A Dynamic Spatial Weight Matrix and Localised STARIMA for Network Modelling

T Cheng, J Wang, J Haworth, BG Heydecker, AHF Chow

Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom.

Telephone: +44 20 7679 2738; Fax: Fax: +44 (0) 20 7679 3042

Email: {tao.cheng; w.jiaqiu; j.haworth; b.heydecker; andy.chow}@ucl.ac.uk

Abstract

Various statistical model specifications for describing spatio-temporal processes have been proposed over the years, including the space-time autoregressive integrated moving average (STARIMA) and its various extensions. These model specifications assume that the correlation in data can be adequately described by parameters that are globally fixed spatially and/or temporally. They are inadequate for cases in which the correlations among data are dynamic and heterogeneous, such as network data. The aim of this paper is to describe autocorrelation in network data with a dynamic spatial weight matrix and a localized STARIMA (LSTARIMA) model that captures the autocorrelation locally (heterogeneity) and dynamically (nonstationarity). The specification is tested with traffic data collected for Central London. The result shows that the performance of estimation and prediction is improved compared with standard STARIMA models that are widely used for space-time modelling.

Introduction

Modelling of spatio-temporal processes requires the use of specialized models that account for its special properties, the most widely studied of which is spatio-temporal *autocorrelation*. Autocorrelation is the tendency for near observations to be more similar than distant observations in space and/or time (see Cliff and Ord 1969; Box and Jenkins 1994). Its presence violates the assumption of stationary; that is, that the data have constant mean and variance. Autocorrelation is inherent in spatio-temporal data and renders traditional statistical techniques such as ordinary least squares (OLS) inefficient. However, to date, little research has been carried out about how autocorrelation varies in a spatio-temporal context, and its implications for space-time modelling (Cheng et al 2011a). This is reflected in the range of models that currently are applied to spatio-temporal data, with structures that are globally fixed both spatially and temporally (Pfeifer and Deutsch 1980; Giacinto 2006; Cheng et al 2011b; Cressie and Wikle 2011).

In autoregressive integrated moving average (ARIMA) models (Tiao and Box 1981) and space-time autoregressive integrated moving average (STARIMA) models (Pfeiffer and Deutsch 1980, 1981), the spatio-temporal series must be transformed to stationarity, usually by differencing. It then can be modelled using a combination of autoregressive (AR) and/or moving average (MA) terms. STARIMA is a generalization of a family of space-time models, with STARMA (Hooper and Hewings 1981), STMA and STAR models (Griffith and Heuvelink 2012) being special cases. The latter, which is more widely applied than STARIMA, does not include a moving average term, and correlation in space and time is captured only by the autoregressive parameters that are fixed globally both spatially and temporally. Elhorst (2001) provides a general framework for identifying autoregressive distributed lag models in space and time.

Spatial panel data models (Elhorst 2003; Elhorst et al. 2010) can be specified to account for autocorrelation in one of two ways: either with a spatial autoregressive process in the error term - a spatial error model (equivalent to a spatial moving average) - or with a spatial autoregressive dependent variable - a spatial lag model. Parameter estimates for panel data models can be heterogeneous in space to account for local variations in autocorrelation properties. More recently, models have been put forward for continuous space-time panels (Oud et al. 2012). In space-time geostatistical models (Gething et al. 2008; Heuvelink and Griffith 2010), the spatio-temporal process is decomposed into a deterministic trend and a space-time stochastic residual, and a space-time covariance function is fitted to the residuals. Then Kriging is used for forecasting (Heuvelink and Griffith 2010). Characterization of the covariance function usually is simplified by assuming spatio-temporal stationarity, which may be restrictive. Furthermore, geostatistical models are less accurate when extrapolating rather than interpolating. Griffith (2010) provides a good overview of the progress to date in statistical space-time modelling.

The assumption of a stationary spatio-temporal process has been shown to be unrealistic in some circumstances. For example, traffic theories say that the current conditions on a section of a road can be influenced by the previous conditions of adjacent road sections in either upstream or downstream directions depending on the degree of congestion (see, for example, Lighthill and Whitham,1955; Richards 1956). In congested conditions, the influence comes mainly from downstream, whereas in free-flowing conditions, the influence comes 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) (Cheng et al 2011a). For example, Cheng et al (2011a) reveal that the spatio-temporal autocorrelation between unit journey times recorded on a 22-link subset of London’s road network is spatially heterogeneous and temporally dynamic. Spatial heterogeneity results from variation in the level of correlation between individual road links across the study area. Temporal dynamics result from changes in the strength of correlation between locations over time, with stronger correlation apparent in the morning peak period compared with the inter-peak and evening peak. Furthermore, the size of the spatial neighbourhood also changes with time, becoming smaller in congested conditions and larger in free flowing conditions. Capturing the most relevant information at any point in time is a problem of determining the instantaneous forecastability of neighbourhood data using a dynamic spatial weight and a dynamic spatial neighbourhood.

Progress has been made in introducing spatial heterogeneity and/or temporal dynamics to space-time models. Giacinto (2006) propose a generalized space-time autoregressive moving average (STARMAG) model that generalizes the STARMA model to include spatially varying AR and MA coefficients. The generalised STARIMA (GSTARIMA) model of Min et al. (2010) also allows the AR and MA parameters to vary by location, and outperforms a standard STARIMA model in terms of forecasting accuracy. Although these methods allow for spatially dynamic parameter estimates, their spatial structure is fixed to an extent, as the size of the spatial neighbourhood considered is the same for each location. They also are fixed temporally. Min and Hu (2009) present a dynamic form of the STARIMA that accounts for temporal dynamics. They replace the traditional distance weighted spatial weight matrix with a temporally dynamic matrix that reflects the current traffic turn ratios observed at each road intersection. The weight matrix can be updated in real time based on current conditions, but the method is limited to intersection-based flow data, and is fixed spatially.

A simpler and more generalizable approach that accounts for both spatial heterogeneity and temporal nonstationarity draws from a set of models, each of which pertains to a certain traffic state. Min and Wynter (2011) devise a multivariate STAR

(MSTAR) model in which the weight matrix is selected from a set of templates that reflect typical traffic states. Average speeds are used to calculate a dynamic spatial neighbourhood based on the number of links that can deliver their traffic to the current location within the forecast horizon. This approach shows impressive forecasting performance for multiple steps ahead, and the authors claim it is scalable to large networks. Ding et al. (2010) propose a similar approach in which the weight matrix varies based on the current level of service (LOS). The approaches of Min et al. and Ding et al. go some way to accounting for the spatial and temporal variability in conditions on a road network. However, calculation of the templates they use is based on historical traffic conditions. Therefore, a natural tendency exists for them to perform better when conditions are close to average conditions. How well they perform when the conditions differ from the typical conditions captured in the templates is unclear.

The aim of this paper is to describe the modeling of dynamic (transient) and heterogeneous autocorrelation in network data with improved traditional models that constitute a generic dynamic model capable of capturing the autocorrelation locally (spatial heterogeneity) and dynamically (temporal nonstationarity), providing an improvement over the traditional space-time series models. We incorporate the concepts of dynamic spatial weights and dynamic spatial neighbourhood using a *dynamic spatial weight matrix* to model the spatial heterogeneity and temporal nonstationarity in network data, which is introduced in next Section.

Dynamic Spatial Weight Matrices

This section describes the construction of a dynamic spatial weight matrix for road network data. The weight matrix has an adjacency and weight structure that is dynamic in time and space, which is updated based on the current traffic condition.

Spatial weight matrix for networks

A spatial weight matrix W encapsulates what we know or hypothesize about the structure and behaviour of some phenomenon over space. When a temporal dimension is included, W also must encapsulate our knowledge or hypotheses about spatial structure and behaviour *over time*. W comprises two components, a spatial adjacency structure and a spatial weighting structure.

Spatial Adjacency

Drawing from graph theory, an arrangement of spatial units may be viewed as a graph $G = (N, E)$ with a set N of n nodes and a set E of e edges joining pairs of nodes. The incidence structure of this graph is defined by the presence or absence of an edge (i,j) linking nodes i and j, and can be represented by an $N \times N$ binary [0,1] adjacency matrix in which non-zero elements signify edges (Peeters and Thomas 2009). Two nodes directly linked by an edge are adjacent and termed first-order spatial neighbours. The adjacency matrix containing all first-order relations between spatial units is termed its first-order adjacency matrix. Second-order spatial neighbours of a node are the first-order neighbours of its first-order neighbours (excluding itself) and so on. By following the paths between nodes in the graph, adjacency matrices $W_1, W_2, \ldots W_k$ of orders up to k can be defined. In networks, an alternative definition of adjacency often is used in which i and j are edges where a variable is observed, and the nodes represent connections between them. Black (1992) proposes this formulation, which has been applied to transport networks (Black and Thomas 1998) and migration flows (Chun 2008) amongst other phenomena, and is the definition we adopt in this study. Because they often are used to represent flows, networks also can include another dimension, namely direction. Transport networks fall somewhere between directed and undirected networks because vehicles can move only in one direction on a carriageway while exerting influence in two directions. Therefore, the influence of upstream and downstream may be different, necessitating a different formulation for the spatial weight of each.

Spatial Weights

Spatial weights w_{ij} is the element of an adjacency matrix W that describes the perceived influence on spatial unit i of its neighbour j. A weight can be chosen in various ways, the simplest of which is the binary scheme previously outlined. Application-specific schemes include the length of shared border or distance between centroids in areal data (Cliff and Ord 1969), the length of road links in network data (Kamarianakis and Prastacos 2005) and the resources of actors in social networks (Leenders 2002). It to Row normalizing a spatial weight matrix (making all rows sum to one) is common practice. However, in some studies, column normalization has been used, allowing the matrix to represent influence exerted by i rather than accepted influence from j (Leenders 2002).

The choice of weighting scheme is non-trivial, and can be very important because different weight matrices often lead to different inferences being drawn, and can introduce bias into an analysis. The effect of this bias has been explored in the spatial literature (see Stetzer 1982, Florax and Rey 1995, Griffith 1996, Griffith and Lagona 1998) and the network literature (Paez et al 2008). In spatio-temporal data, assuming that the relative contributions of the spatial neighbours of a unit remain the same across all times may not be reasonable. For example, weather conditions are spatio-temporally correlated, but the direction of dependence relates to wind direction, which constantly is changing. On road networks, the direction of dependence relates traffic conditions (Chandra and Al-Deek 2008, Cheng et al. 2011a).

