J Geogr Syst (2012) 14:389–413
DOI 10.1007/s10109-011-0149-5

ORIGINAL ARTICLE

Spatio-temporal autocorrelation of road network data

Tao Cheng · James Haworth · Jiaqiu Wang

Received: 19 November 2010 / Accepted: 30 March 2011 / Published online: 16 April 2011

Abstract Modelling autocorrelation structure among space–time observations is crucial in space–time modelling and forecasting. The aim of this research is to examine the spatio-temporal autocorrelation structure of road networks in order to determine likely requirements for building a suitable space–time forecasting model. Exploratory space–time autocorrelation analysis is carried out using journey time data collected on London's road network. Through the use of both global and local autocorrelation measures, the autocorrelation structure of the road network is found to be dynamic and heterogeneous in both space and time. It reveals that a global measure of autocorrelation is not sufficient to explain the network structure. Dynamic and local structures must be accounted for space–time modelling and forecasting. This has broad implications for space–time modelling and network complexity.

Keywords Spatial autocorrelation · Network structure · Space–time autocorrelation · Space–time modelling · Travel time prediction · Network complexity

JEL Classification R41 · C23 · C52

T. Cheng (✉) · J. Haworth · J. Wang
Department of Civil, Environmental and Geomatic Engineering,
University College London, Gower Street, London WC1E 6 BT, UK
e-mail: tao.cheng@ucl.ac.uk

J. Haworth
e-mail: j.haworth@ucl.ac.uk

J. Wang
e-mail: w.jiaiqu@ucl.ac.uk

1 Introduction

In today's data-rich environment, time and location specific data are collected in huge volumes and are often available in real time. The traditional challenges of data sparsity and lack of computational power have been replaced with fresh challenges in data storage, data mining and knowledge discovery. An example of such voluminous data is traffic data. Traffic data are now collected on the urban road systems of many major cities and if put to good use can provide vital information on road network performance to be used in intelligent transportation systems (ITS) and advanced traveller information systems (ATIS). Typical data collected on road systems are traffic flows, journey times and speeds. One of the primary goals of ITS is to forecast future conditions on the road network in order to provide up to date information to travellers.

This has traditionally been accomplished through univariate time series prediction using various techniques such as (seasonal) autoregressive integrated moving average ([S]ARIMA, Williams and Hoel 2003) models, artificial neural networks (Dougherty and Cobbett 1997; van Lint et al. 2005), support vector regression (Wu et al. 2004), Kalman filtering (Liu et al. 2006) and non-parametric regression (Smith et al. 2002) among others. Vlahogianni et al. (2004) provide a good overview. More recently, researchers have begun to develop space–time models for traffic forecasting. Various statistical methods for modelling space–time data have been proposed over the years, including multivariate autoregressive integrated moving average [(M)ARIMA] models, space time autoregressive integrated moving average (STARIMA) models (Pfeifer and Deutsch 1980) and variants (STAR models, STMA models), three-dimensional geostatistical models and spatial panel data models, of which Griffith (2010) provides an overview.

The methodological development of each of these has been motivated by the need to account for *autocorrelation* in spatial (Cliff and Ord 1969) and temporal (Box and Jenkins 1970) data, and they represent a fusion of research from both domains. Correctly identifying and quantifying the extent to which observations are autocorrelated with each other in time and space is a necessity in the statistical modelling of spatial and temporal relationships (Hackney et al. 2007), as to ignore it can result in the underrepresentation of the variability present in data and overassessment of significance implying non-compliance (Patil 2009). All current models assume that the spatio-temporal autocorrelation in data can be adequately described by globally fixed parameters, i.e. the extent to which observations are autocorrelated with each other is fixed in space and/or time. For example, in STARIMA models, space–time processes are stationarized through transformation and differencing and autocorrelation is accounted for in the autoregressive (AR) and moving average (MA) terms. The AR and/or MA orders are fixed globally both spatially and temporally, and a single parameter is estimated for each. Although STARIMA has been employed for traffic flow forecasting (Kamarianakis and Prastacos 2005; Wang et al. 2010), these two assumptions (stationarity) and fixed spatio-temporal neighbourhood are in fact very difficult to be satisfied for dynamic network data such as a transport network.

For instance, the current conditions on a section of road are influenced to some extent by the previous conditions of adjacent road sections in both directions (Chandra and Al-Deek 2008). Shockwave theory has long been used to model the downstream to upstream (backward) progression of queues (Richards 1956), and the movement of vehicles from upstream to downstream (forward) exerts influence in the opposite direction. It follows that in congested conditions the influence will come mainly from downstream whereas in free-flowing conditions the influence will come from upstream. However, the progression from free flowing to congested traffic conditions is continuous, which begs the question; at a given point in the time, how many and how much of the upstream and downstream neighbours of a link contribution to its current condition? Does the extent and the strength of neighbours' contribution change with time? Does this vary spatially across a network?

The answers to these questions have broad implications for space–time modelling of traffic data as representing this relationship too simply may cause poor forecasting results, particularly in the presence of non-recurrent traffic conditions. Given adequate spatial and temporal data granularity, the correlation between locations can be an indicator of the extent and direction of influence between them (Yue and Yeh 2008). On a transportation network, this equates to determining a transfer function of traffic conditions from one road section to the next. Vehicles on a road network do not remain at static points, they move between locations and the time it takes for them to move from location to location depends on the current traffic state. Therefore, the spatial neighbourhood of a link will effectively become larger when traffic conditions are free flow and smaller when they are congested. This changing structure should be accounted for in any space–time traffic model. Yue and Yeh (2008) view this problem as one of determining the forecastability of traffic data, which depends on two factors: the effective data range and the strength of correlation. On a traffic network, the effective data range can be defined as the number of links that can deliver traffic to the current link within a given time interval (usually equal to the forecasting horizon). The strength of correlation can be defined as weights of the influence of those links on the current link.

Recent studies have attempted to capture the dynamics of network process to basic space–time models and can broadly be separated into two categories: those that aim to capture the (dynamic) effective range of spatial neighbourhood (Elhorst 2003; Min et al. 2007; Ding et al. 2011) and those that aim to capture the (dynamic) strength/weight of correlation (Min et al. 2009, 2010). Among others, the most advanced approach is Generalised STARIMA (GSTARIMA) proposed by Min et al. (2010). In this model, spatially heterogeneous autocorrelation structures are captured by allowing the AR and MA parameters to vary by location. The GSTARIMA model outperforms traditional STARIMA in terms of forecasting accuracy. However, visual inspection of the forecasting results suggests a temporal shift in the predictions, indicating that the model is not functioning correctly. Additionally, although the method allows for spatially dynamic parameter estimates, the spatial structure of the model is still fixed to an extent as the size of the spatial neighbourhood considered is the same for each location. Its temporal structure is also fixed.

Although progress has been made to capture spatio-temporal autocorrelation structure effectively which provides the key to accurate forecasting of future traffic

conditions, there is no model that is able to consider dynamic nature of autocorrelation in two aspects: dynamic spatial range and dynamic strength of the correlations. By providing insights into the actual autocorrelation structure of a real road network, a systematic autocorrelation analysis helps to shed light on the requirements for a given space–time model of traffic data. Consequently, the aim of this research is to examine the spatio-temporal autocorrelation structure of networks in order to determine likely requirements for building a suitable space–time forecasting model. Exploratory space–time autocorrelation analysis is carried out using journey time data collected on London's road network. Through the use of both global and local autocorrelation measures, we aim to explore how the autocorrelation structure varies in both time and space. In doing so we aim to reveal whether a global measure of autocorrelation is sufficient to explain the structure in the data or whether more dynamic, local structures must be accounted for.

Different from other researches who use simulations to study autocorrelation of networks (such as Neuman and Mizruchi 2010), we take an empirical approach to study the autocorrelation structure of networks. Furthermore, most existing researches focus on spatial autocorrelation of static networks; here, particular concern is given to the dynamics of autocorrelation structure in time and space. This will also have broad implications for network complexity theory. Understanding the spatial dynamics and their influence on network behaviour and performance has profound impact of urban networks and network theory (Castells 2010; Pflieger and Rozenblat 2010). We acknowledge the achievement of network science in measuring and discovering the heterogeneous topological structure of networks (Newman 2003; Jiang 2007; Xie and Levinson 2007; Xu and Sui 2007). Such analyses are very useful for understanding the influence of topological structure (Farber et al. 2009), but are not directly applicable to space–time modelling and prediction of data collected on them. The emphasis of this paper is not to discuss the complexity of the network topology (though it is very relevant), but the space–time autocorrelation structure of phenomena occurring on the network resulting from such topology given the importance of autocorrelation in space–time prediction. Dynamics of network complexity is a broad issue (Watts and Strogatz 1998), and here, we focus on the aspect of dynamic autocorrelation.

