Australasian Transport Research Forum 2017 Proceedings
27–29 November 2017, Auckland, New Zealand
Publication website: http://www.atrf.info

Spatial variable selection methods for network-wide short-term traffic prediction

Md. Mahmud Hasan, Jiwon Kim, Carlo Prato

School of Civil Engineering, The University of Queensland, Brisbane St Lucia, Australia

Email for correspondence: mahmud.hasan@uq.edu.au

Abstract

This paper proposes statistical approaches to identifying spatial relationships among road links in an urban road network to select predictors for a short term traffic prediction model for a given road link more systematically. For this purpose, two methods i.e. Granger causality test and elastic net regularization techniques have been adopted in this study. Urban road network in Brisbane, Australia is selected as a study site for a case study. One-year traffic flow and speed data from the selected road network have been used for the analysis. The study evaluates the performance of the proposed variable selection methods in terms of the prediction accuracy of short term traffic prediction models constructed based on the variables identified by the methods. This study uses time lagged multiple linear regression method as a short term traffic prediction model. For a given target link, the relevant predictors obtained by Granger causality and elastic net are used separately to build the respective traffic prediction models. It is observed that Granger causality based traffic prediction model provides better prediction accuracy than elastic net based traffic prediction model.

Key words: Traffic prediction, Granger causality, Elastic net, variable selection, time-series.

1. Introduction

One of the crucial tasks in intelligent transport systems is to forecast traffic parameter or condition of a roadway precisely. Practitioners need the forecasted traffic information to improve traffic management and apply control strategies on the roadway if necessary. Future traffic information also aids travel information for the road users. Therefore, it is important to forecast traffic condition accurately. Precise future information of traffic condition can be achieved by developing accurate traffic forecast model. One of the concerns of building traffic model is efficient use of information from a large volume of dataset. It is impractical to use all the information available in high dimensional time series data as it may not improve traffic prediction enough when compared to the increase in computational cost. Therefore, it is necessary to discern the most relevant data from massive time series dataset which is useful to build accurate prediction model. This is also true for short term basis traffic prediction where the traffic conditions of the nearest future interval (i.e. usually less than one hour ahead) of a specific location is forecasted by historic data of response and predictor variables. Taking the most relevant variables as the predictors of response variable will help to build parsimonious or simpler prediction model and improve the accuracy of the prediction model.

In order to predict traffic parameter of a specific section/location of the roadway in short term basis, previous studies used univariate and multivariate approaches of time series analysis and short term prediction modelling. In these studies, different set of predictor variables were used in the prediction models In univariate analysis, historical traffic parameter data, namely the nearest past or moving average data of a target location or the nearest upstream location, is taken as predictors. In multivariate analysis, traffic parameters of nearby locations of target location are considered as the predictors. Univariate modelling approaches include Kalman filtering, nonparametric regression, neural network and autoregressive integrated moving average (ARIMA) whereas multivariate modelling methods include space-time ARIMA and Bayesian networks (Kamarianakis and Prastacos, 2004).

In the univariate approach, traffic information of road links adjacent to the target link is usually omitted. However, utilizing traffic information from other adjacent road links is useful to gain the precise prediction of traffic (Sun and Zhang, 2006). In addition, univariate models are not suitable for a road network where hundreds of road links are available (Kamarianakis and Prastacos, 2004). In comparing univariate and multivariate approach, Chandra and Al Deek (2009) concluded that vector autoregressive (VAR) model provides higher accuracy than popular univariate ARIMA models in predicting traffic flow and speed in a freeway. Similarly, Stathopoulos and Karlaftis (2002) found their proposed multivariate state-space traffic prediction model is more accurate than ARIMA model. On the other hand, multivariate models used in previous researches mostly relied on spatial boundary set up. The studies of Hobeika and Kim (1994), Sun et al. (2006) considered only nearest upstream of target location as spatial boundary for developing traffic flow prediction whereas the study of Chandra and Al-Deek (2009)) selected both nearest upstream and downstream of target location as the spatial boundary. In considering the neighbours of the target location, Stathopoulos and Karlaftis (2002) only selected upstream locations rather than a combination of upstream and downstream locations in an urban arterial. A study by Kamarianakis and Prastacos (2003) used a hierarchical system of two orders of upstream neighbour locations as the spatial boundary of traffic flow prediction in urban road network. Although they found second order neighbours are more useful than first order neighbours to predict target link, they did not apply higher order of neighbour locations. In predicting speed of an urban road network, K-nearest neighbour (k=20) based predictor selection method was used by Smith et al. (2002) in order to finding predictors for a given target location. However, a direct or strong conditional relation may exist between the target link and any distant road link outside the k-nearest neighbourhood, a feature that is not considered in their study. It can be said that previous studies related to traffic prediction model considered a fixed and smaller spatial boundary such as upstream, downstream or nearest neighbour location of a target link as the predictor locations.

This paper proposes a systematic way of selecting predictor road links for a given target road link using spatial dependency structure. Spatial dependency structure includes a given target road link and a set of predictor road links in the road network which are functionally or causally related to the target link. Unlike existing studies, this study does not consider any fixed spatial boundary as the predictor locations. Our approaches initially consider all road links in a study site as the predictor locations and then selects the relevant predictor locations for a given target link by statistical methods. Proposed statistical methods include Granger causality test (Granger, 1969, 1980) and variable selection technique known as elastic net (Zou and Hastie, 2003). Although these two methods are quite new in the area of traffic modelling, these are well as established statistical approach in other areas of research. Granger causality test method identifies directed functional or causal interactions of different variables in the time series data (Seth et al., 2015). Elastic net is useful to improve the prediction performance and detailed analysis of high dimensional massive data (Zou and Zhang, 2009).

The remainder of the paper is organised as follows. Section 2 includes the description of the methodologies of two proposed models such as Granger causality and elastic net. Section 3 consists of the description of short term traffic prediction model. The following section contains the application of the models and detailed discussion about the results obtained. Finally, section 5 concludes the paper.

2. Spatial dependency model for variable selection

2.1. Granger causality based model

This study applies the Granger causality method to find out the spatial relation among the road links in a network. This relation is used to obtain relevant road links that can be used to predict the traffic parameter of a given target road link. Let $\overline{x}_{t-1}$ be the lagged variable of $\overline{x}_t$, $\overline{y}_{t-1}$ be the lagged variable of $\overline{y}_t$ and $\mathbf{A}, \mathbf{B}$ represent vectors of coefficients. The Granger causality method is performed by the following linear regression model:

$$y_t = a_0 + \mathbf{A}.\overline{y}_{t-1} \tag{1}$$

$$y_t = a_0 + \mathbf{A}.\overline{y}_{t-1} + \mathbf{B}.\overline{x}_{t-1} \tag{2}$$

If the second equation becomes a statistically significantly better regression model (i.e. predicts or computes y_t more accurately) than the first equation, it can be said that $\overline{x}_{t-1}$ Granger causes y_t. The statistical test for verifying if it is a better regression is simply an F-test (Arnold et al., 2007).

This study uses vector auto regression (VAR) based Granger causality test. In a VAR, each variable is taken as dependent variable once and the other variables are considered as predictor variables. Multivariate time series is used in VAR model where each variable is computed by lagged values of its own and lagged values of other variables (Zivot and Wang, 2006).