In the next section, we introduce the dynamic spatial weight matrix, which accounts for spatio-temporal nonstationarity by extending existing spatial weight matrix structures in the context of dynamic network processes.

A Dynamic Spatial Weight Matrix

A *dynamic spatial weight matrix* is a spatial weighting scheme that has the flexibility to account for autocorrelation structures that are nonstationary in time and/or heterogeneous in space. This is achieved by incorporating two key concepts: a *dynamic spatial neighbourhood,* which means the size of a spatial neighbourhood can change at each time step; and, a *dynamic spatial weight,* which means that the influence of each neighbour can change at each time step. These two concepts can be incorporated into existing space-time model structures by modifying the spatial weight matrix W at link i with a time varying matrix, $W^{(h,t)}$, where t is a time index, and $h=1,2,\ldots,m$ is the spatial order at time t. In this specification, a spatial neighbourhood is updated by changing the value of h; m is the maximum size of a spatial neighbourhood, and a spatial weight is updated by changing its i,j element in $W^{(h,t)}$, where $i, j = 1,2, \ldots, N$ are the spatial units. Updating these values is application specific, and depends on a researcher's view of the nature of the autocorrelation between locations. To illustrate this point, we use the example of road networks, which are networks of flows where spatio-temporal correlation depends on the traffic state.

A dynamic spatial neighbourhood of road networks

Assuming we have link (road section) based traffic data with sampling time interval Δt (e.g., 5 minutes). The effective spatial neighbourhood $m(t)$ of a link i can be defined as the neighbourhood of links that can deliver their traffic to link i within Δt , based on the information available at time t-1. Links that fall outside the range of Δt cannot have an influence, and hence are excluded. This definition ensures a parsimonious specification of spatial neighbourhood. A spatial neighbourhood is larger in free flowing conditions, and smaller in congested conditions.

A dynamic spatial weight of road networks

Here we assume that prediction is based upon historic travel time, and the dynamic spatial weight between a pair of links is calculated as a function of their relative traffic speeds. If we see a drop in speed on one link, we expect this decrease to translate in to a drop in speed on an adjacent link, and vice versa. For all link pairs (*i*, *j*) that are of spatial lag *h*, the corresponding $w_{ij}^{(h)}$ is defined as

$$w_{ij}^{(h)}(t) = \frac{v_j(t) - v_i(t)}{v_i(t)}, \qquad \text{(if } j \text{ is upstream of } i\text{)} \qquad (1)$$

$$w_{ij}^{(h)}(t) = \frac{v_i(t) - v_j(t)}{v_i(t)} \qquad \text{(if } j \text{ is downstream of } i\text{)} \qquad (2)$$

where $v_i(t)$ and $v_j(t)$ are the respective average traffic speeds on links i and j at time t. The entry w_{ij} takes the value of zero if the spatial lag between i and j is not h.

Here the neighbourhood matrix tends to equilibrate the speed differentials over space. Given a link pair (*i*, *j*), suppose that traffic on link i proceeds at a lower speed than upstream link j (i.e., $v_i(t) < v_j(t)$), which gives $w_{ij} > 0$. The contribution of traffic from link j increases the travel time on link i due to the arrival of higher speed upstream flow. In contrast, suppose that traffic on link i proceeds at a lower speed than downstream link j

(i.e., $v_i(t) < v_j(t)$), which gives $w_{i,j} < 0$. The contribution of traffic from link j decreases the travel time on link i due to its higher speed downstream flow.

A Localized Space-Time Model: LSTARIMA

In this section, we define a new space-time model, the localized space-time autoregressive integrated moving average (LSTARIMA) model. Like the traditional STARIMA model, LSTARIMA makes use of a spatial weight matrix W to model the influence of the spatio-temporal neighbourhood. However, it relaxes the globally fixed parametric structure of STARIMA by allowing the AR and MA parameters to vary by location, which allows it to account for spatial heterogeneity. Furthermore, it accounts for temporal nonstationarity by allowing the size of the spatial neighbourhood to vary with time. In this sense, it is a locally dynamic space-time model.

Model Specification

Let $z(t)$ be an N-dimensional column vector containing the observations $z_i(t)$ on each link i, where $i = 1,2,\ldots,N$, during each time interval t, where $t = 1,2,\ldots,T$. The conventional STARIMA model can be defined as

$$\hat{\mathbf{z}}(t) = \sum_{k=1}^{p}\sum_{h=0}^{m_k} \phi_{kh}\mathbf{W}^{(h)}\mathbf{z}(t-k) - \sum_{l=1}^{q}\sum_{h=0}^{n_l} \theta_{lh}\mathbf{W}^{(h)}\boldsymbol{\varepsilon}(t-l) + \boldsymbol{\varepsilon}(t) \tag{3}$$

The first term in equation (3) is the AR component, while the second term is the MA. The parameters p and q are the AR and MA orders respectively. The term $\boldsymbol{\varepsilon}(\cdot)$ is an N-dimensional column vector of residuals on each link, and h is the spatial order, which represents the order of spatial separation between two locations. The parameters m_k and

n_l are the spatial orders associated with the k^{th} and l^{th} temporally lagged terms in the AR and MA components, respectively. They specify the size of the spatial neighbourhood that could influence the link of interest i within temporal lags k and l. The notation ϕ_{kh} and θ_{lh} are the AR and MA parameters respectively, to be calibrated for the entire network. The matrix $\mathbf{W}^{(h)}$ is an $N \times N$ spatial weight matrix for spatial lag h, containing the set of weights w_{ij} specifying the assumed relationship between i and j (see Kamarianakis and Prastacos 2005, Getis 2009). The number of parameters to be calibrated in equation (3) is $(p \times m_k + q \times n_l)$.

We extend the standard STARIMA model to account for spatial heterogeneity and temporal nonstationarity using the following formulation, which we call localized STARIMA (LSTARIMA):

$$\hat{\mathbf{z}}_i(t) = \sum_{k=1}^{p_i} \sum_{h=0}^{m_k(t-k,i)} \phi_{i,kh} \mathbf{W}^{(h,t-k,i)} \mathbf{z}_i(t-k) - \sum_{l=1}^{q_i} \sum_{h=0}^{n_l(t-l,i)} \theta_{i,lh} \mathbf{W}^{(h,t-l,i)} \varepsilon_i(t-l) + \varepsilon_i(t) \tag{4}$$

where $W^{(h,t-k,i)}$ and $W^{(h,t-l,i)}$ are the elements of dynamic spatial weight matrix $W^{(h,t)}$ pertaining to link i at temporal lags k and l. LSTARIMA has a separate set of AR and MA parameters for each link i, which are stored in $N \times N$ diagonal matrices $\mathbf{\Phi}_{kh}$ and $\mathbf{\Theta}_{lh}$ such that

$$\mathbf{\Phi}_{kh} = diag([\phi_{1,kh}], [\phi_{2,kh}], \dots, [\phi_{N,kh}]) \text{ and } \mathbf{\Theta}_{lh} = diag([\theta_{1,lh}], [\theta_{2,lh}], \dots, [\theta_{N,lh}]) \tag{5}$$

where $[\phi_{i,kh}]$ and $[\theta_{i,lh}]$ are the parameters for each link i (i=1,2,…,N). The number of parameters that needs to be calibrated in equation (4) is $\left(p \times m_{k(t-k,i)} + q \times n_{l(t-l,i)}\right)$, although the whole model needs to calibrate N links of the entire network.

The STARIMA and ARIMA models can be viewed as special cases of the LSTARIMA model. For example, if $\mathrm{p_1 = p_2 = p_N}$ and $\mathrm{q_1 = q_2 = q_N}$ (i.e., p and q are spatially fixed), $m_k(t-1,i) = m_k(t-2,i) = \ldots = m_k(t-k,i)$ and $n_l(t-1,i) = n_l(t-2,i) = \ldots = n_l(t-l,i)$ (i.e., the spatial influence of adjacent links does not change over time), and $[\phi_{1,kh}] = [\phi_{2,kh}] = \cdots = [\phi_{N,kh}]$ and $[\theta_{1,lh}] = [\theta_{2,lh}] = \cdots = [\theta_{N,lh}]$ (i.e., all parameters are the same for all of the links), then LSTARIMA [equation (4)] becomes a STARIMA model [equation (3)]. This is unlikely for road network data, but for other data sets such as annual temperature, spatial autocorrelations may not change rapidly overtime. Moreover, If $m_k(t-k,i)=0$ and $n_l(t-l,i)=0$, (i.e., the adjacent links produce no spatial influence, then the LSTARIMA becomes an ARIMA. This reduction might happen when traffic is flowing freely (or possibly is highly congested), resulting in the speeds on all the links being more or less the same.

In the next sections, we use LSTARIMA(p_i, q_i), ARIMA(p_i, d_i, q_i) and STARIMA (p, d, q) specifications to represent the models because differencing (which is described by parameter d) is not needed for the LSTARIMA model.

Model Calibration

Equation (4) is a time series model that considers the spatial influence of adjacent links in networks. Thus, its parameters can be estimated by means of standard time series calibration algorithms. The procedure of parameter estimation in equation (4) can be regarded as the minimization of the following sum of squared errors function:

$$S(\widehat{\boldsymbol{\beta}}_i) = \sum_{t=1}^{T}\left(\hat{\mathbf{z}}_i(t) - \sum_{k=1}^{p_i}\sum_{h=0}^{m_k(t-k,i)} \phi_{i,kh}\mathbf{W}^{(h,t-k,i)}\mathbf{z}_i(t-k) + \sum_{l=1}^{q_i}\sum_{h=0}^{n_l(t-l,i)} \theta_{i,kh}\mathbf{W}^{(h,t-l,i)}\boldsymbol{\varepsilon}_i(t-l)\right)^2 \tag{6}$$

where T is the number of observations in time, $\mathbf{z}_i(t)$ is the observation vector at time t and link i, $\boldsymbol{\varepsilon}_i(t)$ is the random error vector at time t, and $\widehat{\boldsymbol{\beta}}_i = [\hat{\phi}_{10}, \hat{\phi}_{11}, \cdots, \hat{\phi}_{pm}, \hat{\theta}_{10}, \hat{\theta}_{11}, \cdots, \hat{\theta}_{qn}]$.

