

# **A BAYESIAN TIME-SERIES MODEL FOR SHORT-TERM TRAFFIC FLOW FORECASTING**

by

Bidisha Ghosh<sup>1</sup>, Biswajit Basu<sup>2\*</sup> and Margaret O'Mahony<sup>3</sup>

## **ABSTRACT**

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is one of the popular univariate time-series models in the field of short-term traffic flow forecasting. The parameters of the SARIMA model are commonly estimated using classical (maximum likelihood estimate and/or least square estimate) methods. In this paper, instead of using classical inference the Bayesian method is employed to estimate the parameters of the SARIMA model considered for modelling. In Bayesian analysis, Markov chain Monte Carlo method is used to solve the posterior integration problem in high dimension. Each of the estimated parameters from the Bayesian method has a probability density function conditional to the observed traffic volumes. The forecasts from the Bayesian model can better match the traffic behavior of extreme peaks and rapid fluctuation. Similar to the estimated parameters, each forecast has a probability density curve with the maximum probable value as the point forecast. Individual probability density curves provide a time-varying prediction interval unlike the constant prediction interval from classical inference. The time-series data used for fitting the SARIMA model are obtained from an intersection in Dublin's city centre.

Keywords: Traffic flow, Predictions, Auto-regressive moving-average Models, Bayesian analysis.

---

<sup>1</sup>Graduate Student

<sup>2</sup>Associate Professor

<sup>3</sup>Chair of Civil Engineering

Department of Civil, Structural and Environmental Engineering, Trinity College Dublin, Ireland.

\*Corresponding author, Email: basub@tcd.ie , Tel:+35316082389, Fax:+35316773072

## INTRODUCTION

In any transportation network, the implementation of Intelligent Transportation System (ITS) to provide dynamic traffic control and management requires real-time forecasting of traffic conditions in the near (short-term) future. Short-term traffic forecasting (prediction horizon less than one hour, (Smith et al. 2002)) supports the two most important components of ITS; proactive traffic control required in Advanced Traffic Management Systems (ATMS) and real-time route guidance required in Advanced Traveler Information Systems (ATIS). With the increasing need for developing more efficient ITS based traffic management systems, a considerable amount of research has been focused on short-term traffic forecasting. An extensive review on this subject has been provided by Vlahogianni et al. (2004).

The well-known short-term forecasting methods can be broadly classified into univariate and multivariate approaches. In a univariate approach, traffic variables (such as speed, flow or occupancy etc.) are modeled using observations from any single site, whereas developing a single model to consider several sites is termed a multivariate approach. Univariate models are more common in short-term traffic forecasting literature because they are less complex computationally (Kamarinakis et al. 2003). Both multivariate and univariate models can be developed using different empirical and theoretical techniques (Van Arem et al. 1997). The empirical approaches (non-parametric and parametric) employ statistical methodology and/or heuristic methods for traffic flow forecasting. The non-parametric techniques include non-parametric regression (e.g. Davis and Nihan 1991) and neural networks (e.g. Vythoulkas 1993; Smith and Demetsky 1994; Kirby et al. 1997; Lingras and Mountford 2001; Yin et al. 2002; Lee et al. 2004; Vlahogianni et al. 2005). The parametric techniques include linear and non-linear regression, historical average algorithms (e.g. Smith and Demetsky 1997), smoothing

techniques (e.g. Smith and Demetsky 1997; Williams et al. 1998) and autoregressive linear processes. Of all autoregressive linear processes, the Autoregressive Integrated Moving Average (ARIMA) family of models were first introduced by Ahmed and Cook (1979). Different variations of ARIMA models e.g. simple ARIMA (Levin and Tsao 1980; Nihan and Holmesland 1980; Hamed et al. 1995), ATHENA (Kirby et al. 1997), SARIMA (Williams et al. 1998; Williams 2003; Ghosh et al. 2005), subset ARIMA (Lee and Fambro 1999) are popular in the short-term traffic forecasting literature. In particular, seasonal variations of the ARIMA model (e.g. Williams et al. 2003 and Ghosh et al. 2005) perform better than linear regression, historical average, and simple ARIMA (Chung and Rosalion 2001). Moreover, Smith et al. (2002) has shown that the SARIMA model in situations, where it is applicable, performs better than non-parametric regression. The major criticism towards using ARIMA models focuses on their tendency to concentrate on the mean values and their inability to predict the extremes (Vlahogianni et al. 2004). But SARIMA models are multi-step in nature and once fitted to a particular site, have high accuracy, are computationally cheap and easy to implement in real scenarios. From these considerations, a SARIMA model is chosen to model the traffic flow time-series data in this paper.

The standard SARIMA model is parametric by nature and the estimation of parameters can be done using classical (maximum likelihood estimates or least squares estimates) or Bayesian methods (Box and Jenkins 1976). Unlike classical, Bayesian inference reduces statistical inference to probabilistic inference by defining a joint distribution for both the parameters and the observable data (Neal 1993). Box and Jenkins (1976) proposed Bayesian estimation while developing the general SARIMA model along with other classical methods. Monahan (1983) completed a full Bayesian analysis of the ARIMA models and concluded that the Bayesian

technique holds a distinct advantage over the classical approach as the statistical analysis of ARIMA models is a certain class of non-standard problems where no classical/frequentist approach is widely accepted (Monahan 1983). The use of a *prior* (initial assumed distribution) in Bayesian estimation helps to include more information in the model, which could not be included otherwise. In this paper, a qualitative and quantitative comparison of the Bayesian methods with the classical methods is shown. In addition in this paper the Bayesian estimation of the parameters of the SARIMA model is performed using an innovative simulation procedure.

The use of Bayesian statistics is quite recent in the field of traffic forecasting. Some work has been done using Bayesian networks (using the concept of neural networks) in short-term traffic flow forecasting (Zhang et al. 2004). Hierarchical regression models with Bayesian inference are used in modeling freeway traffic flows considering the variability of parameter values throughout a day due to day-specific idiosyncrasies (Tebaldi 2002). Similar to this concept, instead of a constant parameter value for the parameters in the SARIMA model, a probability distribution (may not be normal as assumed in the case of maximum likelihood estimates) conditional to the observed traffic volume is associated with each parameter and with each point of prediction in the future. The technique eventually generates a prediction interval varying with time, which gives better flexibility towards traffic management in real-time. The inability of the SARIMA models to capture the extreme peaks and to imitate the rapid fluctuations in traffic flow is partly accounted for using the Bayesian technique as the forecasts are conditional on observed traffic flow. The Bayesian method can also be employed in developing a simple incident detection technique.

In this study, univariate traffic flow data from a particular junction in Dublin are fitted to a suitable seasonal ARIMA (SARIMA) time-series model. The model parameters are estimated

using both the classical (maximum likelihood estimate) and the Bayesian methods. A comparison between the forecasts from the two approaches is presented and the additional information obtained from the Bayesian methods in the context of traffic management is discussed.

### SARIMA PROCESS

A simple ARIMA model is made up of three parts, ‘AR’ i.e. autoregressive part, ‘I’ i.e. differencing part, ‘MA’ i.e. moving average part. Differencing is one of the filter/transformation techniques for removing ‘non-stationarity’ from any time-series data (Chatfield, 2004). The equation representing an ARIMA ( $p, d, q$ ) model for a time-series sequence  $Y_t (t = 1, 2, \dots, n)$  is

$$\phi(B)(1-B)^d Y_t = \theta(B)Z_t \quad [1a]$$

$$\text{where, } \phi(B) = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) \quad [1b]$$

$$\text{and } \theta(B) = (1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_q B^q); \quad [1c]$$

$Z_t$  is a white noise sequence;  $B$  is the ‘backshift operator’;

In ARIMA ( $p, d, q$ ),  $p$ ,  $d$  and  $q$  denotes the order of the AR process, the order of differencing and the order of the MA process respectively.

The time-series data containing a seasonal periodic component can be fitted to a SARIMA (seasonal ARIMA) model (Box and Jenkins 1976). The general *multiplicative* seasonal ARIMA model ( $p, d, q$ )( $P, D, Q$ )<sub>s</sub> is as follows

$$\phi(B) \Phi(B^S) (1-B)^d (1-B^S)^D y_t = \theta(B) \Theta(B^S) Z_t \quad [1d]$$

where,  $\Phi, \Theta, P, D, Q$  are the seasonal counterparts of  $\phi, \theta, p, d, q$  respectively and  $S$  denotes the seasonality. Following the three steps of the Box and Jenkins methodology (Box and Jenkins 1976), *identification, estimation and diagnostics checking* the SARIMA models are fitted to stationary or weakly stationary time-series data. In this paper, the estimation of the AR ( $p, P$ ) and

MA ( $q, Q$ ) parameters for fitting a SARIMA model to the traffic flow data are carried out using classical and Bayesian methods.

### **FITTING SARIMA MODEL TO TRAFFIC FLOW DATA**

The traffic flow data used for modelling are obtained from the inductive loop-detectors embedded in the streets of junction TCS 183 in the city-center of Dublin, as a part of the urban traffic control (UTC) data collection system. A map of the junction is given in figure 1. It is a four-legged junction, with one-way traffic on two approaches. The traffic volume passing through Tara Street, measured from the loop-detectors numbered 1, 2, 3, 4 are continuously recorded.