The structure of the paper is as follows. The next section exams the approach to model autocorrelation of network data by using spatial weight matrices. Section 3 presents the global and local measures for space–time autocorrelation. Sections 4 and 5 present an empirical example of a subset of London's road network, which is used to illustrate the dynamic nature of the spatio-temporal autocorrelation structure of the network. The findings are summarized in Sect. 6, and directions for further research are also proposed.

2 Autocorrelation of network data

2.1 Autocorrelation and spatial weight matrices

The study of autocorrelation on networks became popular in the late 1970s and early 1980s and was facilitated by methodological developments in the spatial sciences

both in terms of measuring spatial autocorrelation and modelling spatial processes. A body of research in the new field of network autocorrelation analysis emerged whereby spatial methods were extended to model network data and these have since been applied in different ways to a diverse range of tangible and intangible networks such as social networks (Doreian et al. 1984; Dow et al. 1984; Dow 2007; Dow and Eff 2008; Leenders 2002), migration networks (Black 1992; Chun 2008) and transportation networks (Black and Thomas 1998; Flahaut et al. 2003; Min et al. 2007). The extension of spatial methods to network data is made possible through the modification of spatial weight matrices to account for the structure of networks.

A spatial weight matrix W is an $N * N$ matrix, where N is the number of spatial locations, which is used to incorporate a hierarchical ordering of space in mathematical models. Two non-trivial issues must be tackled when defining a spatial weight matrix. Firstly, the linkage structure must be defined, i.e. which pairs of nodes have edges between them and the direction of dependence. Secondly, the value of the weights w_{ij} must be decided. These issues have been shown to be very important as different weight matrices can lead to different inferences being drawn and can lead to bias in models. The effect of this has been explored in the spatial literature (see Stetzer 1982; Florax and Rey 1995; Griffith 1996; Griffith and Lagona 1998) and the network literature (Mizruchi and Neuman 2008; Páez et al. 2008; Farber et al. 2009; Neuman and Mizruchi 2010). We therefore discuss each of the issues in turn below in the context of road networks.

2.2 Network structure and adjacency matrix

The definition of the linkage structure in a spatial weight matrix is very much dependent on the type of data under study, for instance, in areal data, contiguity may be used (bishop, rook, or queen); in point data, distance bands or nearest neighbours may be used (Getis and Aldstadt (2004) list several alternatives). On intangible networks, the task may be very difficult because linkages need not depend only on spatial proximity but can be based on social, cultural and other factors. For instance, in social networks, the units of analysis are individual social entities linked to other individuals by affective, professional, assistance, and other types of social linkages and deciding the nature of these linkages is a problem for the researcher (Páez et al. 2008). Leenders (2002) discusses the construction of weight matrices in such data.

On road networks, the task is more straightforward if the topology of the network from which the data are collected is well defined. Drawing from graph theory, it is convenient to view a network as a graph $G = (N, E)$ with a set N of n nodes and a set E of edges joining pairs of nodes. The incidence structure of the graph is defined by the presence or absence of an edge (i, j) linking nodes i and j and can be represented by an $N * N$ binary [0, 1] adjacency matrix in which non-zero elements signify edges (Peeters and Thomas 2009). Two nodes directly linked by an edge are termed first-order neighbours. The adjacency matrix containing all first-order relations between the nodes of a graph 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 and so forth. By following the paths

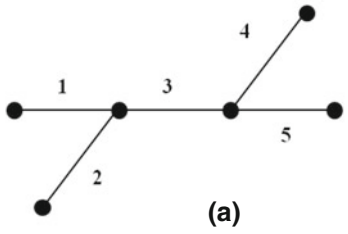
between nodes in the graph, adjacency matrices W_1, W_2,…W_k of orders up to k can be defined.

The incidence structure of networks can be viewed in one of two ways. Either the nodes correspond to measurement locations of a variable with the edges representing communication links between them, as is the norm in the spatial and social networks literature (see, for instance, Dow 2007); or both i and j are edges where a variable is observed and the nodes represent connections between them. The latter formulation was devised by Black (1992) and has been applied to transport networks (Black and Thomas 1998) and migration flows (Chun 2008) among others. We use the latter formulation as we believe it better represents the physical structure of road networks.

In addition to defining the incidence structure, the direction of influence must be defined. Networks can be undirected or directed depending on the supposed direction of influence between edges and different adjacency matrices result from each case. In Fig. 1, as the network is undirected (Fig. 1a), all edges that share a common node are considered to influence each other and this is reflected in the first-order (Fig. 1b) and second-order (Fig. 1c) weight matrices. In a directed network (Fig. 2a), the flow goes in one direction and an edge only influences its edges downstream. Figure 2b and c shows the first- and the second-order weight matrix for this standard directed network case, which is asymmetric since it allows influence only in one direction (from upstream).

Conventionally, undirected graph structure has been adopted for networks in the social and cultural research literature (see for instance, Dow (2007) for cultural data; Chun (2008) for migration). Undirected (Black and Thomas 1998) and directed structure (Kamarianakis and Prastacos 2005) have both been considered for road networks. In our opinion, however, road networks fall somewhere between directed and undirected networks in that the traffic only flows in one direction but the influence between links can occur in both directions. However, a traffic link can only be affected *directly* by its first-order adjacent downstream and upstream traffic, whereas two links flowing into or out of the same link from the same direction will be second-order rather than first-order neighbours.

This can be explained by the example given in Fig. 3 where link 1 and link 2 flow into link 3; if an increase in traffic on link 2 contributes to congestion downstream



(a)

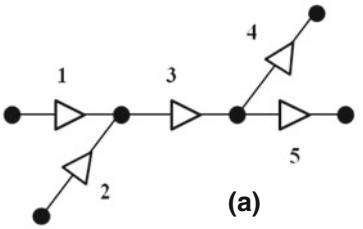
	1	2	3	4	5
1	0	1	1	0	0
2	1	0	1	0	0
3	1	1	0	1	1
4	0	0	1	0	1
5	0	0	1	1	0

(b)

	1	2	3	4	5
1	0	0	0	1	1
2	0	0	0	1	1
3	0	0	0	0	0
4	1	1	0	0	0
5	1	1	0	0	0

(c)

Fig. 1 A simple undirected network and its spatial weight matrices based on connectivity of edges; **a** the undirected network; **b** first-order weight matrix; **c** second-order weight matrix



	1	2	3	4	5
1	0	0	1	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	0	0	0	0	0
5	0	0	0	0	0

(b)

	1	2	3	4	5
1	0	0	0	1	1
2	0	0	0	1	1
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0

(c)

Fig. 2 A simple directed network and its spatial weight matrices based on connectivity of edges; **a** the directed network; **b** first-order weight matrix; **c** second-order weight matrix

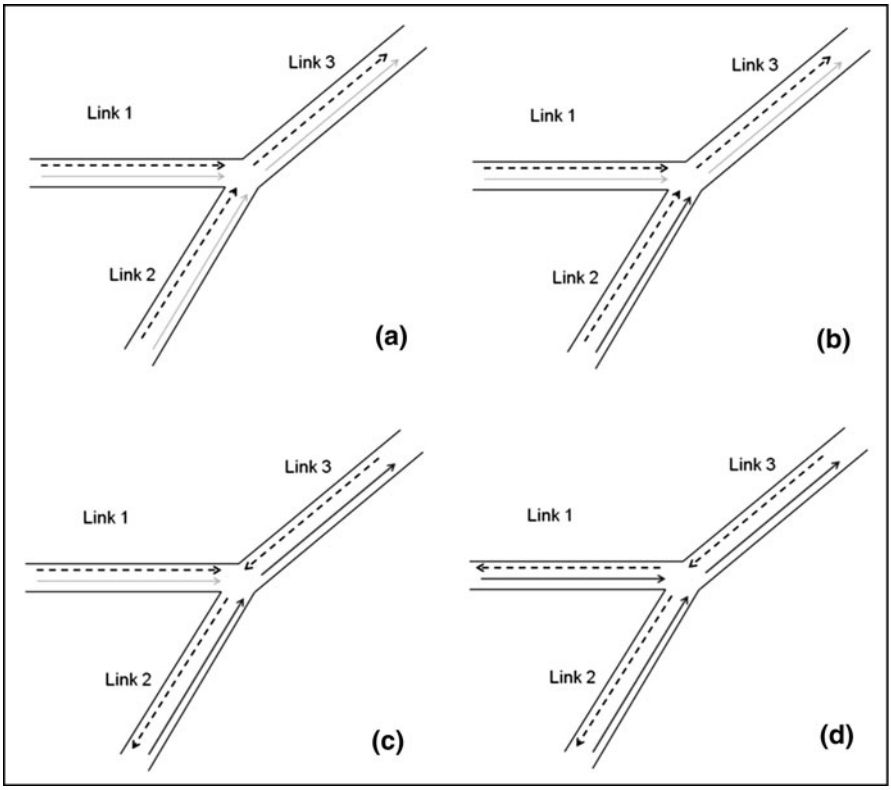
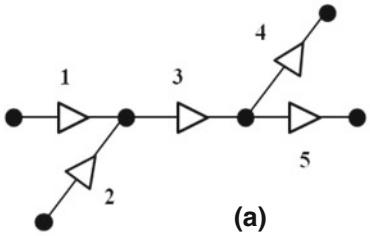


