Graphical LASSO for local spatio-temporal neighbourhood selection

James Haworth[1,2], Tao Cheng[1]

[1]SpaceTimeLab, Department of Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT

[2]j.haworth@ucl.ac.uk

1. Introduction

Nowadays, spatio-temporal datasets are being collected in unprecedented volumes at increasingly fine scales. This creates opportunities to develop more accurate forecasting models, but exposes the weaknesses of traditional methods. The traditional way to incorporate spatial and temporal structure into a forecasting model is to assume that the spatio-temporal relationship can be described by a linear model with global set of parameters, which requires the space-time process to be weakly stationary (Cressie and Wikle, 2011). This is the approach taken in, for example, the space time autoregressive integrated moving average (STARIMA) modelling framework (Pfeifer and Deutsch, 1980). However, such models typically cannot deal with the nonstationary and nonlinear properties of fine grain spatio-temporal datasets, which may be highly dynamic (Kamarianakis and Prastacos, 2005; Cheng et al., 2011).

In response to this, statistical modelling frameworks have been extended to model local and/or dynamic space-time structures recently (Min et al., 2009; Ding et al., 2010; Min and Wynter, 2011; Kamarianakis et al., 2012; Cheng et al., 2014). The improvement that these models gain over global model specifications is striking, and clearly motivates the use of local model structures. In parallel to this, researchers are turning towards less conventional techniques, often with their roots in the machine learning (ML) and data mining communities. Examples of such models include artificial neural networks (ANNs) and support vector machines (SVMs), which have become popular in the environmental sciences due to their ability to model complex, nonlinear spatial/spatio-temporal relationships (Kanevski et al., 2009).

In each case, the problem exists to determine the spatio-temporal neighbourhood (STN) of features that is useful for forecasting the value of a space-time series at a given point in space and time, which consists of spatially and/or temporally lagged observations of the dependent variable in the autoregressive setting. Various feature selection methods exist, including autocorrelation analysis, principal components analysis etc. The simplest approach is to test all combinations of the STN and use the model that minimises the empirical error. However, in high dimensional data this is computationally infeasible, and more sophisticated variable selection methods may aid model selection (Guyon and Elisseeff, 2003). In this study, a recent machine learning method, namely the graphical least adaptive shrinkage and selection operator (GLASSO) is applied to the selection of local STNs. The results are verified using a more traditional cross-correlation (CC) analysis, which demonstrates the validity of the observed relationships.

2. Methodology

2.1. GLASSO

GLASSO is a method for extracting a sparse graph from an inverse covariance matrix using a lasso (L_1) penalty (Friedman et al., 2008). GLASSO is similar in motivation to autocorrelation analysis (see, e.g., Box et al., 1994), but has the advantage that it considers the influence of all the spatio-temporal information simultaneously. Let $\mathbf{X}(n * p)$ denote a set of normally distributed data with n observations, p variables, mean μ and covariance $\mathbf{\Sigma}$. Let $\mathbf{S}$ denote the empirical covariance matrix, and $\mathbf{\Theta}$ the inverse covariance matrix. The maximum likelihood estimate of $\mathbf{\Theta}$ can be found by maximising the following:

$$\log \det(\mathbf{\Theta}) - \mathrm{trace}(\mathbf{S\Theta}) \tag{1}$$

Which has the intuitive solution $\widehat{\mathbf{\Theta}} = \mathbf{S}^{-1}$. An L_1 penalty can be added to the off diagonal elements of $\mathbf{\Theta}$ to give the following maximisation problem:

$$\log \det(\mathbf{\Theta}) - \mathrm{trace}(\mathbf{S\Theta}) - \lambda \mathrm{P}(\mathbf{\Theta}) \tag{2}$$

Where $\mathrm{P}(\mathbf{\Theta})$ is the sum of absolute values of off-diagonal elements of $\mathbf{\Theta}$, and λ is a positive tuning parameter. Solving Eq. 2 with an appropriate value of λ leads to a sparse representation of $\widehat{\mathbf{\Theta}}$, and variables can be estimated as conditionally independent (Rasmussen and Bro, 2012). By considering all variables together, the GLASSO approach reveals information concerning partial correlations between variables. Small partial correlations are considered to be insignificant and are set to zero, as determined by the value of λ.