The data interval is unique to the data collection system of the existing urban traffic control system of any city. The data interval can vary from a few seconds to one hour. Short-term forecasting algorithms applicable to a traffic management system should have a prediction horizon of 15 minutes or less (Smith and Demetsky 1997). Similar to earlier research using SARIMA models for traffic flow modelling, (e.g. Williams et al. 2003 and Ghosh et al. 2005), a 15 minute data interval is used in this paper.

The data used for the modeling were recorded from 6 a.m. on 2<sup>nd</sup> September 2004 to 6 a.m. on 30th September 2004. The data collection period of nearly a month means that two types of periodicity, viz. daily and weekly, can be considered. Weekend travel behavior is quite unlike the travel behavior on weekdays. So, for developing a daily seasonal model the observations during the weekdays are only required to be considered. Williams (1999) showed that the daily models have a better accuracy than the weekly models. Hence, a daily seasonal model excluding weekend travel is selected. The total number of observations excluding weekends is 1920, i.e., in total, 20 days of data are used. The plot of data is shown in figure 2. Fifteen minutes are

considered as unit time index in this plot and consequently 96 observations are plotted over each day.

Stationary or weakly stationary time-series data can only be analyzed using existing time-series models (Chatfield 2004). For weak stationarity, a seasonal difference (over lag 96 i.e. 24 hours) is taken. A correlogram of the differenced data is plotted in figure 3. The correlogram shows that the ACF has larger values at lag 1 and at lag 96. The PACF (partial autocorrelation function) of the centred traffic data (figure 3) is plotted as well. This plot shows the possibility of a seasonal moving average component.

On calculating the error estimates, it can be concluded that a SARIMA (2,0,0)(0,1,1)<sub>96</sub> (Ghosh et al., 2005) is the best-fit model for junction TCS 183. But, there are some other SARIMA models also, which though not the best fit, perform very well compared with the best fit. All of these SARIMA models can capture the daily repetitive nature of traffic flow and the dependence of present traffic conditions on the immediate past. The SARIMA (1,0,0)(0,1,1)<sub>96</sub> model, (which has error estimates varying less than 1% from the best-fit model) is fitted to the traffic flow data obtained from junction TCS 183. This model is chosen instead of the best-fit model in order to ensure that the computations are parsimonious. However, it may be important that if required for a specific situation, this model can be easily extended by addition of a non-seasonal MA part so that it can be more versatile without much additional computation.

The equation used for modeling the centered traffic flow data considered in this paper is

$$(1 - \phi_1 B)(1 - B^{96})Y_t = (1 - \Theta_1 B^{96})\varepsilon_t \quad [2]$$

$$\text{which leads to } Y_t = \phi_1 Y_{t-1} + Y_{t-96} - \phi_1 Y_{t-97} + \varepsilon_t - \Theta_1 \varepsilon_{t-96} \quad [3]$$

where,  $\varepsilon_t \sim N(0, \sigma^2)$ .

The unknown parameters to be estimated are  $\phi_1$ ,  $\Theta_1$  and  $\sigma$  and are represented by a vector  $\xi^T = (\phi_1, \Theta_1, \sigma^2)$ . To estimate  $\xi$ , either a frequentist (classical) or a Bayesian technique is to be used. In frequentist, the classical technique maximum likelihood estimate or least square method is used to estimate the vector  $\xi$  (Box and Jenkins 1976). The exact maximum likelihood estimates are often preferred to least square methods (Chatfield 2004).

### **PARAMETER ESTIMATION USING THE BAYESIAN METHOD**

In the Bayesian approach, all unknown quantities are considered as random variables and uncertainties over those quantities are represented using probability distributions conditional on the available data. While estimating any parameter using classical/frequentist methods, the sampling distribution of the parameter is mostly assumed as normal or Gaussian. This approach is quite crude in the sense that in real situations the sampling distributions of the parameters can be different from normal. With Bayesian analysis, realistic approximations to the sampling distribution are considered and the inferences are reached using generic techniques and the observed data. The basic principle behind Bayesian statistics is as follows. Some prior ideas about any parameter or data set can be obtained from prolonged and detailed observations or by comparing with similar conditions. If these prior ideas are considered as ‘prior beliefs’ in any hypotheses regarding those situations or events, then conditional to the observed data (likelihood) an exact solution ‘posterior’ distribution of the parameter can be obtained (Lee 1997).

*The Likelihood Function* (Box and Jenkins 1976)

For both the frequentist (maximum likelihood estimation) and Bayesian methods for estimation of the vector  $\xi$ , the likelihood function is required to be calculated. Considering the centered traffic flow data diagnostics (mean and quantile-quantile plot) the demand data for each day



when considered over a few weeks seem to be normally distributed. The likelihood function  $L(\xi | Y)$  of  $\xi$  corresponding to a sample of  $n$  observations of  $Y_t$  (where,  $Y$  is vector of  $n$  observations of  $Y_t$ ) is as follows

$$L(\xi | Y) = \sigma^{-n} f(\xi) \exp\left(-\frac{S(\xi)}{2\sigma^2}\right) \quad [4]$$

$$\text{where, } S(\xi) = \sum_{t=-\infty}^n [\varepsilon_t | Y, \xi]^2$$

In equation 4,  $f(\xi)$  is a function of  $\xi$  which is to be calculated based on the inverse of the  $n \times n$  covariance matrix of the  $Y_t$ 's. To ease the computational complexity for calculating  $f(\xi)$  the history of the observed process and the history of the errors are incorporated as in Ravishankar and Ray (1996). In this paper, the history of observations ( $Y_{t-1}, Y_{t-2}, \dots, Y_{t-97}$ ) are available from the SCATS database (urban traffic control system) of past observations and the history of the unknown errors ( $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-96}$ ) (since the model parameters are unknown) is assumed to follow a similar normal distribution as the error  $\varepsilon_t$ . The multivariate ( $n$  variate) normal distribution of the SARIMA process can be represented as

$$\Pi(Y_1, Y_2, Y_3, \dots, Y_n | \xi) = \Pi(Y_1 | Y_2, Y_3, \dots, Y_n, \xi) \Pi(Y_2 | Y_3, \dots, Y_n, \xi) \dots \Pi(Y_n | \xi) \quad [5]$$

In equation 3, the moving average part is assumed to be represented by

$$e_t = \varepsilon_t - \Theta_1 \varepsilon_{t-96} \quad [6]$$

The observations are assumed to have the following multivariate normal distribution,

$$\Pi(Y_1, Y_2, Y_3, \dots, Y_n | \xi) \sim N_n(\mu, \Sigma) \quad [7]$$

where, the mean vector  $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$  and  $\Sigma_{n \times n}$  is the covariance matrix of  $Y_t$ .

The mean values of the distribution are

$$\mu_t = \phi_1 Y_{t-1} + Y_{t-96} - \phi_1 Y_{t-97} \quad t = 1, 2, \dots, n;$$

From Ravishankar and Ray (1996), the likelihood function is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} | Y_{1,2,3,\dots,n}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\frac{1}{2} \sum_{t=1}^n (Y_t - \mu_t)^T \boldsymbol{\Sigma}^{-1} (Y_t - \mu_t)\right\} \quad [8]$$

When the number of observations is greater than 50 the matrix  $\boldsymbol{\Sigma}_{n \times n}$  can be simplified to a scalar form (Box, Jenkins, 1976). Equation 8, when applied to the SARIMA model for the present study, approximates to (considering the simplified form of the matrix  $\boldsymbol{\Sigma}_{n \times n}$ , as  $\tau = \sigma^2 + \Theta_1^2 \sigma^2$ ),

$$L(\mu, \tau | Y_{1,2,3,\dots,n}) = (2\pi\tau)^{-n/2} \exp\left(-\frac{1}{2} \tau^{-1} \sum_{t=1}^n (Y_t - \mu_t)^2\right) \quad [9]$$

$$L(\phi_1, \Theta_1, \sigma^2 | Y_{1,2,3,\dots,n}) \propto \sigma^{-n} (1 + \Theta_1^2)^{-n/2} \exp\left\{-0.5\sigma^{-2} (1 + \Theta_1^2)^{-1} (A + \phi_1 B + \phi_1^2 C)\right\} \quad [10]$$

$$\text{where, } A = \sum_{t=1}^n (Y_t - Y_{t-96})^2 \quad [10a]$$

$$B = -2 \sum_{t=1}^n (Y_t - Y_{t-96})(Y_{t-1} - Y_{t-97}) \quad [10b]$$

$$C = \sum_{t=1}^n (Y_{t-1} - Y_{t-97})^2 \quad [10c]$$

For the priors, there are some restrictions on the range of  $\phi_1, \Theta_1$ . The invertibility region for the model, required by the condition that the roots of  $(1 - \phi_1 B) = 0$  should lie outside the unit circle is defined by the inequality,  $-1 < \phi_1 < 1$ . The condition,  $|\Theta_1| < 1$ , is required to ensure stationarity (Box and Jenkins 1976). The priors considered here for  $\phi_1, \Theta_1$  are uniform priors which give

$$\Pi(\phi_1) = \frac{1}{2} \quad (\text{considering uniform prior from } -1 \text{ to } 1) \quad [11]$$

$$\Pi(\Theta_1) = \frac{1}{2} \quad (\text{considering uniform prior from } -1 \text{ to } 1) \quad [12]$$

and, considering the non-informative prior for  $\sigma^2$ , gives

$$\Pi(\sigma^2) \propto 1/\sigma^2 \quad (\sigma^2 \text{ is always positive}). \quad [13]$$

