Multivariate Short-term Traffic Flow Forecasting using Bayesian Vector Autoregressive Moving Average Model

Tiep Mai
PhD Student
Department of Computer Science and Statistics
Trinity College Dublin
Ireland
E-Mail Address: maik@tcd.ie

Bidisha Ghosh (Corresponding author)
Lecturer
Department of Civil, Structural and Environmental Engineering
Trinity College Dublin
Ireland
E-Mail Address: bghosh@tcd.ie

Simon Wilson
Professor
Department of Computer Science and Statistics
Trinity College Dublin
Ireland
E-Mail Address: swilson@tcd.ie

Word Count
5450 words + 6 Figures + 2 Tables
Abstract (246 words)

ABSTRACT

Short-term Traffic Flow Forecasting (STFF), the process of predicting future traffic conditions based on historical and real-time observations is an essential aspect of Intelligent Transportation Systems (ITS). The existing well-known algorithms used for STFF include time-series analysis based techniques, among which the seasonal Autoregressive Moving Average (ARMA) model is one of the most precise method used in this field. In the existing literature, ARMA model is mostly used in its univariate multiplicative form and the parameters of the model are mostly estimated using a frequentist approach. The effectiveness of STFF in an urban transport network can be fully be realized only in its multivariate form where traffic flow is predicted at multiple sites simultaneously. In this paper, this concept in explored utilizing an Additive Seasonal Vector ARMA (A-SVARMA) model to predict traffic flow in short-term future considering the spatial dependency among multiple sites. The Dynamic Linear Model (DLM) representation of the A-SVARMA model has been used here to reduce the number of latent variables. The parameters of the model have been estimated in a Bayesian inference framework employing a Markov Chain Monte Carlo (MCMC) sampling method. The serial correlation problem of MCMC sampling is relaxed by using marginalization and adaptive MCMC. Multiple variations of A-SVARMA, such as differenced process and mean process, have been studied to identify the most suitable prediction methodology. The efficiency of the proposed prediction algorithm has been evaluated by modelling real-time traffic flow observations available from a certain junction in the city-centre of Dublin.

INTRODUCTION

Intelligent Transportation Systems (ITS) is an emerging concept which has been utilized to improve efficiency and sustainability of existing transportation systems. Short-term traffic forecasting, the process of predicting future traffic conditions in short-term or near-term future, based on current and the past observations is an essential aspect of ITS. In the last decade, considerable research attention has been focused on developing precise, flexible, adaptable and universal short-term prediction algorithms for traffic variable observations. Several parametric and non-parametric techniques have been utilized to develop successful Short-Term Traffic Forecasting (STTF) algorithms (1, 2).

The predominant parametric approach in STTF is time-series analysis techniques. Time-series analysis techniques which are popular in STFF are smoothing techniques (1), Autoregressive linear processes (3) and Kalman filtering (4). Among these, the Autoregressive linear processes are the most developed and well-documented in this field. Ahmed and Cook (5) introduced the Auto-Regressive Moving Average (ARMA) class of models to the traffic flow forecasting literature. The next seminal step was extending the simple ARMA model to a seasonal format and accounting for the daily and/or weekly variability (6, 7). The aforementioned studies on applying ARMA model in developing STTF algorithms mainly focused on univariate structure; traffic data from any single station were modeled. In the last decade, the research attention has been shifted in developing more efficient prediction algorithms through the utilization of multivariate ARMA techniques which can model the spatial dependency and temporal evolution of traffic variables (such as, volume, speed and travel time) simultaneously.

One of the initial multivariate models was developed by Stathopoulos and Karlaftis (8) using state-space methodology. The model provided superior forecasts to equivalent univariate ARMA models. A multivariate ARMA technique called space-time autoregressive integrated moving average (STARIMA) methodology was applied to develop a model to account for the spatial dependency of traffic data in an urban network (9). The spatial dependency of the network were incorporated in the STARIMA model through the use of weighting matrices estimated based on the distances among the data collection points. A new class of time-series model called the Strutural Time-Series Model was used to develop multivariate STTF algorithm; these models outperformed univariate seasonal ARMA models (10). A direct multivariate extension of autoregressive linear processes has been first attempted by Chandra and Al-Deek (11). This model utilizes a Vector Auto-Regressive (VAR) structure for STTF. Freeway traffic speed and volume had been predicted in this study. The model did not consider the correlation of the noise among multiple stations or data collection points as there does not exist a Moving Average (MA) part. Also, the seasonal nature of the traffic data has been modeled by eliminating seasonality through a seasonal difference and not by direct modeling of seasonality in a seasonal ARMA form. In this study the authors compared VAR with other univariate models and concluded that adding correlations among different locations improves the prediction result. Still, the results are restricted to VAR structure which is only a subclass of seasonal Vector ARMA (VARMA) model. In the same year, Multi-Regression Dynamic Model (MDM) was adopted to develop a multivariate algorithm for STTF by Queen and Albers (12). The MDM consists of multiple independent regression equations often represented in Dynamic Linear Model (DLM) form. Similar to the previous study (11), MDM did not include the noise cross-correlation or the MA coefficients and moreover, MDM assumed spatially independent noise, which allows separate statistical inference for each station. Also, the spatial correlation among neighbouring stations evolved contemporaneously in the model and a temporal evolution of spatial cross-correlation was not modelled.

In this paper, a full multivariate extension of the most efficient univariate time-series model i.e. the seasonal ARMA has been proposed. Unlike the past studies this model involves a noise cross-correlation along with a seasonal form. In particular, an Additive Seasonal VARMA (A-SVARMA) model has been developed to predict traffic flow in short-term future in urban signalized arterial networks. A Bayesian framework has been proposed to estimate the parameters of the A-SVARMA model. The inference framework utilizes a Markov chain Monte Carlo (MCMC) sampling method. One serious problem of MCMC is the slow convergence caused by serial correlation. Hence, in order to have a better sampling of MA parameters, marginalization and adaptive MCMC are used. The proposed method has been applied to model traffic volume observations from multiple junctions situated at the city-centre of Dublin, Ireland. The results indicate that the proposed forecasting algorithm is an effective approach in predicting real-time traffic flow at multiple junctions within an urban transport network.

