which we argue is more informative, since the data contains many outliers (sensor errors, highly unpredictable traffic conditions). This can also be seen in the results. For example, Figure 5 shows MAE and Figure 6 shows RMSE for different values of δ parameter. MAE increases as we predict further into

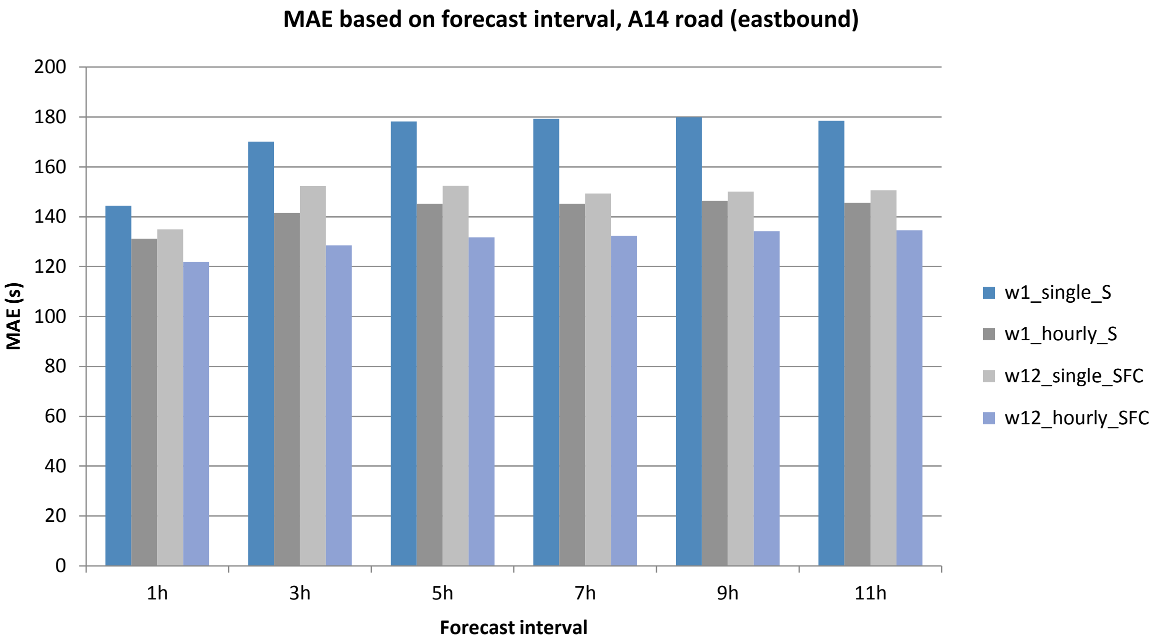

Fig. 5 MAE over different forecast intervals. The error increases as the forecast intervals increase.

the future (increase δ). However, due to several outliers in the data, RMSE results do not share this intuition.

The performance was measured over different road links and different δ and t parameters and over a range of test samples. The default way of aggregating the results is averaging, except in the case of road segments, which are not directly comparable due to varying lengths. When several links are aggregated, as it is in Figure 5, Figure 6 and Figure 7 the errors are first averaged over every link over all parameters that do not appear in the figures and the resulting per-link aggregates are summed over in the end.

When comparing the four models we observe that the model extended with autoregressive, flow and density information consistently leads to significant performance improvements. Figure 5 illustrates how predictive performance decreases as we increase the δ parameter, which is intuitive, since larger forecast intervals correspond to larger uncertainty. Figure 6 measures RMSE of different models based on different choices of δ. Similar performance improvements are observed, though the results are less intuitive when comparing the performances for different δ choices (the worst performance is observed at $\delta = 3h$). A general trend can be observed in all figures: locally linear models (small σ) improve the predictive performance, and including our additional features is beneficial. Figure 7 shows predictive performances while varying the t parameter, where the errors are averaged over different δ values. We observe three main error peaks: 2:00, 8:00 and 17:00. The traffic during the night time is highly variable due to the low number of vehicles in free flow (speed is related to individual driving styles and not bound). This makes predictions at $t = 2{:}00$ less reliable. Degradation at 8:00 and 17:00 is related to unstable traffic conditions at rush hours. Figure 8 illustrates the variations of predictive performance over the 10 longest road segments. Certain segments are significantly harder to model due to their position in the highway system. For example, segment *J1 – M1/M6*, where the A14 joins the motorways M1 and M6 has more variability in the data compared to segment *J29 – J26* which does not have any major crossroads and is located in a rural area.

All experiments indicate that our extended model outperforms the original model proposed in [1]. The improvement is due to additional autoregressive features (historic time window of 12 hours) and addition of new features to the model, namely flow and concentration. As we have seen in Figures 1 -4, all these elements provide significant and complementary contribution to the prediction. Additionally, we have confirmed the findings from [1] that locally linear models provide significant improvement in prediction accuracy.