Hence, the posterior is

$$\Pi(\phi_1, \Theta_1, \sigma^2 | Y_{1,2,3,\dots,n}) \propto \Pi(\phi_1) \Pi(\Theta_1) \Pi(\sigma^2) L(\phi_1, \Theta_1, \sigma^2 | Y_{1,2,3,\dots,n}) \quad [14]$$

$$\text{which yields, } \Pi(\xi | Y_{1,2,3,\dots,n}) \propto \sigma^{-n-2} (1 + \Theta_1^2)^{-n/2} \exp\left\{-0.5\sigma^{-2} (1 + \Theta_1^2)^{-1} (A + \phi_1 B + \phi_1^2 C)\right\} \quad [15]$$

By integrating the other unknown parameters except for the one whose distribution is to be estimated, the ‘marginal (posterior) distributions’ [ $\Pi(\sigma|Y)$ ,  $\Pi(\Theta_1|Y)$ ,  $\Pi(\phi_1|Y)$ ] of each of the unknown parameters can be found. In this paper, to get the marginal distributions of each unknown parameter, the other two unknowns are to be integrated out. So, to find the marginal distribution of the three unknown parameters the following integrations are to be performed.

$$\Pi(\phi_1 | Y) \propto \int_{-1}^1 \int_0^\infty \sigma^{-n-2} (1 + \Theta_1^2)^{-n/2} \exp\left\{-0.5\sigma^{-2} (1 + \Theta_1^2)^{-1} (A + \phi_1 B + \phi_1^2 C)\right\} d\sigma d\Theta_1 \quad [16a]$$

$$\Pi(\Theta_1 | Y) \propto \int_0^\infty \int_{-1}^1 \sigma^{-n-2} (1 + \Theta_1^2)^{-n/2} \exp\left\{-0.5\sigma^{-2} (1 + \Theta_1^2)^{-1} (A + \phi_1 B + \phi_1^2 C)\right\} d\phi_1 d\sigma^2 \quad [16b]$$

$$\Pi(\sigma^2 | Y) \propto \int_{-1}^1 \int_{-1}^1 \sigma^{-n-2} (1 + \Theta_1^2)^{-n/2} \exp\left\{-0.5\sigma^{-2} (1 + \Theta_1^2)^{-1} (A + \phi_1 B + \phi_1^2 C)\right\} d\phi_1 d\Theta_1 \quad [16c]$$

From the posterior distribution again the ‘predictive distribution’ can be obtained using the relation

$$\Pi(Y_{n+1} | Y_1, Y_2, Y_3, \dots, Y_n) = \int L(\xi | Y_{1,2,3,\dots,n}) \Pi(\xi | Y_{1,2,3,\dots,n}) d\xi \quad [16d]$$

### *Simulation of Distributions*

The determination of the posterior distributions often involves integration of complex integrals (mostly integrations of equation 16) with high dimensionality. Numerical integrations are often performed to compute the distributions for which the analytical solution is intractable. However,

numerical integration may lead to too many approximations and may even become intractable for large models. In many high dimensional cases of Bayesian analysis, certain refinements of Monte Carlo integration methods are often used (Carlin 1996). There are different non-iterative and iterative variations of these refinements. *Markov Chain Monte Carlo* (MCMC) is the particular iterative variation of the Monte Carlo method in which the simulated values are not in *iid* but are in a Markov chain. In summary, the goal of the MCMC is, given a target distribution  $\Pi(x)$ , to construct a Markov chain  $\{x_n\}$  whose limiting distribution is  $\Pi(x)$ . There are two popular MCMC algorithms, (i) Gibbs sampler (Geman and Geman 1984) and (ii) Metropolis-Hastings algorithm (Metropolis et al. 1953, Hastings 1970)

Unlike Metropolis-Hastings algorithm, the Gibbs Sampler requires the simulation from the full conditional distributions (equation 16a, b & c in this paper). This kind of sampling is possible only in the case of standard distributions. As the full conditionals in equations 16a, b and c are non-standard in nature, the Metropolis-Hastings algorithm technique is chosen in this paper to simulate the marginal distributions from the posterior distributions.

#### *Metropolis-Hastings Algorithm*

This is an MCMC technique Metropolis-Hastings algorithm and it therefore (Metropolis et al. 1953, Hastings 1970) tries to generate a Markov chain  $x^{(0)}, x^{(1)}, \dots$  having given the ‘target distribution’,  $\Pi(x)$ , as a stationary distribution. Assume that the target distribution  $\Pi(x)$  has a density function  $\pi(x)$  known up to the normalizing constant, denoted by  $\kappa$ . Considering some initial value  $x^{(t)}$ , a proposed simulation  $x^{sim}$  is generated from a candidate generating density  $q(x^{(t)}, x^{sim})$  and then accepted with a probability  $\alpha(x^{(t)}, x^{sim})$ , where,

$$\alpha(u, v) = \begin{cases} \min \left\{ \frac{\pi(v)q(v, u)}{\pi(u)q(u, v)}, 1 \right\} & \text{if } \pi(u)q(u, v) > 0 \\ 1 & \text{if } \pi(u)q(u, v) = 0 \end{cases} \quad [17a]$$

The criteria for acceptance is that the value of  $\alpha(x^{(t)}, x^{sim})$  should be greater than a random number generated from a standard normal distribution. If the proposed value is accepted then  $x^{(t+1)} = x^{sim}$ ; if not, then  $x^{(t+1)} = x^{(t)}$ . Using continuous iterations (until equilibrium or convergence is reached) a series of dependent realizations, forming a Markov chain, are simulated with  $\Pi(x)$  as the equilibrium distribution (Tanner 1996). The normalizing constant,  $\kappa$  is cancelled in the ratio (eq. 17a); hence the knowledge of  $\kappa$  is not needed to implement the algorithm. This makes the algorithm generic in approach. The choice of candidate generating densities influences the simulation performance greatly. Depending on the performance of the Metropolis-Hastings algorithm, the stationary/the equilibrium density are accepted as approximation for the real marginal densities. The candidate values of the parameter  $\phi_l$  is to follow a normal random walk as the proposal distribution is taken as normal centered at the current value with a standard deviation equal to 0.05. Considering this marginal distribution ‘ $\phi_l$ ’ as the ‘target distribution’ and using an arbitrary starting value of  $\phi_l$  (say,  $\phi_{l-in}$ ), a simulated value of  $\phi_l$  (say,  $\phi_{l-sim}$ ) is obtained with  $\phi_{l-sim} \sim N(\phi_{l-in}, \tau_l)$ . The simulated value of  $\phi_{l-sim}$  is accepted following the Metropolis-Hastings’ algorithm. As a symmetric transition function (normal distribution) is assumed, only the Metropolis algorithm is sufficient. According to this algorithm, each simulated value of  $\phi_l$  is accepted with a probability

$$\alpha(\phi_l) = \frac{\Pi(\phi_{l-sim} | \Theta_1, \sigma^2, y_{1,2,3,\dots,n})}{\Pi(\phi_{l-in} | \Theta_1, \sigma^2, y_{1,2,3,\dots,n})} \text{ or } 1 \text{ which is minimum} \quad [17b]$$

with the same acceptance criteria as previously mentioned. For the next simulated value of  $\phi_l$ , instead of the arbitrarily assumed value of  $\phi_{l-in}$ ,  $\phi_{l-sim}$  is used and a new  $\phi_{l-sim}$  is simulated using the Markov chain (normal random walk). The new  $\phi_{l-sim}$  is accepted using the same method as before. In this way 10000 values of the parameter  $\phi_l$  are simulated (figure 4). Two parallel

chains of simulation are shown in the figure. The chains start from different initial values, one from 0.1 and another from 0.8. Both the chains seem to converge towards the same value.

The same type of technique is applied for the other two parameters  $\Theta_1$  and  $\sigma^2$ . Each of them is separately simulated from a normal ‘proposal’ distribution centered at the current value and with standard deviations equal to 0.01 and 0.5 respectively for  $\Theta_1$  and  $\sigma^2$ . But the results obtained [Figure 5] due to this assumption show poor convergence. The reason behind the slow convergence of these two parameters is quite high correlation (correlation coefficient -0.7) between them and that they cannot really be simulated using separate normal random walk proposal distributions.

To avoid this problem the concept called *blocking* or updating multivariate blocks (Carlin 1996) is employed, leading to faster convergence. From Roberts and Sahu (1997), blocking of components leads to faster convergence rates, e.g. the larger the blocks are updated simultaneously the faster the convergence. Though blocking is more computationally demanding than component wise updating, it moves the high correlation to the random vector generator (Roberts & Sahu 1997). To perform blocking, certain methods are used to get an idea of the probable joint density function of the other two parameters. Both the parameters are to be updated using a bi-variate distribution approach.