VARMA MODEL

Time series theory, including VARMA and DLM, is discussed in detail in (13, 14, 15). A brief summary is as follows. The vector of time-series observations is denoted by Y_t $(k \times 1)$ where k is the number of variables observed at time instants $t = 1, 2, ..., n$. A k-variate VARMA(p, q) is considered with $k \times 1$ common mean β and identical independent (iid) Multivariate Normal noise $E_t \sim N(0, \Sigma_e)$:

$$\Phi(B)(Y_t - \beta) = \Theta(B)E_t \tag{1}$$

with

$$\Phi(B) = I - \sum_{l=1}^{p} \phi_l B^l \tag{2}$$

$$\Theta(B) = I + \sum_{l=1}^{q} \theta_l B^l \tag{3}$$

where B is the backshift operator i.e. $B.Y_t = Y_{t-1}$; l is the time lag index. Each element ϕ_l or θ_l is a $k \times k$ matrix. I is the $k \times k$ identity matrix.

VARMA accounts for spatial-temporal dependency between multiple sites, for example the correlation between the traffic observation $Y_{t,i}$ from station i and the traffic observation $Y_{t-1,j}$ from neighbouring station j is modelled. This dependency is defined by the matrix ϕ_l for the spatial and temporal correlation and cross-correlation between all sites. However, using the full matrix is computationally costly. Hence, the dimension and computational cost have been reduced by adding neighbour information. Matrix $\mathcal{SP}_l$ denotes the neighbour dependency; $\mathcal{SP}_l(j, l) = 1$ iff $\phi_l(j, l) \neq 0$. The matrix $\mathcal{ST}_l$ is for θ_l dependency.

In existing literature on STFF, the ARMA class of models in their seasonal form (3) is always expressed in multiplicative form:

$$\Phi_a(B)\Phi_b(B^s)(Y_t - \beta) = \Theta_a(B)\Theta_b(B^s)E_t \tag{4}$$

where s is the seasonal period; $\Phi_a(B)$, $\Theta_a(B)$ are the usual AR, MA polynomials; $\Phi_b(B^s)$, $\Theta_b(B^s)$ are the seasonal AR, MA polynomials. However, the multiplicative form may not be the most appropriate for several reasons. Firstly, matrix multiplication is not commutative; a model with AR part $\Phi_a(B)\Phi_b(B^s)$ is different from a model with $\Phi_b(B^s)\Phi_a(B)$ and so it makes the physical interpretion more ambiguous. The differences between these two models are discussed in more detail by Yozgatligil and Wei (16). Secondly, consider the model with first-order dependency:

$$(I - \phi_{a,1}B)(I - \phi_{b,1}B^s)Y_t = E_t \tag{5}$$

However, the matrix multiplication implies that Y_t has second-order dependency $\phi_{a,1}\phi_{b,1}B^{s+1}$ with Y_{t-s-1}. As the order of spatial dependency may need to be fixed during analysis, this property is not desired. Also, directly using additive form omits the complex and computationally expensive matrix multiplication which is calculated repeatedly in the inference process. Furthermore, the additive form specifies a seasonal VARMA with 2 components, rather than 4 components by multiplicative form. As a result, the serial correlation problem of MCMC sampling which will be described in details later is solved efficiently. So, in this paper, a sparse additive representation of VARMA with a seasonality effect is used.

For the additive form specification, vector $\mathcal{IP}$ is defined as: $\phi_l \neq 0$ iff $l \in \mathcal{IP}$. Hence, $p = max(\mathcal{IP})$. Similarly, there is $\mathcal{IT}$ for the MA coefficients. So, for the VARMA model with seasonal period 10, $\mathcal{IP}$ can be set as $(1, 10)$. The Additive Seasonal VARMA is called by A-SVARMA from now on (The univariate version is called A-SARMA).

An A-SVARMA is then specified with $(\mathcal{IP}, \mathcal{SP}, \mathcal{VP}, \mathcal{IT}, \mathcal{ST}, \mathcal{VT})$ with mean β, noise variance Σ_e; $\mathcal{VP}$ and $\mathcal{VT}$ are the ordered vectors of non-zero elements of ϕ_l and θ_l respectively.

BAYESIAN FRAMEWORK

The parameters of the A-SVARMA model are estimated using Bayesian framework. The fundamentals of ultilizing a Bayesian inference of ARMA models in STFF have been discussed by Ghosh (3). The Bayesian framework for estimation of A-SVARMA model parameters is described in this section.

The A-SVARMA model has been estimated using DLM. The general DLM formula consists of Observation Equation:

$$Y_t = \beta_t + F_t\alpha_t + \nu_t \tag{6}$$

and Evolution Equation:

$$\alpha_t = G_t\alpha_{t-1} + w_t \tag{7}$$

where β_t is a $k \times 1$ vector which can be a time-variant mean in certain models; F_t is a $k \times km$ matrix of known constant; ν_t is the observation noise of k-variate $N(0, \Sigma_\nu)$; α_t is the $km \times 1$ state vector; G_t is the $km \times km$ state evolution matrix and w_t is the state evolution noise of Multivariate Normal distribution $N(0, \Sigma_w)$.

The A-SVARMA(p, q) model in Equation 1 can be represented using the DLM formula by:

$$\begin{aligned}
\beta_t &= \beta \\
F_t &= (I, 0, ..., 0) \\
G_t &= \begin{pmatrix} \phi_1 & I & 0 & \dots & 0 \\ \phi_2 & 0 & I & \dots & 0 \\ \dots & \dots & \dots & \dots & \\ \phi_{m-1} & 0 & 0 & \dots & I \\ \phi_m & 0 & 0 & \dots & 0 \end{pmatrix} \\
v_t &= v = 0 \\
w_t &= HE_t \\
H &= (I, \theta_1, ..., \theta_{m-1})^T
\end{aligned} \tag{8}$$

where $m = max(p, q+1)$; $\phi_i = 0$ for $i > p$; $\theta_i = 0$ for $i > q$; Y_t is the observation; E_t is the iid noise of $N(0, \Sigma_e)$.