Fig. 3 An example of the propagation of traffic conditions at a junction; *arrows* indicate flow direction, *light colour* indicates low flow, *dark colour* indicates high flow **a** free-flowing situation, influence is mainly from upstream **b** flow on link 2 increases **c** Increase in flow on link 2 contributes to congestion on link 3, the direction of influence on links 2 and 3 reverses **d** congestion on link 3 causes congestion on link 1

on link 3, then this congestion may propagate upstream to link 1. Therefore, link 2 and link 1 are not directly influenced by each other, but via link 3 in between—they are not first but second-order spatial neighbours, and both are first-order spatial



	1	2	3	4	5
1	0	0	1	0	0
2	0	0	1	0	0
3	1	1	0	1	1
4	0	0	1	0	0
5	0	0	1	0	0

(b)

	1	2	3	4	5
1	0	1	0	1	1
2	1	0	0	1	1
3	0	0	0	0	0
4	1	1	0	0	1
5	1	1	0	1	0

(c)

Fig. 4 A simplified transport network and its spatial weight matrices based on connectivity of edges; **a** the directed network; **b** first-order weight matrix in the transportation case; **c** second-order weight matrix in the transportation case

neighbours of link 3. This is reflected in Fig. 4c, where link 1 and link 2, and link 4 and link 5, are second-order spatial neighbours. The second-order spatial weight matrix is symmetric like the undirected graph in Fig. 1c, but with different values. Figure 4b shows the first-order weight matrix with influences allowed in both directions, which is symmetric like the undirected network shown in Fig. 1b, but with different values. The implication of this structure is that it accounts for the indirect influence between link 1 and link 2 as second-order spatial neighbours through link 3. Similarly, it accounts for the indirect influence between link 4 and link 5 as second-order spatial neighbours through link 3.

2.3 The spatial weights

Defining the weights w_{ij} of a spatial weight matrix again depends on the type of data under investigation. The simplest weighting system is binary, i.e. $w_{ij} = 1$ if an edge exists between nodes i and j and zero otherwise. Weight matrices defined in this way will often be row standardized to distribute the contribution of each of the nodes in the spatial neighbourhood of a node equally. However, more sophisticated schemes can be defined that assign different weights to different neighbours. For instance, in areal data, the weight may be defined as the length of a shared border divided by the perimeter; in point data, the weight may be some form of distance decay function (Getis and Aldstadt 2004). Readers are again referred to Leenders (2002) for a discussion of weighting schemes in social network data.

On transportation networks, binary (Kamarianakis and Prastacos 2005; De Montis et al. 2011) and distance weighting (Wang et al. 2010) schemes have been employed in the literature to represent the static physical structure of road networks. More sophisticated schemes have also been employed that take into account historical conditions on the road network at different times (Min et al. 2007). Here, we make use of a simple binary weighting scheme with row standardization as it is easy to implement and interpret, and we do not want to introduce assumptions about the relative influence of the spatial neighbourhood at this stage.

3 Measurement of autocorrelations in space–time

3.1 Autocorrelation

Various indices have been devised to quantify autocorrelation in spatial and temporal data, most of which are based on Pearson's familiar product moment correlation coefficient (PMCC) (Soper et al. 1917) which, for two variables X and Y, is defined as their covariance divided by the product of their standard deviations:

$$\rho_{X.Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \tag{1}$$

where μ_X and μ_Y and σ_X and σ_Y are the means and standard deviations of variables X and Y, respectively. The coefficient $\rho_{X.Y}$ is used as a measure of the strength of linear dependence between variables and can fall in the range -1 to 1 with 1 indicating perfect positive correlation, -1 perfect negative correlation and 0 no correlation. Rodgers and Nicewander (1988) provide an overview of some alternative formulations of the correlation coefficient that are not discussed here. Autocorrelation can be measured simply by taking the correlation of a variable with a lagged specification of itself; therefore, temporal autocorrelation can be measured by modifying PMCC to include this lagged specification:

$$\rho_k = \frac{E[(z_t - \mu)(z_{t-k} - \mu)]}{\sigma_z^2} \tag{2}$$

The difference being that the covariance is measured between variable z at time t and variable z at time $t - k$. If the process is stationary, then σ_z^2 can be used as the standard deviation of z, which is assumed to be constant at all times.

Spatial autocorrelation is more complicated than temporal autocorrelation as it can occur in any direction. Moran's I (1950) is an extension of the PMCC to the spatial domain that has been widely used in many spatial applications. It has a local variant I_i that can be used to identify local clusters of spatial autocorrelation (1). Moran's I and other spatial indices such as Geary's (1954) do not consider the temporal dimension so they only provide a snapshot of autocorrelation and do not capture dynamic autocorrelation properties. For this, indices are required that measure spatial and temporal autocorrelation simultaneously.

A number of indices have been devised to this end, Hardisty and Klippel (2010) proposed a spatio-temporal extension of local Moran's I and applied it to an analysis of the 2009 H1N1 flu pandemic. Space–time (semi) variograms have also been proposed (see for example, Griffith and Heuvelink 2009) as well as space–time eigenvector filtering (Griffith 2010). Two indices are borrowed here: the space–time autocorrelation function (ST-ACF) that measures global space–time autocorrelation and the cross-correlation function (CCF) that measures local space–time autocorrelation between two locations. These indices are extensions of the temporal autocorrelation function (Eq. 2) and are selected as they are easily interpretable and have a practical application in established space–time modelling frameworks.

3.2 The global measure

The ST-ACF measures the N^2 cross-covariances between all possible pairs of locations lagged in both time and space (Pfeifer and Deutsch 1980). Given the weighted lth order spatial neighbours of any spatial location at time t and the weighted kth order spatial neighbours of the same spatial location s time lags in the future, the space–time cross-covariance can be given as:

$$\gamma_{lk}(s) = E\left\{ \frac{\left[W^{(l)}z(t)\right]' \left[W^{(k)}z(t+s)\right]}{N} \right\} \tag{3}$$

where N is the number of spatial locations, $W^{(l)}$ and $W^{(k)}$ are the $N * N$ spatial weight matrices at spatial orders l, and k $z(t)$ is the $N * 1$ vector of observations z at time t, $z(t+s)$ is the $N * 1$ vector of observations z at time $t+s$ and the symbol $'$ denotes matrix transposition. Based on Eq. 3, the ST-ACF can be defined as:

$$\rho_{lk}(s) = \frac{\gamma_{lk}(s)}{\left[\gamma_{ll}(0)\gamma_{kk}(0)\right]^{1/2}} \tag{4}$$

ST-ACF has been used in STARIMA to calibrate the order of moving average (MA), which define the range of spatial neighbourhoods that contribute to the current location at a specific time lag (Pfeifer and Deutsch 1980). The MA orders are fixed globally both spatially and temporally, and a single parameter is estimated for it in practical application such as by Kamarianakis and Prastacos (2005).

3.3 The local measure

The cross-correlation function (CCF) (see, for example, Box and Jenkins 1970) treats two time series as a bivariate stochastic process and measures the cross-covariance coefficients between each series at specified lags. It provides a measure of the similarity between two time series. Given two time series X and Y, the CCF at lag k is given as:

$$\rho_{xy}(k) = \frac{E[(x_t - \mu_x)(y_{t+k} - \mu_y)]}{\sigma_x \sigma_y} \quad k = 0, \pm 1, \pm 2, \pm \ldots \tag{5}$$

The CCF is a lagged specification of PMCC that measures cross-correlations in both directions, as denoted by subscript k; therefore, the temporal lag at which the CCF peaks can be used to determine a transfer function between two series. A simple way to interpret the CCF is by taking its squared value ρ^2 to give the coefficient of determination (CoD). Multiplying this value by 100 gives the percentage of variance that two series share at a given time lag.

Yue and Yeh (2008) demonstrate through an empirical example that the cross-correlation function (CCF) can be used to determine the spatio-temporal relationship between a road link and its neighbours. This is, however, dependent on sufficient spatial and temporal resolution in the data. A peak at lag zero indicates that the current resolution does not capture the direction of influence of one location

on another, but behave very similarly at the same time. In transport case, this usually happens when the network is highly congested in the morning peak.