The original expression of the posterior was given in equation 16. The parameter  $\phi_l$  has to be integrated out from the expression of equation 15, to get a tentative idea of the bi-variate proposal kernel. However, since analytical integration is not possible within the bounded intervals of [-1,1], the integral is approximated by changing the limits from [-1,1] to  $(-\infty, \infty)$  for the variables. The joint distribution has a form (integrating out the parameter  $\phi_l$  )

$$p^*(\Theta_1, \sigma^2) \propto \sigma^{-T+1} (1 + \Theta_1^2)^{\frac{1-T}{2}} \exp\left[-\left(\frac{AC-B^2}{2C}\right) \sigma^2 (1 + \Theta_1^2)^{-1}\right] \quad [18]$$

Given  $\Theta_1$ , the conditional distribution of  $\sigma^2$  is an inverse gamma distribution with parameters,

$$\alpha = T/2 \quad [19a]$$

$$\beta = \left[ \frac{(AC-B^2)/2C}{(1 + \Theta_1^2)} \right] \quad [19b]$$

Thus, simulation from the bi-variate proposal kernel (equation 18) is possible if a suitable approximation of the marginal of  $\Theta_1$  can be obtained. The marginal of  $\Theta_1$  (obtained by integrating out  $\phi_1$  and  $\sigma^2$ ) is given by,

$$F(\Theta_1) = \frac{\int_{-\infty}^{\Theta_1} (\sqrt{1+u^2}) du}{2.2956} \quad [20]$$

Simulation from this density is done by using the inverse transform method. Given that  $F(\Theta_1) \sim U(0,1)$ , and solving equation 20 for  $\Theta_1$ , values of  $\Theta_1$  are simulated. As,  $\Theta_1$  and  $\sigma^2$  are correlated, they are to be simulated from a joint distribution  $p^*(\Theta_1, \sigma^2)$  to facilitate 'block updating'. Using the simulated value of  $\Theta_1$  from equation 22, a value of  $\sigma^2$  is simulated from the inverse gamma distribution described in equation 19. Both values are accepted with a probability,

$$p(\Theta_1, \sigma^2) = \frac{\Pi(\Theta_{1\_sim}, \sigma_{sim}^2 | \phi_1, y_{1,2,3,\dots,n}) q(\Theta_{1\_in}, \sigma_{in}^2)}{\Pi(\Theta_{1\_in}, \sigma_{in}^2 | \phi_1, y_{1,2,3,\dots,n}) q(\Theta_{1\_sim}, \sigma_{sim}^2)} \quad [21]$$

where,

$$q(\Theta_1, \sigma^2) = \frac{\exp\left(-\left(\frac{AC-B^2}{2C}\right) \sigma^{-2} (1 + \Theta_1^2)^{-1}\right)}{\sigma^{T+1} (1 + \Theta_1^2)^{\frac{T-1}{2}}} \quad [22]$$

in accordance with the same acceptance criteria as previously mentioned. Following the same procedure as described for  $\phi_l$ , 10000 values of both parameters are simulated [figure 6 and 7]. The values for  $\Theta_1$  and  $\sigma^2$  are simulated from two sets of initial values as performed earlier in the case of simulation of  $\phi_l$ . The chains are seen to converge satisfactorily for both the parameters. Using the simulated values of  $\xi$ , 10000 values of  $Y_{n+l}$  are simulated from  $\Pi(Y_{n+1} | Y_1, Y_2, Y_3, \dots, Y_n)$  as described in equation 16d.

Traffic volumes are predicted from 6:30 a.m. to 12 p.m. on 30<sup>th</sup> September, 2004 with the proposed Bayesian SARIMA model with the estimated parameters. The period with twenty-three data points covers the morning peak hours of traffic. The densities of the first nine points of forecasts along with the observed data are plotted in nine subplots of figure 8. In the subplots, the vertical axis shows the probability density of each point in the range. The horizontal axis specifies the range of the forecast point which ideally varies from  $-\infty$  to  $+\infty$ . The density functions in the subplots vary considerably with different mean and variance values which can be obtained using Bayesian inference. As in the case of the original observations no uncertainty is involved, they are represented by a vertical line in each subplot.

### **PARAMETER ESTIMATION USING THE CLASSICAL METHOD**

To compare the Bayesian estimates with the results from the classical method, the estimation of parameters using the classical technique is carried out. The exact maximum likelihood method (Fuller 1996) is used for estimating  $\xi$  of the SARIMA (1,0,0)(0,1,1)<sub>6</sub> model. The non-seasonal AR1 parameter  $\phi_1$  has a value 0.4215 and the seasonal MA parameter  $\Theta_1$  has a value of 0.8215. According to the classical theories, the parameters are supposed to have normal densities. The normal density plots of  $\phi_1$  and  $\Theta_1$  are shown in figure 9(a) and (b). The standard deviation  $\sigma$  of the SARIMA process is 43.5.



## COMPARISON OF BAYESIAN AND CLASSICAL INFERENCE

In figures 9(a) and (b), the density estimates of the SARIMA parameters from Bayesian inference are plotted along with the normal density plots from classical inference. In both subplots, the parameter density obtained from the Bayesian estimate is not exactly normal. The variability of parameters is considered in Bayesian inference while forecasting. But in the case of classical forecasting the standard error associated with the parameter is ignored and the mean of the density plot is used only as a constant.

The five and half hour traffic volume predictions from the Bayesian and classical inference are plotted along with the original observations in figure 10 and figure 11 respectively. The absolute percentage error (APE) of the predictions from the Bayesian inference is 5.4%, whereas the APE for classical forecasts is 5.1%. Though the overall error estimates are comparable, only the Bayesian inference can capture the extreme morning peak.

The point forecasts in figure 11 vary within a range of 50 cars/15min, whereas in figure 10 the same varies within a range of 100cars/15min. Hence, the Bayesian forecasts have a bigger range of variability compared to the classical forecasts which are nearly flat in nature. The Bayesian predictions can more closely match the rapid variability of the real observations unlike the classical forecasts.

Both the plots in figure 10 and figure 11 include two other lines specifying the upper and lower limit of prediction. In figure 10, this prediction interval is the 95% confidence interval. The Bayesian analogue of classical/frequentist confidence interval is a credible set (Carlin 1996). A 95% credibility interval means that there is a probability 0.95 that the forecast/parameter lies within that particular interval at that point in time. Whereas a 95% confidence interval in a

classical case means that if the same procedure as used for constructing the interval is repeated several times then in 95% of the cases the range of the intervals will include the true value of the forecast/parameter. The credibility set is more rational as it is not based on the asymptotic theory.

As seen in figure 8, each prediction of Bayesian inference has a distribution of its own. So, each future point has a different variance. In figure 12, the plot of the standard deviation from Bayesian inference varies continuously with time accounting for the happening of the events (Bayesian update) whereas the same from classical inference is constant over time failing to account for the change in the conditional probability of a forecast.

## **CONCLUSIONS**

A Bayesian SARIMA time-series model has been proposed in this paper for short-term traffic flow forecasting. In the Bayesian inference, a MCMC simulation has been proposed to estimate the distributions of the parameters of the model. The study in this paper shows that the Bayesian inference of SARIMA models provides a more rational technique towards short-term traffic flow prediction compared to the commonly applied classical inference. Bayesian inference is conditional on the observed traffic volumes and generates an individual probability density curve for each of the SARIMA model parameters as well as for each point of prediction. Instead of using a constant parameter value estimated by classical methods, the variability of the parameters conditional on the traffic flow observations can be considered in Bayesian inference. Forecasts from the Bayesian approach can better model the traffic behavior in reality with rapid fluctuations and extreme peaks. In real traffic flow, uncertainties associated with traffic forecasts over different times of the day should vary depending on the congestion and other factors (such

as weather). To account for this, the prediction interval from Bayesian inference varies with time unlike the classical inference where point forecasts and a constant confidence interval are obtained based on normal distribution assumption. This provides more realistic information to the traffic planners and transport network managers and is important for efficient and optimal traffic management. The Bayesian inference can also be applied for incident detection by identifying occurrence outside a defined credibility interval and will be considered by the authors for future research.

### **Acknowledgements**

The authors would like to thank Dr. Sourabh Bhattacharya for his helpful discussions and advices on this paper. The research work is funded under the Programme for Research in Third-Level Institutions (PRTL), administered by the HEA.

## REFERENCES

- Ahmed, M. S. and Cook, A. R. (1979). "Analysis of freeway traffic time-series data by using Box–Jenkins techniques". *Transportation Research Record*, 722, 1–9.
- Box, G. E. P. and G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.
- Carlin, B.P. and Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, New York.
- Chatfield, C. (2004). *The Analysis of Time Series: An Introduction, Sixth Edition*, Chapman and Hall/CRC, London.
- Chung, E. and Rosalion, N. (2001). "Short term traffic flow prediction". *Proceedings of the 24<sup>th</sup> Australian Transportation Research Forum*, Hobart, Tasmania.
- Davis, G. & Nihan, N. (1991). "Nonparametric Regression and Short-Term Freeway Traffic Forecasting". *J. of Transportation Eng.*, 117, 178-188.
- Fuller, W.A. (1996). *Introduction to Statistical Time Series (Second Edition)*. John Wiley and Sons, Inc., New York.
- Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions and Bayesian restoration of images". *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Ghosh, B., Basu, B. and O'Mahony, M.M. (2005). "Time-series modelling for forecasting vehicular traffic flow in Dublin". 85<sup>th</sup> Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.
- Hamed, M.M., Al-Masaeid, H.R., Bani Said, Z.M. (1995). "Short-term prediction of traffic volume in urban arterials". *J. of Transportation Eng.*, 121(3), 249–254.