2.2. Local spatio-temporal GLASSO

GLASSO has been used for feature selection in the context of traffic flow forecasting (Gao et al., 2011). In their approach, Gao et al. use GLASSO to define the STN of SCOOT (split cycle offset optimisation technique) flow detectors. The resulting STN is used to produce forecasts with an ANN model. The assumption is made that the spatial and temporal composition of the STN is fixed in time. However, this is not necessarily a realistic assumption. The size and composition of the STN may vary with traffic states. Recent developments in the literature have attempted to address this, either by drawing from a set of models that each pertain to a certain traffic state (Stathopoulos and Karlaftis, 2003; Min and Wynter, 2011; Kamarianakis et al. 2012) or by changing the composition of the STN based on the prevailing traffic condition (Min et al., 2009; Ding et al. 2010; Cheng et al., 2013).

In this study, GLASSO is used to determine the time-varying STN in the context of traffic forecasting. First, let $d = 1,2,\ldots,D$ denote an index of the regular seasonal component in the data (i.e. the number of observations collected in a single day), let k denote the maximum spatial extent (lag) of the STN, and let m denote the maximum temporal extent (lag) of the STN. Let $z = \{z(s,t) | s \in S, t \in T\}$ denote a space-time series in spatial domain S and temporal interval T, collected at times $t = 1,2,\ldots,T$ and locations $s = 1,2,\ldots,N$. GLASSO is used to estimate a sparse graph centred on each time point of the index D. Therefore, D sparse graphs are estimated from D covariance matrices with structure shown in figure 1.

$$
\mathbf{S}_{\boldsymbol{d}} = \begin{pmatrix}
Z_{s_0t}, Z_{s_0t} & Z_{s_0t}, Z_{s_0t-1} & \cdots & Z_{s_0t}, Z_{s_0t-m} & Z_{s_0t}, Z_{s_1t} & Z_{s_0t}, Z_{s_1t-1} & \cdots & Z_{s_0t}, Z_{s_1t-m} & \cdots & Z_{s_0t}, Z_{s_kt-m} \\
Z_{s_0t-1}, Z_{s_0t} & Z_{s_0t-1}, Z_{s_0t-1} & \cdots & Z_{s_0t}, Z_{s_0t-m} & Z_{s_0t}, Z_{s_1t} & Z_{s_0t}, Z_{s_1t-1} & \cdots & Z_{s_0t}, Z_{s_1t-m} & \cdots & Z_{s_0t}, Z_{s_kt-m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
Z_{s_0t-m}, Z_{s_0t} & Z_{s_0t-m}, Z_{s_0t-1} & \cdots & Z_{s_0t}, Z_{s_0t-m} & Z_{s_0t}, Z_{s_1t} & Z_{s_0t}, Z_{s_1t-1} & \cdots & Z_{s_0t}, Z_{s_1t-m} & \cdots & Z_{s_0t}, Z_{s_kt-m} \\
Z_{s_1t}, Z_{s_0t} & Z_{s_1t}, Z_{s_0t-1} & \cdots & Z_{s_1t}, Z_{s_0t-m} & Z_{s_1t}, Z_{s_1t} & Z_{s_1t}, Z_{s_1t-1} & \cdots & Z_{s_1t}, Z_{s_1t-m} & \cdots & Z_{s_1t}, Z_{s_kt-m} \\
Z_{s_1t-1}, Z_{s_0t} & Z_{s_1t-1}, Z_{s_0t-1} & \cdots & Z_{s_1t-1}, Z_{s_0t-m} & Z_{s_1t-1}, Z_{s_1t} & Z_{s_1t-1}, Z_{s_1t-1} & \cdots & Z_{s_1t-1}, Z_{s_1t-m} & \cdots & Z_{s_1t-1}, Z_{s_kt-m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
Z_{s_1t-m}, Z_{s_0t} & Z_{s_1t-m}, Z_{s_0t-1} & \cdots & Z_{s_1t-m}, Z_{s_0t-m} & Z_{s_1t-m}, Z_{s_1t} & Z_{s_1t-m}, Z_{s_1t-1} & \cdots & Z_{s_1t-m}, Z_{s_1t-m} & \cdots & Z_{s_1t-m}, Z_{s_kt-m} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
Z_{s_kt-m}, Z_{s_0t} & Z_{s_kt-m}, Z_{s_0t-1} & \cdots & Z_{s_kt-m}, Z_{s_0t-m} & Z_{s_kt-m}, Z_{s_1t} & Z_{s_kt-m}, Z_{s_1t-1} & \cdots & Z_{s_kt-m}, Z_{s_1t-m} & \cdots & Z_{s_kt-m}, Z_{s_kt-m}
\end{pmatrix}
$$