Let $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})$ be a $(n\times1)$ vector of time series variables. The basic p-lag vector VAR model can be written as

$$\mathbf{Y}_t = c + \Pi_1\mathbf{Y}_{t-1} + \Pi_2\mathbf{Y}_{t-2} + \cdots + \Pi_p\mathbf{Y}_{t-p} + \varepsilon_t \tag{3}$$

where Π_i are $(n\times n)$ coefficient matrices, c is the intercept, $t = 1, \dots, T$ and ε_t is the error term. The detailed explanation of the VAR model is provided in the study of Hasan and Kim (2016).

Model selection criteria such as the Akaike Information Criterion (AIC) and Schwarz-Bayesian Information Criterion (BIC) are used to determine the time lag order p. In order to find the suitable time lag order, at first, the VAR model is fitted with each of the lag orders $p = 0, \dots, p_{max}$ and the corresponding value of the model selection criterion is calculated. Then, the actual time lag order can be computed by the scores of BIC and AIC which are computed as:

$$\mathrm{BIC} = -2\,l(\widehat{\theta}) + m \log n \tag{4}$$

$$\mathrm{AIC} = -2\,l(\widehat{\Theta}) + 2m \tag{5}$$

where $l(\widehat{\theta})$is the maximum log-likelihood as a function of the vector of parameter estimates$(\widehat{\theta})$, m is the number of parameters in the model, and *n* is the number of observations. BIC or AIC score decreases as the likelihood increases and the number of parameter decreases. The lag order with the lowest AIC or BIC score is considered as the actual lag order for modelling (Cottrell and Lucchetti, 2016). Since one of the objectives of this study is to reduce the number of parameters in the model, the lowest lag order between AIC and BIC is selected for the study.

After developing the VAR model, the Granger causality test is performed by hypothesis testing of F test with zero restriction. In this test, a given time series is considered as Granger cause of another time series if at least one value in the coefficient vector is found to be non-zero by statistical significance test (Bahadori and Liu, 2012). The F test statistic is computed as

$$F_0 = \frac{\frac{SSR_r - SSR_{ur}}{q}}{\frac{SSR_{ur}}{n-(k+1)}} \tag{6}$$

where SSR_r is the sum of the squared residuals of the restricted model (model excluding testing predictor) and SSR_{ur} is the sum of the squared residuals of the unrestricted model (model including all predictors), *q* is the number of restrictions or the number of coefficients being jointly tested,n is the number of observations, and *k* is the number of independent variables in the unrestricted model (Blackwell, 2008).

Then F_0value is compared with the critical value of F with significance value 0.01. If the F_0 value is higher than the critical value (obtained from the F distribution table), then it rejects the null hypothesis and hence the testing time series does not Granger-cause target time series. Otherwise, the alternative hypothesis prevails and this means that the testing time series Granger causes the target time series.

2.2. Elastic net based model

For a massive dataset, it is essential to select the number of relevant variables before developing the model. This can be done by adopting regularization techniques such as ridge, lasso and elastic net. As the performance of elastic net is better than the other two regularization techniques in context of high dimensional data analysis and correlated group of predictors selection (Zou and Hastie, 2005; Zou and Zhang, 2009, Li and Lin, 2010), elastic net is selected for this study. Elastic net is performed by the following process.

Consider a linear regression model, $\mathbf{y} = \beta_0 + \beta\mathbf{X}$ with n observations and p predictors. Let $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ be the response and $\mathbf{X} = (x_1, x_2, \ldots, x_p)$ be the model matrix, where x_1, x_2 and x_p are the predictors. Each of these predictors includes n number of observation value. This can be written as $x_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T$ where j=1, 2,…,p. It is assumed that the predictors are standardized.

$$\sum_{i=1}^{n} x_{ij} = 0, \quad \sum_{i=1}^{n} x^2{}_{ij} = 1 \text{ , where j=1, 2,…,p} \tag{7}$$

For any fixed non-negative penalty parameters λ_1 and λ_2, the elastic net loss function is

$$L(\lambda_1, \lambda_2, \beta) = |y - \beta_0 - X\beta|^2 + \lambda_2|\beta|^2 + \lambda_1|\beta|_1 \tag{8}$$

Where,

$|\beta|^2 = \sum_{j=1}^{p} \beta_j^2$, $|\beta|_1 = \sum_{j=1}^{p} |\beta_j|$

The elastic net estimator $\widehat{\beta}$ is the minimizer of the equation

$$\widehat{\beta} = argmin_{\beta} \ \{L(\lambda_1, \lambda_2, \beta)\} \tag{9}$$

This procedure can be written as a penalised least squares method.

Let $\alpha = \frac{\lambda_2}{\lambda_1+\lambda_2}$ and $1-\alpha = \frac{\lambda_1}{\lambda_1+\lambda_2}$ then equation 8 can be written as

$$L(\lambda_1, \lambda_2, \beta) = |y - \beta_0 - X\beta|^2 + (\lambda_1 + \lambda_2)[\alpha|\beta|^2 + (1-\alpha)|\beta|_1] \tag{10}$$

Let $(\lambda_1 + \lambda_2) = \lambda$, then equation 10 can be written as

$$L(\lambda_1, \lambda_2, \beta) = |y - \beta_0 - X\beta|^2 + \lambda(\alpha|\beta|^2 + (1-\alpha)|\beta|_1) \tag{11}$$

Solving $\widehat{\beta}$ in equation 9 is equivalent to the following optimization problem:

$$\widehat{\beta} = argmin_{\beta} \ | y - \beta_0 - X\beta \ |^2, \text{ subject to } \lambda(\alpha|\beta|^2 + (1-\alpha)|\beta|_1) <= t \tag{12}$$

We call the function $\lambda(1-\alpha)|\beta|_1 + \alpha|\beta|^2$ the elastic net penalty which is a convex combination of the lasso and ridge penalty (Zou and Hastie, 2005). Elastic net penalty is controlled by two tunning parameters, α and λ . α value can vary between 0 and 1 where 0 corresponds to the lasso and 1 corresponds to the ridge regression. If the predictors are correlated in groups, α=0.5 tends to select the group in or out of the model (Hastie and Qian, 2014). We select the value of α is 0.5 in order to avoid the proximity to both criteria and emphasise the correlated group selection. On the other hand, λ value is selected by cross validation. In order to avoid intensive computation, a grid of values for λ is first specified (Li and Lin, 2010). The grid of λ values can be (0, 0.001, 0.1, 1, 10, 100, 1000, 10000). By using α and each of the λ values, equation 13 can be solved by using cyclical coordinate descent algorithm (Friedman and Hastie, 2008). Cyclical coordinate descent algorithm successively optimizes the objective function over each parameter by keeping other parameters fixed, and cycles repeatedly until convergence (Hastie and Qian, 2014). For each of the combination of α and λ, mean squared error of argmin_{β} $|y - \beta_0 - X\beta|^2$ is computed. Then, over the range λ values, two values of λ i.e. λ(min) and λ(1se) are found out. λ(min) and λ(1se) provide minimum mean cross-validated error and 1-standard error of the minimum mean cross-validated error respectively (Hastie and Qian, 2014). Any value between the range of λ(min) and λ(1se) can be used as the value of tunning parameter λ.

After the regularization, a number of predictors whose coefficients are reduced to 0 are removed from the model. On the other hand, correlated predictors with a reduced coefficients' value stay in the model. Through this process, the elastic net can provide a parsimonious model. As the selected predictors are highly correlated with response variable, the result can be used for the development of a new approach of spatial dependency model.

