Spatiotemporal Prediction Using Hierarchical Bayesian Modeling

Taghreed Alghamdi*, Khalid Elgazzar*, Taysseer Sharaf§
*Ontario Tech University, Oshawa, ON, Canada
§University of Michigan-Dearborn, Dearborn, MI, USA
taghreed.alghamdi@ontariotechu.net, khalid.elgazzar@ontariotechu.ca, tsharaf@umich.edu

***Abstract*—Hierarchical Bayesian models (HBM) are powerful tools that can be used for spatiotemporal analysis. The hierarchy feature associated with Bayesian modeling enhances the accuracy and precision of spatiotemporal predictions. This paper leverages the hierarchy of Bayesian models using the Gaussian process to predict long-term traffic status in urban settings. The Gaussian process is used with different covariance matrices: exponential, Gaussian, spherical, and Matérn to capture the spatial correlation. Performance evaluation on traffic data shows that the exponential covariance yields the best precision in spatial analysis with the Gaussian process, while the Gaussian covariance outperforms the others in temporal forecasting.**

***Index Terms*—Uncertainty, hierarchical Bayesian, spatiotemporal, Gaussian processes.**

I. Introduction

Recently, traffic flow modeling has attracted great attention in intelligent transportation and traffic management fields, whereby modeling historical and recent traffic data can describe and predict traffic status. This can serve the traffic management sector to improve the transportation network infrastructures and support real-time decision making. The availability of enormous traffic data plays an essential role in developing highly efficient models for traffic flow estimation and prediction. In addition, recent statistical studies show the growing needs to control traffic congestion and reduce its negative impact, especially in urban settings where congestion can cause significant delays and accidents. INRIX in 2017 [1] estimated the total number of hours lost in peak travel times in Los Angeles as 102 hours per driver. In the same report, Boston ranked the highest in waiting time during driving with an average of 14% due to traffic congestion. Similarly, London and Paris have the highest traffic congestion in Europe, with drivers spending an average of 56 hours in traffic at peak travel times. These studies have also revealed that traffic congestion is a major concern in large cities due to the significant overload on public transportation services and limited road occupancy, especially on routes leading to busy business districts and city centers [2].

In order to better understand the roots of traffic congestion, a robust model must be developed. However, the study of traffic data modeling has encountered obstacles in recent years due to a number of factors such as data reliability, data classification, and data integrity. Therefore, these models suffer high inaccuracy compared to real-time data measurements. In addition, different approaches to traffic modeling have different structure complexity to target various criteria. Consequently, traffic modeling has been associated with complex processes in terms of implementation and data processing.

Modeling the traffic flow of massive traffic data requires comprehensive analysis and a robust methodology to estimate the uncertainty and detect the variance effects on various outcomes. This inspired many studies and researchers to develop methods that minimize uncertainty in predicted outcomes. Bayesian methods offer a statistical framework for computing the uncertainty using probability [3]. The principle of Bayesian theory is based on updating subjective beliefs in light of new data until the uncertainty of the hypothesis is minimal. As more data arrives, new beliefs are formed and the hypothesis becomes less uncertain.

Limitations and gaps in the traffic modeling studies encouraged extensive research to be carried out in order to reduce the uncertainty of the traffic flow prediction. Recent developments in the Bayesian modeling approach propose a hierarchical structure where the model is built in multiple levels. Each level is implemented through a number of iterations using the Markov Chain Monte Carlo (MCMC) algorithm in order to define the *prior*, the *joint likelihood of model parameters*, and the *joint posterior*. The three levels (aka. sub-models) as described in the literature build the HBM model's functions. Bakar and Sahu [4] developed a hierarchical Bayesian approach using three sub-models: data model, process model, and parameter model in its hierarchy. The results of their research support the space-time and air-pollution pair datasets to predict the daily 8-hour maximum ozone concentration. The performance evaluation of their study using GP with the Matérn covariance matrix shows high prediction accuracy with the used data. Utilizing their approach with different covariance matrices to predict the traffic flow data should make an essential contribution to the domain of spatiotemporal analysis. [4].