Figure 1 – Diagram of covariance matrix

The matrix $\mathbf{S}_d$ contains the covariance between observations recorded at time of day d at location 0, and the lagged observations up to temporal lag m and spatial lag k, for all days T/D (assuming complete data for all days). The first row of the sparse graph created from $\mathbf{S}_d$ is used to define the STN of location 0 at d. Spatial and temporal lags with nonzero coefficients are included in the STN.

3. Data description and experimental procedure

The data are unit travel times (seconds/metre) collected using automatic number plate recognition (ANPR) on London's road network as part of Transport for London's (TfL's) London Congestion Analysis Project (LCAP). Observations are made every 5 minutes. The data used here are collected between 6AM and 9PM (180 observations per day, so $D = 180$) on 274 consecutive days, beginning in January 2011. The time varying mean is subtracted from the data prior to analysis. For illustrative purposes, a single road link was chosen for analysis, which is link 442, shown in figure 2 in red. Details of the link are shown in table 1. Three values of λ are tested: 0.001, 0.005 and 0.01. k is set to 10 and m is set to 3. All experiments are carried out using R (R Core Team, 2014) with the glasso package (Friedman et al, 2011).

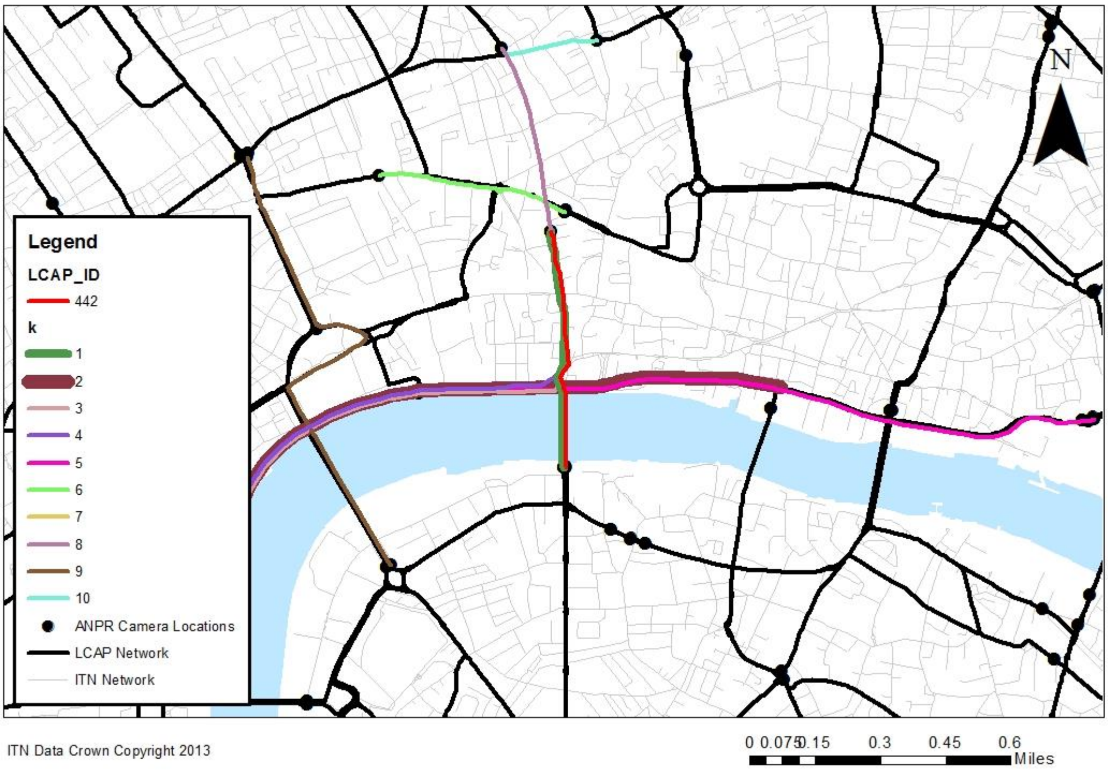

Figure 2 – Map of the study area: The link coloured in red is the test link. The other coloured links are its ten nearest neighbours in terms of network distance (calculated from link midpoint).

Table 1 – Details of link 442

LCAP ID	Direction	LENGTH (metres)	Name
442	South	899.4	A201 Farringdon / A201 Blackfriars Bridge SEB

4. Results

Figures 3 – a)-c) show the nonzero coefficients at the three values of λ for link 442. Larger values of λ result in fewer variables being selected. A number of points can be noted from the results. Firstly, those members of the STN with non-zero coefficients are not necessarily those that are closest in terms of network distance. For example, the first and second nearest neighbours of link 442 have no non-zero coefficients when the value of λ is set to 0.005 or 0.01. Interestingly, the third and eighth nearest neighbours provide the most meaningful information. This is because the overlapping spatial layout of the network complicates the proximity measure; GLASSO automatically accounts for this. Secondly, the coefficients with nonzero values, and hence the composition of the STN, changes throughout the day. Examining figure 3 b) and c), it can be seen that their coefficients are only nonzero during the AM and PM peak periods. The implication of this is that the size of the STN increases as congestion increases. This motivates the use of temporally local model forms.

