Modelling Dynamic Space-Time Autocorrelations of Urban Transport Network

T. Cheng, J. Wang, J. Harworth, B.G. Heydecker, A.H.F. Chow

Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom.
Telephone: +44 207 679 2738
Email: tao.cheng@ucl.ac.uk

1. Introduction

Various methods for modelling space-time data have been proposed over the years, including multivariate autoregressive integrated moving average (ARIMA) models and its extension space time autoregressive integrated moving average (STARIMA) models (Pfeiffer and Deutsch, 1980). In these time series models, autocorrelation is accounted for in the autoregressive and moving average terms. Parameter estimates are fixed globally both spatially and temporally. The models assume that the correlation in data can be adequately described by such globally set parameters, but this may not be the case if the correlation between data is dynamic, which it is likely to be on road transport networks (Cheng et al, 2011). For instance, traffic theories say that the current conditions on a section of road are influenced to some extent by the previous conditions of adjacent road sections along both upstream and downstream directions (see for example, Lighthill and Whitham, 1955; Richards, 1956). In congested conditions, the influence will come mainly from downstream whereas in free flowing conditions the influence will come from upstream. On a road network comprising hundreds or thousands of links, such spatio-temporal autocorrelation structure is dynamic (in time) and heterogeneous (in space). Yue and Yeh (2008) show the correlation between locations on a road network determines the forecast ability of a space-time model. This fact has been recognised in previous studies that achieve improved results by incorporating a dynamic structure in their weighting systems (Min et al, 2009; 2010; Min and Wynter, 2011). The aim of this study is to model dynamic autocorrelations of road transport network data By modifying tradition model to a generic dynamic model which capture the autocorrelation locally (heterogeneity) and dynamically (dynamic state of the network) over the traditional time series models. The proposed model is tested with traffic data collected from Central London. The result shows that the performance of estimation and prediction is improved on average through the proposed modifications.

2. A Localised Dynamic Space-Time Model - NSTARIMA

STARIMA model considers the observation at location i during time interval t to be a weighted linear combination of observations in its spatial neighbours at previous time intervals. Consider that a road network, in which measurements (e.g. speeds, journey times, etc) are collected on N links over a time period T. Let $\mathbf{z}(t)$ be an N-dimensional column vector containing the observations $z_i(t)$ on each link i, where i = 1,2,…,N, during each time interval t, where t = 1,2,…,T. STARIMA model states that

$$\hat{\mathbf{z}}(t) = \sum_{k=1}^{p}\sum_{h=0}^{m_k}\varphi_{kh}\mathbf{W}^{(h)}\mathbf{z}(t-k) - \sum_{l=1}^{q}\sum_{h=0}^{n_l}\theta_{lh}\mathbf{W}^{(h)}\boldsymbol{\varepsilon}(t-l), \tag{1}$$

in which $\hat{\mathbf{z}}(t)$ is a N-dimensional column vector of predictions on all links i at time t. The first term in the equation is the autoregressive (AR) component, while the second term is the moving average (MA). The term, $\boldsymbol{\varepsilon}(\cdot)$, is a N-dimensional column vector of residual on each link. The spatial lag (h) represents the spatial distance between two locations. The spatial orders associated with each k^{th} or l^{th} temporally lagged term in AR and MA components are respectively m_k and n_l. The spatial order specifies the spatial extent that could have an effect on the link of interest i within the temporal lags of k and l. The notation φ_{kh} and θ_{lh} are the model parameters to be calibrated. The matrix $\mathbf{W}^{(h)}$ is an N x N spatial weight matrix for spatial lag h. This spatial weight matrix $\mathbf{W}^{(h)}$ contains the set of weights w_{ij} specifying the degree of spatial correlation between links i and j (see Kamarianakis and Prastacos, 2005; Getis, 2009).

We identify several deficiencies of the above STARIMA model for traffic modelling and propose a new dynamic time series model – which we call NSTARIMA - that includes several new features. Details are discussed below.