Equation (6) presents a nonlinear least squares minimisation problem because $\boldsymbol{\varepsilon}_i(t)$ is required for the calibration of the MA parameter θ_{lh}, but is unknown a priori. Thus, $\boldsymbol{\varepsilon}_i(t)$ must be estimated first in order to determine ϕ_{kh} and θ_{lh}. Furnishing appropriate starting values is important in order to ensure convergence of the optimization procedure. Although trial-and-error can be used to implement the parameter optimization process, it cannot guarantee convergence. Moreover, an exhaustive search is needed, which is time-consuming.

Hannan and Rissanen (1982) demonstrated that the Hannan-Rissanen (HR) algorithm is an effective approach for parameter optimization of time series models. It makes use of the residuals of a high order AR model to feed the $\varepsilon_i(t)$ as initialized values. Then, parameters ϕ_{kh} and $\boldsymbol{\theta}_{lh}$ are calibrated using the linear least squares method. The methodology has been proven valid in their work, and has been broadly accepted in practice. Here we use the same procedure for LSTARIMA model calibration by considering $\varepsilon_i(t)$ to be an independent random variable after the spatio-temporal auto-correlation has been fully modelled by the dynamic weight matrix. This result is verified by testing for spatial, temporal and spatio-temporal autocorrelation in the predictive residuals in the subsequent case study.

Although Box and Jenkins's algorithm (Box and Jenkins, 1970, p. 498-505) was proposed earlier than the HR algorithm, it doesn't provide the details of how to furnish appropriate initial values (epsilon) in order to ensure convergence of the optimization procedure of model calibration. Given the implementation method and mathematical proof has been provided (Hannan and Rissanen, 1982), HR algorithm is common used in practice, and is chosen here to calibrate our model.

A Case Study

The LSTARIMA model is designed specifically to deal with highly heterogeneous and nonstationary spatio-temporal (network) processes, an example of which is road traffic. On traffic networks, the spatio-temporal relationship between observations recorded at detector locations is dependent on the traffic state, which is constantly changing. In this section, we use an empirical example to demonstrate the model building procedure for a LSTARIMA model in the context of travel time prediction on London's road network.

The study area and road traffic network data

The case study area is identical to that used in Cheng et al. (2011a). It consists of 22 road links located in central London, UK (Fig. 1a). The topological representation of the test network can be seen in Fig. 1b. The average length of road links in the test network is 1.4km, with minimum and maximum lengths of 0.473km and 3.85kmm respectively.

The data are travel times collected using automatic number plate recognition (ANPR) technology by Transport for London (TfL) as part of their London Congestion Analysis Project (LCAP). The raw observations are 5-minute aggregated travel times (in seconds). The dataset spans 166 successive days (Monday to Sunday), from 24th May, 2010 to 5th Nov, 2010, which were selected after discussions with Transport for London (TfL).

To reduce noise in the data, only the period between 6 am and 9pm is used because capture rates often are low overnight.

The traffic pattern differs between weekends and working days, and within the working week it differs from Monday to Friday. Wednesday is considered to be a neutral day during the week. Strictly speaking, traffic is different everyday (even for a single link), even between Saturday and Sunday. A recent investigation into the autocorrelation structure of traffic data in a study by Cheng et al. (2011a) demonstrates that the autocorrelation of the traffic road network is nonstationary in both time and space, revealing that the data violate the assumption of spatial homogeneity and temporal stationarity of the STARIMA model. Many ways exist to deal with these kinds of differences in a traffic study. For example, modelling each day of the week separately, or grouping them into working days and weekends. Differencing either daily and/or weekly normally is required to transform the time series into stationary data so that it can be accommodated by an ARIMA or STARIMA model. The proposed LSTARIMA model tackles this challenge by using a dynamic weight matrix, which requires no such transformation.

To train and predict the LSTARIMA model, the dataset was separated into two subsets: the training set (24th May–24th Oct, 2010, 154 days) for calibrating the model parameters; and, the testing set (25th Oct-5th Nov, 2010, 12 days) for evaluating the prediction performance of the model. The raw travel time data have been converted to unit travel times (seconds/kilometre) to allow comparability between travel times on links of differing lengths.

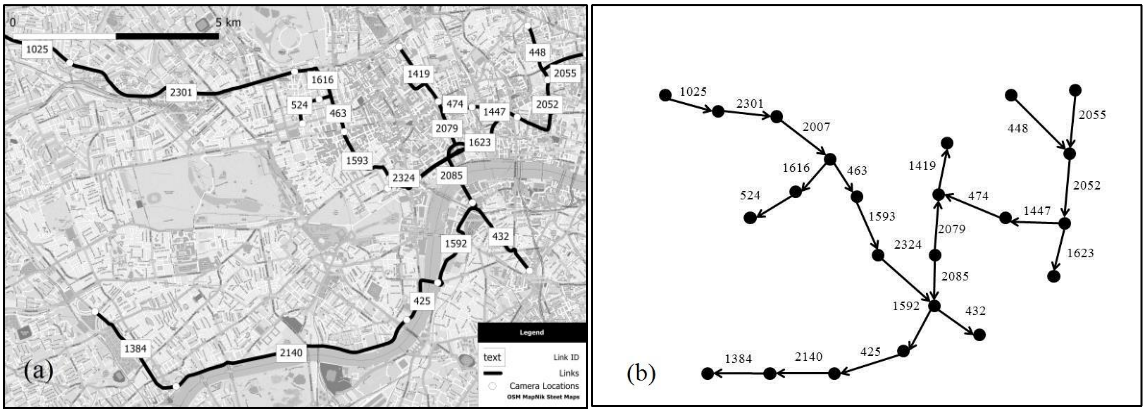


Figure 1. Selected road network in central London. (a) spatial location of selected links in central London, (b) network diagram of links; arrows represent traffic flow direction, numbers are link IDs (Cheng et al. 2011a).

Construction of a dynamic spatial weight matrix

As discussed previously, the impact of the spatial neighbourhood can be modelled using a dynamic spatial weight matrix. The following three steps achieve this purpose: 1) Model the adjacency structure; 2) Determine the dynamic spatial order; and, 3) Calculate the dynamic spatial weights.

Step 1: build a spatial adjacency matrix

The first step is to build a spatial adjacency matrix based on the topological structure of the network, which appears in Figure 1. Spatial adjacency matrices of spatial order up to 3 were constructed using the method described in this paper. Figure 2 shows the sketch map of $W^{(1)}$, where 1 indicates that two links are spatially adjacent, and "-" (zeros have been replaced with "-" for clarity of presentation) that they are not. This matrix is used in both the STARIMA and LSTARIMA model specifications. Several nodes in the adjacency matrix have entries of zero (in the respective column/row). This is the border effect. The 2nd-order of adjacency of these nodes is much more connected. When this adjacency definition is applied to a large network, such boundary effects are not substantial.

$W^{(1)}$	425	432	448	463	474	524	1025	1384	1419	1447	1592	1593	1616	1623	2007	2052	2055	2079	2085	2140	2301	2324
425	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-
432	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
448	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
463	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
474	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
524	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1025	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-
1384	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1419	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1447	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1592	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1593	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
1616	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1623	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2007	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-
2052	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-
2055	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-
2079	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-
2085	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-
2140	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2301	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-
2324	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	-

Figure 2 Sketch map of spatial adjacency matrix $W^{(1)}$

Step 2: determine the dynamic spatial order

The second step is to determine the dynamic spatial order for each link in the network using the method described in this paper. Figures 3 (a) and (b) show sketch maps of the dynamic spatial order of L463 between 06:00 and 21:00 on 24 May, 2010. Figure 3(a) reveals that, in most cases, L463 only can deliver its traffic to its first-order spatial neighbour within the forecasting horizon, and rarely reaches its second-order spatial neighbour. In contrast, Figure 3(b) portrays the situation upstream, revealing that, in most cases, L2301 (a second-order spatial neighbour of L463) can deliver its traffic to L463. Investigation of the dynamic spatial order for all 22 links reveals considerable variation across links, with a maximum spatial order of three.

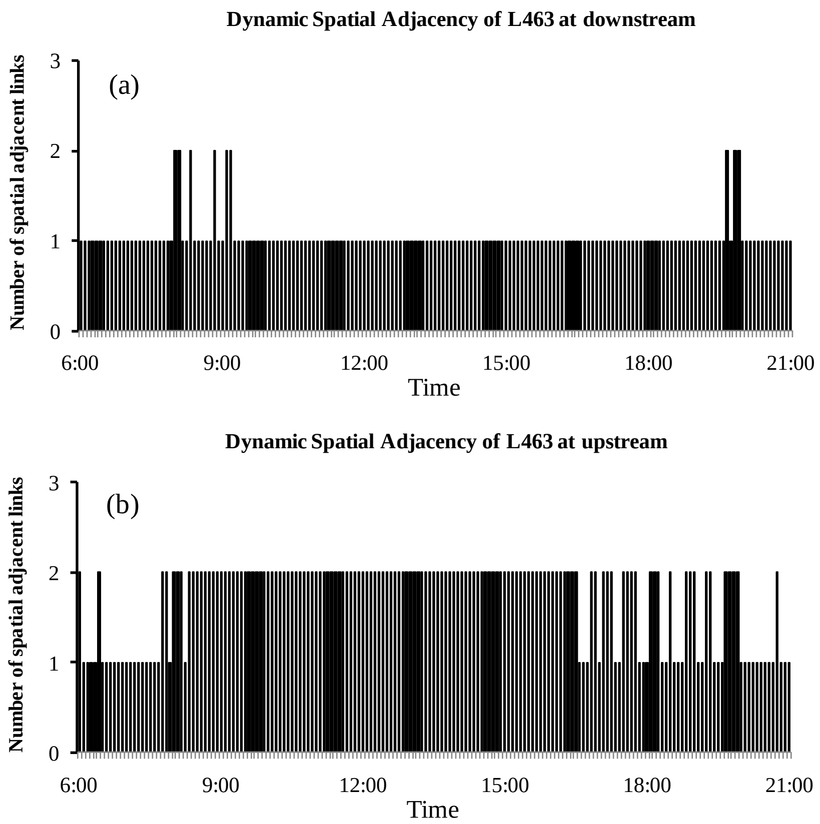


Figure 3. Sketch map of dynamic spatial adjacency in L463 downstream (a) and upstream (b).