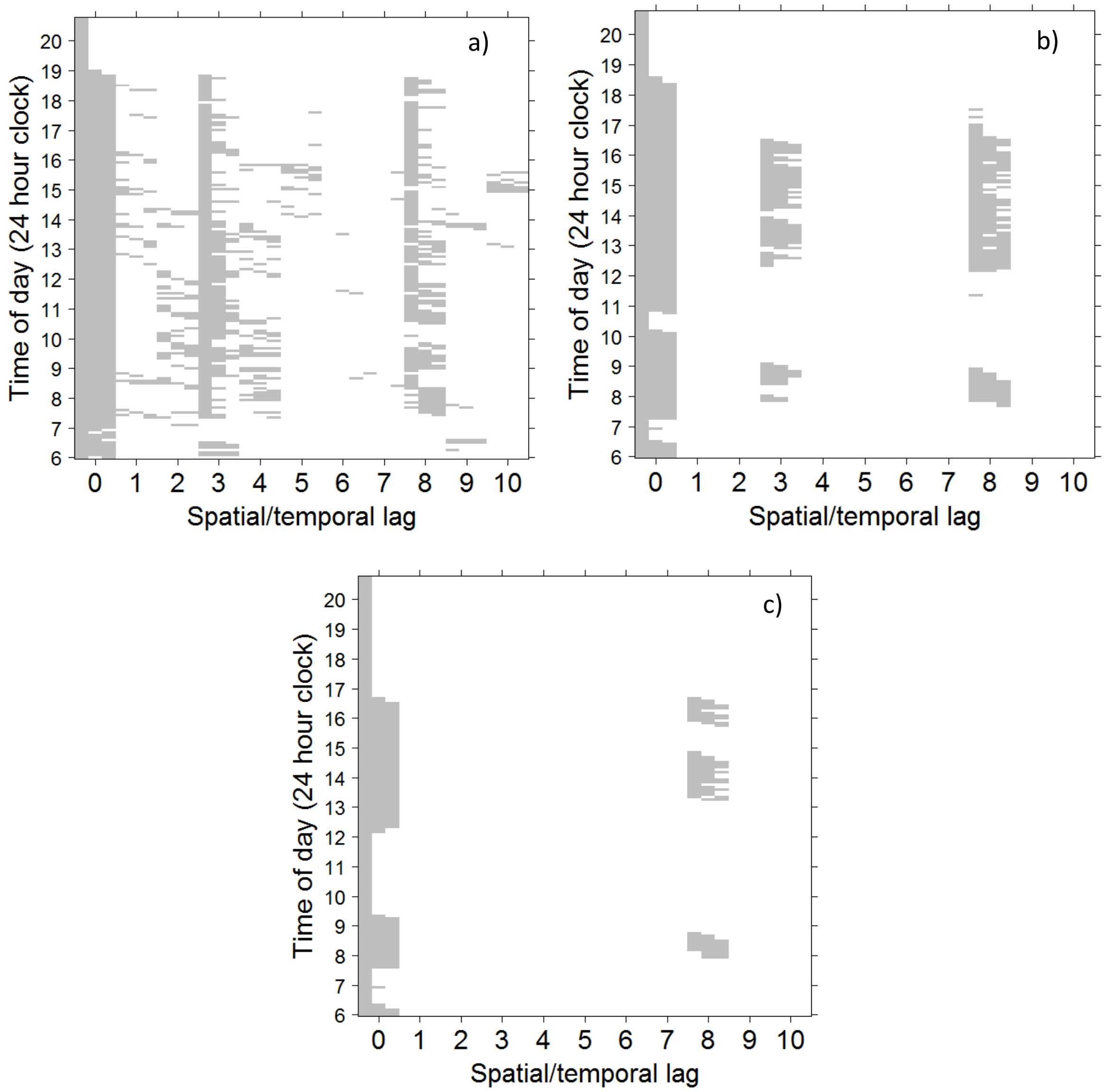

Figure 3 – Nonzero GLASSO coefficients λ values of a) 0.001, b) 0.005 and c) 0.01. On the x-axis, the first 3 cells are temporal lags 1,2 and 3 of spatial lag 1, and so on.

5. Comparison autocorrelation analysis

To assess the validity of the observed results, the pairwise cross correlation (CC) is calculated between link 442 and its 10 nearest neighbours. The cross-correlation function is an extension of the Pearson coefficient to bivariate data (see, e.g. Kendall and Ord, 1990). It measures the correlation between observations of two series separated by time lags. To assess the direction of dependence, the CCF is usually calculated in both directions, i.e. CC of series X with series Y and series Y with series X. The temporal lag at which the CC peaks can be used to determine a transfer function, provided it is nonzero. A peak at lag zero indicates contemporaneous correlation. A first difference is applied before the CC is calculated to ensure temporal stationarity.