3. Time-series model for short term traffic prediction

Based on the assumption that the prediction accuracy of the traffic parameter of a particular target link is improved by including only those links that are statistically related to the target link, this study proposes a Granger causality method and an elastic net method that aim to automatically identify a set of predictor links for each target link.

In order to evaluate the effectiveness of predictor selection of the proposed methods, a simple traffic prediction model is built by using multiple linear regression. The prediction model is developed on the basis of short term prediction, specifically next time interval ahead. So, the relevant predictors are in 1-lagged time interval than response variable. From the Granger causality and the elastic net based models, the correlated and most relevant predictors for each target link can be found. By using these predictors, we can develop multiple linear regression for each target link. The prediction model can be written as:

$$\mathrm{Y_t} = \mathrm{c} + \text{ß}_1 \mathrm{x}_{1,t-1} + \text{ß}_2 \mathrm{x}_{2,t-1} + \text{ß}_3 \mathrm{x}_{3,t-1} + \ldots\ldots\ldots + \text{ß}_n \mathrm{x}_{n,t-1} \quad (13)$$

where Y_t is the response variable, $\mathrm{x}_{1,t-1}, \mathrm{x}_{2,t-1}, \mathrm{x}_{3,t-1} \ldots . \mathrm{x}_{n,t-1}$ are the relevant predictors at 1-lagged time than response variable and $\text{ß}_1, \text{ß}_2, \text{ß}_3 \ldots \text{ß}_n$ are the coefficients.

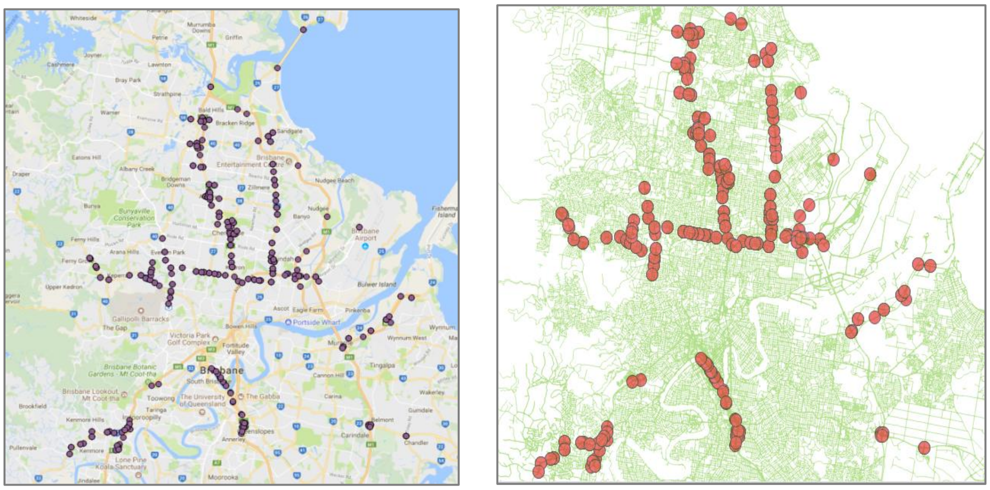
For the estimation and validation of the prediction model, traffic data is separated into a training and a testing dataset. In this study, first 80% of total data are taken as a training set and the remaining 20% of the total data are considered as a testing dataset. The multiple linear regression based prediction model is developed on the training dataset and the model accuracy is evaluated on the testing dataset.

4. Case Study

4.1. Study area

For this study, the urban road network in Brisbane, Australia, is selected as the test-bed. The study site has 454 road links that include freeways, highways, arterials and minor road links. Figure 1 shows the study site. The dot points in the following figures represent the road links in the selected network.

Figure 1: Selected test bed and road links



(a) Google Maps view (b) GIS shape file view

4.2. Data

In this study, traffic flow and average speed are selected as traffic parameters. Traffic data of Brisbane are obtained from Queensland Department of Transport and Main Road (DTMR) through the Public Traffic Data System (PTDS). Data were collected from loop detectors installed on state-controlled roads. Traffic data contains 15-minute interval traffic flow, average speed and occupancy data of 1659 road links in South-East Queensland for the duration of 365 days from January 1, 2014 to December 31, 2014. The road network selected for this study includes 454 road links out of total 1659 road links of entire South-East Queensland. Therefore, this study is based on traffic flow and speed of 454 road links of the road network. The total number of observations of the continuous time series data of each road link is 35,040. It is noted that sometimes the detectors are not working properly or sometimes there are missing values. For these missing values, a multiple imputation technique (Sterne et al., 2009) is applied to fill out the data used for modelling.

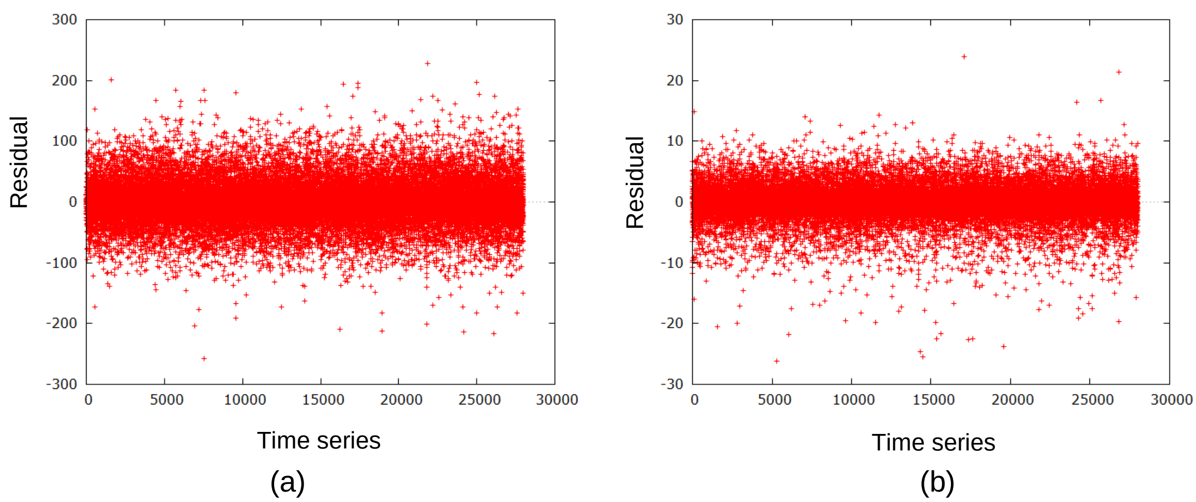
4.3. Model building and estimation

4.3.1. Granger causality based spatial dependency model

This study develops the VAR based Granger causality model for two traffic parameters: flow and speed. This model is built by using the software package Gretl (Cottrell and Lucchetti, 2016). The traffic parameter value of each road link is considered as a target variable once and the lagged value of the remaining variables along with the lagged value of the target are taken as predictors. As 15-minute interval traffic data is used in this study, data has four lags in an hour. Thus, the maximum time lag order is equal to 4. The actual lag order selection by AIC and BIC can be different. It is found that both AIC and BIC provide 1 lag order for flow cases. However, in speed cases, AIC provide 3 lag order and BIC provide 1 lag order. Since one of the objectives of this study is to reduce the number of parameters in the model, the lowest lag order between AIC and BIC is selected for the study. Therefore, BIC score based lag order is selected as it provides the lowest lag order for both flow and speed cases.