When measuring autocorrelation, it is important that the method employed provides an unbiased indicator of the true autocorrelation in the data. Olden and Neff (2001)used Monte Carlo simulation to develop formulas to quantify the bias in cross-correlation measures as a function of sample size, true correlation between series and the number of time lags examined. They examined series of length up to 100 observations and found that the bias introduced decreases with series length. Although no literature yet exists on bias introduced by the ST-ACF, we consider the series length in this study ($\gg$100) sufficiently long as to minimize the bias in the autocorrelation calculations. Additionally, in the context of forecasting, we believe that the misspecification of linkage (adjacency) structure (spatial orders) could be corrected in cross-validation so that only valid orders are included in the models. That's why it is important to have validation test in space–time modelling (Box and Jenkins 1970).

In the case study, the global (ST-ACF) and local (CCF) measures are applied to journey time data in central London in order to gain an understanding of the complexity of spatio-temporal autocorrelations on road networks.

4 Empirical example

4.1 The LCAP data and test network

The London Congestion Analysis Project (LCAP) network is a system of automatic number plate recognition (ANPR) cameras maintained by Transport for London (TfL) that collect journey time information on London's road network (see Fig. 5). The cameras operate in pairs; number plates are read as vehicles pass each camera and the time taken between passing the first and second cameras is recorded. These raw journey time observations are averaged over a 5-min period to give the journey time data (in seconds) used in this study. The ANPR camera network was originally installed to help enforce London's congestion charging scheme so collecting journey time data is a secondary use of the ANPR system; as such it was not designed with this in mind. Therefore, there are a number of issues that must be considered:

- Link lengths (distance between the first and the second cameras) are heterogeneous. Links vary in length from 207.7 m to 15.5 km.
- The incidence structure of the ANPR network does not fully mimic that of the real road network so many junctions may be included in one link and many minor roads are omitted due to the limit of setting up the cameras.
- The links do not form a complete network, and many links are not connected to each other or may overlap on the same road section in the same direction.
- Data quality depends on capture rates; during the night time period journey time observations can be subject to error due to few vehicles on the road.

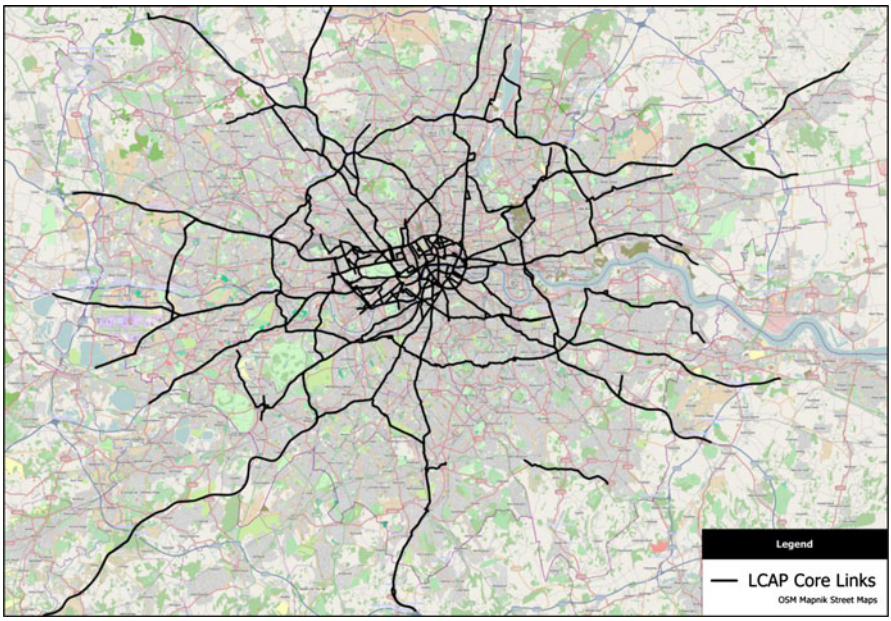


Fig. 5 Spatial extent of the LCAP core links network in London (Open Street Map Mapnik base map)

Although data are available for the whole ANPR network, the aforementioned reasons make a network-wide study infeasible at this stage. Therefore, a subsection of the network is chosen (Fig. 6a). The test network comprises 22 links in central London and was selected as its incidence structure can be well defined as in Fig. 6b. It has variable link lengths, ranging from 473.4 m to 3.85 km with an average length of 1.4 km.

After conversations with experts from TfL, data for 33 consecutive Tuesdays from 6 January 2009 to 18 August 2009 were selected. Tuesday data only are chosen as the behaviour of traffic that is known to be different on different days of the week, and Tuesday is presumed to be close to an 'average' weekday, separate from the influences of weekend traffic patterns. Nine days were removed (days 2–6 and days 19–22) as the data for several links were replaced with profile data for the whole day. Since the data are obtained at 5-min interval, there are 288 observations per day. Therefore, taking into account the omitted data, the space–time series comprises $24 * 22 * 288 = 152{,}064$ observations. Due to the variable link lengths, it is decided to convert the raw journey times into a relative measure of link performance. Either speed or unit journey time (UJT, inverse of speed) would be appropriate; in this case, UJT is preferred as it can be interpreted as a measure of delay/congestion and hence is useful to traffic managers. The average unit journey time between 7 am and 7 pm on a selection of links over these 24 Tuesdays is presented in Fig. 7, which shows a very dynamic and heterogeneous traffic situation, i.e. some links are smooth while others are congested.

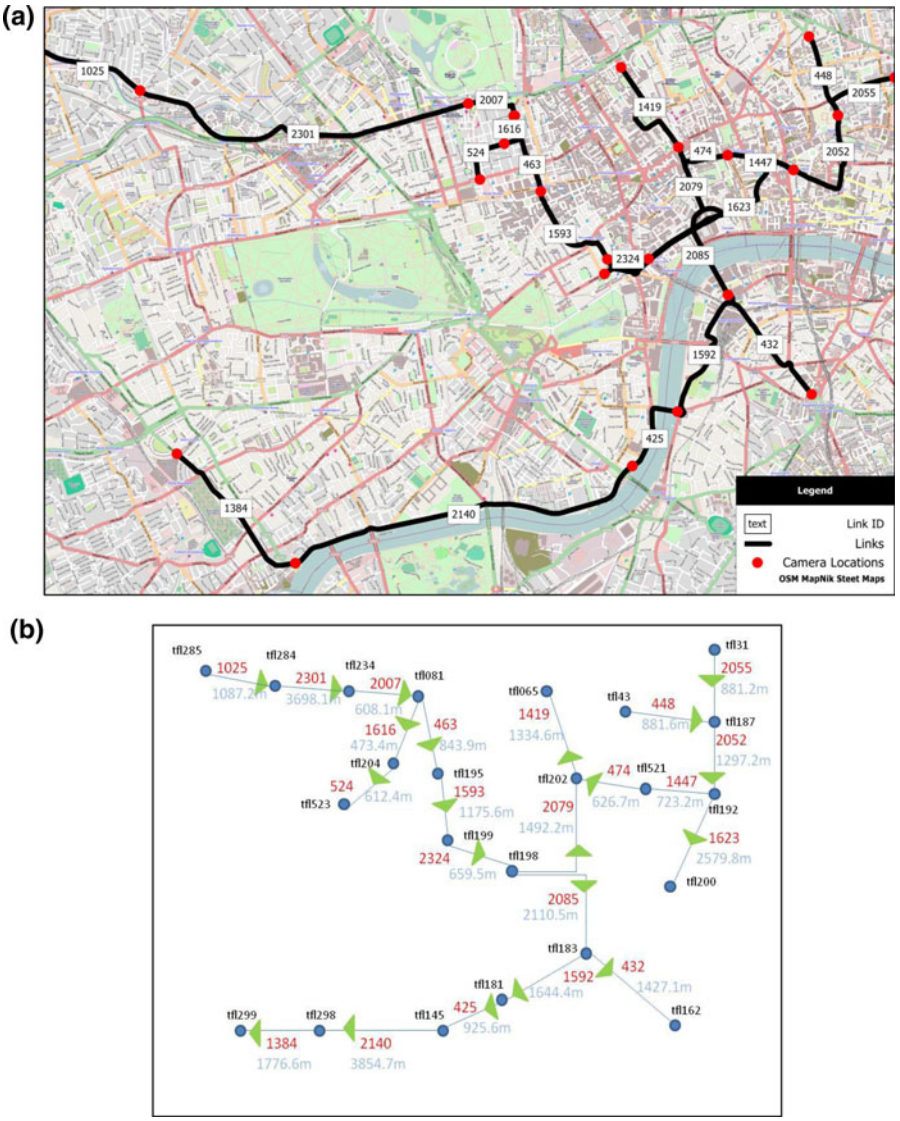


Fig. 6 Selected Road Network in Central London **a** spatial location of selected links in central London; **b** network diagram of links, *arrows* represent traffic flow direction, integer numbers are link IDs, numbers starting with tfl are ANPR camera IDs and numbers ending with m are link lengths in meters

