Date

Title

Tiep Mai
PhD Researcher
Department of Computer Science and Statistics
Trinity College Dublin
Email: maik@tcd.ie

Simon Wilson
Professor
Department of Computer Science and Statistics
Trinity College Dublin
Email: swilson@tcd.ie

Bidisha Ghosh
Assistant Professor
Department of Civil, Structural and Environmental Engineering
Trinity College Dublin
Email: bghosh@tcd.ie

Number of words: [max 5000]
Number of figures:

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 methods used in this field. 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 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.

[Abstract: approximately word limit: 150. Needs shortening]

Keywords: Multivariate Seasonal ARMA; Short-term traffic flow forecasting.

I. Introduction

Intelligent Transportation Systems (ITS) are emerging concept which have been utilized to improve the efficiency and sustainability of existing transportation systems. Short-term traffic flow forecasting, the process of predicting future traffic conditions in the short-term, based on current and 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 Flow Forecasting (STFF) algorithms \cite{ref:Smith2002, ref:Vlahogianni2004}.

The predominant parametric approach in STFF is time-series analysis techniques. Time-series analysis techniques which are popular in STFF are smoothing techniques \cite{ref:Smith2002}, autoregressive linear processes \cite{ref:Ghosh2007} and Kalman filtering \cite{ref:Okutani1984}. Among these, the autoregressive linear processes are the most developed and well-documented in this field. Ahmed and Cook \cite{ref:Ahmed1979} 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 \cite{ref:Williams2003, ref:Ghosh2005}. The aforementioned studies on applying ARMA model in developing STFF 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 \cite{ref:Stathopoulos2003} 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 \cite{ref:Kamarianakis2005}. 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 Structural Time-Series Model was used to develop a multivariate STFF algorithm; these models outperformed univariate seasonal ARMA models \cite{ref:Ghosh2009}. A direct multivariate extension of autoregressive linear processes has been first attempted by Chandra and Al-Deek \cite{ref:Chandra2009}. This model utilizes a Vector Auto-Regressive (VAR) structure for STFF. 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 is no Moving Average (MA) part. Also, the seasonal nature of 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 the VAR structure which is only a subclass of the seasonal Vector ARMA (VARMA) model. In the same year, Multi-Regression Dynamic Model (MDM) was adopted to develop a multivariate algorithm for STFF by Queen and Albers \cite{ref:Queen2009}. The MDM consists of multiple independent regression equations often represented in Dynamic Linear Model (DLM) form. Similar to the previous study \cite{ref:Chandra2009}, 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 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.

II. VARMA Model

Time series theory, including VARMA and DLM, is discussed in detail in \cite{ref:Box1994,ref:Shumway2006, ref:Prado2010}. A brief summary is as follows. The vector of time-series observations is denoted by $\square \times 1$ Y_t where $\square$ is the number of variables observed at time instants $\square = 1,2,\ldots,\square$. A $\square$-variate VARMA$(\square,\square)$ is considered with $\square \times 1$ common mean $\square$ and identical independent (iid) Multivariate Normal noise $\square_{\square} \sim \square(0,\square_{\square})$:

$\Phi(\square)(\square_{\square} - \square) = \Theta(\square)\square_{\square}$	(1)

where B is the backshift operator; $\Phi(\square) = \square - \sum_{\square=1}^{\square} \square_{\square}\square^{\square}$; $\Theta(\square) = \square - \sum_{\square=1}^{\square} \square_{\square}\square^{\square}$; $\square$ is the lag index; Each element $\square_{\square}$ or $\square_{\square}$ is a $\square \times 1$ matrix. $\square$ is the $\square \times 1$ identity matrix.

VARMA accounts for spatial-temporal dependency between multiple sites, for example the correlation between the traffic observation $\square_{\square,\square}$ from station $\square$ and the traffic observation $\square_{\square-1,\square}$ from neighbouring station $\square$ is modelled. This dependency is defined by the matrix $\square_{\square}$ 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 $\square\square_{\square}$ denotes the neighbour dependency; $\square\square_{\square}(\square,\square) = 1 \Leftrightarrow \square_{\square}(\square,\square) \neq 0$. The matrix $\square\square_{\square}$ is for $\square_{\square}$ dependency.