Residual vs. Time series plots of traffic flow and speed based VAR represent that errors are not in the form of positive or negative correlation rather these are independent. Figure 2 shows the Residual vs. Time series plot of one road link as an example. No trend can be identified from Figure 2 which means residuals are not auto correlated. After building the VAR model, we have total 454 regression models, namely one for each target link with their 1- lagged predictors. Then the F statistics is applied in each of the regression models to find out a set of Granger causal links for each target link.

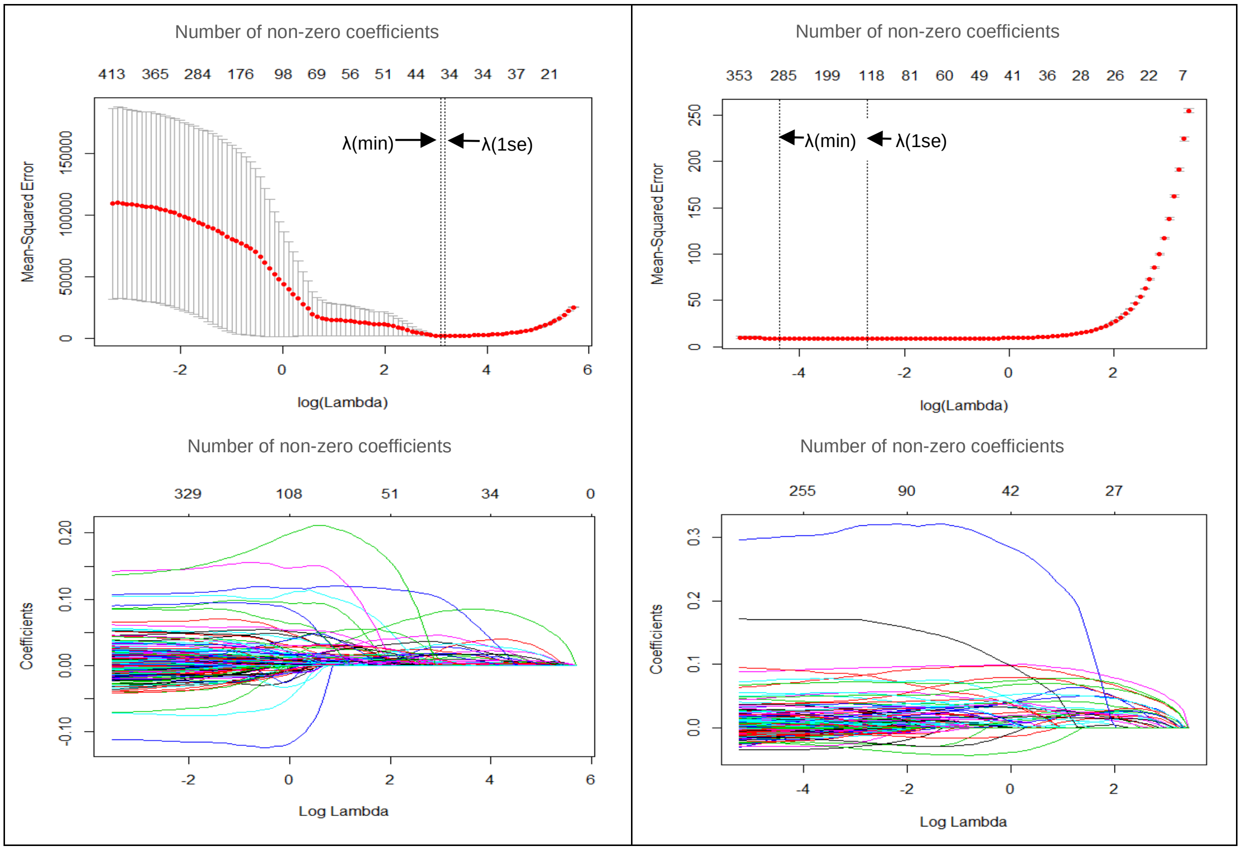
Figure 2: Residuals vs. Time series for (a) flow and (b) speed based VAR



4.3.2. Elastic net based spatial dependency model

This case study develops the elastic net based spatial dependent model for two traffic parameters i.e. flow and speed. Like for the VAR model, the traffic parameter value of each road link is considered as a target variable once and the lagged value of the remaining variables along with the lagged value of the target are taken as predictors in elastic net. So we have total 454 models to be regularized by elastic net in each traffic parameter analysis. This model is built and regularized by using the 'glmnet' package (Hastie and Qian, 2014) of the software R. The time lag order selection is same as used in the VAR based Granger causality model which is 1 lag order for both flow and speed based model. According to equation 13, two tuning parameters need to be considered for elastic net which are α and λ. As mentioned before, we selected the value of α as 0.5. We considered λ(1se) value as other tuning parameter since it offers more regularized model i.e. less number of predictors. Figure 3 shows the effect of λ(min) and λ(1se) to shrink the coefficients and select the correlated variables. Here, one link has been taken as an example to show the concept.

Figure 3: Selection of correlated predictors by elastic net for (a) flow and (b) speed based model



(a) Flow based model (b) Speed based model

After the regularization, the coefficients of a number of predictors are reduced to zero whereas coefficients of some predictors are found to be non-zero. The variables which have zero coefficients value are considered as uncorrelated with the target link whereas the variables which have non-zero coefficients are taken as relevant predictors of the target link. Hence, these variables are removed as the irrelevant variables. The rest other variables are taken as the correlated with target link and thus they stay in the model. By this way, a set of predictors for each target links are found which are considered as elastic net links for the target link.

4.3.3. Time-series models for short term traffic prediction

In this case study, a time-series model is developed by taking each link as the target link and its 1-lagged value of relevant predictors. The predictors are based on three different scenarios in order to find the most accurate methods of variable selection for short term (i.e. 15 minute ahead) traffic prediction. Therefore, we have total 454 regression models for each scenario. Given a target link,

- Scenario 1: Prediction model includes those links identified by the Granger causality.
- Scenario 2: Prediction model includes those links identified by the elastic net.
- Scenario 3: Prediction model includes all links of the selected road network.

Each of these scenarios is tested for two traffic measures i.e. flow and speed separately for each target link. Prediction accuracies of these three scenarios are evaluated by the root mean square error (RMSE). RMSE is calculated as:

$$\text{RMSE}= \sqrt{\sum_{i=1}^{n}(x_t - \hat{x_t})^2} \qquad (14)$$

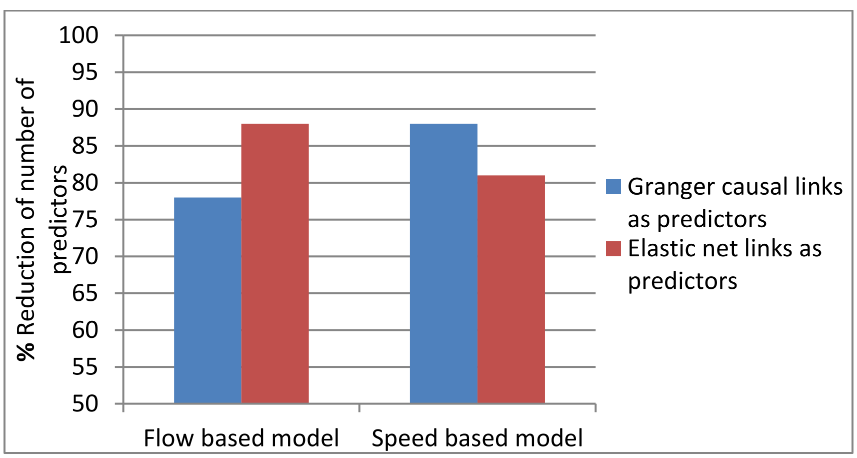
Where x_t is the observed value of traffic parameter of target link at time t, $\hat{x_t}$ is the predicted valueof traffic parameter of target link at time t and N is the number of observations (Clark, 2003).