Step 3: calculate dynamic spatial weights

The third step is to calculate dynamic spatial weights by means of equation (1), which is the difference of the average speeds of two adjacent links divided by the speed of the target link (i.e., L463). Here, the average speed of each link is the reciprocal of unit travel time. Table 1 shows a snapshot of the dynamic spatial weight matrix of L463 for 7:00 am to 9:00am on 24 May, 2010. The weights, which are updated every 5 minutes, represent the strength of spatial influence of adjacent links on the target link. In this case, L463 involves three upstream links (L2301, L2007, and L1616), and two downstream links (L1593 and L2324). Table 1 shows that the weights vary across time, reflecting the dynamics of the spatial influence of the road traffic network.

Table 1. Snapshot of the dynamic spatial weight of L463

	Upstream links			Downstream links	
Spatial order *h*	**1st**	**2nd**		**1st**	**2nd**
Temporal order k	L2007 (len=0.6km)	L2301 (len=3.7km)	L1616 (len=0.5km)	L1593 (len=1.2km)	L2324 (len=0.6km)
7:00	0	8.71	0.21	-1.82	0
7:05	0	8.80	-0.25	-3.13	0
7:10	0	2.31	-0.07	0.00	0
7:15	0	10.29	0.87	-1.42	0
7:20	0	6.39	1.56	-0.29	0
7:25	0	3.70	2.00	0.34	0
7:30	0	2.92	0.21	0.40	0
7:35	0	3.57	1.18	0.06	0
7:40	0	3.00	0.83	-0.70	0
7:45	0	1.87	0.25	-0.23	0
7:50	0.37	0	0.24	-0.45	0
7:55	-0.14	0	0.32	0.21	0
8:00	0.06	0	0.03	-0.04	0
8:05	0.00	1.53	0.98	0.07	0
8:10	0.12	0	0.91	-0.18	0
8:15	0.00	1.96	0.54	0.00	-0.08
8:20	0.25	0	0.32	0.04	0
8:25	0.36	0	0.51	0.20	0
8:30	0.25	0	0.43	-0.36	0
8:35	0.20	0	0.49	-0.06	0
8:40	0.24	0	0.65	-0.05	0
8:45	0.38	0	0.26	-0.76	0
8:50	0	2.32	0.62	-0.57	0
8:55	0	3.52	1.27	-1.69	0
9:00	0	3.81	1.21	-0.97	0

Model training

The LSTARIMA model is trained using the HR calibration algorithm (Hannan and Rissanen, 1982). One model must be trained for each AR order p and MA order q. This training can be time consuming, so the search range must be constrained to realistic values. In this case, p and q are varied from 1 to 5 because we expect the correlation to decline after 5 time lags (30 minutes at 5 minute aggregation intervals) based on spatio-temporal autocorrelation analysis. Then the best performing model is chosen from these possibilities.

Results

This section summarizes the experimental results. First, we introduce three models which we use to benchmark the model. Then we examine the results in terms of predictive accuracy, model structure and residual autocorrelation analysis.

Benchmark Models

To assess the performance of the LSTARIMA model, we compare its accuracy with three other models: a naïve model, the ARIMA model, and the STARIMA model.

The naïve model, also called a *random walk model*, has the following form (Thomakos and Guerard 2004):

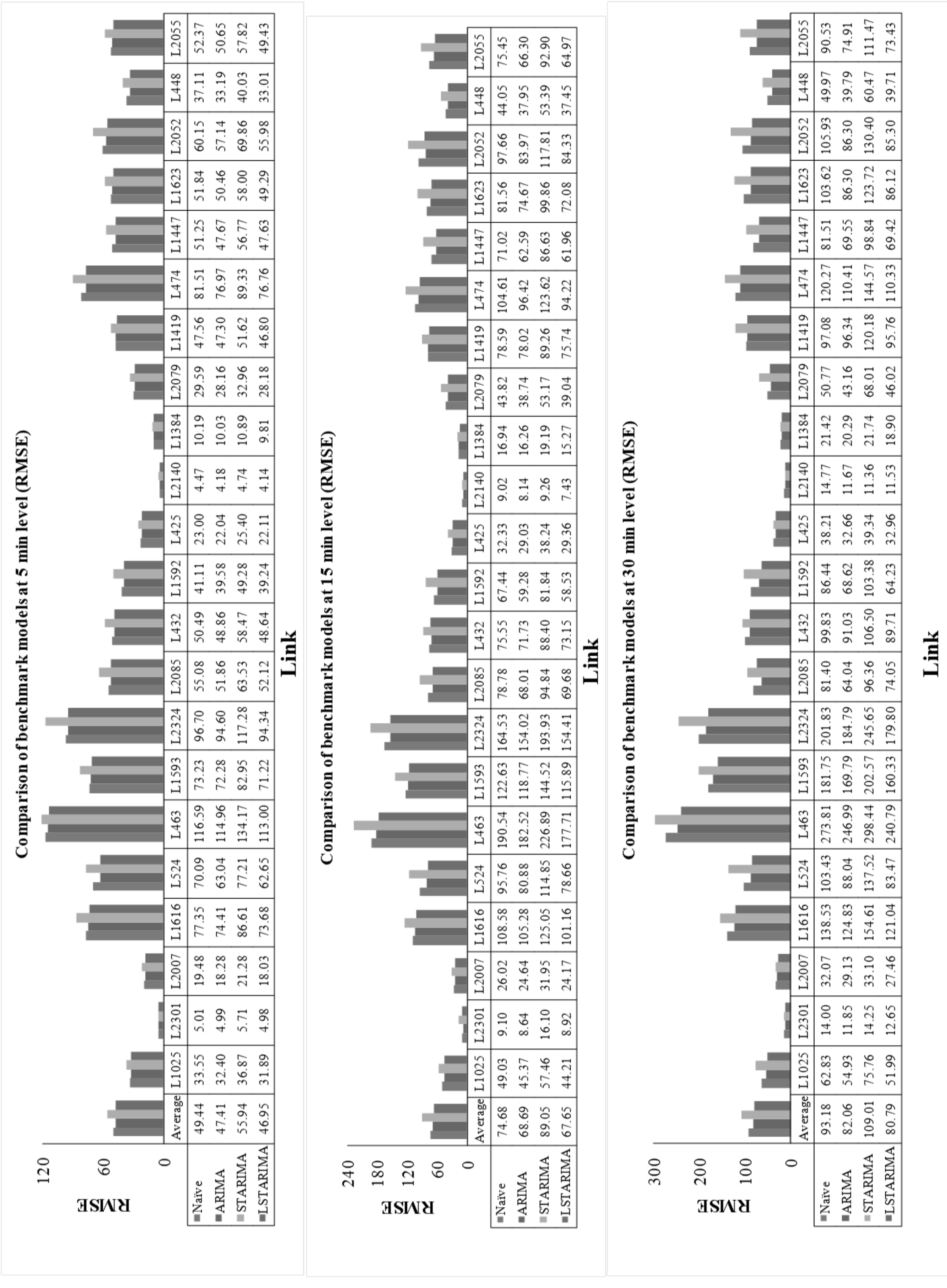
$$\hat{z}(t) = z(t-1) + \varepsilon_t \tag{7}$$

where $z(t-1)$ is an historical observation at time $t-1$; $\hat{z}(t)$ is the predicted value at time t; and e_t is a zero-mean residual. The forecast of the naïve model is just the previously observed data point in the series (nothing changes). This model is the simplest forecasting model such that the minimum requirement for any more complicated model is to outperform it.

The ARIMA and STARIMA models are two popular models that can be used for benchmarking. Both models have proved reliable in the traffic forecasting context (Thomakos and Guerard 2004, Kamarianakis and Prastacos 2005, Min and Wynter 2011). For consistency, the HR algorithm that is used to train the LSTARIMA model also is used to train the ARIMA and STARIMA models. Again, AR and MA orders of 1 to 5 are tested. For the STARIMA model, spatial orders of 1 to 3 are fixed globally and tested. After an extensive search, the preferred model is found to be a STARIMA (4, 0, 3) at spatial order 2. Because the LSTARIMA model extends the principles of the ARIMA and STARIMA models, it should outperform both models in the majority of cases in order to be preferable.

Prediction Accuracy

To assess the prediction accuracy of each of the models, the root mean squared error (RMSE) index is used. Figure 4(a) shows a bar graph of RMSE of the best performing models for each link at 5 minute intervals. The more sophisticated models perform only slightly better than the naïve model at this aggregate level. The traffic in the test dataset generally does not change much during 5-minute time increments, making the naïve model a strong predictor.



	Average	L1025	L2301	L2007	L1616	L524	L463	L1593	L2324	L2085	L432	L1592	L425	L2140	L1384	L2079	L1419	L474	L1447	L1623	L2052	L448	L2055
Naïve	49.44	33.55	5.01	19.48	77.35	70.09	116.59	73.23	96.70	55.08	50.49	41.11	23.00	4.47	10.19	29.59	47.56	81.51	51.25	51.84	60.15	37.11	52.37
ARIMA	47.41	32.40	4.99	18.28	74.41	63.04	114.96	72.28	94.60	51.86	48.86	39.58	22.04	4.18	10.03	28.16	47.30	76.97	47.67	50.46	57.14	33.19	50.65
STARIMA	55.94	36.87	5.71	21.28	86.61	77.21	134.17	82.95	117.28	63.53	58.47	49.28	25.40	4.74	10.89	32.96	51.62	89.33	56.77	58.00	69.86	40.03	57.82
LSTARIMA	46.95	31.89	4.98	18.03	73.68	62.65	113.00	71.22	94.34	52.12	48.64	39.24	22.11	4.14	9.81	28.18	46.80	76.76	47.63	49.29	55.98	33.01	49.43

	Average	L1025	L2301	L2007	L1616	L524	L463	L1593	L2324	L2085	L432	L1592	L425	L2140	L1384	L2079	L1419	L474	L1447	L1623	L2052	L448	L2055
Naïve	74.68	49.03	9.10	26.02	108.58	95.76	190.54	122.63	164.53	78.78	75.55	67.44	32.33	9.02	16.94	43.82	78.59	104.61	71.02	81.56	97.66	44.05	75.45
ARIMA	68.69	45.37	8.64	24.64	105.28	80.88	182.52	118.77	154.02	68.01	71.73	59.28	29.03	8.14	16.26	38.74	78.02	96.42	62.59	74.67	83.97	37.95	66.30
STARIMA	89.05	57.46	16.10	31.95	125.05	114.85	226.89	144.52	193.93	94.84	88.40	81.84	38.24	9.26	19.19	53.17	89.26	123.62	86.63	99.86	117.81	53.39	92.90
LSTARIMA	67.65	44.21	8.92	24.17	101.16	78.66	177.71	115.89	154.41	69.68	73.15	58.53	29.36	7.43	15.27	39.04	75.74	94.22	61.96	72.08	84.33	37.45	64.97