In existing literature on STFF, the VARMA class of models in their seasonal form \cite{ref:Ghosh2007} is usually expressed in multiplicative form.

$\Phi_a(\square)\Phi_b(\square^{\square})(\square_{\square} - \square) = \Theta_a(\square)\Theta_b(\square^{\square})\square_{\square}$	(2)

where $\square$ is the seasonal period; $\Phi_a(\square)$, $\Theta_a(\square)$ are the usual AR, MA polynomials; $\Phi_b(\square^{\square})$, $\Theta_b(\square^{\square})$ are the seasonal AR, MA polynomials. However, the multiplicative form may not be the most appropriate for several reasons. Matrix multiplication is not commutative; a model with AR part $\Phi_a(\square)\Phi_b(\square^{\square})$ is different from a model with $\Phi_b(\square^{\square})\Phi_a(\square)$ and so it makes the physical interpretion more ambiguous. The differences between these two models are discussed in more detail by Yozgatligil and Wei \cite{ref:Yozgatligil2009}. 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 is solved efficiently by using marginalization and adaptive MCMC. So, in this paper, a sparse additive representation of VARMA with a seasonality effect is used. For example, the VAR model with normal effect at lag 1 and seasonal effect at lag 96 is specified by $\square_1$ and $\square_{96}$ ($\square_{\square} = 0$ for $\square \neq 1{,}96$). The Additive Seasonal VARMA is called by A-SVARMA from now on. $\square\square$ and $\square\square$ are the ordered vectors of non-zero elements of $\square_{\square}$ and $\square_{\square}$ respectively.

III. Bayesian Inference and MCMC sampling

In this paper, the A-SVARMA model has been estimated using DLM. By using the DLM representation, the number of initial random variables is reduced from $\square \times (\square + \square)$ to $\square \times \square\square\square(\square, \square + 1)$ \cite{ref:Chib1994, TRB paper}. Define the initial state vector of DLM representation of A-SVARMA by $\square_0$. For Bayesian inference, the prior of parameters $\square\square, \square\square, \square_{\square}, \square, \square_0$ is defined as follows:

$\square(\square\square, \square\square, \square, \square\square, \square_0) =$	$\square(\square\square \vert \square_{\square\square}, \square_{\square\square}^{-1})\square(\square\square \vert \square_{\square\square}, \square_{\square\square}^{-1})\square(\square \vert \square_{\square}, \square_{\square}^{-1})$	(3)
	$\square\square(\square_{\square} \vert \square_{\square}, \Psi_{\square})\square(\square_0 \vert \square_{\square}, \square_{\square}^{-1})$	

where $\square\square(.\vert\square, \Psi)$ denotes the inversed Wishart distribution with degree $\square$ and inverse scale matrix Ψ; $\square(.\vert\square, \square^{-1})$ denotes the Multivariate Normal distribution with mean $\square$ and precision matrix $\square$. These priors are chosen for both their flexibility in modelling prior opinion about the parameters' values, and mathematical convenience as they are in a form that permits closed-form computation of some posterior distributions.

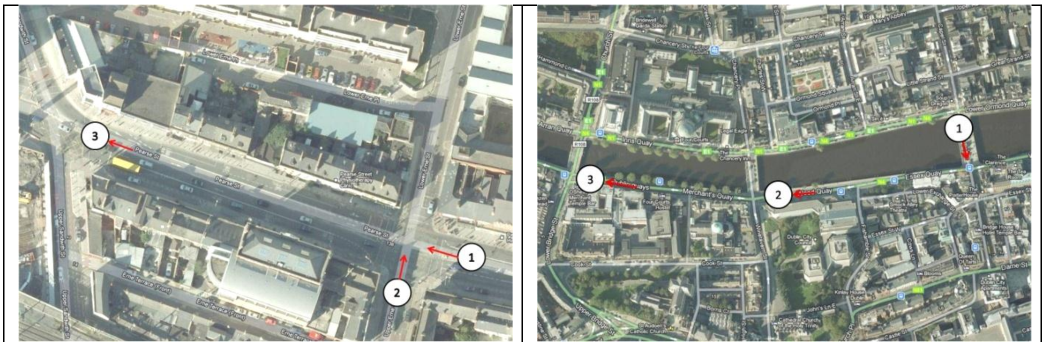
MCMC sampling is used to implement the Bayesian method. Gibbs Sampling \cite{ref:Ghosh2007} is used for each parameter block. We refer to \cite{TRB paper} for details of this method, which we note is computationally intensive but a well-established approach for Bayesian inference in time series models.