4.2 Experimental procedure

To examine how the autocorrelation structure changes throughout the day, the data are divided into three distinct time periods: AM peak (7:00–10:00), interpeak (10:00–16:00) and PM peak (16:00–19:00), as defined by TfL. It has been widely accepted in transport studies that traffic behaves differently in each time period. Due

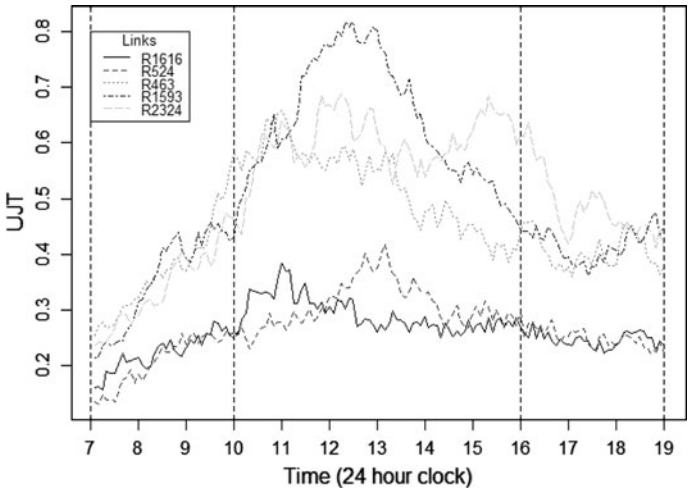


Fig. 7 Average unit journey time profiles over 24 Tuesdays from 6 January 2009 to 18 August 2009 (9 weeks are omitted)

to the length of the links, it is considered unnecessary to examine spatial orders higher than three as they will be outside the forecasting horizon of 5 min. Therefore, only three spatial weight matrices W_1, W_2 and W_3 are defined. The matrices are defined based upon the network structure graph, which allows traffic flowing in one direction and influence between links occurring in both directions, i.e. the immediate upstream/downstream links are first-order spatial neighbours such as links R2007 and R1616 which have a value 1 in W_1 as shown in Fig. 8, but link R1616 and link R463 are only second-order spatial neighbour which has a value 1 in W_2.

The analysis is threefold; firstly, the global STACF is calculated at spatial orders zero, one, two and three for each of the time periods. Secondly, the pairwise cross-correlations between individual links on a subset of the network are calculated and compared with the global pattern. We then focus the cross-correlation analysis on a subset of the network for a directional analysis. Finally, we examine the spatial pattern of correlation in each of the time periods.

5 Experimental results

5.1 Global space–time autocorrelations

Figure 9 shows the ST-ACF between all links in the AM peak at spatial orders zero to three. The horizontal axis shows the temporal lag, and the vertical axis shows the value of the ST-ACF. Under the assumption of space–time stationarity, the ST-ACF should be insignificant at all lags greater than zero, and hence, significantly positive values indicate the presence of space–time autocorrelation. A cyclic pattern indicates the presence of a seasonal pattern in the data. There is a clear seasonal pattern present at all spatial lags with a period of 36, corresponding to the length of

R1025	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R2301	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R2007	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R1616	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R524	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R463	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R1593	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
R2324	0	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0
R2085	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
R432	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
R1592	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
R425	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0
R2140	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
R1384	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
R2079	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
R1419	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
R474	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
R1447	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0
R1623	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
R2052	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
R448	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
R2055	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Fig. 8 First-order adjacency matrix for the test network (column names correspond to row names)

the peak period. This indicates concurrent increases and decreases in unit journey times as the AM rush hour traffic builds and then recedes towards the interpeak period, which is to be expected. At spatial order zero (equivalent to the average temporal autocorrelation), the strength of autocorrelation is very strong and shows very slow decay. This is indicative of non-stationarity in the data and, combined with the seasonal component, provides evidence of strong non-seasonal and seasonal dependency. At spatial order one, the strength of autocorrelation decreases. However, the periodic component is still evident indicating some spatio-temporal dependency between locations. This pattern is replicated at spatial orders two and three but the autocorrelations are insignificant at the 95% confidence level at these orders.

The interpeak period (Fig. 10) shows a similar pattern to the AM peak at spatial order zero, displaying strong significant positive autocorrelation with very slow decay and a seasonal component corresponding to the length of the period (72 observations). However, the amplitude of the seasonal component is lower. Theoretically, the interpeak period should capture a traffic state separating the two peak periods characterized by a return to free-flowing conditions and therefore should not display seasonal dependency. The presence of such dependency indicates that the arbitrary division of the data into three time periods based on accepted definitions is not sufficient to isolate the specific traffic states on the network.

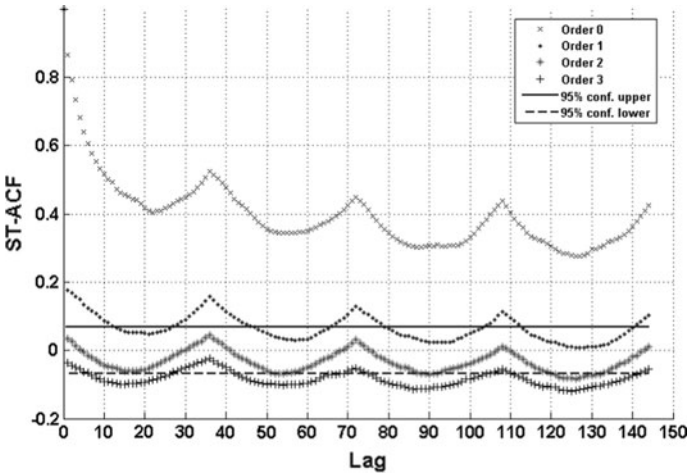


Fig. 9 ST-ACF for the AM peak at spatial orders 0, 1, 2 and 3

At spatial order one, the pattern of significant autocorrelation remains, but the strength of autocorrelation decreases and the seasonal component is less apparent. The strength of the ST-ACF is slightly higher than in the AM peak period. Autocorrelations do become insignificant at spatial order two but then become significantly negative at spatial order three, again in contrast to the AM peak.

In the PM peak period (Fig. 11), the autocorrelations are again strong, significant and positive at spatial order zero. However, the strength of the seasonal component is less than in the AM peak and interpeak periods. This is contrary to expectations and suggests that the definition of the PM peak period (16:00–19:00) does not isolate a particular traffic state. The presence of a stronger seasonal component in the interpeak period suggests that the PM peak period may begin earlier than the threshold time and may also end later. Therefore, the transition from free flowing to congested and back to free-flowing conditions is not captured within this period. At spatial orders one to three, the PM peak follows the same pattern as the interpeak.

Examining the global ST-ACF at each of the time periods reveals two main findings. Firstly, the strength and pattern of spatio-temporal autocorrelation are *temporally dynamic*, reflecting the different traffic patterns that occur in each period. Secondly, the pattern of space–time autocorrelation remains broadly similar but *decreases in strength at increasing spatial orders* in each of the time periods. The implication of this is that the global effective data range can be determined from the ST-ACF as the spatial order where autocorrelations become statistically insignificant. In each of the time periods, this occurs at spatial order two. Intuitively, this makes sense given the high average length of links on the network (the mean journey time of the network is four and a half minutes, and the temporal resolution is 5 min). The global ST-ACF reveals much about the global space–time autocorrelation properties of the data but does not reveal anything about the local properties. In the following section, the CCF is applied to examine the local