	Average	L1025	L2301	L2007	L1616	L524	L463	L1593	L2324	L2085	L432	L1592	L425	L2140	L1384	L2079	L1419	L474	L1447	L1623	L2052	L448	L2055
Naïve	93.18	62.83	14.00	32.07	138.53	103.43	273.81	181.75	201.83	81.40	99.83	86.44	38.21	14.77	21.42	50.77	97.08	120.27	81.51	103.62	105.93	49.97	90.53
ARIMA	82.06	54.93	11.85	29.13	124.83	88.04	246.99	169.79	184.79	64.04	91.03	68.62	32.66	11.67	20.29	43.16	96.34	110.41	69.55	86.30	86.30	39.79	74.91
STARIMA	109.01	75.76	14.25	33.10	154.61	137.52	298.44	202.57	245.65	96.36	106.50	103.38	39.34	11.36	21.74	68.01	120.18	144.57	98.84	123.72	130.40	60.47	111.47
LSTARIMA	80.79	51.99	12.65	27.46	121.04	83.47	240.79	160.33	179.80	74.05	89.71	64.23	32.96	11.53	18.90	46.02	95.76	110.33	69.42	86.12	85.30	39.71	73.43

Figure 4. Predictive accuracy (RMSE) of LSTARIMA vis-á-vis the benchmark models at a) 5-minute b) 15-minute and c) 30-minute intervals.

However, we do observe that the LSTARIMA model has the lowest average RMSE compared with the other models for this time interval. The second best performing model is the ARIMA model, followed by the naïve model. The most surprising result is that the STARIMA model performs worse than the naïve model. The STARIMA model uses global parameter coeffeicients and a fixed spatial adjacency structure to explain spatio-temporal autocorrelation, and thus can not capture the spatio-temporal nonstationarity and heterogeneity of travel time on the road traffic network, *even after differencing*. Figures 4 b) and c) show the bar graphs of RMSE at 15 minute and 30 minute intervals. The results are similar to those for the 5-minute interval, with the LSTARIMA being the best performing model. However, the naïve model shows a sharp decrease in performance at these levels of aggregation.

To assess the differences in predictive performance of the various models, a pairwise *F*-test was carried out with their residuals. An *F*-test (Snedecor and Cochran, 1989) is a test of statistical significance of the ratio of two sample variances, which is used here to test if the residual variance of one model is significantly less than that of another (one-tailed test). Because the residuals of the fitted models have zero mean, their variances can be used as a measure of model performance, with smaller values being preferred. The ratio of the variances of the residuals of a pair of models is used to calculate the *F* value using the following expression:

$$F = S_1^2/S_2^2 \tag{8}$$

where S_1^2 and S_2^2 are the error variances. The quotient of equation (8) would have the *F* distribution if the error terms were independent in time and space; failing this, Pitman's *t* test (Pitman, 1939) can be used, though in the present case the results from the more straightforward but approximate *F* test are so strong as to be conclusive. In this case, we test the null hypotheisis that "*model 1*" (i.e., LSTARIMA) is no better than "*model 2*" (i.e., STARIMA). This is a one-tailed (left side) *F*-test because the alternative hypothesis is that model 1 performs better as indicated by a smaller error variance with an F value less than 1. The significance level is set to 0.05. If the probability value of the *F*-test is less than 0.05 there is sufficient evidence to reject the null hypothesis that model 1 (i.e. LSTARIMA) performs no better than model 2 (i.e. STARIMA).

The *F*-test of residuals between LSTARIMA and STARIMA shows LSTARIMA performs better than STARIMA on all 22 links at 5-minute, 15-minute and 30-minute intervals (with p-values less than 0.05). The same tests are applied to other group comparisions (i.e., LSTARIMA and the naïve, and LSTARIMA and ARIMA). The *F*-test results for LSTARIMA and the naïve model also indicate that LSTARIMA performs significantly better. Although the comparision between LSTARIMA and ARIMA does not indicate a significant difference in predictive accuracy, LSTARIMA is superior to ARIMA in terms of RMSE link by link, achieving better accuracy on 22, 18 and 12 links at the 5-, 15- and 30-minute intervals, respectively.

Model Structure

For the ARIMA and LSTARIMA models, one model must be built for each link, whereas only one STARIMA model is required for the entire network. Both ARIMA and STARIMA involve preprocessing as part of model fitting to remove trend and cyclical patterns. The strategy adopted here involves testing combinations of differencing and logarithmic transformations. Examples of the preferred models includes ARIMA(3, 0, 1) with a logarithm transformation for L463; and, ARIMA(2, 0, 1) and ARIMA(4, 0, 2) without a transformation for links L1593 and L2324, respectively. Table 2 summarizes the AR and MA orders for each of the best models across the test network.

Table 2 AR and MA orders (*p* and *q*) of the LSTARIMA and ARIMA models for all 22 road links

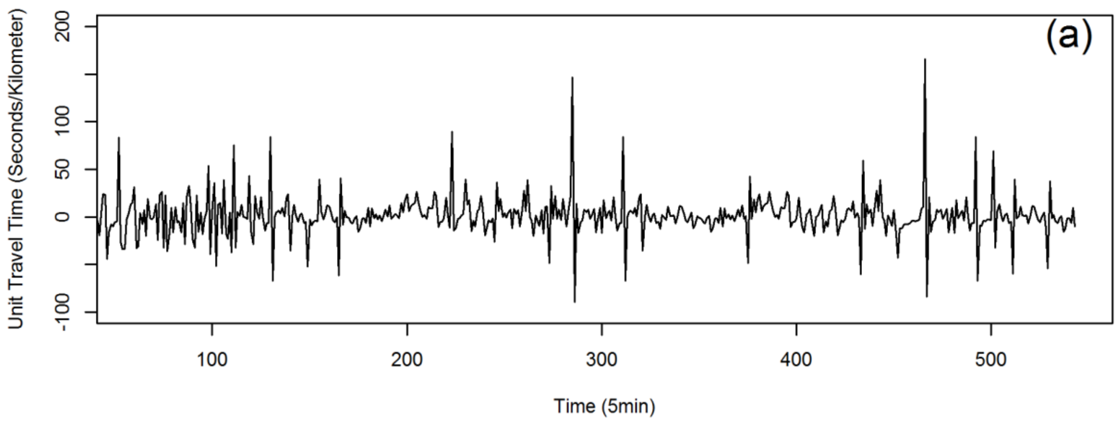
LSTARIMA											
Pars	L1025	L2301	L2007	L1616	L524	L463	L1593	L2324	L2085	L432	L1592
p	1	2	3	2	1	4	2	2	3	3	2
q	1	3	4	4	1	2	1	2	4	4	1
Pars	L425	L2140	L1384	L2079	L1419	L474	L1447	L1623	L2052	L448	L2055
p	3	2	3	2	4	1	1	1	1	1	1
q	2	3	4	1	1	1	2	2	3	4	1
ARIMA											
Pars	L1025	L2301	L2007	L1616	L524	L463	L1593	L2324	L2085	L432	L1592
p	4	2	4	1	3	3	2	4	3	3	2
d	0	0	2	0	0	0	0	0	0	0	1
q	1	3	2	4	1	1	1	2	4	2	1
Pars	L425	L2140	L1384	L2079	L1419	L474	L1447	L1623	L2052	L448	L2055
p	4	4	3	3	3	2	4	4	4	2	4
d	2	1	0	1	0	0	0	1	0	0	0
q	1	3	2	1	2	2	1	1	1	1	1

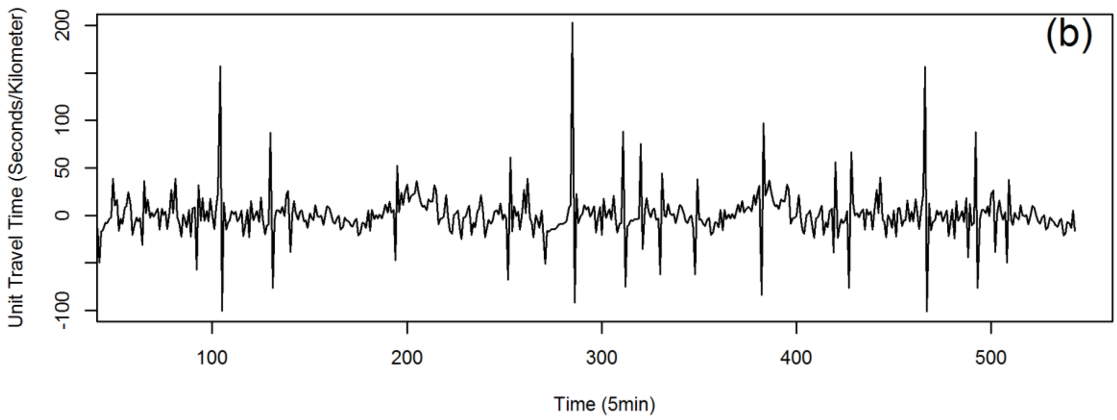
Table 2 reveals that different *p* and *q* values exists for LSTARIMA and ARIMA models on the same links. For example, at link L1025, ARIMA (4, 0, 1) achieves the best accuracy, whereas the preferred LSTARIMA model is LSTARIMA (1,0,1). This discrepancy might be caused by the spatial contribution of adjacent links in the LSTARIMA specification. Statistical results for the 22 road links show that generally a simpler LSTARIMA model with smaller *p* and *q* values can perform better than a complicated ARIMA model with larger *p* and *q* values. The best STARIMA model is STARIMA(4, 0, 3), which also has larger *p* and *q* values than most LSTARIMA models.