2.1 Spatial orders

Traditional STARIMA model considers the spatial orders to be fixed and preset for the associated temporal lag. It may not be appropriate for traffic modelling as the spatial influences vary under different traffic conditions due to different speeds encountered (Min et al., 2008). This study relaxes such assumption and considers the spatial orders to be dynamic and dependent on traffic state. Given the model updating time interval (Δt), the spatial order $m_k(t)$ at time t associated with temporal lag k is determined as

$$m_k(t) = \arg\min_{m}\left\{ m \,\middle|\, \sum_{i_0=i}^{i-m} z_{i_0}(t)L(i_0) > k\Delta t \right\}, \tag{2}$$

where $L(i_0)$ is the length of the intermediate link i_0 between the link of interest i and the spatial extent m. Essentially, $m_k(t)$ returns the number of links that traffic can proceed toward the point of interest i in a time period of $k\Delta t$.

2.2 Spatial weight matrix

The spatial weight matrix ($\mathbf{W}^{(h)}$) is usually regarded as the physical distances between the corresponding locations. In road traffic setting, the correlation of traffic at two locations does not only depend on the spatial distance, but also on the traffic conditions. We propose a novel spatial weight matrix which takes the traffic states into account. For a link pair (i, j), the corresponding element in the spatial weight matrix is defined as

$$w_{ij} = \frac{v_j(t) - v_i(t)}{D_{ij}}, \tag{3}$$

where $v_i(t)$ and $v_j(t)$ are the respective average speeds on links i and j during time interval t; D_{ij} is the distance between i and j. The speed $v_i(t)$ is defined to be zero if no data is observed on the link during time t. The spatial weight matrix derived using (4) is time-varying and traffic state dependent.

2.3 Model formulation

We formulate our new time series model – NSTARIMA - as

$$\hat{\mathbf{z}}(t) = \sum_{k=1}^{p} \sum_{h=0}^{m_k(t-k,i)} \boldsymbol{\varphi}_{kh} \mathbf{W}^{(h,t-k,i)} \mathbf{z}_i(t-k) - \sum_{l=1}^{q} \sum_{h=0}^{n_l(t-l,i)} \boldsymbol{\theta}_{lh} \mathbf{W}^{(h,t-l,i)} \boldsymbol{\varepsilon}_i(t-l). \tag{4}$$

Original STARIMA model is specified by a single global set of parameters (φ_{kh}, θ_{kh}) for the entire network. In this new model, the model parameters are N x N diagonal matrices ($\boldsymbol{\varphi}_{kh}$ and $\boldsymbol{\theta}_{kl}$):

$$\boldsymbol{\varphi}_{kh} = diag([\varphi_{kh}]_1, [\varphi_{kh}]_2, ..., [\varphi_{kh}]_N) \text{ and } \boldsymbol{\theta}_{lh} = diag([\theta_{lh}]_1, [\theta_{lh}]_2, ..., [\theta_{lh}]_N), \tag{5}$$

where $[\varphi_{kh}]_i$ and $[\theta_{lh}]_i$ are the parameters for each link i. It is noted that the NSTARIMA model covers the STARIMA and ARIMA models as special cases.

3. Case Study

The test network, which comprises 22 links in Central London, is selected for this study as shown in Figure 1 with arrows showing the directions of traffic. It has variable link lengths, ranging from 473.4m to 3.85km with an average length of 1.4km. Journey times of vehicles across the network are measured by Automatic Number Plate Recognition (ANPR) system which is operated by Transport for London (TfL). The raw journey time data are aggregated into 5-minute averages.

After discussing with TfL, data from 16 Feb 2009 to 30 Mar 2009 (43 days in total) are selected for the case study. The dataset is divided into two sets. The first 36 days are used for calibration which determines the temporal orders (p, q) and the model parameters ($\boldsymbol{\varphi}_{kh}$ and $\boldsymbol{\theta}_{kl}$). The remaining 7 days are used for validation which compares the predictions made by the calibrated model and the actual observations.

The experiment consists of three stages: identification, calibration, and validation.

- *Identification* refers to the determination of temporal orders – autoregressive (p) and moving average (q) – in the time series model.
- Given the temporal orders, the model parameters are the determined in the *calibration* step by using a least square error approach. This study compares three

different time series models: ARIMA, original STARIMA and modified STARIMA.

- Finally, in *validation*, predictions made by the calibrated models are compared with the actual observations.