At each scenario, RMSE for each target link prediction is computed. Then, RMSE value of all predicted target links are summed up and averaged. The average value is then used to compare the prediction accuracies of three scenarios. The scenario which produces lowest RMSE is considered as better prediction model.

4.4. Results and discussion

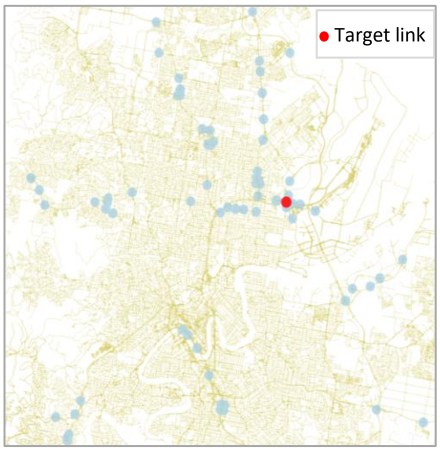
The Granger causality and elastic net are used to minimise the irrelevant predictors and only select the relevant predictors for a given target link in a road network in order to build accurate traffic forecast model. Therefore, a number of predictors are expected to be reduced by both methods. In the context of reduction of the number of predictors for forecast model, it can be seen from Figure 4 that the Granger causality and elastic net act differently when flow and speed are the variables considered. When looking at flow, on average, elastic net requires less predictor than the Granger causality, whereas when looking at speed, the Granger causality reduces the number of relevant predictors more than the elastic net. However, both methods are capable to reduce more than 75% of the total number of predictors when looking at both flow and speed.

Figure 4: Reduction of number of predictors by Granger causality and elastic net

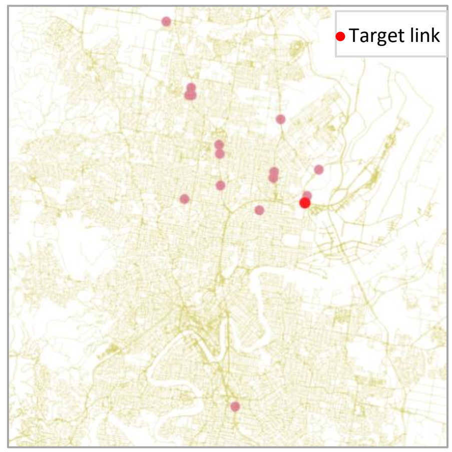


To understand spatial patterns of predictor links, the locations of the predictor links identified by each method for a single target link are presented in Figure 5. One observation is that some predictor links are located far from the target link. This indicates that a link can provide information that improves the prediction of a target link even if the link is not physically connected to the target link and, thus, the traditional approach of only including the upstream and downstream links as predictors may not be the best way of constructing prediction models.

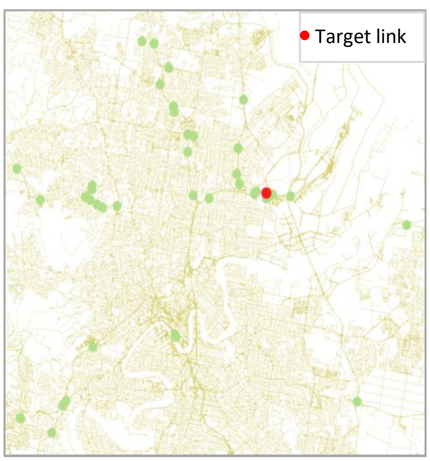
Figure 5: Spatial pattern of target link and the selected predictors by Granger causality and elastic net



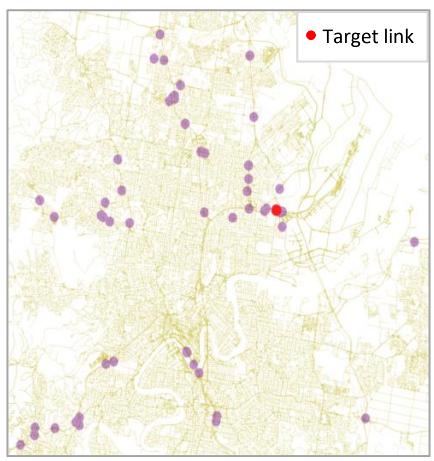
(a) Flow based Granger causal links



(b) Flow based Elastic net links



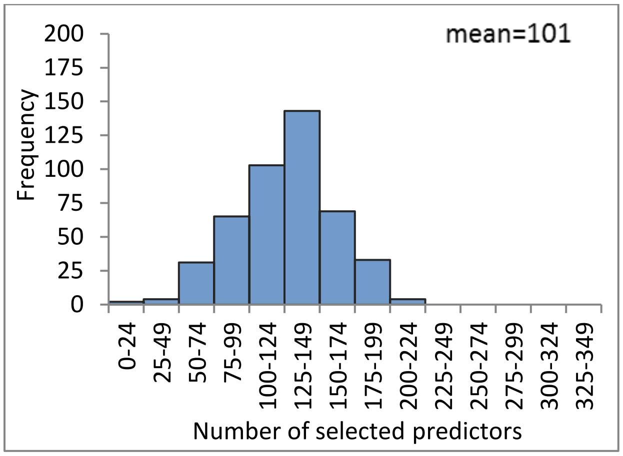
(c) Speed based Granger causal links

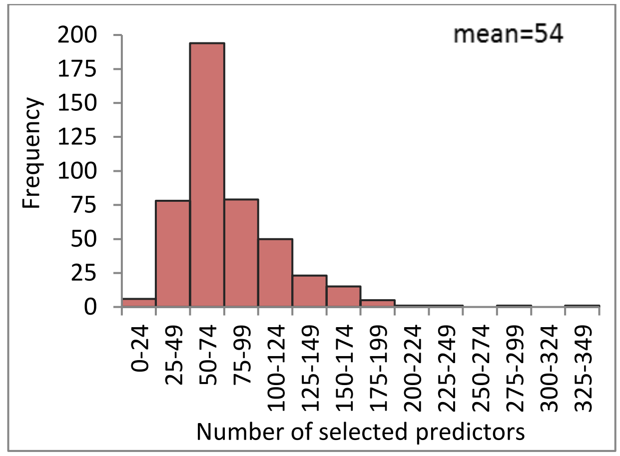


(d) Speed based Elastic net links

The following histograms (Figure 6) show the distribution of the selected predictors into different categories. The number of relevant predictors selected by a spatial model is divided into 19 categories where each category consists of 25 predictors. It can be said that all histograms show right skewness because of significant reduction of predictors obtained by both variable selection methods. For flow based elastic net and speed based Granger causality model, it can be observed that the number of selected predictors are mostly into 50-99 range per link. However, these are the cases where more reductions of predictors are observed in Figure 4. On the other hand, very high frequency of any category cannot be observed in flow based Granger causality and speed based elastic net model. The average selected predictors of flow based Granger causality are higher than speed based Granger causality model. On the other hand, the average selected predictors of speed based elastic net models are higher than flow based elastic net model.