Residual Autocorrelation Analysis

Pure predictive accuracy is not the only measure of a good spatio-temporal model. The model also must be able to capture the dynamics of the underlying spatio-temporal process and account for spatio-temporal autocorrelation and nonstationarity in data. One way of evaluating this is to test the predictive residuals of a model for autocorrelation. If a model truly captures the underlying process, then its residuals should be independent and identically distributed (i.i.d.) random variables. To assess this, we examine the temporal, spatial, and spatio-temporal autocorrelation in the residuals of the LSTARIMA, STARIMA and ARIMA models. Strictly speaking, we do not think ACFs and Moran's I are good indicators for space-time data because, unlike the ST-ACF, they cannot capture the interaction between the spatial and temporal dimensions. However, they are presented here for reference because most temporal and/or spatial models use these indicators.

Figures 5 (a)-(c) show snapshots of the residual profile (3 days from 25 October — 27 October, 2010) on L1025 for each of the models at a 5-minute interval. Visual inspection suggests a cyclical pattern in the residuals of the ARIMA (Fig 5b)) and STARIMA (Fig. 5c)) models that is stronger than one in the LSTARIMA residuals. The largest residual of the LSTARIMA model (around 150 seconds) is smaller than the other two (around 200 seconds). autoregressive conditional heteroskedastic (ARCH) processes exist in the data (stochastic volatility in the variances) that are very common in series of high frequency, such as the traffic data here.





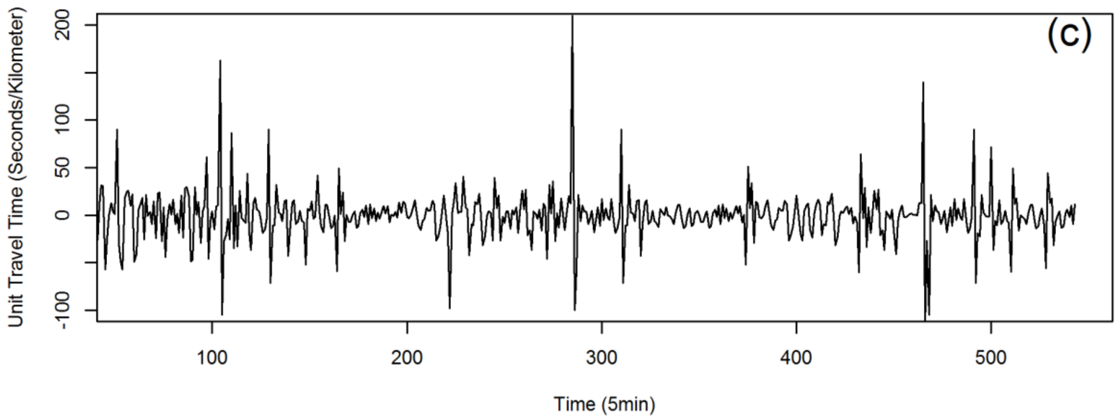


Figure 5. Snapshot of profile of residuals for three models (3 days from 25th Oct —27th Oct, 2010) on L1025; (a) LSTARIMA; (b)ARIMA; (c) STARIMA.

Temporal autocorrelation

We investigate the statistical significance of this pattern using the temporal autocorrelation function (ACF). Figures 6 (a)-(c) show the ACF plots of the entire residual series of each model.

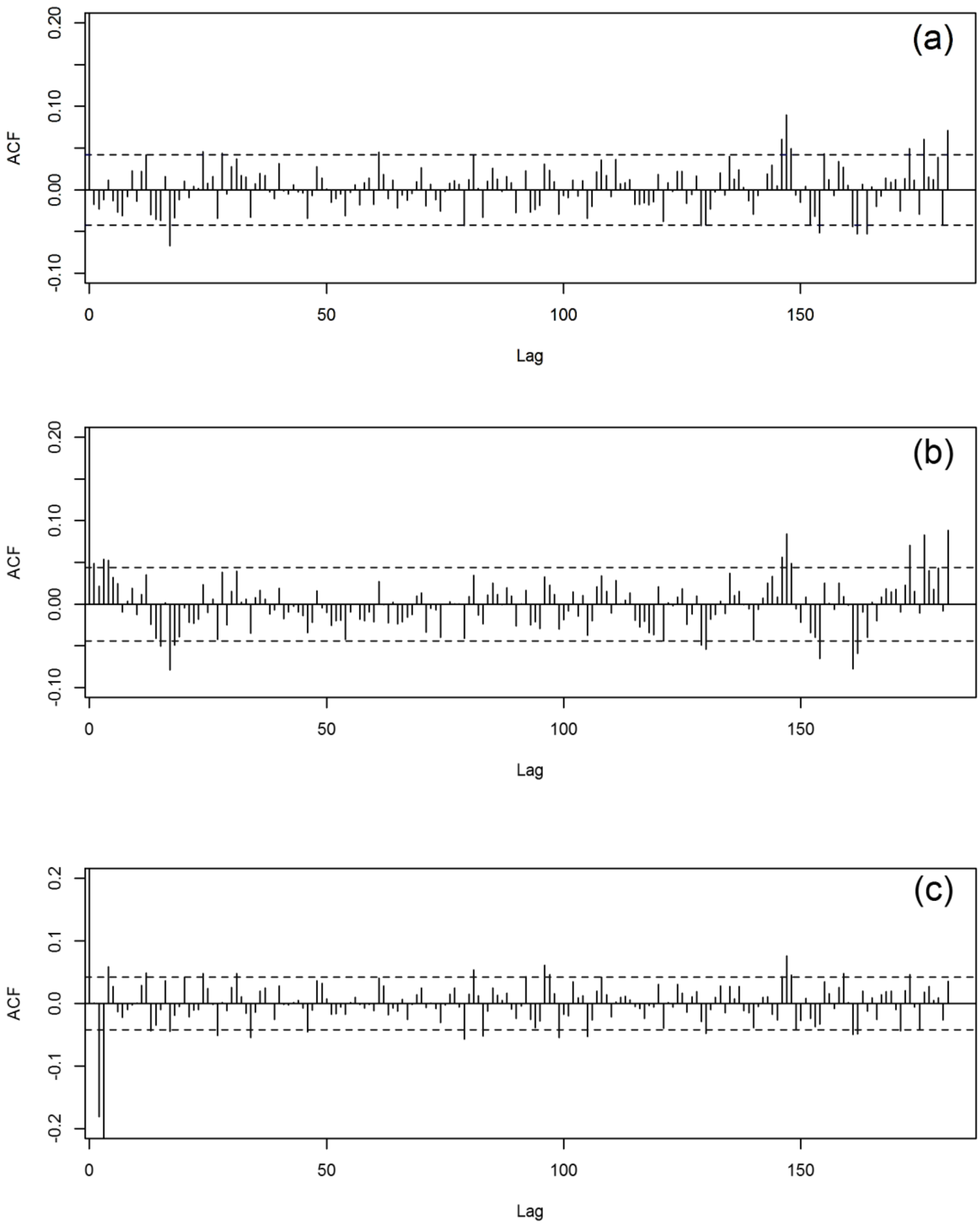
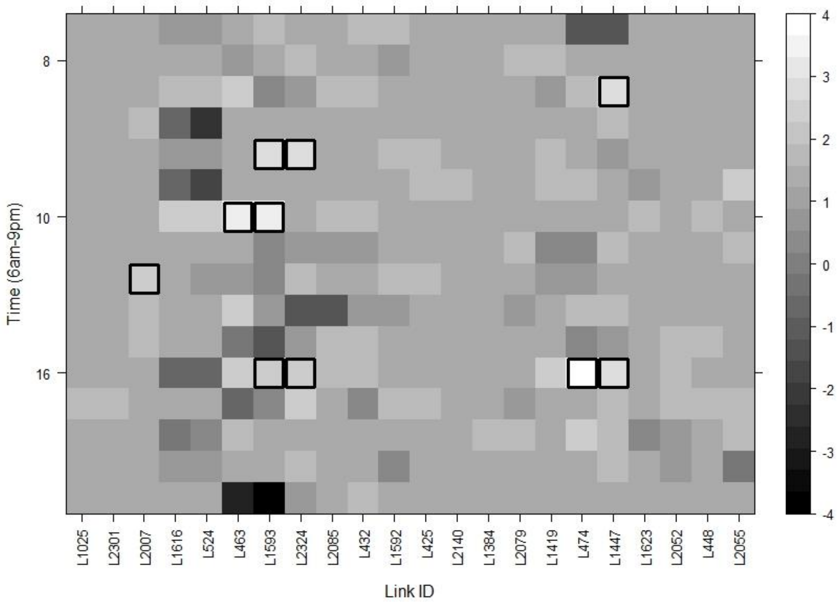


Figure 6. ACF analysis of residual series of the three models on L1025; (a) LSTARIMA; (b) ARIMA; (c) STARIMA.

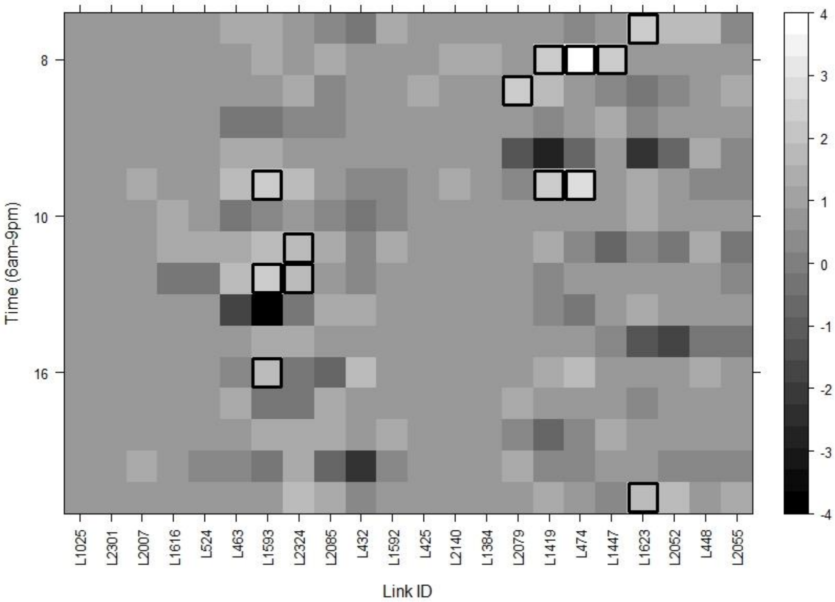
The most striking observation is that the ACF plot of the STARIMA model (Fig. 6c) shows significant autocorrelation up to temporal lag 3 (2 negative correlations, 1 positive correlation). Similar results are observed when examining the residuals of the STARIMA model on other links, indicating that the STARIMA specification cannot fully capture the temporal autocorrelation in the network. This problem is caused by the global parametric structure of the STARIMA model, and has been avoided in the LSTARIMA (and ARIMA) model because it relaxes the assumption of fixed temporal dependence for all locations, allowing the autoregressive parameters p and moving average parameters q to change across space (links). The ACF plots for the LSTARIMA and ARIMA models appear broadly similar, although the ARIMA model has more significant spikes, indicating that the LSTARIMA model performs slightly better in dealing with temporal autocorrelation. Although the coefficients keep on changing from positive to negative, or vice versa, there is no regular wave-shaped pattern. This outcome indicates that the ARIMA model, and the LSTARIMA model in particular, can account for most of the temporal autocorrelation in the data. However, all three models exhibit significant residual autocorrelation at a lag of one day (180), indicating that the daily cyclical pattern is not fully accounted for. Although this feature does not affect the performance of the models, it will be addressed in further research.