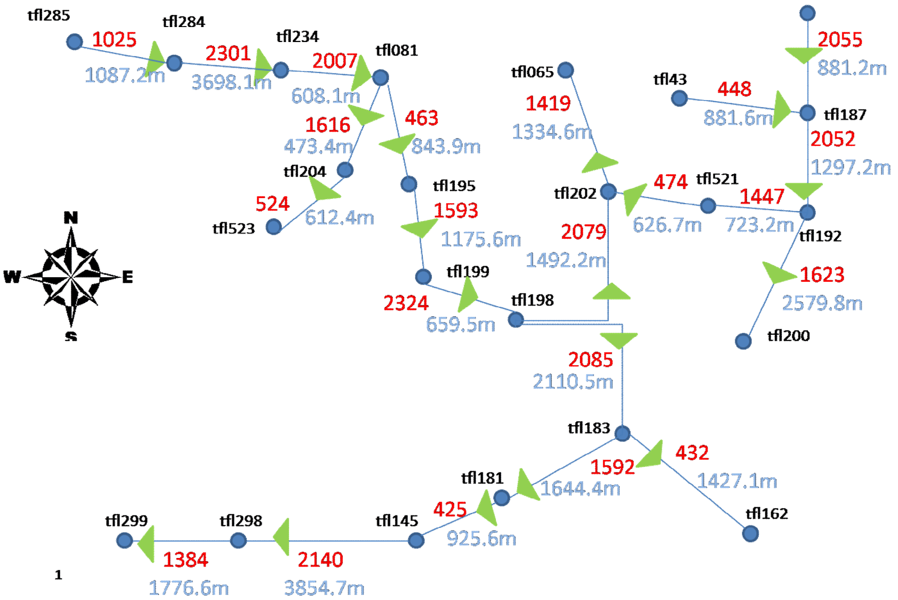


Figure 1 London test network (Cheng et al., 2011)

Figures 2 and 3 respectively show the R-square and root mean square error (RMSE) of the 22 links, which are arranged in ascending order of lengths. Results show that there is no single model dominates the others. However, if we summarise average R-square and RMSE of all links, it shows that NSTARIMA outperforms traditional STARIMA and ARIMA model.

Figure 4 shows the predictions from 12:00 to 16:00 on 30 Mar. Overall, original STARIMA has the worst average prediction results as the heterogeneity and dynamics of the urban road network cannot be well captured (Cheng et al, 2011). However, the NSTARIMA outperforms the other models on average.

4. Conclusions

This paper proposes a new space-time series model – NSTARIMA - for road traffic modelling. The proposed model is tested with journey time data obtained from the Automatic Number Plate Recognition (ANPR) system in Central London. Results show the average prediction accuracy of the NSTARIMA is better than traditional STARIMA and ARIMA model. This indicates that the new NSTARIMA can capture heterogeneity and dynamics of road traffic by modifying the original STARIMA as proposed. Given travel time is an important index for measuring transport system performance, the work reported here will contribute to the literature of traffic analysis and management.

5. Acknowledgements

The authors would like to thank Transport for London for providing the journey time data. This research is carried out under the STANDARD project, which is sponsored by the UK Engineering and Physical Sciences Research Council (EP/G023212/1).

6. References

Cheng, T., Haworth, J., Wang, J. (2011) Some implications of complexity in network spatio-temporal autocorrelation structures, *Journal of Geographical Systems* (accepted).

Getis, A. (2009) Spatial Weight Matrices. *Geographical Analysis*, 41(4): 404-410.

Kamarianakis, Y. and Prastacos, P. (2005) Space-time modelling of traffic flow. *Computers and Geosciences*, 31: 119 -133.

Lighthill, M.J., Whitham, J.B., 1955, On kinematic waves: I. Flow movement in long rivers. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society (London)*, A229: 281-345.

Min, X., Hu, J., Chen, Q. , Zhang, T. , and Zhang, Y, 2009, Short term traffic flow forecasting of urban network based on dynamic STARIMA model. *Proceedings of 12th International IEEE Conference on Intelligent Transportation Systems. (ITSC):* 1–6.

Min, X., Hu, J., and Zhang, Z, 2010, Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. *Proceedings of 13th International IEEE Conference on Intelligent Transportation Systems (ITSC),* 19-22 September: 1535-1540.

Pfeifer, PE and Deutsch, SJ, 1980, A three-stage iterative procedure for space-time modelling. *Technometrics*, 22(1): 35-47.