The DLM representation helps to reduce the number of initial random variables from $k(p+q)$ to $k\ max(p, q+1)$. For the Bayesian inference, the prior of parameters $(\mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0)$ is defined as follows:

$$\begin{aligned}
p(\mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0) &= p(\mathcal{VP})p(\mathcal{VT})p(\beta)p(\Sigma_e)p(\alpha_0) \\
&= N(\mathcal{VP}|\mu_{\mathcal{VP}}, Q_{\mathcal{VP}}^{-1})N(\mathcal{VT}|\mu_{\mathcal{VT}}, Q_{\mathcal{VT}}^{-1})N(\beta|\mu_\beta, Q_\beta^{-1}) \\
&\quad IW(\Sigma_e|m_\Sigma, \Psi_\Sigma)N(\alpha_0|\mu_\alpha, Q_\alpha^{-1})
\end{aligned} \tag{9}$$

where $IW(.|m, \Psi)$ denotes the inversed Wishart distribution with degree m and inverse scale matrix Ψ; $N(.|\mu, Q^{-1})$ denotes the Multivariate Normal distribution with mean μ and precision matrix Q.

The conditional likelihood is:

$$p(Y_{1:n}|\mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0) = \prod_t p(Y_t|Y_{1:(t-1)}, \mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0) \tag{10}$$

The closed-form conditional posterior of $\mathcal{VP}, \beta, \Sigma_e, \alpha_0$ is obtained by the conjugate prior and some transformations. The transformation and conditional posterior is discussed in Appendix. The sampling $\mathcal{VP}, \beta, \Sigma_e, \alpha_0$ can be done by using standard distributions. However, sampling $\mathcal{VT}$ is complex and suffers from the serial correlation as its posterior density is intractable. Hence, MCMC sampling technique is used.

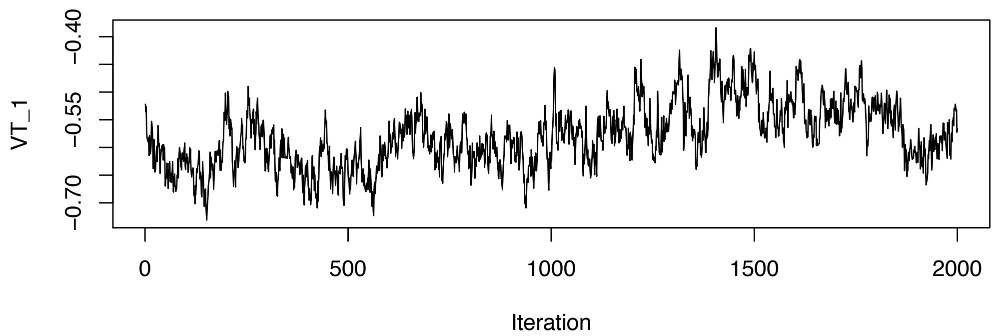
MCMC Sampling Scheme

(Please refer to the Appendix for the definition of all new variables introduced in this section.)

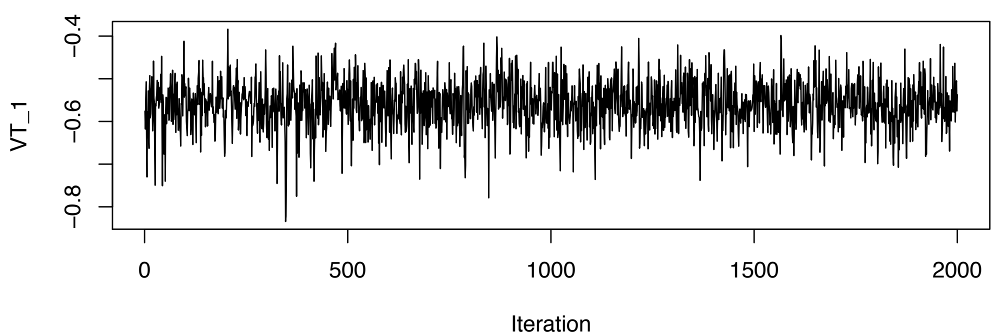
MCMC sampling is used to realize the Bayesian framework. The conditional posteriors of $\mathcal{VP}$, β, Σ_e, α_0 are inferred in Appendix. Gibbs Sampling (3) is used for each parameter block. The overall sampling scheme is as follows:

a) Sample $p(\beta|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \Sigma_e, \alpha_0)$ from $N(\beta|\mu'_\beta, Q'^{-1}_\beta)$

b) Sample $p(\Sigma_e|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \beta, \alpha_0)$ from $IW(\Sigma_e|m'_\Sigma, \Psi'_\Sigma)$



(a) Simple MCMC



(b) Adaptive MCMC with marginalization

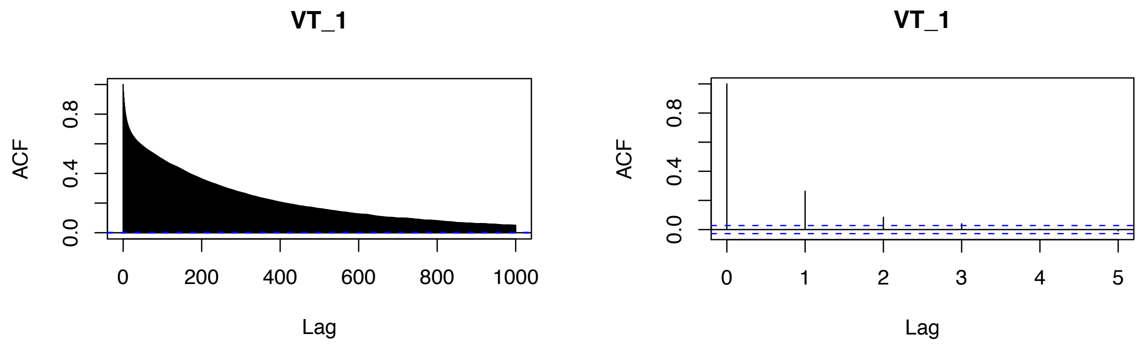
FIGURE 1: Trace of 2000 consecutive $\mathcal{VT}_1$ samples