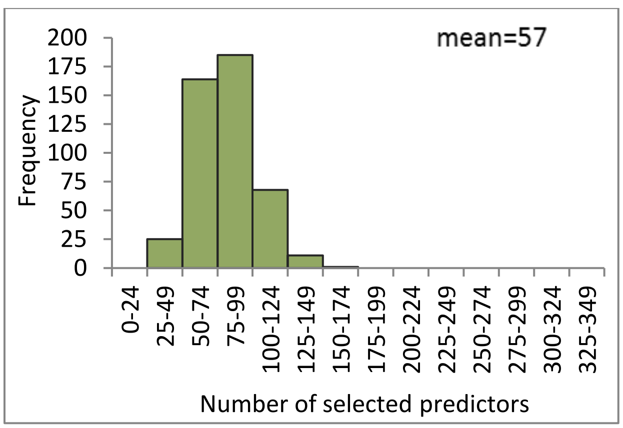
Figure 6: Distribution of number of selected predictors by Granger causality and elastic net

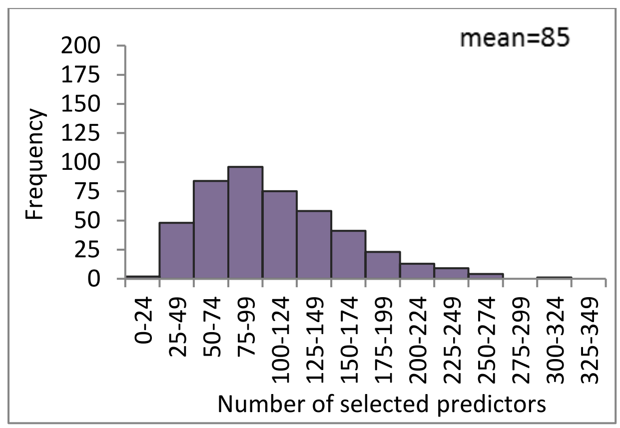




(a) Flow based Granger causal links

(b) Flow based Elastic net



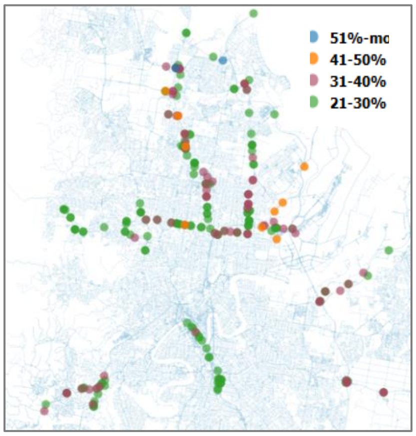


(c) Speed based Granger causal links

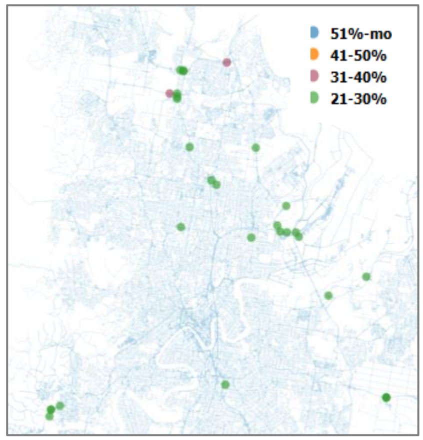
(d) Speed based Elastic net links

This study also analysed the importance of each road link as a predictor, which is defined as the number of time the road link is selected as the relevant predictor by the proposed spatial model. Results are presented according to six categories (0-10%, 11-20%, 21-30%, 31-40%, 41-50%, 51%-more) corresponding to the percentage of times that each link is selected as the predictor. Figure 7 illustrates the locations of the links which are selected more than 20% times (i.e. four categories: 21-30%, 31-40%, 41-50%, 51%-more). It can be observed that the links of each category are different in Granger causality and elastic net based spatial model. This means that the importance of a link as a predictor varies with the type of spatial models and the parameters that were considered (i.e., flow or speed). The histograms in Figure 8 represent the distribution of the number of links in each category. It can be said that in most models, variables are tend to be selected less than 30% time and most frequently 11-20% times as the predictors. The information obtained from both figures is useful to understand where the important predictors are located for developing a short term traffic flow prediction model.

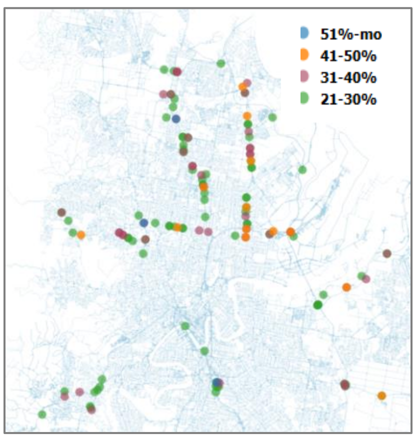
Figure 7: Locations of the links which are selected as the relevant predictors more than 20% times



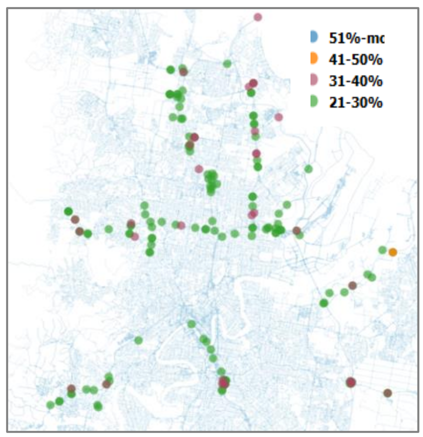
(a) Flow based Granger causal links



(b) Flow based Elastic net links

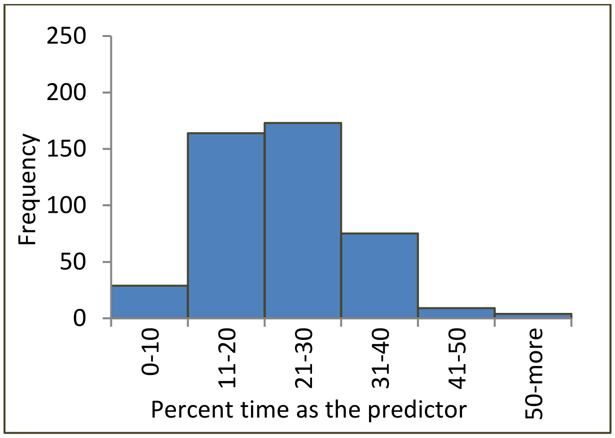


(c) Speed based Granger causal links

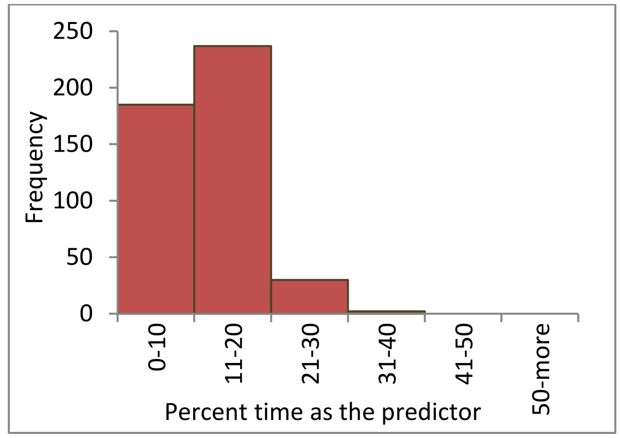


(d) Speed based Elastic net links

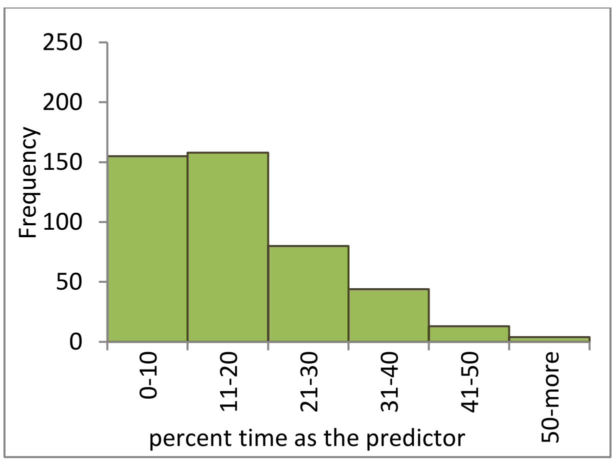
Figure 8: Distribution of number of times (in percentage) the links are selected as the relevant predictors in different spatial dependent models