Richards, P.I., 1956, Shockwaves on the highway. Operations Research 4: 42-51.

Yue, Y. and Gar-On Yeh, A., 2008, Spatiotemporal traffic-flow dependency and short-term traffic forecasting. *Environment and Planning B: Planning and Design*, 35: 762-771.

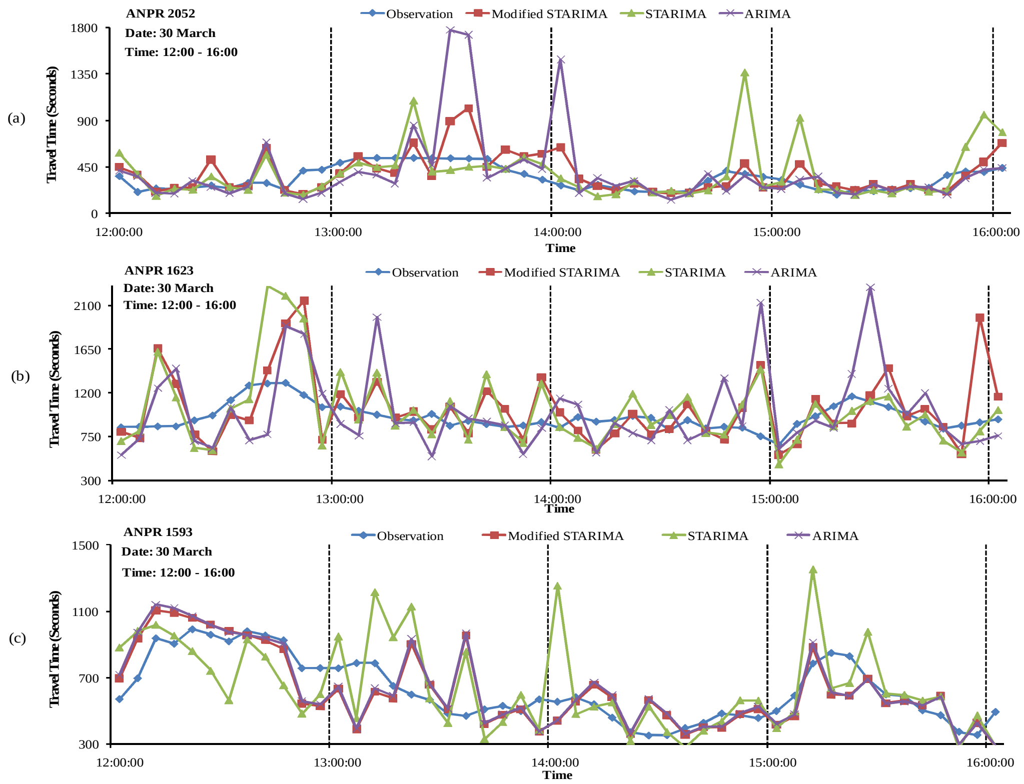
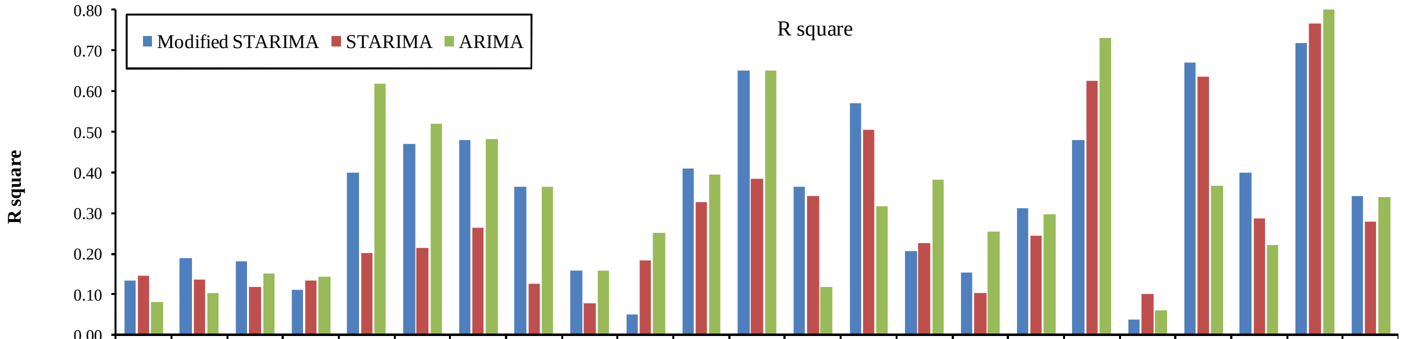
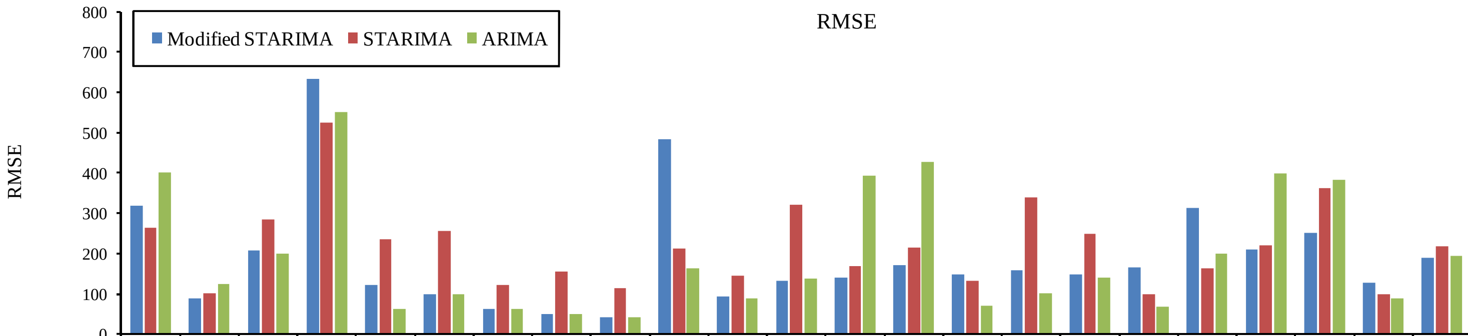