c) Sample $p(\alpha_0|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e)$ from $N(\alpha_0|\mu'_\alpha, Q'^{-1}_\alpha)$

d) Sample $p(\mathcal{VP}, \mathcal{VT}|Y_{1:n}, \beta, \Sigma_e, \alpha_0)$

Steps a-c) are straightforward. For Step d), the following simple MCMC method is used. Firstly, $p(\mathcal{VP}|Y_{1:n}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0)$ is sampled from $N(\mathcal{VP}|\mu'_{\mathcal{VP}}, Q'^{-1}_{\mathcal{VP}})$. Then, a Metropolis-Hastings (3) is used with random walk for $p(\mathcal{VT}|Y_{1:n}, \mathcal{VP}, \beta, \Sigma_e, \alpha_0)$. The MCMC trace and autocorrelation plots of such a method with simulated data are in Figures 1(a) and 2(a). It can be observed from the figures, this method suffers from the serial correlation of $(\mathcal{VT}, \mathcal{VP})$. The serial correlation problem restricts the MCMC update and results in the slow MCMC convergence.

To solve the serial correlation problem, the following method is proposed: the posterior $p(\mathcal{VP}, \mathcal{VT}|Y_{1:n}, \beta, \Sigma_e, \alpha_0)$ is analytically marginalized with respect to $\mathcal{VP}$. Then, adaptive MCMC is used with the numerical evaluation of $q(\mathcal{VT}) = p(\mathcal{VT}|Y_{1:n}, \beta, \Sigma_e, \alpha_0)$. Consequently, the MCMC convergence is significantly faster. Figures1(b) and 2(b) show the trace and autocorrelation of this scheme which is much better than ones of simple MCMC.



(a) Simple MCMC (b) Adaptive MCMC with marginalization

FIGURE 2: Autocorrelation plot of $\mathcal{VT}_1$

APPLICATION

Data

The proposed Bayesian A-SVARMA methodology has been evaluated by modelling traffic volume observations from a busy thoroughfare in the city-centre of Dublin in Ireland. Three sites had been chosen for traffic data collection and the map of the chosen section of the transport network is provided in Figure 3. As seen in the figure, all the modelled sites are situated on/near Pearse Street which is one of the busiest roads in Dublin City.

The chosen sites will be referred as Stn 1, Stn 2 and Stn 3 from now on in this paper. The three sites are located in two signalized traffic intersections. To avoid unnecessary complexity in matrix and equation representation in an illustrative example, no other junctions were considered in this application of the A-SVARMA model. Both these junctions are four-legged cross-sections, with two-way traffic on all approaches barring the east-bound approach in Stn 3. Stn 1 and Stn 2 are two separate approaches on the second junction. These two approaches receive green time in separate phases. Both the junctions have three-phase signals with turning protection on Pearse Street. Both Stn 1 and Stn 3 have three lanes each and Stn 2 has two lanes. The data obtained collectively from all detectors on any approach, is used for the modelling. As the weekend traffic dynamics is very much unlike the traffic dynamics in the weekdays, the modelling is essentially carried out on the data observed during weekdays. Since, the used data set does not contain any missing data, no special treatment for missing data is required to be utilized here.

The data used for modelling was recorded from 22nd July 2010 to 3rd September 2010. 15-minute aggregate traffic volume observations were used in modelling purposes. Total 32 days of data from weekdays were used. The traffic observations from first 26 days were used for fitting and the rest of data was used for model evaluation. In this section, the multivariate traffic time-series is denoted as, $Z_t(1:k)$, where k is the number of stations at which traffic data were measured. k equals to 3 in this application and consequently, $Z_t(i)$ is the traffic volume time-series from Stn i. Stn 1 and Stn 2 are chosen in such a way that the observation from these two sites at do not have any obvious spatial correlation. The observations from both these stations are spatially correlated with observations at Stn 3. However, these visual observations were not utilized or implemented in the development of the matrices $\mathcal{VP}$ and $\mathcal{VT}$.

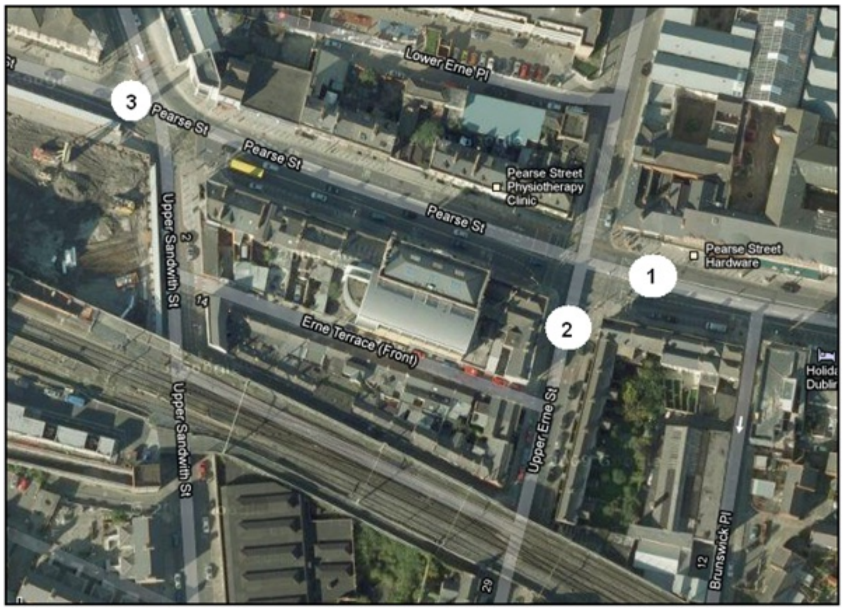


FIGURE 3: Map of the chosen junctions

The traffic volume observations from the abovementioned junctions form non-stationary time-series datasets (3). The initial step in analyzing this multivariate time-series dataset is to eliminate the seasonality and the trend of Z_t through filtering and/or transformations. Stationarity or weak stationarity was attempted through removal of trend and seasonal patterns of the traffic data. In this paper, multiple variations of A-SVARMA and A-SARMA have been developed and fitted to identify the most suitable one. In the next section, the modelling variations are discussed in further detail.