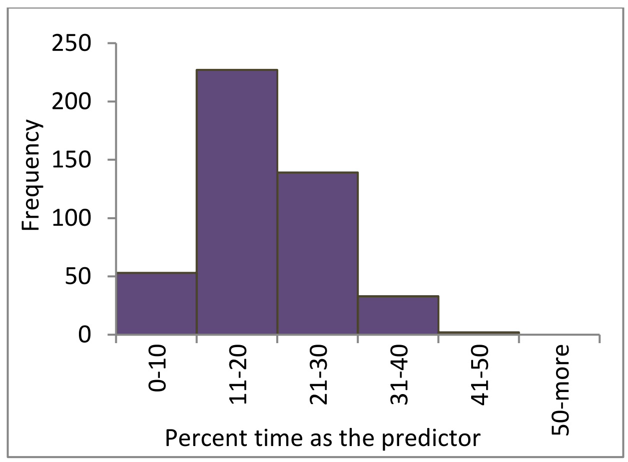
(a) Flow based Granger causal links



(b) Flow based Elastic net links



(c) Speed based Granger causal links

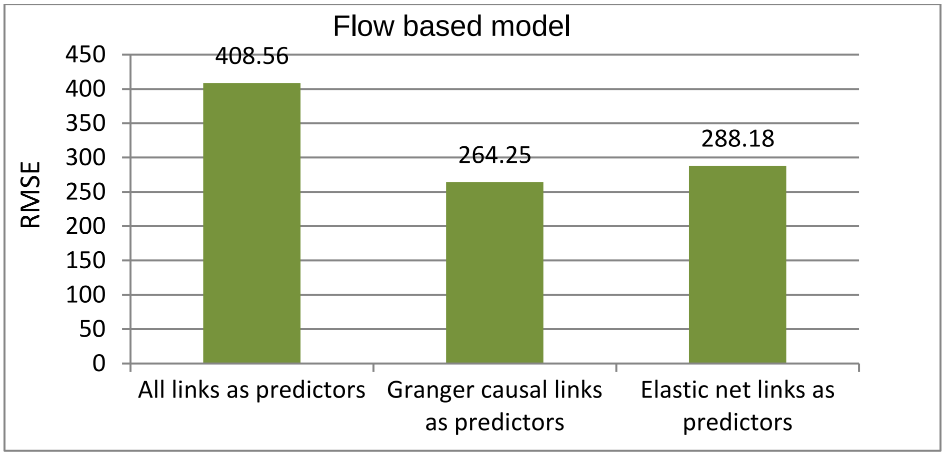


(d) Speed based Elastic net links

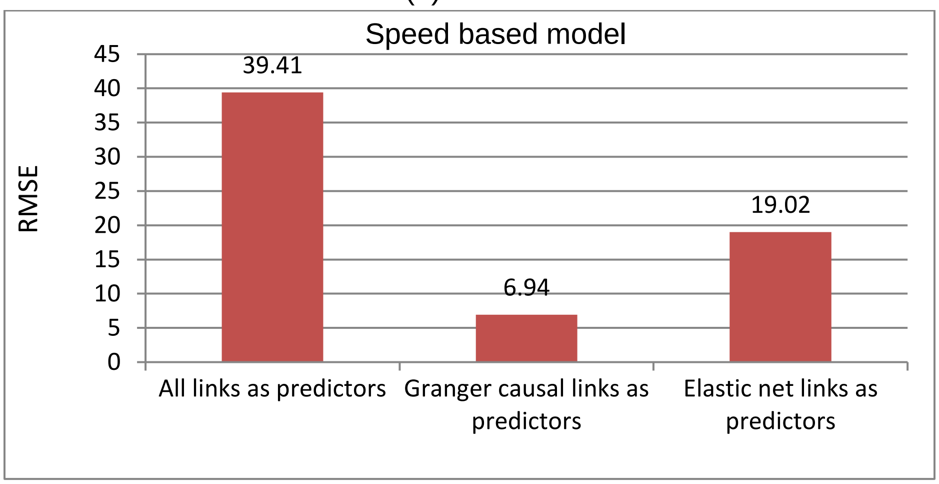
Figure 9 shows the root mean square error of six different models. It can be observed that the prediction model based on all the links of the road network provides a higher RMSE, which means a higher prediction error when compared to the Granger causal links and elastic net links based prediction models. Therefore, the selected relevant predictors have the potential to forecast accurately the traffic parameters of the target link. Moreover, the exclusion of irrelevant variables in the prediction model is effective.

Furthermore, compared to the Elastic net links based prediction model, the Granger causal links based prediction model provides better results when considering either traffic flow or speed as the traffic parameter to be measured. Interestingly, according to Figure 9, the Granger causality model selects more accurate predictors which can reduce RMSE value approximately three times less than that of elastic net links and more than five times less than that of all links based model.

Figure 9: Comparison of average RMSE of (a) flow (b) speed based models



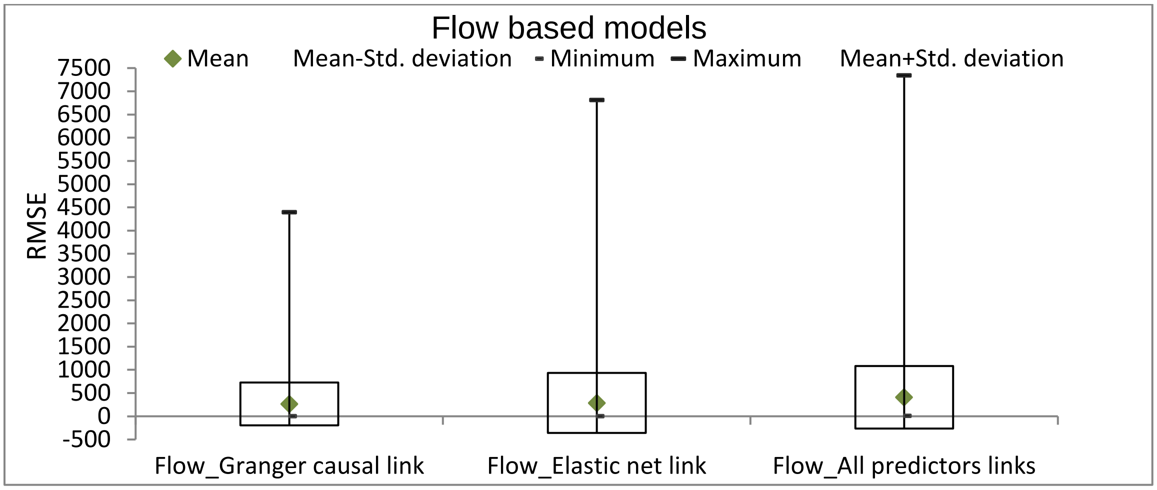
(a)