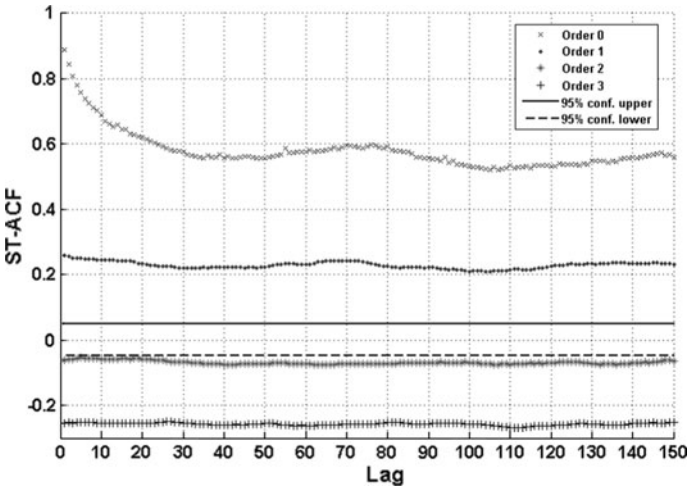


Fig. 10 ST-ACF for the interpeak at spatial orders 0, 1, 2 and 3

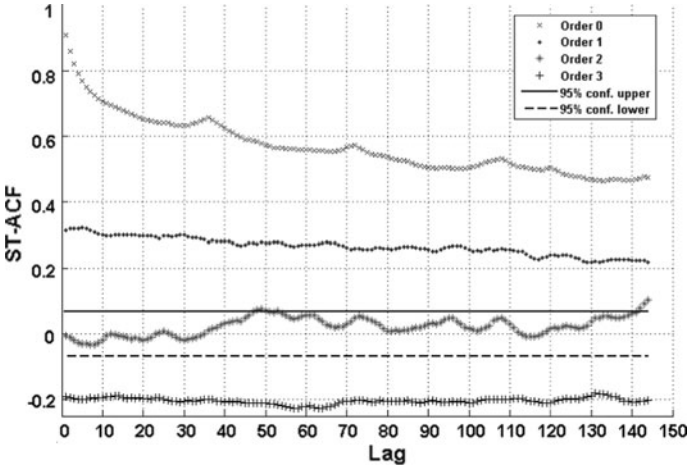


Fig. 11 ST-ACF for the PM peak at spatial orders 0, 1, 2 and 3

cross-correlations between pairs of links. First, the overall picture is summarized and then, a subset of the network is examined for a more detailed analysis.

5.2 Local cross-correlations

Examining the CCF between pairs of links gives a much more local view of the autocorrelation structure of the road network. Determining if and how the local view differs from the global view provided by the ST-ACF is important from a modelling perspective as it gives an idea as to whether the autocorrelation structure is global or local and whether global autocorrelation parameters are likely to be effective in a

space–time model. Like the ST-ACF, the CCF is also a temporally global indicator. Therefore, we first examine the CCF for the whole day and then for the AM peak, interpeak and PM peak periods, respectively, to see how the autocorrelation structure varies locally, both spatially and temporally. On each of the figures, temporal lag zero is shown at the centre. Positive lags indicate the influence of downstream links on upstream links, while negative lags indicate upstream influence on downstream. The vertical axis shows the value of the CCF, and the 95% significance levels are shown as horizontal red lines. Although the CCF was calculated between all pairs of links for all orders, we only show the first-order CCF for 3 pairs of links for conciseness.

In terms of the daily pattern (Fig. 12), considerable heterogeneity in the strength of the CCF between links is apparent. For instance, at temporal order zero and spatial order one, the value of the CCF ranges from nearly zero to 0.50. This means that some links are highly correlated with their neighbours, while others show no significant correlation at all. This is particularly unexpected at spatial order one given that the links are physically adjacent. The corresponding ST-ACF indicates significant positive space–time autocorrelation, which is true for some links but not for others. The same phenomenon is present at spatial orders 2 and 3, although the overall strength of correlation falls, which is to be expected.

Between time periods, the results of the local CCF analysis mirror the results of the global analysis in that the strength and pattern of correlation is different in each period. This is masked by the temporally global CCF. Although the peak value of the CCF is similar in each of the time periods; 0.51, 0.57 and 0.46 in the AM (Fig. 13), inter (Fig. 14) and PM (Fig. 15) peaks, respectively, at spatial order one, it occurs between different pairs of links in each case. In the AM peak the peak correlation is between links R1593 and R2324, in the interpeak it is between links R2140 and R1384 and in the PM peak it occurs between links R425 and R2140. Although the peak occurs in the same area of the network in the interpeak and PM peak periods, it is in a different area in the AM peak, indicating that the *autocorrelation structure is changing both spatially and temporally*.

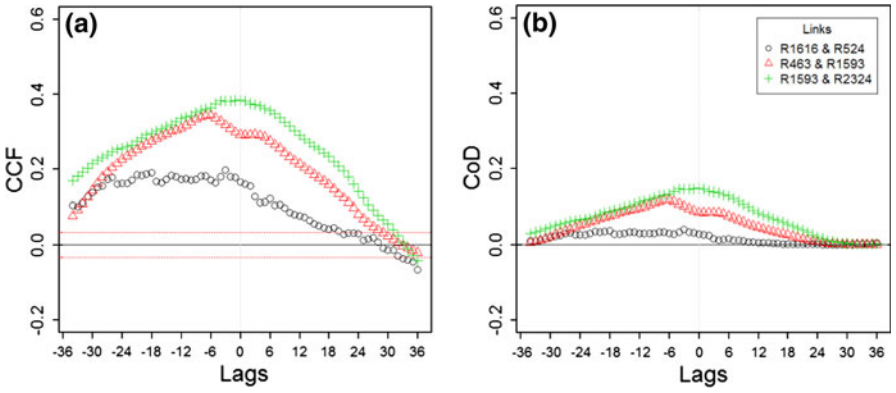


Fig. 12 CCF (**a**) and CoD (**b**) for daytime period

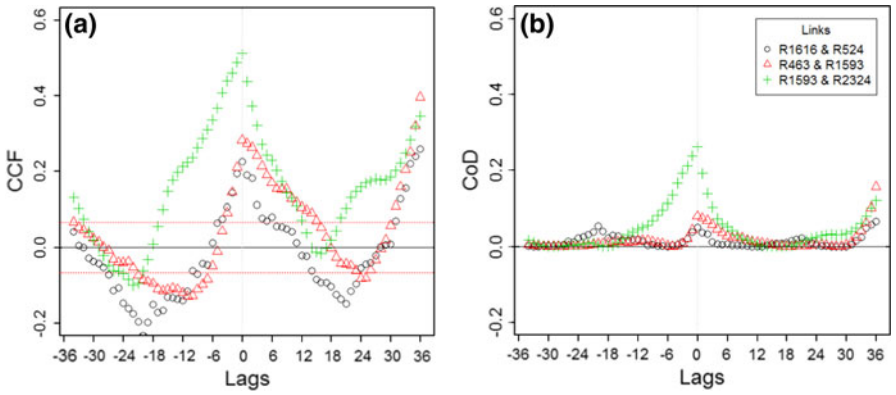


Fig. 13 CCF (**a**) and CoD (**b**) for AM peak period

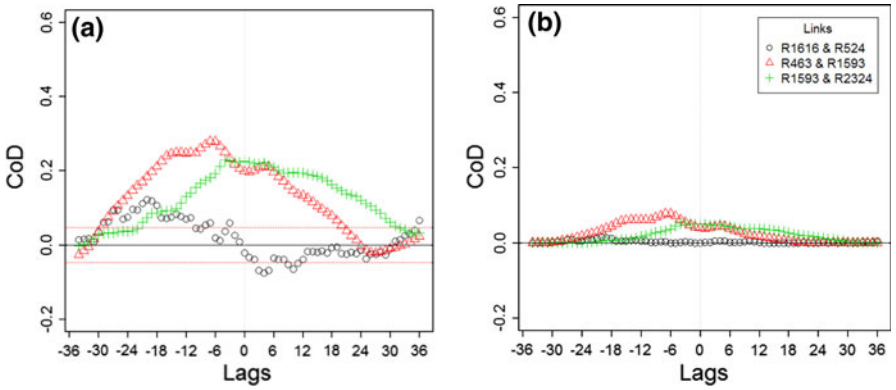


Fig. 14 CCF (**a**) and CoD (**b**) for interpeak period

Generally, the strength of correlation and periodic component is higher in the AM peak period than the interpeak and PM peak periods, although there is wide spatial variation between pairs of links. In the AM peak at spatial order one, the general pattern of the CCF matches that of the global ST-ACF with a strong periodic component of length 36 apparent between many pairs of links. However, there is considerable spatial heterogeneity in the strength of autocorrelation, and some links display insignificant cross-correlations. The situation is reversed at spatial order two; although the ST-ACF indicates insignificant global spatio-temporal autocorrelation at this order, the CCF reveals *significant local spatio-temporal autocorrelation between some pairs of links*.

Insignificant ST-ACF values are indicative of a stationary space–time process; however, these results demonstrate that local significant autocorrelation may be masked by the global indicator. The impacts on modelling may be substantial and difficult to detect. For instance, if a STARIMA model is built based on the assumption of stationarity as indicated by the ST-ACF but the space–time process is

in fact non-stationary, then modelling results are likely to be poor. The presence of significant autocorrelation extends to spatial order three but the strength of the pattern is diminished.