- Hastings, W.K. (1970). "Monte Carlo sampling methods using Markov chains and their application". *Biometrika*, 57, 97-100.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). "Equation of State Calculations by Fast Computing Machines". *J. of Chemical Physics*, 21, 1087-1092.
- Monahan, J.F. (1983). "Fully Bayesian analysis of ARMA time series models". *J. of Econometrics*, 21, 307-331.
- Neal, R.M. (1993). "Probabilistic inference using Markov Chain Monte Carlo methods". *Technical Report* CRG-TR-93-1
- Nihan, N.L. and Holmesland, K.O. (1980). "Use of the Box and Jenkins time series technique in traffic forecasting". *Transportation*, 9, 125-143.
- Kamarianakis, Y. and Prastakos, P. (2003). "Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches". 82nd Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.
- Kirby, H.R., Watson, S.M. and Dougherty, M.S. (1997). "Should we use neural network or statistical models for short-term motorway traffic forecasting?". *Int. J. of Forecasting*, 13, 43-50.
- Lee, D.H., Zheng, W.Z. and Shi, Q.X. (2004). "Short-term freeway traffic flow prediction using a combined neural network model". 84<sup>th</sup> Annual Meeting of Transportation Research Board (CD-ROM), TRB, Washington, D. C.
- Lee, P.M. (1997). *Bayesian Statistics: An Introduction (Second edition)*. Arnold, A member of the Hodder Headline group, London.
- Lee, S., Fambro, D.B. (1999). "Application of subset autoregressive moving average model for short-term freeway traffic volume forecasting". *Transportation Research Record* 1678, 179-188.

- Levin, M., Tsao, Y.D. (1980). "On forecasting freeway occupancies and volumes". *Transportation Research Record* 773, 47–49.
- Lingras, P., Mountford, P. (2001). "Time delay neural networks designed using genetic algorithms for short-term inter-city traffic forecasting", IEA/AIE 2001, LNAI 2070, 290–299.
- Smith, B.L., Demetsky, M.J. (1994). "Short-term traffic flow prediction: neural network approach". *Transportation Research Record* 1453, 98–104.
- Ravishanker, N. and Ray, B.K. (1997). "Bayesian analysis of vector ARMA models using Gibbs sampling. *J. of Forecasting*, 16, 177-194.
- Roberts, G.O. and Sahu, S.K. (1997). "Updating schemes, correlation structure, blocking and parameterization of Gibbs sampler". *J. of Royal Society, Series B (methodological)*, 59 (2), 291-317.
- Smith, B. L. and Demetsky, M. J. (1997). "Traffic flow forecasting: comparison of modelling approaches". *J. of Transportation Eng.*, 123(4), 261–266.
- 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), 257–321.
- Tanner, M.A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions (Third Edition)*, Springer-Verlag, New York.
- Tebaldi, C., West, M. and Karr, A. F. (2002). "Statistical analysis of freeway traffic flows". *J. of Forecasting*, 21, 39-68.
- Van Arem, B., Kirby, H.R., Van Der Vlist, M.J.M. and Whittaker, J.C. (1997). "Recent advances and applications in the field of short-term traffic forecasting". *Int. J. of Forecasting*, 13, 1-12.

- Vlahogianni, E.I., Golias, J.C. and Karlaftis, M.G. (2004). "Short-term forecasting: Overview of objectives and methods". *Transport Reviews* 24 (5), 533-557.
- Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C. (2005). "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach". *Transportation Research Part C: Emerging Technologies*, 13(2), 211–234.
- Vythoulkas, P.C. (1993). "Alternative approaches to short-term traffic forecasting for use in driver information systems". *Transportation and Traffic Theory, Proceedings of the 12th International Symposium on Traffic Flow Theory and Transportation*, Berkley, C.F. Daganzo (Ed.).
- Williams, B.M., Durvasula, P.K., Brown, D.E. (1998). "Urban Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models". *Transportation Research Record* 1644, 132–144.
- Williams, B.M. (1999). "Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process". PhD dissertation. University of Virginia, Charlottesville, Va.
- Williams, B.M. and Hoel, L.A. (2003). "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results". *J. of Transportation Eng.*, 129(6), 664-672.
- Yin, H.B, Wong, S.C., Xu, J.M. and Wong, C.K. (2002). "Urban traffic flow prediction using a fuzzy-neural approach". *Transportation Research: Part C*, 10, 85-98.
- Zhang, C., Sun, S., Yu, G. (2004). "A Bayesian network approach to time series forecasting of short-term traffic flows", *Proceedings, the 7th IEEE International Conference on Intelligent Transportation Systems (ITSC2004)*, Washington, D.C., 216-221.

## LIST OF FIGURES:

1. Diagram of the junction TCS 183.
2. Time series plot of the traffic data.
3. Correlogram plot of the centered traffic data.
4. Simulation of values of  $\phi_I$
5. Simulation for the values of  $\Theta_1$  and  $\sigma^2$  generated from individual normal distributions respectively.
6. Simulation of values of  $\Theta_1$ .
7. Simulation of values of  $\sigma^2$ .
8. Predictive distributions for 25 future data points/forecasts.
9. (A) Bayesian and normal density (from classical estimate) plots for  $\phi_I$ .  
(B) Bayesian and normal density (from classical estimate) plots for  $\Theta_1$ .
10. Forecasts from Bayesian inference.
11. Forecasts from classical/frequentist analysis.
12. Variance from Bayesian and classical inference.