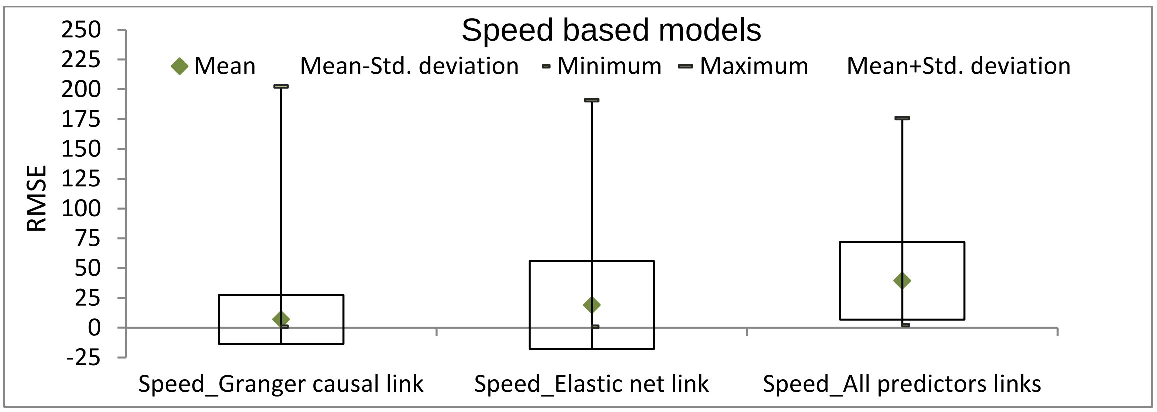


(b)

From the figure above, we can only see the average value of RMSE. The following figure (Figure 10) illustrates an overview of the variation of RMSE value for 454 predictors in each six models. It can be said that in both flow and speed cases, Granger causal links based model has less variation compared to other two models. However, the maximum value is much higher in all models because of having much higher RMSE value in some road links. These outlier values of RMSE also increase the average RMSE value in each model.

Figure 10: Overview of the distribution of RMSE for (a) Flow based (b) speed based models.





5. Conclusion

This study proposes statistical methods such as Granger causality and elastic net to identify a set of road links that are correlated or functionally related to a given target road link in an urban road network. Both Granger causality and elastic net significantly reduce the number of predictors for forecasting traffic parameter value of a target link. Short term traffic prediction models are developed by using the predictors obtained by these two methods. The prediction accuracies of these models are evaluated and compared with another traffic prediction model which has been built using all variables as the predictors. The comparison shows that both Granger causality and elastic net links based prediction models provide better accuracy than the model with all predictors. Also, in both flow and speed cases, model developed by Granger causal links shows less prediction error than model developed by elastic net links. These findings show that Granger causality method is efficient in reducing the irrelevant variables by systematically identifying the most relevant variables for short term traffic prediction model. In our future research, an approach to combining these two methods will be considered to implement the combined effect of Granger causality and elastic net on identifying predictors for short term traffic prediction. This paper also identifies the hierarchy of the importance of road links as the predictors as well as the location of these important predictors. This can be useful if all road links cannot be considered to build the prediction model due to having unavailable or malfunctioned detectors at many locations in the road network. If the target location has malfunctioning detector, importance predictors in the network can be used to predict traffic parameter in the target location.

6. References

1. Arnold, A., Liu, Y. and Abe, N., 2007, August. Temporal causal modeling with graphical Granger methods. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 66-75). ACM.

2. Bahadori, M.T. and Liu, Y., 2012, October. On Causality Inference in Time Series.In *2012 AAAI Fall Symposium Series*.

3. Blackwell, M., 2008. Multiple hypothesis testing: The F-test. Matt Blackwell Research, 2008.https://www.mattblackwell.org/files/teaching/ftests.pdf . Accessed June 15, 2016.

4. Chandra, S.R. and H. Al-Deek, 2009. Predictions of freeway traffic speeds and volumes using vector autoregressive models.Journal of intelligent Transportation Systems, 13 (2), pp. 53-72.

5. Clark, S., 2003. Traffic prediction using multivariate nonparametric regression. Journal of transportation engineering, Series 2003, 129:2 (161).

6. Cottrell, A. and Lucchetti, R., 2016. GNU regression, econometrics and time-series library. *Computer software. Retrieved from http://gretl.sourceforge.net*. Accessed June 8, 2016.

7. Friedman, J., Hastie, T. and Tibshirani, R., 2010. Regularization paths for generalised linear models via coordinate descent. Journal of statistical software, 33(1), p.1.

8. Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pp.424-438.

9. Granger, C.W., 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, *2*, pp.329-352.

10. Hasan. M.M. and Kim,J., 2016, November. Analysing functional connectivity and causal dependence in road traffic networks with Granger causality.In Australasian Transport Research Forum (ATRF), 38th, 2016, Melbourne, Victoria, Australia.

11. Hastie, T. and Qian, J., 2014. Glmnet vignette. Retrieve from http://www.web.stanford.edu/~hastie/Papers/Glmnet_Vignette.pdf . Accessed September 20, 2016.

12. Hobeika, A. G., and C.K. Kim, C.K., 1994, August. Traffic-flow-prediction systems based on upstream traffic.In Vehicle Navigation and information Systems Conference, 1994. Proceedings, 1994 (pp. 345-350). IEEE.

13. Kamarianakis, Y. and Prastacos, P., 2003. Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches. *Transportation Research Record: Journal of the Transportation Research Board*, (1857), pp.74-84.

14. Kamarianakis, Y. and Prastacos, P., 2004. Space–time modeling of traffic flow. Computers & Geosciences, 31(2), pp.119-133.

15. Li, Q. and Lin, N., 2010. The Bayesian elastic net. *Bayesian Analysis*, *5*(1), pp.151-170.

16. Sun, S., Zhang, C. and Yu, G., 2006, March. A Bayesian Network Approach to traffic flow Forecastig, IEEE, Trans. On Intell. Transp. Syst., Vol 7, no. 1, pp. 124-132.

17. Seth, A.K., Barrett, A.B. and Barnett, L., 2015. Granger causality analysis in neuroscience and neuroimaging. *The Journal of Neuroscience*, *35*(8), pp.3293-3297

18. Smith, B.L., Williams, B.M. and Oswald, R.K., 2002. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *10*(4), pp.303-321.

19. Stathopoulos, A., and Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, *11*(2), 2003, pp.121-135.

20. Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M. and Carpenter, J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj, 338, p.b 2393.

21. Zivot, E. and Wang, J., 2006. Vector autoregressive models for multivariate time series. *Modeling Financial Time Series with S-PLUS®*, pp.385-429.

22. Zou, H. and Hastie, T., 2003. Regression shrinkage and selection via the elastic net, with applications to microarrays. *Journal of the Royal Statistical Society: Series B. v67*, pp.301-320.

23. Zou H, Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B 2005, 67:301-320.

24. Zou H, Zhang, H.H., 2009. On the adaptive elastic-net with a diverging number of parameters: The annals of Statistics, 37:1733-1751.