V. RELATED WORK

Travel time prediction studies can be categorised based on what data sources were used, the spatial and temporal resolution of the data and the modelling methodology adopted.

There are three basic types of data used in travel time prediction studies: probe vehicle tracking data obtained by GPS[6]; fixed vehicle detectors such as traffic loops[1]; and using incident data (such as accident data)[7]. The data used in our work was acquired by a network of inductive traffic loop sensors.

The data in our study is sampled every ten minutes, which can be considered as a low frequency when compared to 30 seconds[8], which we consider as high frequency data. High frequency data is suitable to study phenomena like traffic congestion, which can form or end in short time periods. The spatial resolution of the data used in this study is heterogeneous:

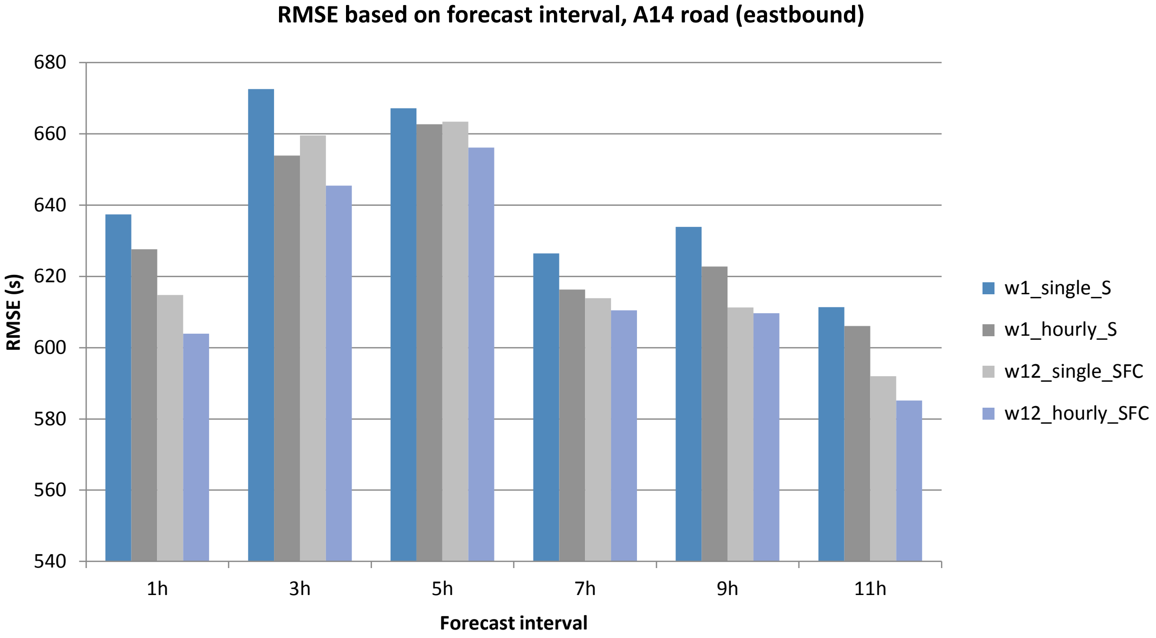

Fig. 6 RMSE over different forecast intervals. The results are less intuitive (δ=3h,δ=5h exhibit higher errors) due to outliers.

the lengths of the road segments that we model span from a typical few hundreds of meters to 10km in the longest case. Similar data was used in [7], whereas many other studies are based on homogeneous, finer grained road segments[9] . The heterogeneity of the road segments plays an important role when the models are evaluated: relative errors become less informative (50% error on a 100 m road segment is not comparable to a 50% error on a 10 km segment), leading us to focus on absolute errors.

Some of the most popular approaches to modelling include: neural networks[10][11], support vector machines[12][13], clustering[14][15], nearest neighbours[17], time series stochastic models[18], local linear models[1][2], etc. Our work