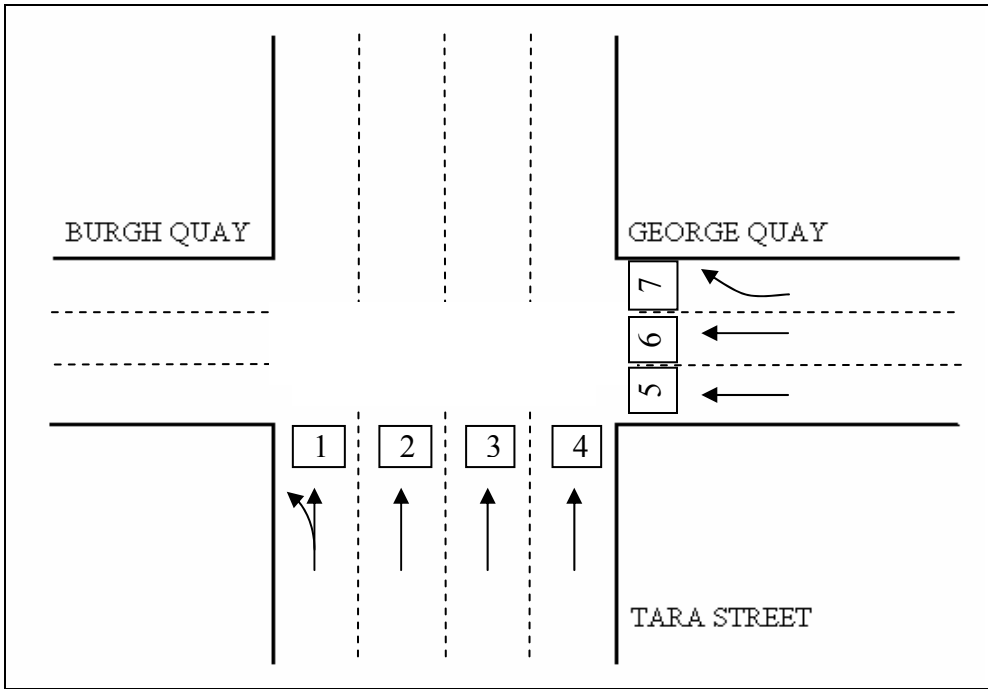


FIGURE 1

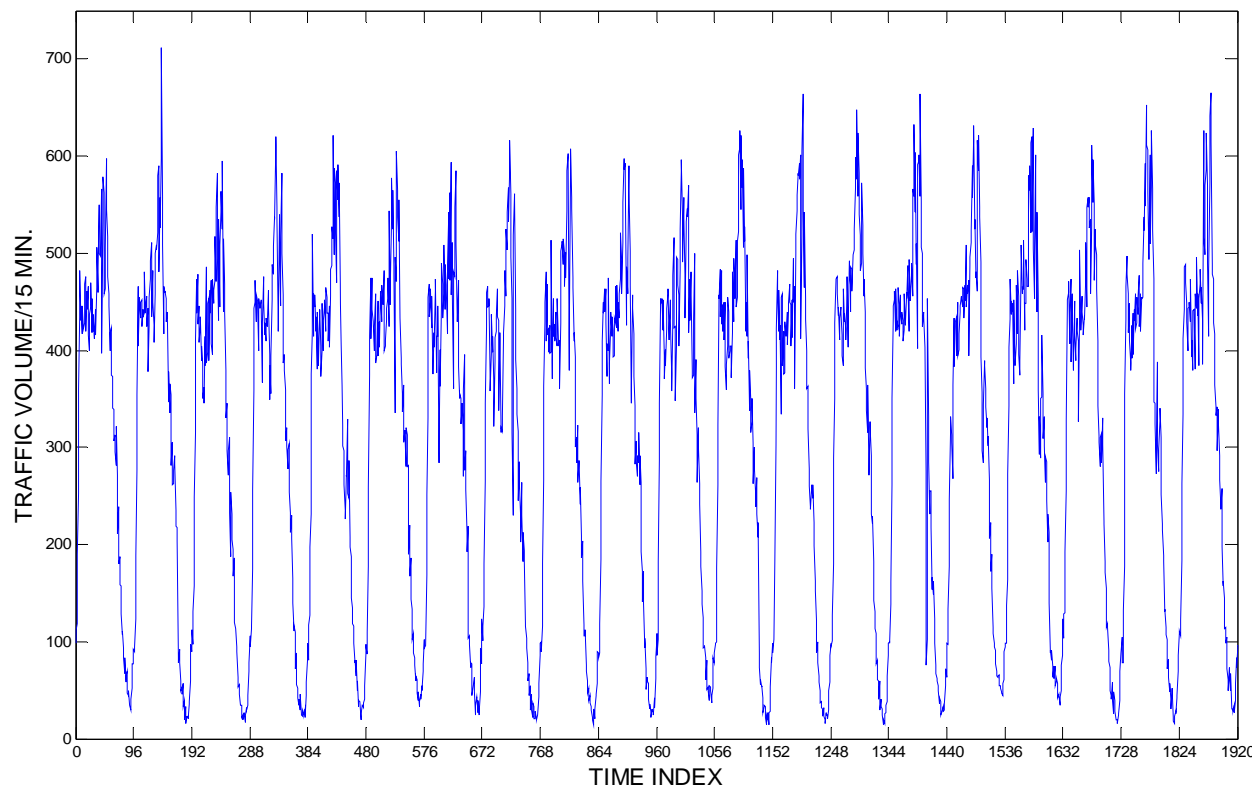


FIGURE 2

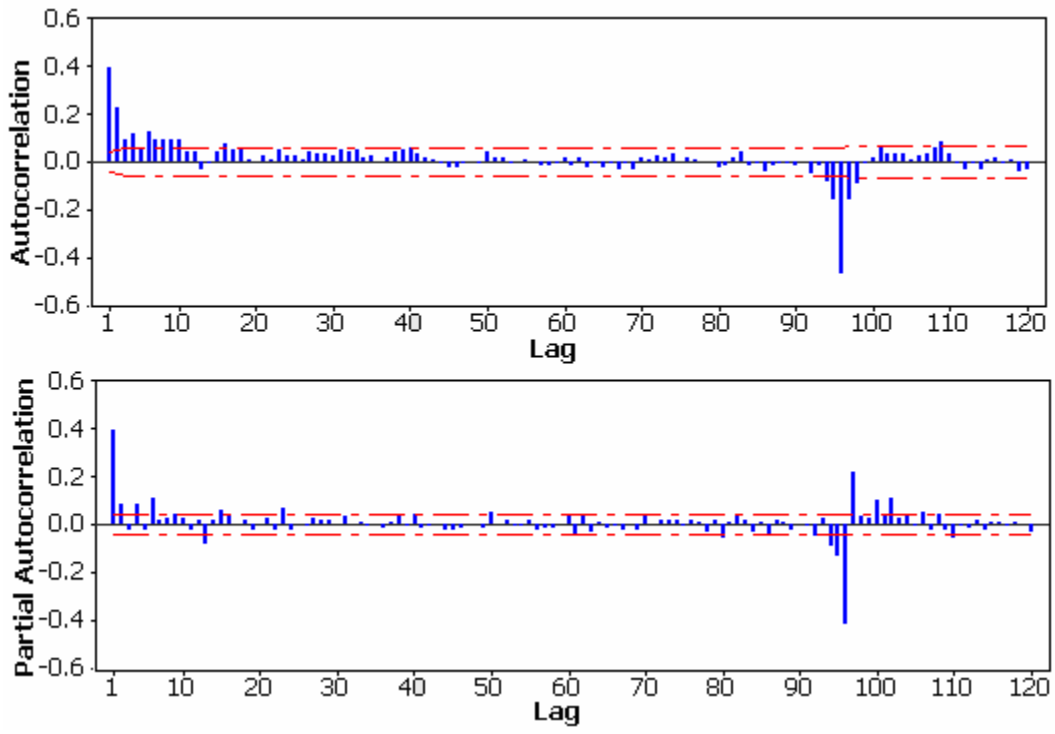


FIGURE 3

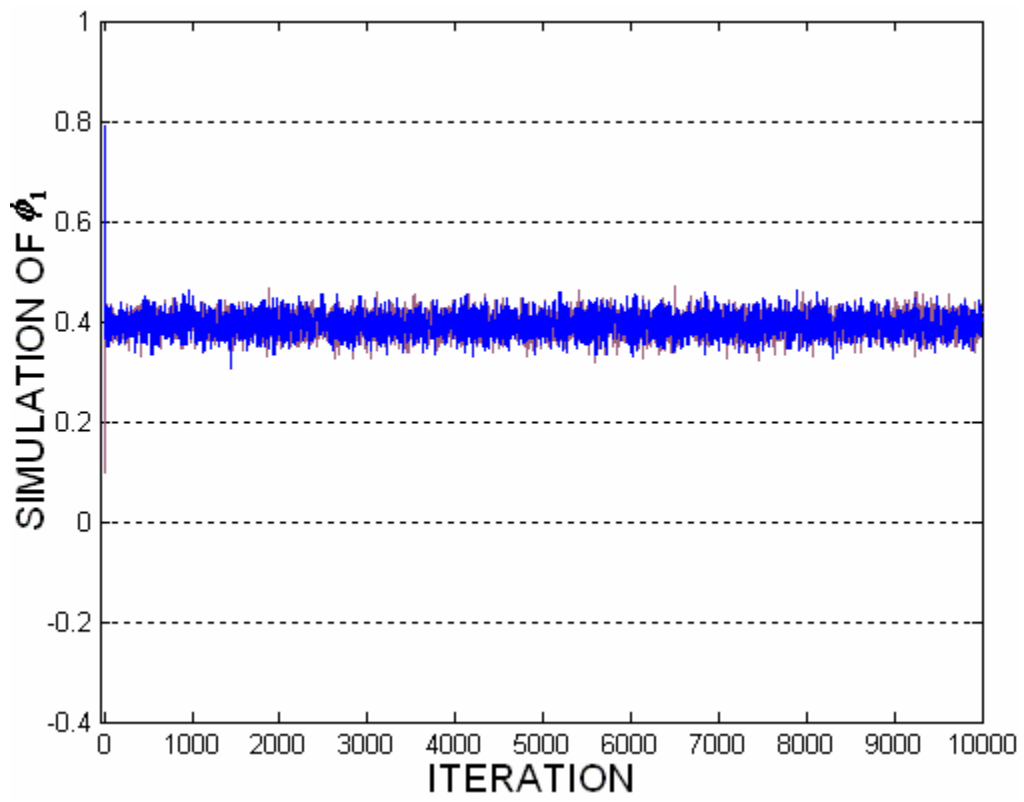


FIGURE 4

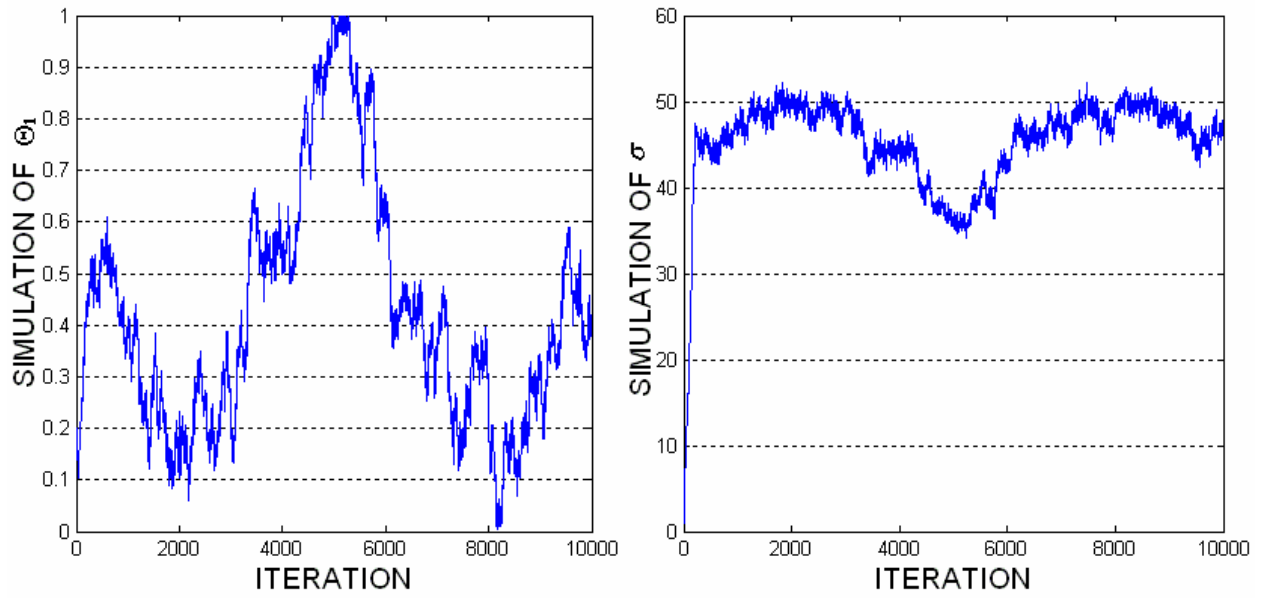