Comment [h1]: Simon, we remove the MCMC part as it is technical for this journal ? If so, for the same reason, do you think whether we should add the DLM-VARMA conversion or not.

Comment [SW2]: I think that we should not add the DLM-VARMA conversion either – perhaps we can just add a reference?

IV. Short-term Traffic Flow Forecasting

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. Two datasets corresponding to two chosen sections of Dublin traffic map were considered as in Figure 1. In each dataset, three sites had been chosen for traffic data collection. As seen in the figure, the modelled sites are situated on/near Pearse Street and the South Quays of River Liffey which are some of the busiest roads in Dublin City.



(a) Section 1 (b) Section 2

Figure 1: Map of the chosen junctions

For both datasets, the chosen sites will be referred as Stn 1, Stn 2 and Stn 3 from now on in this paper. 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, although in practice the methods can be extended to parts of the road network with many more junctions. In dataset 1, the three sites are located in two signalized traffic intersections. 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.

Dataset 1 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

[Description for dataset 2]

Comment [h3]:

Dataset 2 consists of 25 weekdays recorded from 11st June 2008 to 7th July 2008. Originally, it is 5-minute data but 2 consecutive data points are aggregated to avoid the traffic light effect (Hence, in each day, there are 288 data points each of which is a 10-minute aggregate traffic flow). The first 20-day data is used for fitting.

In this section, the multivariate traffic time-series is denoted as, □□(1: □), where □ is the number of stations at which traffic data were measured. □ equals to 3 for both datasets and □□(□), is the traffic volume time-series from Stn □. A usual trace plot of traffic flow of Stn 3 of dataset 1 is in Figure 2.

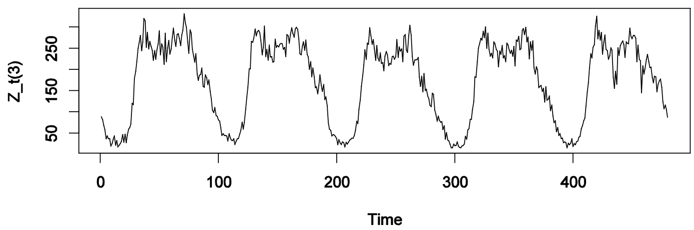


Figure 2: 5-day traffic flow of □□(□) of dataset □

The traffic volume observations from the abovementioned junctions form non-stationary time-series datasets \cite{ref:Ghosh2007}. The initial step in analyzing this multivariate time-series dataset is to eliminate the seasonality and the trend of □□ through filtering and/or transformations. Stationarity or weak stationarity was attempted through removal of trend and seasonal patterns of the traffic data. In most of the previous models \cite{ref:Ghosh2007} [add more refs], a differencing or seasonal differencing has been performed to eliminate the daily periodicity in the traffic data. Hence, these models correspond to the class of ARIMA models. In this study, the within-day variability of the traffic data is considered by using the time-varying within-day mean □□(□):

$$
\begin{aligned}
□_{□}(□) &= □_{□}(□) + □_{□}(□) \ \text{ for } □ = 1,2,3 \\
□_{□}(□) &= □_{□+□}(□)
\end{aligned} \tag{4}
$$

where □ is the daily period. $□_{□}(□)$ is the empirical mean/average which varies with the time of the day and is obtained from fitting the past traffic observations. Then, $□_{□}$ is modeled as A-SVARMA.

In 15-minute dataset 1 with daily period $□ = 96$, 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. The correlations are modeled in the first normal lag and first daily lag. Hence, there are 4 AR, MA matrices $□_1$, $□_{96}$, $□_1$ and $□_{96}$with the same neighbour dependency structure:

$$
□□_{□} = □□_{□} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \tag{5}
$$

Notice that with the above structure, the variability at Stn 3 is dependent on the variability of Stn 1 and Stn 2.