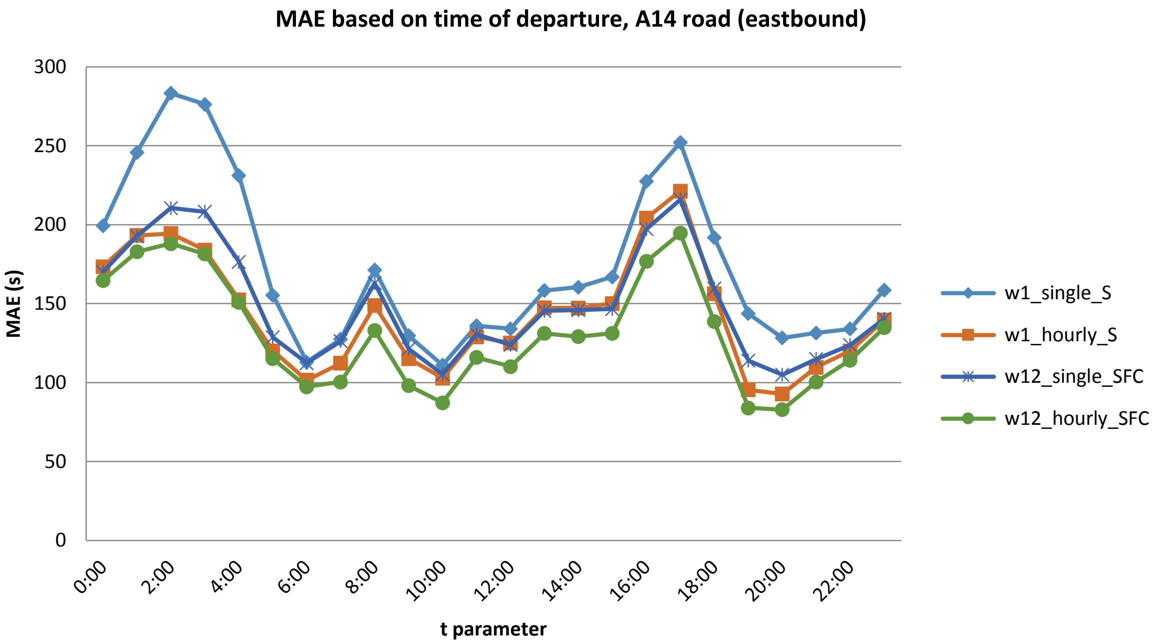

Fig. 7MAE based on different t parameters. We observe three error peaks during 2AM, morning and afternoon rush hours.

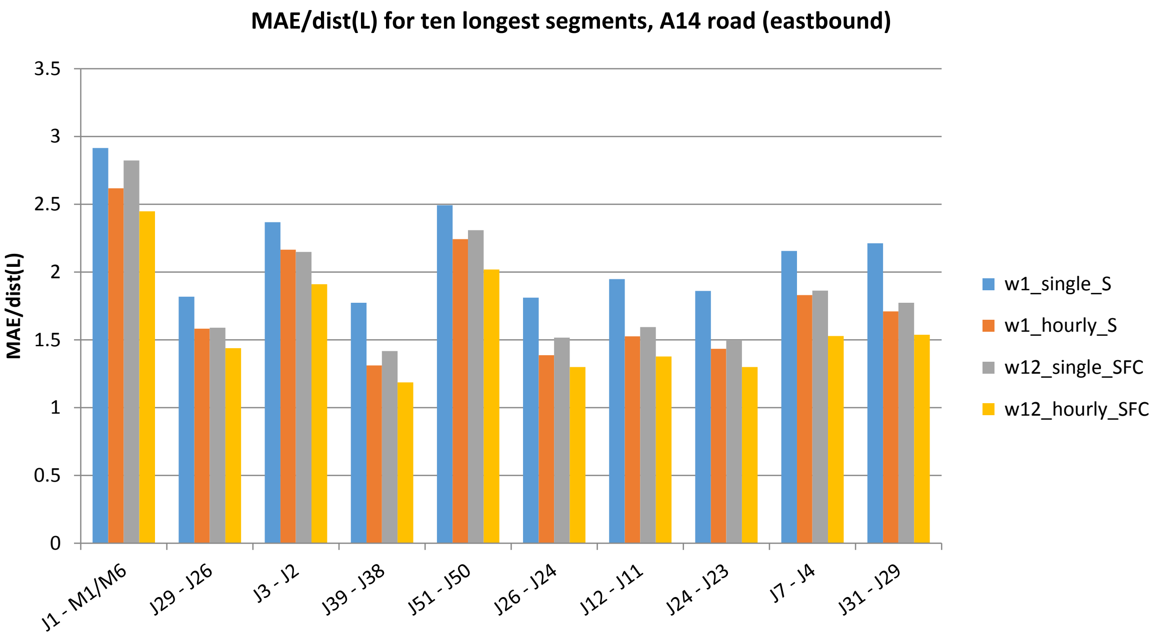

Fig. 8 MAE for different road segments (normalised over segment lengths). The first two sets of columns represent the two longest links (9km). The first segment (J1 - M1/M6) exhibits higher unpredictability.

is closely related to[1], since both works are based on the local linear models. The main difference lies in the feature sets: we used higher order autoregressive features based on three sources of data (flow, density and velocity) as opposed to [1] where the authors constructed an autoregressive model of order one based solely on velocity.

VI. CONCLUSIONS

We explored the impact on the predictive performance of a simple local linear regression when additional sources of traffic information are used (flow, concentration, and higher order autoregression). We find that the additional features can lead to improved performance and we have also confirmed that local linear modelling is preferable to global linear modelling for the task of travel time prediction.