FIGURE 5

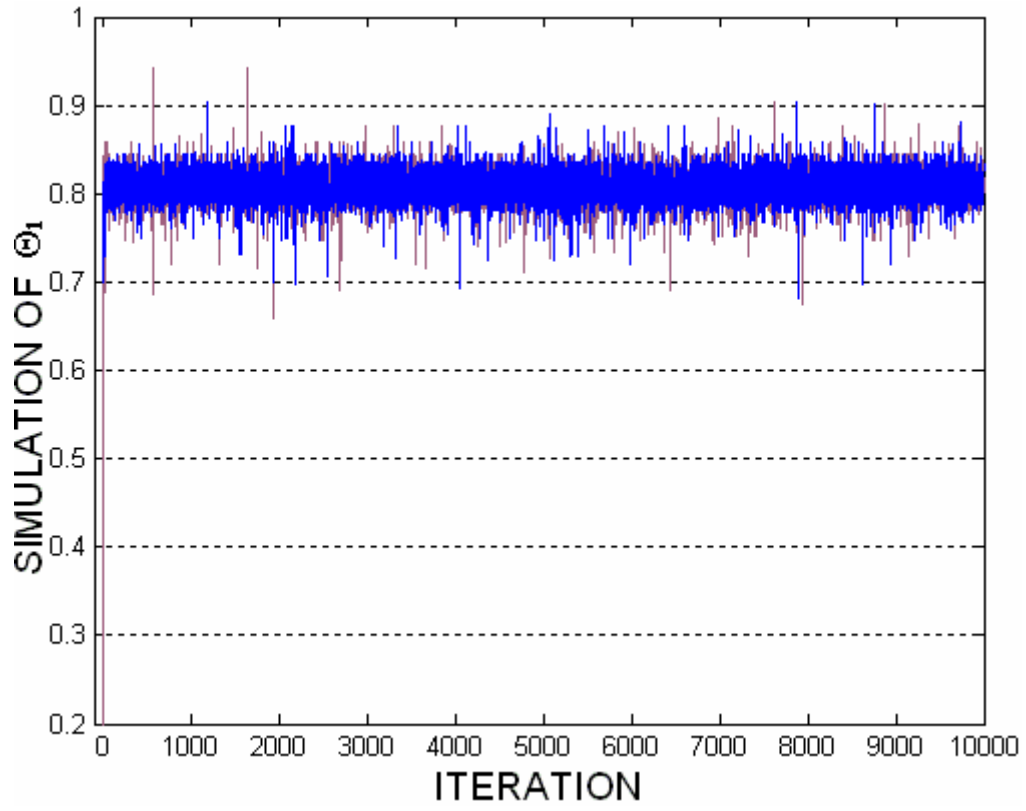


FIGURE 6

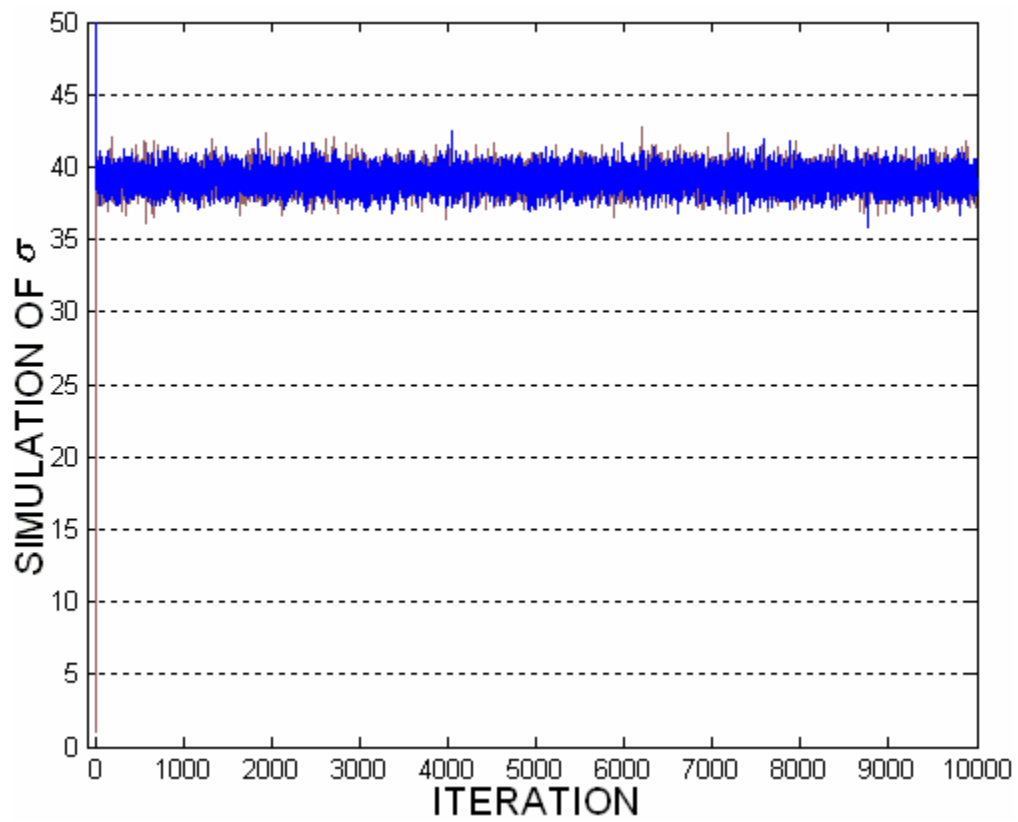


FIGURE 7

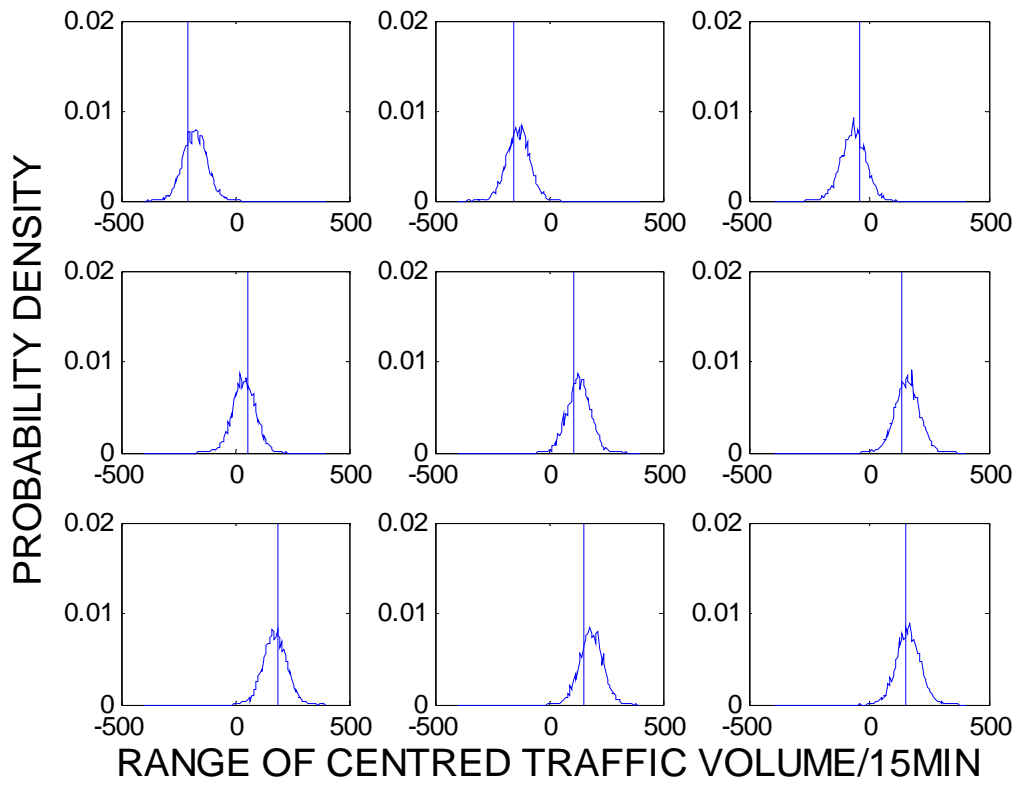


FIGURE 8



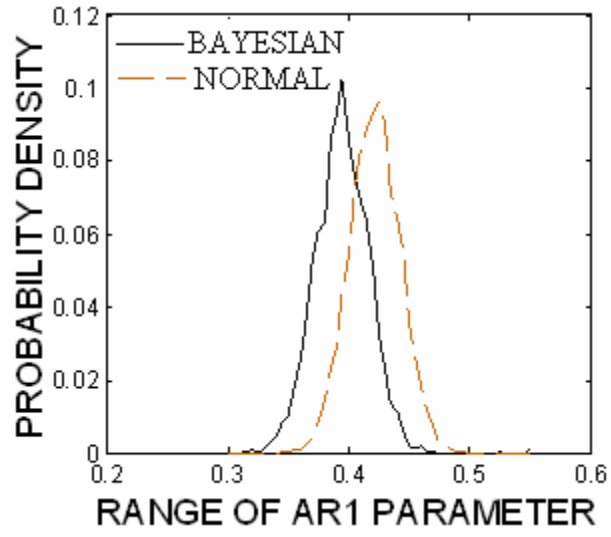


FIGURE 9(A)