Spatial autocorrelation

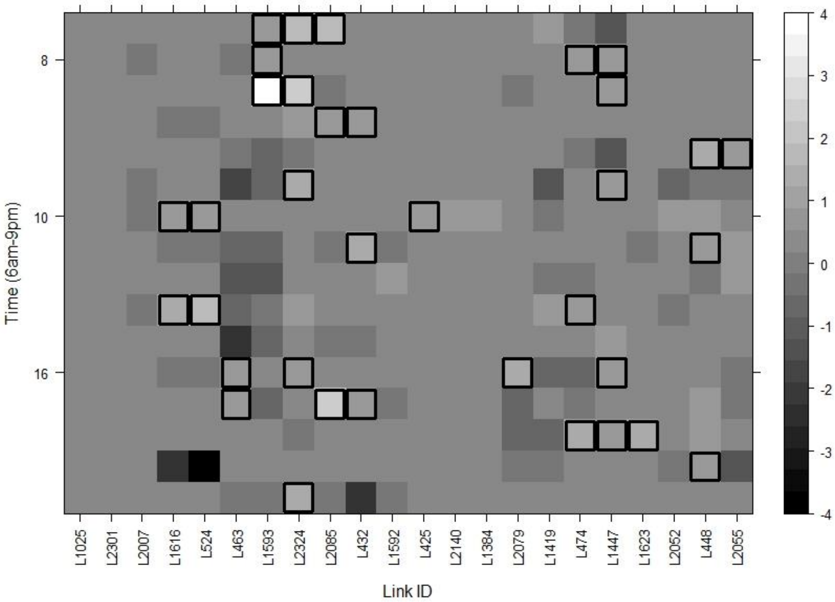
Spatial autocorrelation is tested for by calculating the value of local Moran's I [abbreviated to LISA (local indicator of spatial autocorrelation), after Anselin 1995] for each hour between.6:00am and 9:00pm on 25 October, 2010. Figures 7(a)-(c) show level plots of the LISA of all 22 links at spatial order 1.. In general, Figures 7 (a) to (c) reveal that positive and negative autocorrelation is staggered over space and time, and the minority of LISA values are statistically significant, indicating that the residuals of the three models seem to approximate a random distribution in space. However, the number of statistically significant cells increases from Figure 7(a) to (c). Specifically, there are 35 statistically significant cell values in Figure 7(c), indicating that the spatial autocorrelation at those times and positions (link) cannot be fully captured by the STARIMA specification. The LSTARIMA model has the fewest cells that are statistically significant - only 10 compared with 13 for ARIMA - indicating that the LSTARIMA model is the most effective model for capturing spatial autocorrelation.



(a)



(b)



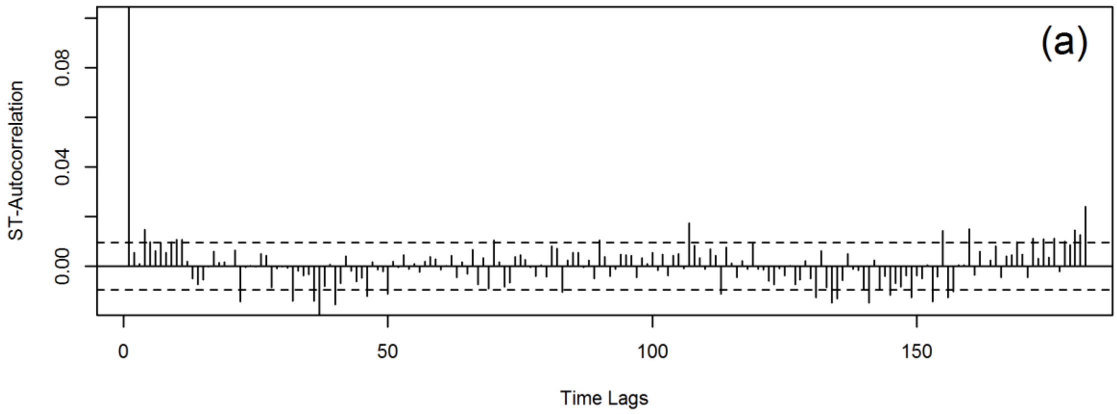
(c)

Figure 7. Map of local indicator of spatial autocorrelation (LISA) of all 22 link at spatial order one on 25th Oct, 2010 for the three models, where the x-axis is the link ID and the y-axis is time (6:00am-9:00pm); (a) LSTARIMA; (b) ARIMA; (c) STARIMA. Grayscale is the strength of LISA; cells with black borders indicate statistically significant LISA values (p-value < 0.05)

Spatio-temporal autocorrelation

Finally, we evaluate the residual spatio-temporal autocorrelation for each of the models. The indicator we use is the spatio-temporal autocorrelation function (STACF, Martin and

Oeppen 1975). Figures 8(a)-(c) show the STACF plots for the three models at spatial order one and the 5-minute interval. Strong, significant spatio-temporal autocorrelation is observed at temporal lags 1 to 3 in the residuals of the STARIMA model. The STACF plots of the LSTARIMA and ARIMA models are comparatively closer to a random distribution over space and time. However, the ARIMA model displays moderately higher spatio-temporal autocorrelation at temporal lags 1 to 10.



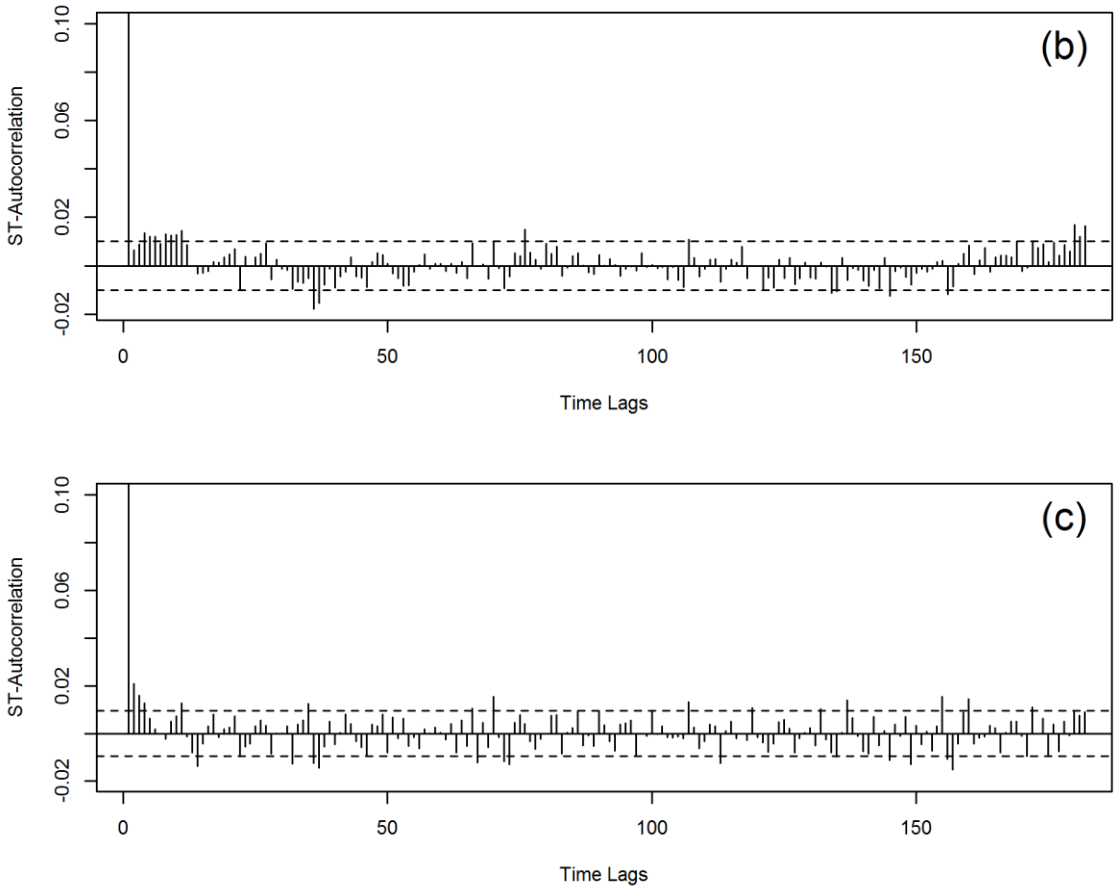


Figure 8. STACF plots for three models at first order spatial weight matrix; (a) LSTARIMA; (b) ARIMA; (c) STARIMA.

Summary of Results

In summary, the *F*-test results confirm that the LSTARIMA model achieves the best overall performance across all timeframes, followed closely by the ARIMA model. The STARIMA model does not perform as well and only performs better than the naïve model when the temporal interval of prediction increased to 30 minutes. The LSTARIMA appears to be a better model than the STARIMA because of greater parameter flexibility (dynamic spatial neighbourhood and dynamic spatial weight). Although the F-test indicates that the LSTARIMA and ARIMA models are equivalent, higher predictive accuracy for individual links is obtained with the LSTARIMA model. Although the principle of ARIMA seems simpler than LSTARIMA, the structure of the ARIMA model is more complicated than that for the LSTARIMA model because larger p and q values are obtained for the estimated ARIMA models.