Model Variations

Three variations of A-SVARMA model were fitted to the traffic volume observations using the aforementioned Bayesian using the aforementioned Bayesian inference framework. Traffic volume observations from signalized intersections in the city-centre of Dublin had been modelled previously by the authors (3, 7, 10). The previous studies showed that there exists a daily seasonality in the weekday traffic volume datasets. The seasonality induces non-stationarity in the traffic volume time-series. As mentioned in the previous section, stationarity is a key condition for time-series analysis and 3 different variations of A-SVARMA utilizes 3 different approaches to attain stationarity in the traffic volume time-series.

In the previous studies (3), a seasonal differencing has been performed to eliminate the daily periodicity in the traffic data. The first variation of the A-SVARMA utilizes this in an attempt to attain stationarity. This model is named the SD model. Similar to the previous studies, a 15-minute aggregate traffic data has been modeled in this paper and hence, for a daily model the season $s = 96$. In a SD model,

$$Y_{SD,t}(i) = \nabla_s Z_t(i) = Z_t(i) - Z_{t-s}(i) \qquad i = 1, 2, 3 \tag{11}$$

The autocorrelation plot of $Y_{SD,t}(3)$ is shown in Figure 5(a). $Y_{SD,t}(1:3)$ is further modelled using Equations 6 to 10.

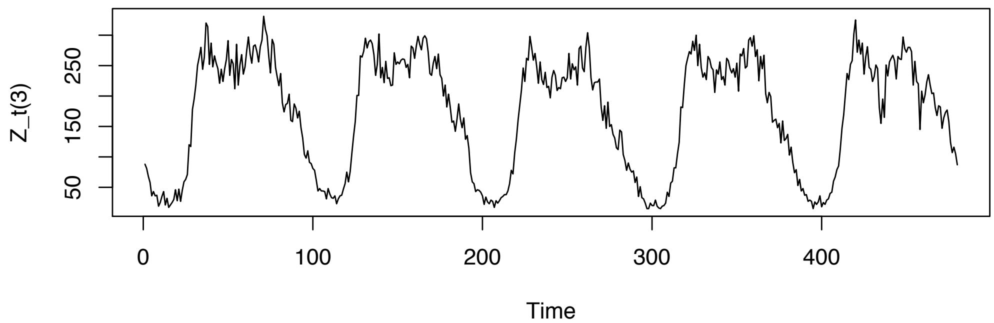
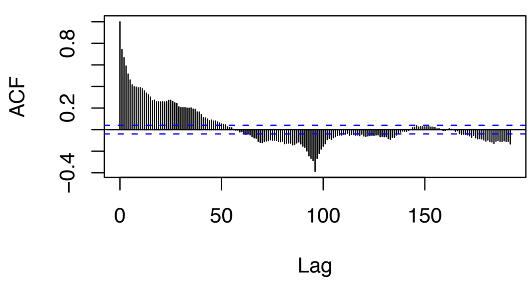
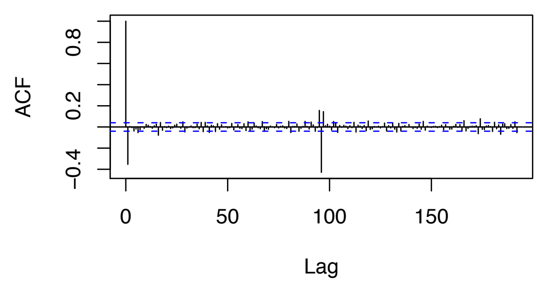


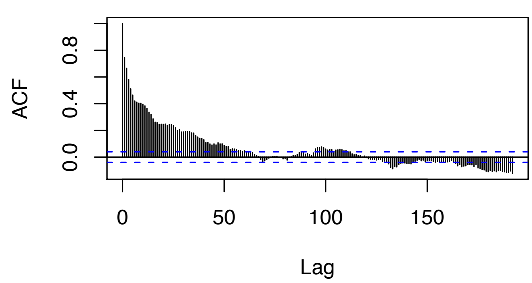
FIGURE 4: 5-day traffic flow of $Z_t(3)$



(a) $Y_{SD,t}(3)$



(b) $Y_{DSD,t}(3)$



(c) $Y_{MP,t}(3)$

FIGURE 5: Autocorrelation plot of (a) $Y_{SD,t}(3)$, (b) $Y_{DSD,t}(3)$ and (c) $Y_{MP,t}(3)$

The second variation of the A-SVARMA model studied utilizes a first and a seasonal difference to attempt a stationary behavior. The model is named as DSD model and follows an equation:

$$Y_{DSD,t}(i) = \nabla\nabla_s Z_t(i) = (Z_t(i) - Z_{t-s}(i)) - (Z_{t-1}(i) - Z_{t-s-1}(i)) \qquad i = 1, 2, 3 \tag{12}$$

The autocorrelation of $Y_{DSD,t}(3)$ is shown in Figure 5(b). The plot is definitely a faster decaying one than SD model and shows improved indications of stationary behavior in the modeled time-series data. $Y_{DSD,t}(i)$ is further modelled using Equations 6 to 10.

The third variation of the A-SVARMA model is slightly different from the previous models as it does not attempt to attain stationarity through differencing at the initial step. Rather the model is developed considering the within day variability of the traffic data which is modeled as:

$$\begin{aligned} Z_t(i) &= \mu_{MP_t}(i) + Y_{MP,t}(i) \qquad i = 1, 2, 3 \\ \mu_{MP_t}(i) &= \mu_{MP_{t+s}}(i) \end{aligned} \tag{13}$$

$\mu_{MP_t}(i)$ is the empirical mean/average which varies with the time of the day and is obtained from fitting the past traffic observations. This model has been named as the MP model. The autocorrelation plot of $Y_{MP,t}(i)$ in Figure 5(c) indicates nearly stationary behavior similar to Figure 5(a). $Y_{MP,t}(i)$ is further modeled using Equations 6 to 10.