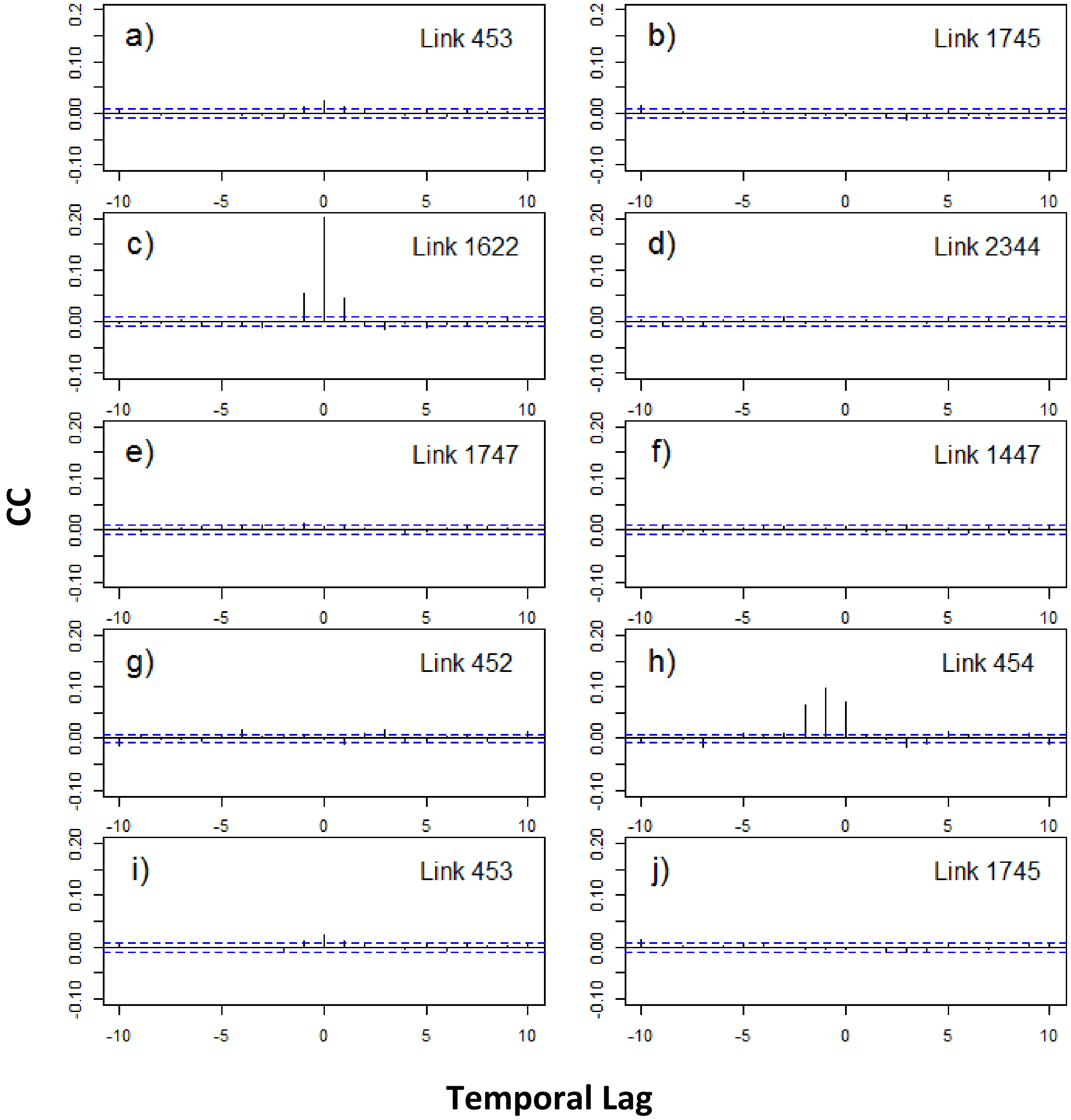

Figure 4 – Empirical cross-correlations between UTT at link 442 and each of its neighbours: a) is the first neighbour and j) is the 10th neighbour bottom right etc. Blue lines are 95% confidence interval

Figure 4 shows the results of the cross-correlation analysis. On the plots, lag zero is at the centre. The blue lines represent the 95% confidence interval for significant CC. It can be seen that most of the links (1,2,4,5,6,7,9,10) exhibit negligible CC with link 442. There are a few statistically significant spikes but the actual value of the CC coefficient is too low to indicate that these represent true statistical relationships. The spikes are more due to the fact that the long series length leads to a very low threshold for statistical significance. The exceptions to this pattern are links 3 and 8, which display significant, positive CC with link 442 at lags -1, 0 and 1. The peak CC is contemporaneous on link 1622 but occurs at lag -1 on link 8. Although the level of CC is moderate, the results suggest directional dependence between links 442 and 454. This is reflected in the results of GLASSO, in which link 454 is the only link with nonzero coefficients when $\lambda = 0.01$. Link 1622's coefficients shrink to zero because the contemporaneous CC provides comparatively less information, given information about link 442.