For dataset 2 ($□ = 288$), there are $□_1$, $□_{288}$, $□_1$ and $□_{288}$with the neighbour dependency structure:

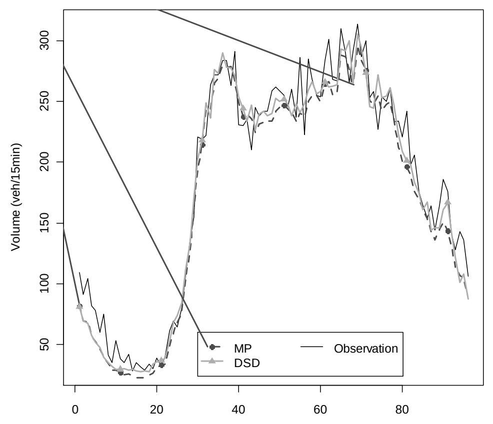
$$
□□_{□} = □□_{□} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \tag{6}
$$

The above model in which A-SVARMA and time-varying within day mean is used is notated by MP for short. In the next section, MP model is compared with a model, named DSD, in which differencing and daily differencing is used in conjunction with univariate ARMA. Model DSD, fitted separately in each site, is similar to ARIMA$(□, □, □) \times (□, □, □)$.

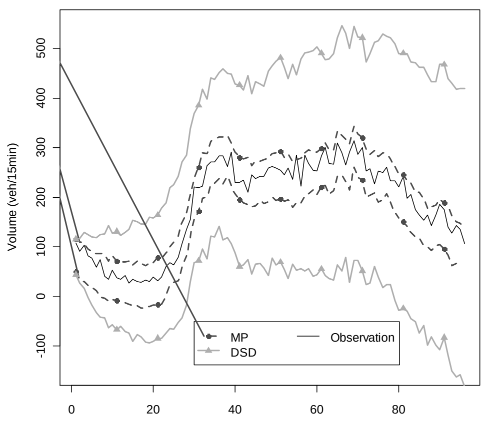
V. Result

In this section, the comparison between the MP model (A-SVARMA model with time-varying mean) and the DSD model which belongs to popular ARIMA class is made. For illustration, all the plots are for dataset 1; the summary result of both datasets is shown at the end.

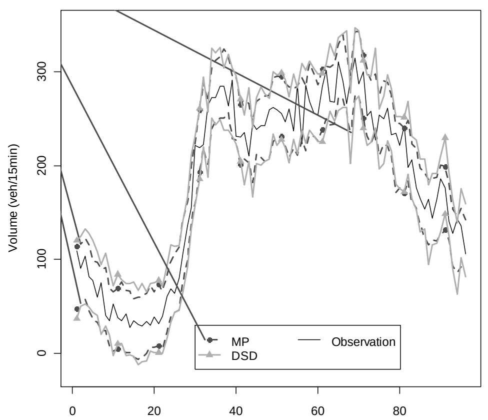
The 1-day ahead mean prediction and 90% credible interval of station Stn 3 of both models, MP and DSD, are in Figures 3(a) and 3(b). As in Figure 3(b), both intervals contain the true observation but the credible prediction interval of DSD model explodes over time and is of little use; the bandwidth of MP model is small and hence more informative. Figure 3(c) shows the 90% credible interval of 1-step ahead prediction for a whole day of Stn 3 (The 1-step prediction is not shown as it is hard to see the difference visually). Both MP and DSD bandwidths are quite small in this case. Notice that the positive value constraint is not imposed as in most ~~of~~ other papers.



(a) 1-day ahead prediction



(b) 90% credible interval of 1-day ahead prediction



(c) 90% credible interval of 1-step ahead prediction

Figure 3: Prediction plot of dataset □ - Stn □

It is hard to compare the predictive results in Figure 3(a) visually. Hence, the following Mean Accumulated Error (MAE) for □-steps ahead forecasts is used for prediction analysis:

$\square_{\square,\square} =$	$\frac{1}{\square}\sum_{\square=1}^{\square} \lvert \square_{\square+\square}(\square) - \hat{\square}_{\square,\square,\square+\square} \rvert$	(7)
$\square_{\square} =$	$\frac{1}{\square}\sum_{\square=\square+1}^{\square+\square} \square_{\square,\square}$	

where □ is the number of the starting prediction point ($□ = 576$ for dataset 1) ; $1 \leq □ \leq □$. The prediction accuracy is expressed in MAE values in vehicles/15 minutes for data 1. In Figure 4, plots of the MAE values in vehicles/15 minutes are plotted against the steps of forecast □. Tables 1 and 2 summarize the MAE values for both 2 datasets; $□□□_{□}$ is the standard deviation of the residuals,$□_{□}$, of each model; for DSD model, $□_{□} = \nabla\nabla_{□}□_{□} = (□_{□} - □_{□-□}) - (□_{□-1} - □_{□-□-1})$. It is clearly seen that the prediction of MP model is better than that of DSD model.

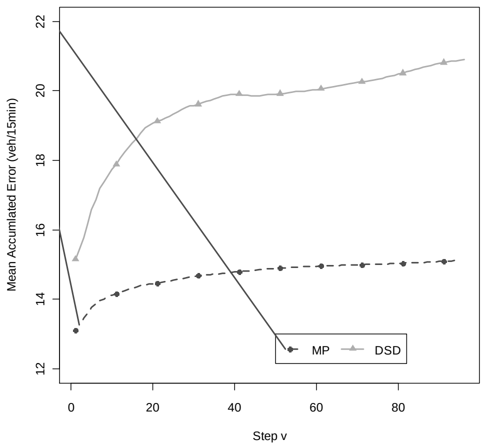


Figure 4: Mean Accumulated Error of dataset □ - Stn □

	Stn	1				2				3			
		□□□$_{□}$	$\square_{1}$	$\square_{20}$	$\square_{96}$	□□□$_{□}$	$\square_{1}$	$\square_{20}$	$\square_{96}$	□□□$_{□}$	$\square_{1}$	$\square_{20}$	$\square_{96}$
Model	MP	9.26	4.88	5.29	5.57	13.86	8.42	8.67	8.76	29.64	13.11	14.45	15.12
	DSD	11.94	5.78	6.63	7.79	20.42	10.02	11.57	12.05	31.32	15.15	19.07	20.90

Table 1: Mean Accumulated Error of dataset □ (Measurement is in □□□/□□□□□)

	Stn	1				2				3			
		□□□$_{□}$	$\square_{1}$	$\square_{100}$	$\square_{288}$	□□□$_{□}$	$\square_{1}$	$\square_{100}$	$\square_{288}$	□□□$_{□}$	$\square_{1}$	$\square_{100}$	$\square_{288}$
Model	MP	19.22	8.85	13.86	13.72	19.55	8.30	13.32	13.54	21.21	9.28	15.32	15.70
	DSD	20.11	10.89	20.38	21.55	20.14	10.40	19.14	20.72	21.42	11.73	20.27	22.28

Table 2: Mean Accumulated Error of dataset □ (Measurement is in □□□/□□□□□)

VI. Discussion

Fitting these models, while of some interest as a way to summarise traffic behaviour, is more usefully seen as a tool to allow for better management of urban traffic networks, where the short-term forecast is used to derive an optimal signal control policy for the network in the immediate future.

Various approaches have been proposed, most recently methods from the autonomous computing literature such as reinforcement learning e.g. Cahill and Dusparic (2009). Bayesian decision theory (Lindley, 1985) methods have great appeal here as they naturally link to Bayesian model fitting methods and are appropriate for decision making where there is uncertainty about the forecast. These methods require a cost function to be defined e.g. total journey time through the network, and determine a policy that minimizes the expected cost.

The vector model approaches can be particularly useful in situations where unusual traffic patterns occur, such as during road or lane closure, for which little historic data will be available. They not only provide forecasts of how traffic will enter the area of unusual pattern from unaffected areas outside it but, since they have learnt about the correlations between junction flows, may still produce reasonable forecasts within the area.

However, it still may be more appropriate to model such an area with a different model, and this will form the basis of future work.

[Tiep – do you have a reference to similar models to the one that we are proposing or are there none? Do we want to say anything more about these models, given that we have not implemented them yet?]

Comment [SW4]: Tiep, so let's not reference any similar model.
Yes, let's move the first 2 paragraphs to the Intro and remove the last 2.
For this journal a flowchart is a good idea and we can certainly use it in the PhD thesis as well.