The A-SVARMA structure for the three multivariate models (SD, DSD, MP) are the same:

$$\begin{aligned} \mathcal{IP} &= \mathcal{IT} = (1, 96) \\ \mathcal{SP}_l &= \mathcal{ST}_l = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \end{aligned} \tag{14}$$

where $l = 1$ or 96. Notice that with such $\mathcal{SP}_l$ and $\mathcal{ST}_l$, the transformed $Y_{M,t}(3)$ (at Stn 3) is dependent on $Y_{M,t-l}(1)$ and $Y_{M,t-l}(2)$ (at Stn 1 and Stn 2). For comparison purposes, along with each variation of the A-SVARMA model, a univariate model of similar ARMA structure has been fitted to the traffic volume data $Z_t(3)$. The univariate variation of SD model is named as SD-U which is essentially a univariate SARIMA model with seasonal differencing. The univariate variation of DSD model is named as DSD-U which is a univariate SARIMA model with both first and seasonal differencing. The MP-U is the univariate variation of MP model. The Equation 14 for univariate ARMA models changes slightly; $\mathcal{IP}$ and $\mathcal{IT}$ are still (1,96), but $\mathcal{SP}_l = \mathcal{ST}_l = 1$. All six models are used for forecasting short-term traffic volume and the results are discussed in the following subsection.

Result

All univariate and multivariate models are compared based on their predictive performances while forecasting traffic data at Stn 3. For any model M, at any time instant t, the v-step ahead forecasts are denoted by $\widehat{Y}_{M,t,t+v}(i)$. $\widehat{Z}_{M,t,t+v}(i)$ can be generated by performing inverse operations on $\widehat{Y}_{M,t,t+v}(i)$. The models have been used to generate s-step ahead or one-day ahead forecasts. The

TABLE 1: Summary of all stations for multivariate models

		Model		
Stn		DSD	SD	MP
1	$Std_Y(veh/15min)$	11.94	13.20	9.26
	$E_{M,1}(veh/15min)$	5.78	5.41	4.88
	$E_{M,20}(veh/15min)$	6.64	5.73	5.29
	$E_{M,96}(veh/15min)$	7.71	5.83	5.57
2	$Std_Y(veh/15min)$	20.42	18.94	13.86
	$E_{M,1}(veh/15min)$	10.02	9.05	8.42
	$E_{M,20}(veh/15min)$	11.58	9.12	8.67
	$E_{M,96}(veh/15min)$	12.38	9.17	8.76
3	$Std_Y(veh/15min)$	31.32	40.75	29.64
	$E_{M,1}(veh/15min)$	15.00	14.79	13.11
	$E_{M,20}(veh/15min)$	18.88	15.86	14.45
	$E_{M,96}(veh/15min)$	20.75	16.25	15.12

TABLE 2: Summary for univariate models

		Model		
Stn		DSD-U	SD-U	MP-U
3	$Std_Y(veh/15min)$	31.32	40.75	29.64
	$E_{M,1}(veh/15min)$	15.15	15.40	13.23
	$E_{M,20}(veh/15min)$	19.07	16.06	14.48
	$E_{M,96}(veh/15min)$	20.90	16.71	15.22

prediction error is defined as the Mean Accumulated Error (MAE). MAE for model M for v-steps ahead forecasts is:

$$
\begin{aligned}
E_{M,v,t} &= \frac{1}{v}\sum_{l=1}^{v}(|Z_{t+l}(i) - \widehat{Z}_{M,t,t+l}(i)|) \\
E_{M,v} &= \frac{1}{T}\sum_{t=n+1}^{n+T} E_{M,v,t}
\end{aligned}
\tag{15}
$$

where, $T = 576$ for this study as all the models are used for predictions over six days; $1 \leq v \leq s$. The prediction $\widehat{Z}_{M,t,t+l}$ has been marginalized by the posterior density. Hence, $E_{M,v}$ can be estimated by MCMC samples. The prediction accuracy of the SD, DSD and MP are represented in Table 1. The prediction accuracy is expressed in MAE values in vehicles/15 minutes. In Table 2, the predictive performance of the univariate models while used for forecasting traffic volume at Stn 3 are tabulated in the form of MAE values. In Figure 6, plots of the MAE values in vehicles/15 minutes are plotted against the steps of forecast for Stn 3 for all six models. In the figure, the MAE values increase with steps of forecast. However the rate of this increase is different for different variations of A-SVARMA model.

In general, multivariate models give better prediction than univariate models and the MP