The residuals from the three models are not fully independent (in time and in space). This implies that pre-processing for the ARIMA and STARIMA models may be inappropriate. This possibility supports our previous supposition in Cheng et al. (2011a) that the extraction of a globally stationary spatio-temporal process through differencing may not be realistic. Although some pre-processing for the LSTARIMA model also might be needed, its predictive residuals display almost negligible temporal, spatial and spatio-temporal autocorrelation, suggesting that preprocessing seems unnecessary. Therefore, we think pre-processing is not needed for the LSTARIMA model, especially for practical purposes, because finding an adequate transformation given the results for the ARIMA and STARIMA models is very difficult. No data pre-processing is a great advantage of the LSTARIMA specification, as is its simpler structure (low p and q values) compared with the ARIMA and STARIMA specifications.

Discussion and Conclusions

Existing space-time models such as the STARIMA one are designed to account for the effects of autocorrelation in spatio-temporal data. Their global parametric structure does not equip them to deal with spatial heterogeneity and temporal nonstationarity, even though these are characteristic of many spatio-temporal datasets, with networks of traffic flows being just one example. As we demonstrate in this paper, extracting a stationary spatio-temporal process from highly nonstationary data can be unrealistic, which limits the efficacy of existing global models. This article addresses the shortcomings of such models by introducing a new concept; the dynamic spatial weight matrix. This weight matrix comprises the following three components that enable local dynamics to be incorporated into established model structures: an adjacency structure; a dynamic spatial neighbourhood; and a dynamic spatial weight..

We demonstrate application of the dynamic spatial weight matrix by extending the STARIMA model specification to create a new space-time model—the LSTARIMA model – which is aimed at improving the ability of the STARIMA model to cope with the dynamics and heterogeneity of spatio-temporal data on networks. This new model specification relaxes the assumption of fixed temporal dependence for all locations, allowing the autoregressive parameters p and moving average parameters q to change across space. By introducing space-varying $\mathbf{\Phi}_{kh}$ and $\boldsymbol{\theta}_{kl}$ coefficients for each location, spatial heterogeneity is substantially accounted for in the specification. This model framework is generic, and we demonstrate that the ARIMA and STARIMA models are special cases of it.

We examine the efficacy of the LSTARIMA model with the case of road traffic networks in London. Unlike the traditional (ST)ARIMA models, our experiment demonstrates that the LSTARIMA model works well without the need for data pre-processing (e.g., a logarithmic transformation and differencing) used widely with (ST)ARIMA models, suggesting that the LSTARIMA model is capable of dealing with traffic time series better than the (ST)ARIMA model. With smaller p and q values, the LSTARIMA model also has a simpler structure than the (ST)ARIMA model. This results from the innovative dynamic spatial weight matrix used in the LSTARIMA specification, which allows it to capture unstable traffic states in a space-time series.

Although we only examine the case of road traffic networks in this study, one of the key strengths of our methodology is its modularity. Each of the three components of the dynamic spatial weight matrix (i.e., an adjacency structure, a dynamic spatial neighbourhood, and a dynamic spatial weight) can be modified according to the application of interest.

For example, within the transport setting, this model could be used to predict traffic flows as well as speeds by replacing the speed based dynamic spatial weight with a flow based weight. This substitution could be linked to, for example, Kinematic wave theory (Lighthill and Witham 1955). Modification of the spatial adjacency structure and the dynamic spatial neighbourhood may not even be needed in this case. Furthermore, the method could be applied to flows on other networks, such as the internet, where the heterogeneity may result from global and local shifts in usage. This application would require domain specific knowledge to be incorporated into each of the three elements of the dynamic spatial weight matrix, but would not change the overall model structure. Moreover, the method is not limited to networks, and a range of other spatio-temporal processes exist that may benefit from its application, such as environmental monitoring and house price forecasting. In this latter case, the update of dynamic spatial weights may be driven by knowledge of exogenous factors in the wider (local) economy. As an aside, the dynamic spatial weight matrix framework is not dependent on the LSTARIMA estimator algorithm used here.

The algorithm can be seen as a *fourth* module, and there is scope to replace the linear model used here with nonlinear models from the field of machine learning such as kernel-based approaches (Wang et al 2012). Combining this with Geographically Weighted Regression (GWR) could lead to a localized dynamic GWR; i.e., a spatio-temporal GWR (STGWR). This will be the direction for our future research.

In conclusion, the results presented here demonstrate that the combination of a dynamic spatial weight matrix and the LSTARIMA model specification is a viable approach for space-time modelling of highly dynamic, heterogeneous network processes. Further studies will be carried out on a city-wide scale to demonstrate the effectiveness of the approach in a real time, applied setting. Furthermore, the general framework presented here lends itself to applications in a wide range of network and spatial processes. All that is required is to define an application-specific adjacency structure, dynamic spatial weight matrix, and dynamic spatial neighbourhood that can replace the traffic-specific formulations given here.

Acknowledgements

The authors thank Transport for London for providing the journey time data. The research for this paper was carried out under the STANDARD project, sponsored by the UK Engineering and Physical Sciences Research Council (EP/G023212/1). Comments from the editor and the anonymous reviewers are greatly appreciated, and improved the readability of the paper. The authors bear sole responsibility for any mistakes that may remain.

References

Anselin, L (1995) Local Indicators of Spatial Association—LISA. *Geographical Analysis*, 27, 93-115.

Black, WR and I Thomas (1998) Accidents on Belgium's motorways: a network autocorrelation analysis. *Journal of Transport Geography*, 6(1), 23–31.

Box, GEP, GM Jenkins and G Reinsel (1970) Time series analysis, forecasting and control. San Francisco: Holden Day.

Chandra, SR and H Al-Deek (2009) Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems*, 13(2), 53-72

Cheng T, JQ Wang and X Li (2011) A Hybrid Framework for Space-Time Modeling of Environmental Data. *Geographical Analysis*, 43(2), 188-210.

Cheng, T, Haworth, J and JQ Wang (2011) Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*. DOI: 10.1007/s10109-011-0149-5.

Cressie N, and CK Wikle (2011), *Statistics for Spatio-Temporal Data.*Wiley.

Chun, Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 10(4),317-344.

Cliff, A and J Ord (1969) The Problem of Spatial Autocorrelation. In *London Papers in Regional Science*, edited by A. Scott. London, Pion, 25-55.

Ding, Q, X Wang, X Zhang and Z Sun (2011) Forecasting Traffic Volume with Space-Time ARIMA Model. *Advanced Materials Research*, 156-157, 979-983.

Elhorst, JP (2001) Dynamic Models in Space and Time. *Geographical Analysis*, 33, 119–140.Elhorst, JP (2003) Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review*, 26(3), 244-268.

Elhorst, JP, G Piras and G Arbia (2010) Growth and Convergence in a Multiregional Model with Space–Time Dynamics. *Geographical Analysi*s, 42, 338–355.

Florax, RJGM and S Rey (1995) The impact of misspecified spatial structure in linear regression models. In *New Directions in Spatial Econometrics*. 111–135, edited by L Anselin and RJGM Florax. Berlin: Springer-Verlag

Gething, PW, AM Noor, PW Gikandi, SI Hay, MS Nixon, RW Snow and PM Atkinson (2008) Developing Geostatistical Space–Time Models to Predict Outpatient

Treatment Burdens from Incomplete National Data. *Geographical Analysis*, 40, 167–188.

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

Giacinto, VD (2006) A Generalized Space-Time Model with an Application to Regional Unemployment Analysis in Italy. *International Regional Science Review*. 29(2): 159-198.

Griffith, DA (1996) Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In *Practical Handbook of Spatial Statistics*, 82–148, edited by SL Arlinghaus, DA Griffith, WD Drake and JD Nystuen. Boca Raton, FL: CRC Press.

Griffith, DA (2010) Modeling spatio-temporal relationships: retrospect and prospect. *Geographical Analysis*, 12(2), 111-123.

Griffith, DA and GB Heuvelink (2009) Deriving Space–time Variograms from Space–time Autoregressive (STAR) Model Specifications. In *StatGIS 2009 conference, Milos, Greece, June*.

Griffith, DA and F Lagona (1998) On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference*, 69(1), 153–174.

Hannan, EJ and J Rissanen (1982) Recursive estimation of mixed autoregressive-moving average order, *Biometrika*, 69, 81-94.

Heuvelink, GBM and DA Griffith (2010) Space–Time Geostatistics for Geography: A Case Study of Radiation Monitoring Across Parts of Germany. *Geographical Analysis*, 42(2), 161-179.

Hooper, PM and GJD Hewings (1981) Some Properties of Space-Time Processes. *Geographical Analysis*, 13, 203–223.

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

Leenders, RT (2002) The specification of weight structures in network autocorrelation models of social influence. *SOM Research Reports*, No. 02B09.

Lighthill, MJ and JB Whitham (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.

Martin, RL and JE Oeppen (1975) The Identification of Regional Forecasting Models Using Space: Time Correlation Functions. *Transactions of the Institute of British Geographers*, 66, 95–118.

Min, X, J Hu, Q Chen, T Zhang and Y Zhang (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, W and L Wynter (2011) Real-time road traffic prediction with spatio-temporal correlations. *Transport Research Part C*, 19, 606-616.

Oud, JHL, H Folmer, R Patuelli and P Nijkamp (2012) Continuous-Time Modeling with Spatial Dependence. *Geographical Analysis*, 44, 29–46.

Páez, A, Scott, DM and E Volz (2008) Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality. *Social Networks*, 30(4),309-317.

Peeters, D and I Thomas (2009) Network Autocorrelation. *Geographical Analysis*, 41(4),436-443.

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

Pfeifer, PE and SJ Deutsch (1981) Seasonal Space-Time ARIMA Modeling. *Geographical Analysis*, 13, 117–133.

Pitman, EJG. (1939). A note on normal correlation. *Biometrika*, *31*, 9-12.

Richards, PI (1956) Shock Waves on the Highway. *Operations Research*, 4(1), 42-51.

Stetzer, F (1982) Specifying weights in spatial forecasting models—the results of some experiments. *Environment and Planning A*, 14(5),571–584.

Snedecor, GW and WG Cochran (1989) Statistical Methods, 8th Edition. Iowa State University Press.

Tiao, GC and GEP Box (1981) Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376), 802–816.

Thomakos, DD and JB Guerard (2004) Naïve, ARIMA, nonparametric, transfer function and VAR models: A comparison of forcasting performance. *International Journal of Forecasting*, 20, 53- 67.

Wang, JQ, T Cheng and J Haworth (2012) Space-Time Kernels in ISPRS book series: *Advances in Geo-Spatial Information Science*, 25-31, Taylor and Francis, CRC Press.