In the interpeak and PM peak periods, similar issues are apparent whereby the cross-correlations broadly reflect the global pattern but display considerable heterogeneity that the global indicator cannot account for. The implication of this is that although the global indicator can capture the global autocorrelation structures quite well it is inadequate in describing the local, dynamic autocorrelation structure in the data. This goes some way towards explaining why space–time models of traffic data based on global autocorrelation parameters exhibit unsatisfactory performance.

5.3 Directional analysis

Thus far, we have neglected to discuss the direction of influence between neighbouring links, and this issue is treated here based on the pairs of links shown in Figs. 12, 13, 14. In the daytime period (Fig. 12), the asymmetry in the CCF is immediately apparent, with the peak CCF being seen at negative lags between links R1616 and R524 and R463 and R1593. The peak CCF between R1593 and R2324 occurs at lag zero but the pattern of decay is similar to the other pairs of links. This demonstrates that over the course of a day, the direction of influence between these links is predominantly upstream to downstream, which is not captured in the global measure.

Examining the CCFs for each of the daytime periods separately provides additional insights into the results. In the AM peak period (Fig. 13), each pair of links displays a similar pattern to one another. The peak CCF occurs at lag zero, and there is sharp decay at subsequent positive and negative lags, meaning that neighbouring links are most correlated with each other contemporaneously. This means that the direction of influence between links cannot be determined from the data in its current form; this is in contrast to the daily pattern. The strength of the periodic component is also apparent representing a recurrent daily traffic pattern, which reflects the global pattern captured by the ST-ACF. However, the strength of the CCF varies between pairs of links, with the most correlated being links R1593 and R2324.

The situation in the interpeak period (Fig. 14) is very different to the AM peak and appears to reflect the daily picture in terms of the patterns that are evident. All pairs of links display peak values of the CCF at negative lags indicating upstream to downstream propagation of conditions with varying time lags. In the PM peak, however (Fig. 15), the CCF between two pairs of links, R1616 and R524, and R463 and R1593 shows no clear pattern and fluctuates approximately around zero. Although significant cross-correlations are observed between R1616 and R524, their pattern appears random and the CCF between R463 and R1593 is insignificant at all lags. The implication of this is that there is little or no spatial dependency between these links in the PM peak. However, links R1593 and R2324 display an interesting result. Compared with the interpeak period, the direction of influence has reversed, with the influence now coming from downstream. It is likely that *this shift in*

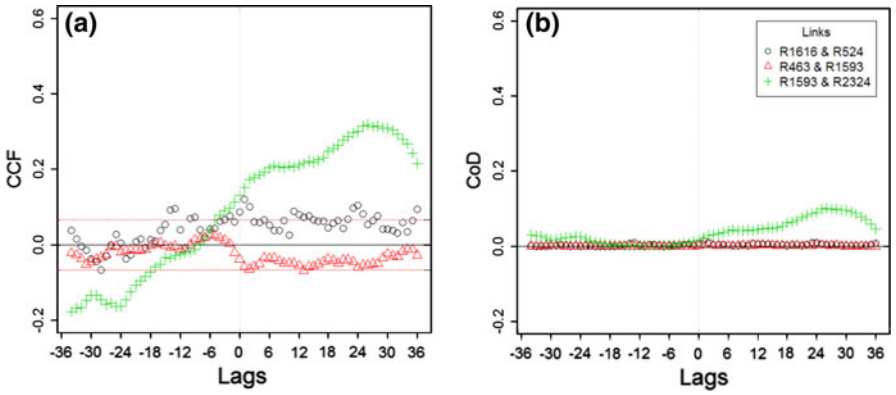


Fig. 15 CCF (**a**) and CoD (**b**) for PM peak period

direction will be different depending on the characteristics of links, for instance inbound and outbound links.

Also, surprising is the low overall level of correlation between locations. Examining the CoD between the pairs of links (Figs. 12b, 13b, 14b, 15b), the shared variance between first-order adjacent links on the subnetwork is a maximum of 7.8% in the interpeak period and 10.1% in the PM peak period. This rises to 26.1% in the AM peak period. This also serves to highlight the increased strength of spatio-temporal dependency in the AM peak.

5.4 Spatial pattern of correlation

The preceding section showed that the cross-correlations vary greatly between pairs of links, but how do these correlations vary spatially, and can any pattern be observed?

Figure 15 shows how the spatio-temporal dependency structure varies geographically in time. The network is more correlated in the AM peak than the interpeak and PM peak; however, there is a noticeable divide to the north-west where the cross-correlations between neighbours become weakly negative. The strongest positive correlations occur to the south and north-east of the network. This pattern is largely borne out in the interpeak and PM peak periods, although the negative correlation extends further east with the links to the south remaining positively correlated with their first-order neighbours. The magnitude of the negative correlations also increases from the interpeak to the PM peak. The reasons for this spatial pattern are not apparent and require further investigation (Fig. 16)

6 Discussion and conclusions

The space–time autocorrelation structure of London's road network has been shown to be dynamic in time and differs with traffic states. This is evident when observing

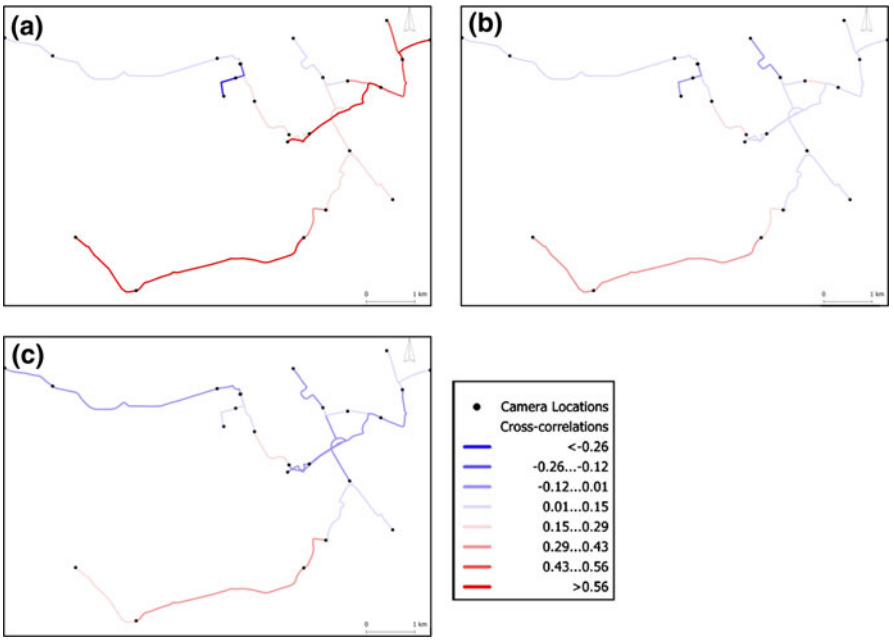


Fig. 16 Average CCF between links and their first-order neighbours at temporal lag zero in **a** the AM peak; **b** interpeak; and **c** PM peak

the differences between the AM peak, interpeak and PM peak periods. However, although the AM peak traffic state appears reasonably well defined, this is not the case in the interpeak and PM peak periods. The complex interactions and demand patterns of urban traffic mean that spatially contemporaneous traffic states are not evident to a large extent. The influence of links on their neighbours is local and varies widely both in space and time; there is also evidence of changing direction of influence that is not captured in the global autocorrelation measure.

The finding from the empirical case study shows that the spatio-temporal network autocorrelation structure is dynamic in time and heterogeneous in space, which is a direct reflection of dynamics and heterogeneity of network complexity. As mentioned in the introduction, network science has gained progress in measuring and discovering the topology structure of networks, which is more a measure of the heterogeneity than the dynamics of network. They are useful for finding most vulnerable or important links in the network. However, most existing finding are based on static network or using static data or simulated data. The study of network dynamics and their influence on network behaviour and performance is far from mature. Our research presented here is an attempt for such purpose.

These findings have significant implications for space–time modelling of traffic (and possibly other) data on networks. Firstly, the complexity of the spatio-temporal autocorrelation structure means that extracting a stationary space–time process from the data is likely to be difficult, especially in real time. This means that models such as STARIMA that rely on such assumptions being fulfilled are likely to have low

predictive power. Additionally, the global parameter estimates that STARIMA and other models rely on are likely to be insufficient to explain all the local variations in the data; they assume a globally stationary space–time autocorrelation structure exists, when in reality it may not. The recent effort to this problem that has been examined in the literature (see for instance, Elhorst 2003; Min et al. 2007, 2009, 2010) is to create separate models for different traffic states but this relies on those traffic states being identical and identifiable as those in the pre-defined state models. For the reasons given elsewhere, this is likely to be difficult.