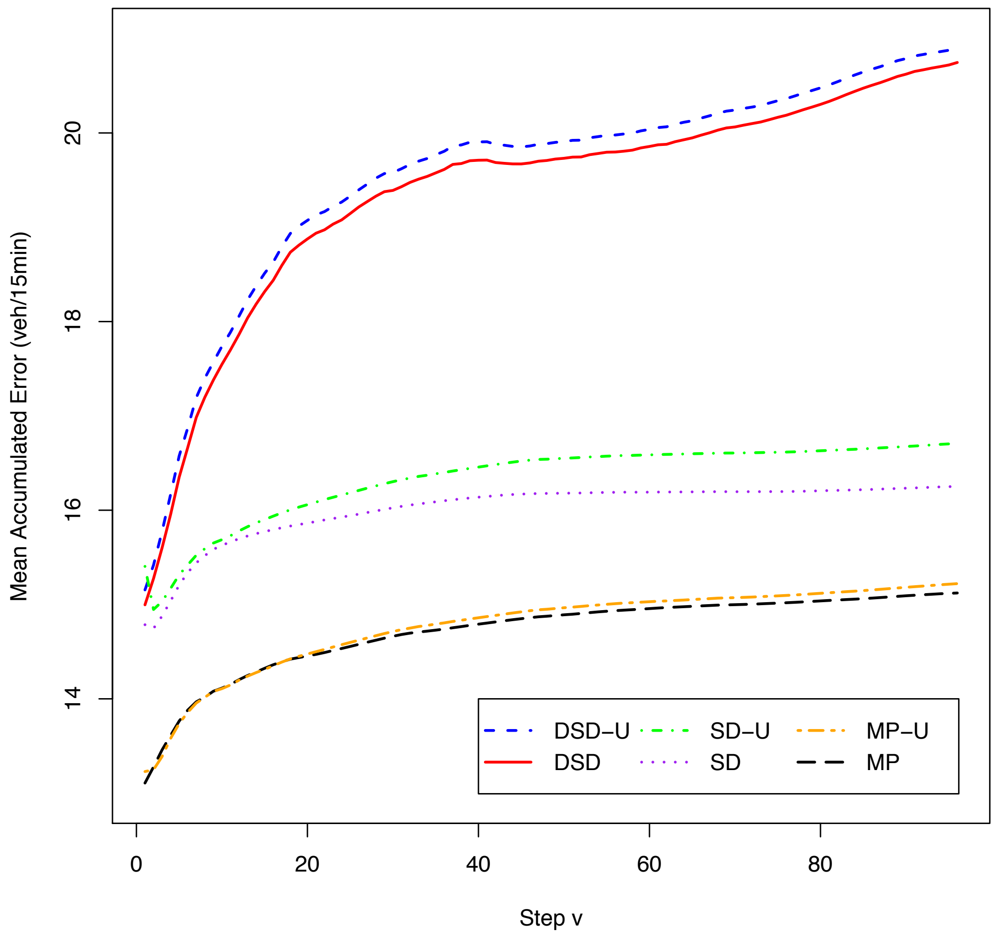


FIGURE 6: Accumulated error

model provides the best forecast among all. However, the univariate variation in this case, the MP-U model, provides only slightly inferior prediction results. This is due to the fact that 15-minute aggregate data were considered in developing the models. As a car can travel a considerable distance within a 15-minute time interval, the spatial correlation among the stations are not as high as can be seen from high resolution traffic observations. It is expected that for a dataset of higher resolution the impact of multivariate modeling will be more significant. Among all three variations the predictions from the DSD model and its univariate counterpart is the worst. The reason lies in the utilization of the first difference operator and its inverse operation. The prediction is,

$$\widehat{Z}_{DSD,t,t+l}(i) = Z_{t+l-s}(i) + (\widehat{Z}_{DSD,t,t+l-1}(i) - Z_{t+l-s-1}(i)) + \widehat{Y}_{DSD,t,t+l}(i) \quad (16)$$

The prediction error at each step of the forecast is carried to the next step when performing the inverse difference operation for converting back to the original traffic flow. This process is the main reason behind high MAE values for DSD and DSD-U models. As seen in Figure 6, the prediction of models DSD and DSD-U is quite precise at step 1 but the errors quickly increase from step 2. This fact remains true for every prediction model using first difference operator. The SD and SD-U models produce stable forecasts. The standard deviation (Std_Y) of $Y_{MP,t}$ is smaller than one of $Y_{SD,t}$ and this is the main reason behind the comparatively higher prediction accuracy of MP model.

CONCLUSION

In this paper, for the first time a seasonal ARMA model has been used to develop STTF algorithm in a multivariate paradigm. A Bayesian inference framework for estimating the parameters of A-SVARMA has been developed for the STFF algorithm. The Bayesian estimation provides flexibility to introduce expert knowledge in the model with the use of prior densities. MCMC sampling is applied to realize the Bayesian estimations. In such sampling method, marginalization and adaptive MCMC are proposed to solve the problem of serial correlation. As a result, the MCMC sampling converges much faster.

The proposed A-SVARMA model is also the first attempt in modeling correlation of spatial noise (MA components) among multiple traffic data collection points or junctions for STFF. The model proves that there exists such spatial correlation and it is beneficial to model such behavior to improve the applicability, efficiency and robustness of STFF algorithms.

The study compares three variations of A-SVARMA model For illustrative purposes, real-time traffic data from a small network in the city Dublin, Ireland had been considered and the traffic volume observations from the network has been modeled successfully using the proposed variations of A-SVARMA model. A time-variant mean process model (MP) provides the best prediction accuracy. This model outperforms all other multivariate and univariate models.

ACKNOWLEDGEMENT

This work is funded by the STATICA project, a Principal Investigator program of Science Foundation Ireland, Grant number 08/IN.1/I1879.

REFERENCES

1. Brian L. Smith, Billy M. Williams, and R. Keith Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10(4):303 – 321, 2002.

2. Eleni I. Vlahogianni, John C. Golias, and Matthew G. Karlaftis. Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews: A Transnational Transdisciplinary Journal*, 24(5):533 – 557, 2004.

3. Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Bayesian time-series model for short-term traffic flow forecasting. *Journal of Transportation Engineering*, 133(3):180–189, 2007.

4. Iwao Okutani and Yorgos J. Stephanedes. Dynamic prediction of traffic volume through kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1):1 – 11, 1984.

5. M S Ahmed and A R Cook. Analysis of freeway traffic time-series data by using box-jenkins techniques. *Transportation Research Record No. 722, Urban Systems Operations*, pages 1 – 9, 1979.

6. Billy M. Williams and Lester A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 129(6):664–672, 2003.

7. Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Time-series modelling for forecasting vehicular traffic flow in dublin. In *84th Transportation Research Board Conference, National Academies, Washington D.C*, pages 1–22, 2005.

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

9. Yiannis Kamarianakis and Poulicos Prastacos. Space-time modeling of traffic flow. *Computers & Geosciences*, 31(2):119 – 133, 2005. Geospatial Research in Europe: AGILE 2003.

10. Bidisha Ghosh, Biswajit Basu, and Margaret O'Mahony. Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE-Transactions on Intelligent Transportation Systems*, 10:246–254, June 2009. ISSN 1524-9050.

11. Srinivasa Ravi Chandra and Haitham Al-Deek. Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 13(2):53 – 72, 2009.

12. Catriona Queen and Casper Albers. Intervention and causality: Forecasting traffic flows using a dynamic bayesian network. *Journal of the American Statistical Association*, 104(486):669–681, June 2009.