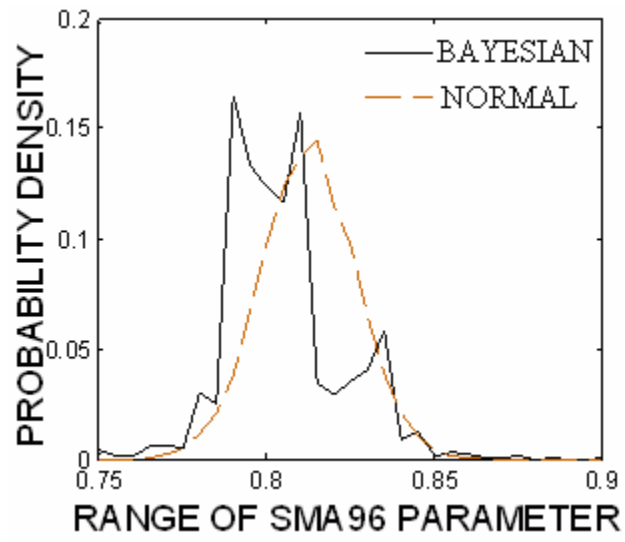


FIGURE 9(B)

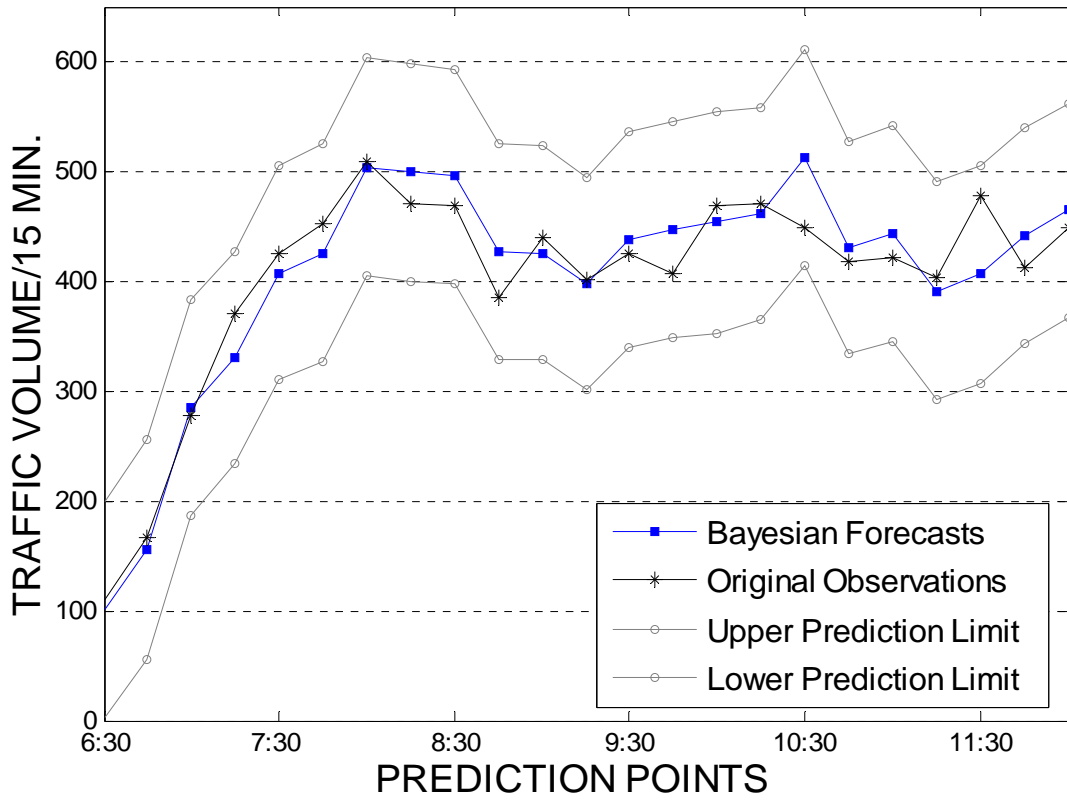


FIGURE 10

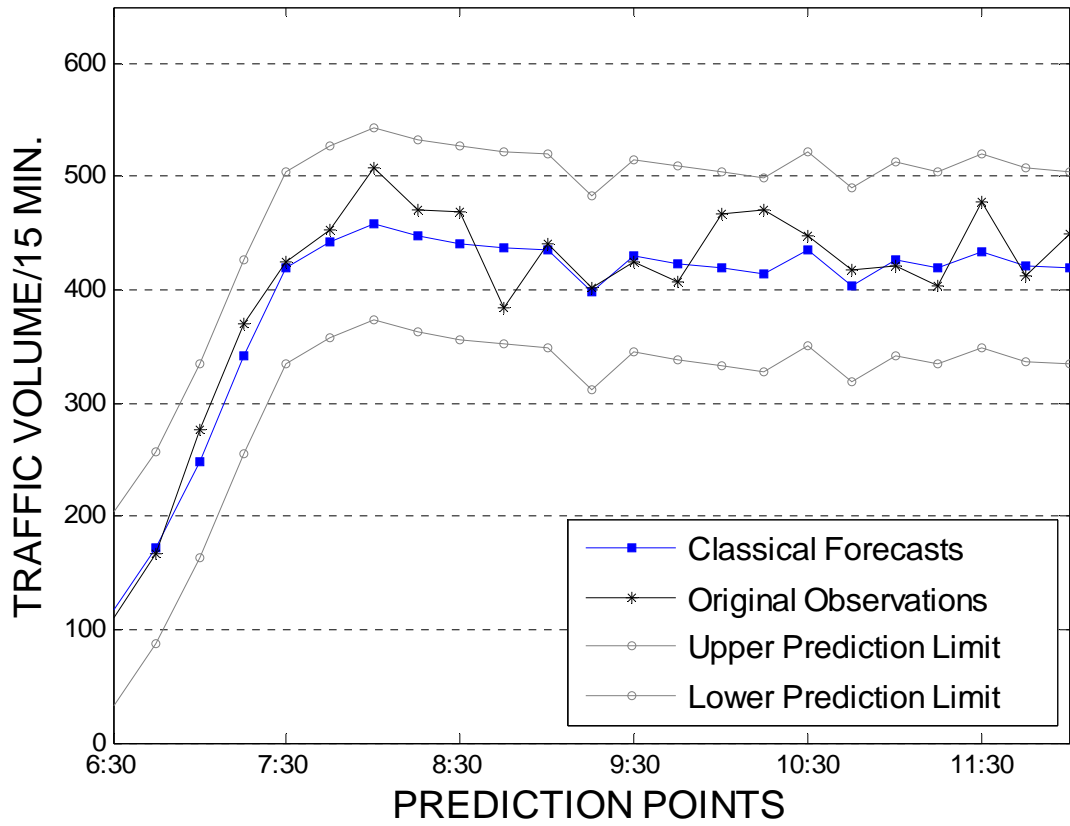


FIGURE 11

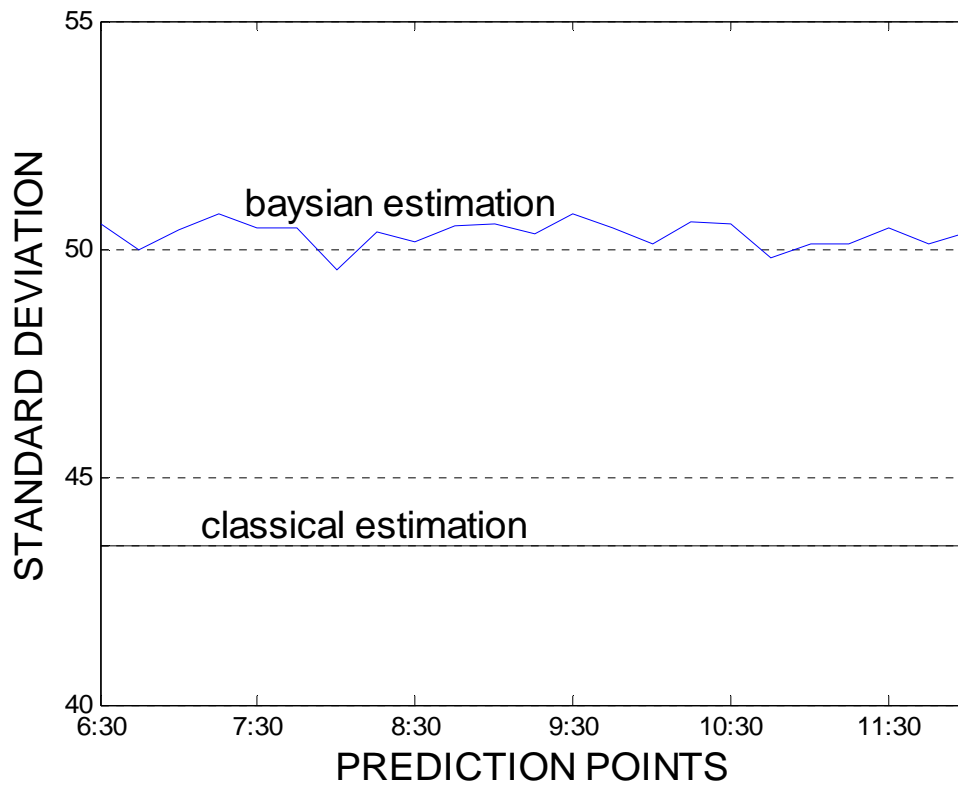


FIGURE 12