In this paper, we apply the HBM approach in the traffic domain and use the GP as the parameter model. The estimation of model parameters is carried out using Bayesian inference. To obtain an accurate spatial prediction and temporal forecasting, we test four different spatial correlation matrices: *exponential, Gaussian, spherical,* and *Matérn*. Constructing the temporal forecasting is based on two different units of time: *day* and *month*. We use a dataset collected by the Chicago Transit

Authority (CTA) about bus traffic counts to apply our model and conduct a spatiotemporal traffic prediction.

II. Literature Review

Traffic modeling has been an active area of research that has gained a lot of traction over the past few years due to advancements on different modeling techniques. A considerable amount of literature has been published on spatiotemporal modeling using Bayesian approach, however, these models are not performed on large space-time data [5]. Also, most of the literature on hierarchical Bayesian models is centered around the usage of this approach in specific applications such as environment, healthcare, and finance [6], [7]. Chen et al. [8] used S-Kriging for short-term traffic speed prediction on freeways excluding the temporal component. Lu et al. [9] also used S-Kriging as a form of Bayesian inference to study the concentration of PM10 in air at different neighboring locations and perform spatial prediction.

Much of the available literature on methods that are Bayesian-based have been developed to accurately predict the values from joint space-time data such as Gaussian process regression methods [10]. However, the latter method estimates model parameters by approximating the posterior distribution rather than using Markov Chain Monte Carlo sampling. Methods based on approximating the posterior provides a poor estimation of model parameters. The computational advantages of using Markov Chain Monte Carlo in Bayesian inference appears in the estimation of the likelihood, especially when having many integrals [11].

Some authors have been mainly interested in Non-Bayesian models which are not process-based such as the Generalized Additive Models [12]. These models are implemented based on a dynamic relationship between the value and the spatial points coordinates, yet, these models are not process-based which means that these models do not integrate random collection of points of spatiotemporal processes. Consequently, it impacts the spatiotemporal autocorrelation.

Brogan et al. [13] apply hierarchical Bayesian models to study engine noise and estimate the noise characteristics and the separation filters. The thesis of this study indicates that using Gibbs sampler with their given prior in their study about engine cycles led to a longer computational time to verify the hypotheses they used, thus, they conclude that the influence of the observed prior on the MCMC algorithm is significant.

Bakar et al. [4] propose a spatiotemporal model to predict the ozone concentration level in New York City. The model is implemented in their spTimer R package to use one of the three models; GP, autoregressive (AR), or GPP. The Gibbs sampler is used to estimate the likelihood functions. We employ the same framework in the traffic domain to better understand the traffic behavior using hierarchical Bayesian models. Although our study did not apply all three different models on the traffic data, it applies four different covariance matrices using the GP model. The findings of the performance of the covariance matrices conclude that covariance matrices perform differently based on different characteristics of the dataset.

III. Methodology and Data

We can summarize the fundamental concept of Bayesian theory into three keywords: *prior*, *likelihood*, and *posterior*. The prior is an initial belief to begin with based on the current information and can be updated when new information arrives. The likelihood is the joint distribution of the data given the model parameters β, and σ. Model parameters can be found after updating the prior. Lastly, the posterior is the conditional probability distribution of our dependent variable θ which depends on and data and the prior. The posterior is computed as the product of the prior and likelihood [11].

Generally speaking, the Bayesian inference can be performed as follows: (1) define the prior empirical probability distribution or the assumptions for the hypothesis; (2) compute the marginal likelihood probability of the data using a sampler (e.g., MCMC) to generate random samples from the probability distribution [14]. In each sample, we calculate the marginal likelihood probability, which contains all the relevant information to evaluate the evidence. However, estimating the marginal likelihood typically is a difficult task because we have to integrate all model parameters; (3) determine the posterior, which is the probability distribution of a particular value of the parameter after having seen the whole data [15]. The Bayesian inference theory is formally expressed as:

$$posterior \propto \ prior \ \times \ likelihood$$

A. Hierarchical Bayesian Modeling

The hierarchical Bayesian model can be structured as three levels of probabilistic models [4]: the data model, the process model, and the parameters model. These three levels (or stages) can be represented as follows:

$$First\ level\ [data \mid process, parameter_{data}]$$
$$Second\ level\ [\ process \mid parameter_{process}]$$
$$Third\ level\ [\ parameter_{data},\ parameter_{process}]$$