13. George Edward Pelham Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

14. Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples (Springer Texts in Statistics)*. Springer, 2nd edition, 2006.

15. Raquel Prado and Mike West. *Time Series: Modeling, Computation, and Inference*. Chapman & Hall, 2010.

16. Ceylan Yozgatligil and William W. S. Wei. Representation of multiplicative seasonal vector autoregressive moving average models. *The American Statistician*, 63(4):328–334, 2009.

17. Siddhartha Chib and Edward Greenberg. Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, 64(1-2):183 – 206, 1994.

APPENDIX

Transformation

The first 3 theorems of this section are from univariate ARMA analysis of (17). The results are extended to VARMA and sumarized here. We use the DLM representation in the Section Bayesian Framework for the sparse VARMA.

Theorem 1. *For $t > 0$, define:*

$$\Phi(B,t) = I - \sum_{i=1}^{(t-1)} \phi_i B^i \tag{17}$$

$$\Theta(B,t) = I + \sum_{i=1}^{(t-1)} \theta_i B^i \tag{18}$$

with $\phi_i = 0$ for $i > p$; $\theta_i = 0$ for $i > q$; $\alpha_{i,t} = 0$ for $i > m$. Then:

$$\Phi(B,t)(Y_t - \beta) = \phi_t \alpha_{1,0} + \alpha_{t+1,0} + \Theta(B,t)E_t \tag{19}$$

Let $X_t = Y_t - \beta$. Using Equation 19, some results follows:

Theorem 2. *For $t = 1, ..., n$, define:*

$$\mathcal{V}_{1,t} = Y_t - \sum_{i=1}^{(t-1)} \phi_i Y_{t-i} - \sum_{i=1}^{(t-1)} \theta_i \mathcal{V}_{1,t-i} - \phi_t \alpha_{1,0} - \alpha_{t+1,0} \tag{20}$$

$$\mathcal{W}_{1,t} = I - \sum_{i=1}^{(t-1)} \phi_i - \sum_{i=1}^{(t-1)} \theta_i \mathcal{W}_{1,t-i} \tag{21}$$

then:

$$\mathcal{V}_{1,t} - \mathcal{W}_{1,t}\beta = E_t \tag{22}$$

for every $t = 1, ..., n$

From Equations 9 and 22, the conditional posteriors of β and Σ_e are:

$$p(\beta|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \Sigma_e, \alpha_0) = N(\beta|\mu'_\beta, Q'^{-1}_\beta) \tag{23}$$

and:

$$p(\Sigma_e|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \beta, \alpha_0) = IW(\Sigma_e|m'_\Sigma, \Psi'_\Sigma) \tag{24}$$

where $\mu'_\beta, Q'_\beta, m'_\Sigma, \Psi'_\Sigma$ are calculated from $\mu_\beta, Q_\beta, m_\Sigma, \Psi_\Sigma, \mathcal{V}_{1,t}, \mathcal{W}_{1,t}$.

Theorem 3. *For* $t = 1, ..., n$, *define:*

$$\mathcal{V}_{2,t} = X_t - \sum_{i=1}^{(t-1)} \theta_i V_{2,t-i} - \alpha_{t+1,0} \tag{25}$$

$$\mathcal{U}_{2,t} = \sum_{i=1}^{(t-1)} \phi_i X_{t-i} + \phi_t \alpha_{1,0} \tag{26}$$

$$\mathcal{W}_{2,t}\mathcal{VP} = \mathcal{U}_{2,t} - (\sum_{i=1}^{(t-1)} \theta_i \mathcal{W}_{2,t-i})\mathcal{VP} \tag{27}$$

then:

$$\mathcal{V}_{2,t} - \mathcal{W}_{2,t}\mathcal{VP} = E_t \tag{28}$$

for every $t = 1, ..., n$

From Equations 9 and 28, the conditional posterior of $\mathcal{VP}$ is :

$$p(\mathcal{VP}|Y_{1:n}, \mathcal{VT}, \beta, \Sigma_e, \alpha_0) = N(\mathcal{VP}|\mu'_{\mathcal{VP}}, Q'^{-1}_{\mathcal{VP}}) \tag{29}$$

where $\mu'_{\mathcal{VP}}, Q'_{\mathcal{VP}}$ are calculated from $\mu_{\mathcal{VP}}, Q_{\mathcal{VP}}, \mathcal{V}_{2,t}, \mathcal{W}_{2,t}$.

In (17), Kalman smoothing is used to sample α_0. However such a method involves a state evolution matrix G and a matrix inverse of size $km \times km$. For a sparse matrix G in seasonal VARMA, the method is costly. So, here we use a transformation similar to above methods:

Theorem 4. *For* $t = 1, ..., n$, *define:*

$$\mathcal{V}_{3,t} = X_t - \sum_{i=1}^{(t-1)} \phi_i X_{t-i} - \sum_{i=1}^{(t-1)} \theta_i \mathcal{V}_{3,t-i} \tag{30}$$

$$\mathcal{W}_{3,t} = \mathcal{U}_{3,t} - \sum_{i=1}^{(t-1)} \theta_i \mathcal{W}_{3,t-i} \tag{31}$$

where matrix $k \times km$ $\mathcal{U}_{3,t}$ = $(\mathcal{U}_{3,t,1}, ..., \mathcal{U}_{3,t,m})$ *and* $\mathcal{U}_{3,t,1} = \phi_t$; $\mathcal{U}_{3,t,t+1} = I_k$ *for* $t + 1 \leq m$. *Then:*

$$\mathcal{V}_{3,t} - \mathcal{W}_{3,t}\alpha_0 = E_t \tag{32}$$

for every $t = 1, ..., n$

Finally, the conditional posterior of α, is obtained from Equations 9 and 32:

$$p(\alpha_0|Y_{1:n}, \mathcal{VP}, \mathcal{VT}, \beta, \Sigma_e) = N(\alpha_0|\mu'_\alpha, Q'^{-1}_\alpha) \tag{33}$$

where μ'_α, Q'_α are calculated from $\mu_\alpha, Q_\alpha, \mathcal{V}_{3,t}, \mathcal{W}_{3,t}$.