GLASSO has a number of advantages over using CC alone. Firstly, CC is a global indicator that measures the average pairwise CC between links at all times. Therefore, it is not able to quantify how the relationship between links changes over the course of the day. Secondly, CC measures correlations directly without accounting for correlation between variables. In time series analysis, the partial autocorrelation function is used to account for this, but this requires specification of the ordering of the variables. GLASSO considers all variables together.

6. Conclusion and future work

In this study, the use of GLASSO for the identification of local STNs for use in spatio-temporal models has been investigated. The STNs determined by GLASSO can be used as input variables to a variety of statistical and machine learning models. This entails a model structure that varies with time of day. A local kernel based model or a graphical model could be used to achieve this type of structure. Furthermore, it would be simple to apply the method to larger time periods following a similar approach to Kamarianakis et al. (2012), which would enable the creation of a model that varies with traffic states. Future work will involve the development of forecasting models based on the STNs defined here. The appropriate values of λ can be determined through cross-validation, where the neighbourhood that produces the lowest error is selected.

Reference List

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time series analysis: forecasting and control. Prentice Hall.

Cheng, T., Haworth, J., Wang, J., 2011. Spatio-temporal autocorrelation of road network data. J. Geogr. Syst. 14, 389–413. doi:10.1007/s10109-011-0149-5

Cheng, T., Wang, J., Haworth, J., Heydecker, B., Chow, A., 2014. A Dynamic Spatial Weight Matrix and Localized Space–Time Autoregressive Integrated Moving Average for Network Modeling. Geogr. Anal. 46, 75–97. doi:10.1111/gean.12026

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons.

Ding, Q.Y., Wang, X.F., Zhang, X.Y., Sun, Z.Q., 2010. Forecasting Traffic Volume with Space-Time ARIMA Model. Adv. Mater. Res. 156-157, 979–983. doi:10.4028/www.scientific.net/AMR.156-157.979

Friedman, J., Hastie, T., Tibshirani, R., 2008. Sparse inverse covariance estimation with the graphical lasso. Biostat. Oxf. Engl. 9, 432–441. doi:10.1093/biostatistics/kxm045

Gao, Y., Sun, S., Shi, D., 2011. Network-scale traffic modeling and forecasting with graphical lasso. Adv. Neural Networks–ISNN 2011 151–158.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Kamarianakis, Y., Prastacos, P., 2005. Space-time modeling of traffic flow. Comput. Geosci. 31, 119–133. doi:10.1016/j.cageo.2004.05.012

Kamarianakis, Y., Shen, W., Wynter, L., 2012. Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. Appl. Stoch. Models Bus. Ind. 28, 297–315. doi:10.1002/asmb.1937

Kanevski, M., Timonin, V., Pozdnukhov, A., 2009. Machine Learning for Spatial Environmental Data: Theory, Applications, and Software, Har/Cdr. ed. EFPL Press.

Kendall, M.G., Ord, J.K., 1990. Time Series. E. Arnold.

Min, W., Wynter, L., 2011. Real-time road traffic prediction with spatio-temporal correlations. Transp. Res. Part C Emerg. Technol. 19, 606–616. doi:10.1016/j.trc.2010.10.002

Min, X., Hu, J., Chen, Q., Zhang, T., Zhang, Y., 2009. Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. IEEE, pp. 1–6. doi:10.1109/ITSC.2009.5309741

Pfeifer, P.E., Deutsch, S.J., 1980. A Three-Stage Iterative Procedure for Space-Time Modelling. TECHNOMETRICS 22, 35–47.

Rasmussen, M.A., Bro, R., 2012. A tutorial on the Lasso approach to sparse modeling. Chemom. Intell. Lab. Syst. 119, 21–31. doi:10.1016/j.chemolab.2012.10.003

Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. Transp. Res. Part C Emerg. Technol. 11, 121–135. doi:10.1016/S0968-090X(03)00004-4