Comment [h5]: Simon, we don't have the specific model yet. The discussed model shares some similarities with Queen's MDM model. But it is hard to have a short comparison here.

How about moving the first 2 paragraphs of this section to section Introduction and omit the last 2 paragraphs of future model ?

Also, do you think we should add the interaction flowchart of forecasting module, decision module and data feedback ? If so, I could add a flowchart to the first 2 paragraphs.

[Link to the decision module]
[Future work: forecasting of unusual pattern?]

VII. Conclusion

In this paper, 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 modelling 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 the MP model (which comprises A-SVARMA model and time-varying mean process) with the DSD model, a member of univariate ARIMA class. 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 have been modelled successfully. Conclusively, the time-variant mean process model (MP) provides the better prediction accuracy than DSD model.

References

Comment [h6]:

[1] B. L. Smith, B. M. Williams, and R. K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting. Transportation Research Part C: Emerging Technologies, 2002.
[2] E. I. Vlahogianni, John C. Golias, andMatthew G. Karlaftis, Short-term traffic forecasting: Overview of objectives and methods. Transport Reviews: A Transnational Transdisciplinary Journal, Vol. 24(5), pp. 533 – 557, 2004.
[3] B. Ghosh, B. Basu, and M. O'Mahony, Bayesian time-series model for short-term traffic flow forecasting. Journal of Transportation Engineering, Vol 133(3),pp. 180–189, 2007.
[4] I. Okutani and Y. J. Stephanedes, Dynamic prediction of traffic volume through kalman filtering theory. Transportation Research Part B: Methodological, Vol 18(1), pp. 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, pp. 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, Vol 129(6), pp. 664–672, 2003.
[7] B. Ghosh, B. Basu, and M. O'Mahony, Time-series modelling for forecasting vehicular traffic flow in dublin. In 84th Transportation Research Board Conference National Academies, Washington D.C, pp. 1–22, 2005.
[8] A. Stathopoulos and M. G. Karlaftis, A multivariate state space approach for urban traffic flow modeling and prediction. Transportation Research Part C: Emerging Technologies, Vol 11(2), pp. 121 – 135, 2003.
[9] Y. Kamarianakis and P. Prastacos, Space-time modeling of traffic flow. Computers & Geosciences, Vol 31(2):119 – 133, 2005.

[10] B. Ghosh, B. Basu, and M. O'Mahony, Multivariate short-term traffic flow forecasting using time-series analysis. IEEE-Transactions on Intelligent Transportation Systems, Vol 10, pp. 246–254, 2009.

[11] S. R. Chandra and H. Al-Deek, Predictions of Freeway Traffic Speeds and Volumes Using Vector Autoregressive Models. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Vol 13(2), pp. 53 – 72, 2009.

[12] C. Queen and C. Albers, Intervention and causality: Forecasting traffic flows using a dynamic B~~b~~ayesian network. Journal of the American Statistical Association, Vol 104(486), pp. 669–681, 2009.

[13] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control. Prentice Hall PTR, Upper Saddle River, NJ, USA, 3rd edition, 1994.

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

[15] R. Prado and M. West, Time Series: Modeling, Computation, and Inference. Chapman & Hall, 2010.

[16] C. Yozgatligil and W. W. S. Wei, Representation of multiplicative seasonal vector autoregressive moving average models. The American Statistician, Vol 63(4), pp.328–334, 2009.

[17] S. Chib and E. Greenberg, Bayes inference in regression models with ~~arma~~ ARMA(p, q) errors. Journal of Econometrics, Vol 64(1-2), pp. 183 – 206, 1994

[18] V. Cahill and I. Dusparic, Distributed W~~w~~-learning: Multi-policy optimization in self-organizing systems., *3rd IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, *San Francisco*, *14-September*, IEEE, pp. 20 –29,2009.

[19] D. Lindley, Making Decisions, second edition. London: Wiley, 1985.

[20] T. Mai, B. Ghosh and S. Wilson, Multivariate Short-termTraffic Flow Forecasting using Bayesian Vector Autoregressive Moving Average Model, *Proceedings* of the 91st *Annual Meeting* of the *Transportation Research Board, 2012 (accepted).*