In our experiments we observe that rush hours and night time are more difficult for prediction. The errors during the night time are less important, since they are related to low traffic density and high variation in vehicle speed (free flow), which is not relevant to routing applications, since a static model of average speed based on link lengths is often sufficient when traffic is in free flow.

In this work we modelled the spatio-temporal distribution of travel times separately for each link. An interesting direction for future work is to build more global models by extending the feature vectors of each link by information measured by its neighbours.

ACKNOWLEDGMENT

The work reported here was carried out as part of the Stride Internet of Things project[1] using the Stride information hub[2]. The authors acknowledge the European Commission for its support and partial funding and the partners of the research project: FP7-318452 MobiS. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 636160-2. This publication only reflects the author's view and the European Union is not liable for any use that may be made of the information contained therein.

REFERENCES

[1] John Rice and Erik van Zwet, "A simple and effective method for predicting travel times on freeways" in Proc. 4th IEEE Conference Intelligent Transportation Systems, 2001, pp. 229–234.

[2] Gibbens RJ, Saatci Y. Data, modelling and inference in road traffic networks. Philos Trans A Math Phys Eng Sci. 2008 Jun 13;366(1872):1907-19. doi: 10.1098/rsta.2008.0020.

[3] Holland, P. W., and R. E. Welsch. "Robust Regression Using Iteratively Reweighted Least-Squares." Communications in Statistics: Theory and Methods, A6, 1977, pp. 813–827.

[4] Greenshields, B. D., D. Schapiro, and E. L. Ericksen. Traffic Performance at Urban Street Intersections, Technical Report No. 1, Yale Bureau of Highway Traffic, Yale University, New Haven, Conn., 1947.

[5] Transportation Research Circular, Number E-C149, June 2011, 75 Years of the Fundamental Diagram for Traffic Flow Theory

[6] Mei Chen and Steven I. J. Chien, "Dynamic freeway travel time prediction using probe vehicle data: Link-based vs. path-base," in Transportation Research Record 1768. Washington, DC: TRB, Nat. Res. Council, 2001, pp. 157–161.

[1] Stride was funded by the UK government's innovation agency, InnovativeUK

[2] See http://www.stride-project.com/

[7] Wei, Chien-Hung, and Ying Lee. "Development of freeway travel time forecasting models by integrating different sources of traffic data." Vehicular Technology, IEEE Transactions on 56.6 (2007): 3682-3694.

[8] Benjamin Coifman, “Estimating travel times and vehicle trajectories on freeways using dual loop detectors,” Transp. Res., vol. 36A, no. 4, pp. 351–364, 2002.

[9] Xiaoyan Zhang and John. A. Rice, “Short-term travel time prediction,” Transport Research, vol. 11C, no. 3/4, pp. 187–210, 2003.

[10] Van Lint, J. W. C., S. P. Hoogendoorn, and Henk J. van Zuylen. "Accurate freeway travel time prediction with state-space neural networks under missing data."Transportation Research Part C: Emerging Technologies 13.5 (2005): 347-369.

[11] Lee, Ying. "Freeway travel time forecast using artifical neural networks with cluster method." Information Fusion, 2009. FUSION'09. 12th International Conference on. IEEE, 2009.

[12] Wu, Chun-Hsin, Jan-Ming Ho, and D. T. Lee. "Travel-time prediction with support vector regression." Intelligent Transportation Systems, IEEE Transactions on 5.4 (2004): 276-281.

[13] Vanajakshi, Lelitha, and Laurence R. Rilett. "Support vector machine technique for the short term prediction of travel time." Intelligent Vehicles Symposium, 2007 IEEE. IEEE, 2007.

[14] Asamer, Johannes, and Kashif Din. "Prediction of Velocities on Motorways by k-Means Clustering." Artificial Intelligence, 2008. MICAI'08. Seventh Mexican International Conference on. IEEE, 2008.

[15] Nath, Rudra Pratap Deb, Nihad Karim Chowdhury, and Masaki Aono. "A Naïve Clustering Approach in Travel Time Prediction." INTERNATIONAL JOURNAL OF DATA ENGINEERING (IJDE): 62.

[16] Sun, Hongyu, et al. "Use of local linear regression model for short-term traffic forecasting." Transportation Research Record: Journal of the Transportation Research Board 1836.1 (2003): 143-150.

[17] Liu, Xiaobo, Steven I. Chien, and Mei Chen. "An adaptive model for highway travel time prediction." Journal of Advanced Transportation (2012).

[18] Chowdhury, Nihad Karim, et al. "Development of an effective travel time prediction method using modified moving average approach." Knowledge-Based and Intelligent Information and Engineering Systems. Springer Berlin Heidelberg, 2009. 130-138.