Balancing model complexity and parsimony is vital in building an effective forecasting model that is capable of operating on real-time data and capturing the dynamic autocorrelation structure that has been identified. The key to this is leveraging the forecastability of the data in the model structure in order to capture the dynamics of the network processes. Hence, a model is needed that accounts for the instantaneous forecastability of traffic data. It is the opinion of the authors that this can be achieved by defining a *dynamic spatial neighbourhood* based on the instantaneous data range and a *dynamic spatial weight* based on the strength of local autocorrelation. These can be incorporated into a model through the use of a dynamic spatial weight matrix. The implementation of such a model will be the focus of future research.

Acknowledgments 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 under Research Grant EP/G023212/1. The support from Chinese NSF (40830530) is acknowledged. The authors are grateful to three anonymous reviewers and the editor for their valuable suggestions.

References

Anselin L (1995) Local indicators of spatial association: LISA. Geogr Anal 27(2):1–25

Black WR (1992) Network autocorrelation in transportation network and flow systems. Geogr Anal 24(3):207–222

Black WR, Thomas I (1998) Accidents on Belgium's motorways: a network autocorrelation analysis. J Transp Geogr 6(1):23–31

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

Castells M (2010) Globalisation, networking, urbanisation: reflections on the spatial dynamics of the Information Age. Urban Stud 47(13):2737–2745

Chandra S, Al-Deek H (2008) Cross-correlation analysis and multivariate prediction of spatial time series of freeway traffic speeds. Transp Res Rec 2061:64–76

Chun Y (2008) Modeling network autocorrelation within migration flows by eigenvector spatial filtering. J Geogr Syst 10(4):317–344

Cliff AD, Ord JK (1969) The problem of spatial autocorrelation. In: Scott AJ (ed) Lond Pap in Reg Sci. Pion, London, pp 25–55

De Montis A, Caschili S, Chessa A (2011) Time evolution of complex networks: commuting systems in insular Italy. J. Geogr Syst 13(1):49–65

Ding Q, Wang X, Zhang X, Sun Z (2011) Forecasting traffic volume with space-time ARIMA model. Adv Mater Res 156–157:979–983

Doreian P, Teuter K, Wang C (1984) Network autocorrelation models: some Monte Carlo results. Sociol Methods Res 13(2):155–200

Dougherty MS, Cobbett MR (1997) Short-term inter-urban traffic forecasts using neural networks. Int J Forecast 13(1):21–31

Dow MM (2007) Galton's problem as multiple network autocorrelation effects: cultural trait transmission and ecological constraint. Cross-Cult Res 41(4):336–363
Dow MM, Eff EA (2008) Global, regional, and local network autocorrelation in the standard cross-cultural sample. Cross-Cult Res 42(2):148–171
Dow MM, Burton ML, White DR, Reitz KP (1984) Galton's problem as network autocorrelation. Am Ethnolog 11(4):754–770
Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3):244–268
Farber S, Páez A, Volz E (2009) Topology and dependency tests in spatial and network autoregressive models. Geogr Anal 41(2):158–180
Flahaut B, Mouchart M, San Martin E, Thomas I (2003) The local spatial autocorrelation and the kernel method for identifying black zones: a comparative approach. Accid Anal Prev 35(6):991–1004
Florax RJGM, Rey S (1995) The impact of misspecified spatial structure in linear regression models. In: Anselin L, Florax RJGM (eds) New Dir in Spat Econom. Springer-Verlag, Berlin, pp 111–135
Geary RC (1954) The contiguity ratio and statistical mapping. Inc Stat 5(3):115–145
Getis A, Aldstadt J (2004) Constructing the spatial weights matrix using a local statistic. Geogr Anal 36(2):90–104
Griffith DA (1996) Some guidelines for specifying the geographic weights matrix contained in spatial statistical models. In: Arlinghaus SL, Griffith DA, Drake WD, Nystuen JD (eds) Pract Handb of Spat Stat. CRC Press, Boca Raton, FL, pp 82–148
Griffith DA (2010) Modeling spatio-temporal relationships: retrospect and prospect. J Geogr Syst 12(2):111–123
Griffith DA, Heuvelink GB (2009) Deriving space–time variograms from space–time autoregressive (STAR) model specifications. In StatGIS 2009 conference, Milos, Greece, June
Griffith DA, Lagona F (1998) On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. J Stat Plan Inference 69(1):153–174
Hackney JK, Bernard M, Bindra S, Axhausen KW (2007) Predicting road system speeds using spatial structure variables and network characteristics. J. Geogr Sys 9(4):397–417
Hardisty F, Klippel A (2010) Analysing spatio-temporal autocorrelation with LISTA-Viz. Int J Geogr Inf Sci 24(10):1515–1526
Jiang B (2007) A topological pattern of urban street networks: universality and peculiarity. Physica A 384(2):647–655
Kamarianakis Y, Prastacos P (2005) Space-time modeling of traffic flow. Comput Geosci 31(2):119–133
Leenders RT (2002) The specification of weight structures in network autocorrelation models of social influence. SOM rep ser No. 02B09
Liu H, van Zuylen HJ, van Lint H, Salomons M (2006) Predicting urban arterial travel time with state-space neural networks and kalman filters. Transp Res Rec J Transp Res Board 1968(1):99–108
Min W, Wynter L, Amemiya Y (2007) Road traffic prediction with spatio-temporal correlations. In: Proceedings of the sixth trienn symp on transp anal, phuket Island, Thailand, June 2007
Min X, Hu J, Chen Q, Zhang T, Zhang Y (2009) Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. In: Proceedings of the 12th international IEEE conference on intelligent transportation systems, St. Louis, Missouri, USA, 3–7 Oct 2009
Min X, Hu J, Zhang Z (2010) Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. In: Proceedings of the 13th international IEEE conference on intelligent transportation systems, 19–22 Sept 2010, pp 1535–1540
Mizruchi MS, Neuman EJ (2008) The effect of density on the level of bias in the network autocorrelation model. Soc Netw 30:190–200
Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23
Neuman EJ, Mizruchi MS (2010) Structure and bias in the network autocorrelation model. Soc Netw 32(4):290–300
Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256
Olden JD, Neff BD (2001) Cross-correlation bias in lag analysis of aquatic time series. Marine Biol 138(5):1063–1070
Páez A, Scott DM, Volz E (2008) Weight matrices for social influence analysis: an investigation of measurement errors and their effect on model identification and estimation quality. Soc Netw 30(4):309–317
Patil GP (2009) Impacts and Wider Impacts on Statistics (of Cliff and Ord's 1969 article on Spatial Autocorrelation). Geogr Anal 41(4):430–435

Peeters D, Thomas I (2009) Network autocorrelation. Geogr Anal 41(4):436–443
Pfeifer PE, Deutsch SJ (1980) A three-stage iterative procedure for space-time modelling. Technometrics 22(1):35–47
Pflieger G, Rozenblat C (2010) Introduction. Urban networks and network theory: the city as the connector of multiple networks. Urban Stud 47(13):2723–2735
Richards PI (1956) Shock waves on the highway. Oper Res 4(1):42–51
Rodgers JL, Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. Am Stat 42(1):59–66
Smith BL, Williams BM, Keith Oswald R (2002) Comparison of parametric and nonparametric models for traffic flow forecasting. Transp Res Part C Emerg Technol 10(4):303–321
Soper HE, Young AW, Cave BM, Lee A, Pearson K (1917) On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A co-operative study. Biometrika 11(4):328–413
Stetzer F (1982) Specifying weights in spatial forecasting models: the results of some experiments. Env and Plan A 14(5):571–584
van Lint JWC, Hoogendoorn SP, van Zuylen HJ (2005) Accurate freeway travel time prediction with state-space neural networks under missing data. Transp Res Part C Emerg Technol 13(5–6):347–369
Vlahogianni EI, Golias JC, Karlaftis MG (2004) Short-term traffic forecasting: overview of objectives and methods. Transp Rev A Transnatl Transdisciplinary J 24(5):533–557
Wang J, Cheng T, Heydecker BG, Haworth J (2010) STARIMA for journey time prediction in London. In: Heydecker BG (ed) Proceedings of the 5th IMA conference on math in transp
Watts DJ, Strogatz SH (1998) Collective dynamics of "small-world" networks. Nature 393(6684): 440–442
Williams BM, Hoel LA (2003) Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J Transp Eng ASCE 129(6):664–672
Wu C, Ho J, Lee D (2004) Travel-time prediction with support vector regression. IEEE Trans Intell Transp Sys 5(4):276–281
Xie F, Levinson D (2007) Measuring the structure of road networks. Geogr Anal 39(3):336–356
Xu Z, Sui DZ (2007) Small-world characteristics on transportation networks: a perspective from network autocorrelation. J Geogr Syst 9(2):189–205
Yue Y, Yeh AGO (2008) Spatiotemporal traffic-flow dependency and short-term traffic forecasting. Environ Plan B 35(5):762–771