Figure 4. Prediction plots of three links 2052 (a), 1623 (b), and 1593 (c) at 12:00 - 16:00 on 30 Mar using three different models NSTARIMA, STARIMA, and ARIMA



Link id	1616	2007	524	474	2324	1447	463	2055	448	425	1025	1593	2052	1419	432	2079	1592	1384	2085	1623	2301	2140	
Length of link	473m	608m	612m	627m	660m	723m	844m	881m	882m	926m	1087m	1176m	1297m	1335m	1427m	1492m	1644m	1777m	2111m	2580m	3698m	3855m	average
Modified STARIMA	0.1332	0.1886	0.1804	0.1119	0.3982	0.4708	0.4792	0.3635	0.1577	0.0506	0.4082	0.6501	0.3642	0.5703	0.2059	0.1532	0.3113	0.4802	0.0386	0.6701	0.4006	0.7182	0.3411
STARIMA	0.1449	0.1349	0.1191	0.1339	0.2001	0.2146	0.2634	0.1262	0.0785	0.1842	0.3277	0.3832	0.3411	0.5054	0.2257	0.1044	0.2449	0.6248	0.0997	0.6361	0.2867	0.7645	0.2793
ARIMA	0.0812	0.1029	0.1518	0.1429	0.6175	0.5187	0.4829	0.3635	0.1574	0.2526	0.3945	0.6495	0.1179	0.3172	0.3809	0.2542	0.2975	0.7298	0.0610	0.3668	0.2205	0.8003	0.3392

Figure 2. R-square comparison of three different models



Link id	1616	2007	524	474	2324	1447	463	2055	448	425	1025	1593	2052	1419	432	2079	1592	1384	2085	1623	2301	2140	
Length of link	473m	608m	612m	627m	660m	723m	844m	881m	882m	926m	1087m	1176m	1297m	1335m	1427m	1492m	1644m	1777m	2111m	2580m	3698m	3855m	average
Modified STARIMA	319	87	208	634	122	99	62	49	43	483	94	133	140	171	148	157	148	165	314	211	251	126	189
STARIMA	264	100	285	524	237	257	123	154	113	212	145	322	167	216	131	338	248	98	162	221	362	98	217
ARIMA	402	125	199	550	63	100	62	49	43	163	88	138	392	426	72	102	140	68	200	399	383	88	193

Figure 3. RMSE comparison of three different models