In the first level, we obtain the data model according to a certain process $Y_{(s_{ij};t)}$ and some errors $\epsilon_{l(s_{ij};t)}$ that are assumed to be independently normally-distributed ($\epsilon_i \sim N(0, \sigma^2)$). The data model is described as shown below:

$$Z_{(s_{ij};t)} =_{(s_{ij};t)} + \epsilon_l(s_{ij};t) \tag{1}$$

The process model in Equation 2, captures the relationship of the underlying nature expressed by the data. The process model can be one of three; Gaussian process GP, Auto-Regressive AR, or Gaussian Predictive Process GPP. In Equation 2, the process $Y_{l(s_{ij};t)}$, is expressed by a Gaussian process $\mu_{(s_{ij};t)}$ plus some errors $\eta_{(s_{ij};t)}$.

$$Y_{(s_{ij};t)} = \mu_{(s_{ij};t)} + \eta_{(s_{ij};t)} \tag{2}$$

The third level of the hierarchical modeling defines the model parameters. According to Equations 1 and 2 , these parameters are: the variance of $\epsilon_{l(s_{ij};t)}$, the variance of $\eta_{(s_{ij};t)}$,

the coefficients of the GP, and ϕ which defines the spatial correlation.

Generically, Equation 3 represents the structure of estimating the model parameters of the hierarchical model using Bayesian inference. Starting with the Gaussian process, $Y_{(s_{ij};t)}$ which follows a normal distribution conditioning on the model parameters θ_i. While θ_i is a vector containing σ_η, β_{GP}, and ϕ. The prior distribution of θ_i conditioning on ϕ where ϕ will have a prior distribution that follows a gamma or uniform distribution.

$$\begin{aligned} y_i &\sim p(y|\theta_i) \\ \theta_i &\sim p(\theta|\phi) \\ \phi &\sim p(\phi) \end{aligned} \tag{3}$$

$$p(\theta, \phi|y) \propto p(y|\theta, \phi) \tag{4}$$

From Equation 4, the posterior distribution is proportional to the likelihood being conditional on both θ, and ϕ multiplied by the prior for θ and ϕ. Based on the conditional independence rules and knowing that our data is independent of ϕ, if we know θ, we can take the joint distribution and break it down into conditional distribution $p(\theta|\phi)$ and multiply it by the distribution of $p(\phi)$. Equation 5 formalizes these steps, which is derived from Equation 4.

$$p(\theta, \phi|y) \propto p(y|\theta, \phi) = p(y|\theta)p(\theta|\phi)p(\phi) \tag{5}$$

1) Gaussian Processes Model: The GP includes the temporal effect as well as the spatial effect denoted in Equation 6 which captures the space-time relationship. [16].

$$[Y(s,t) : s \in D_s,\ t \in D_t] \tag{6}$$

Where Y is the value of the traffic flow at location s in time t, and D_s is a vector of spatial coordinates (s_i, s_j), and $i, j = 0, \ldots, m$, where m is the total number of locations. For the temporal component, we have two time components denoted by l and t that represent the short time component, and the long time component, respectively. In our dataset we only use the day and month at the observed spatial points. The dataset has 28 locations, so to decompose Y in Equation 6 where the spatial process at fixed time t_0 and the temporal process at fixed spatial point s_0 denoted by Equation 7, and Equation 8, respectively [16]:

$$Y_{t_0} = (Y_{t_0}(s_0), \ldots, Y_{t_0}(s_0 + 28\Delta))' \tag{7}$$

$$Y_{s_0} = (Y_{t_0}(s_0), \ldots, Y_{t_0+28}(s_0)' \tag{8}$$

And by combining Y_{t_0} and Y_{s_0} as follows:

$$\begin{aligned} Y_{t_0(s_i)}|Y_{t_0(s_j)} : j \neq i \sim Gau((\phi_{t_0}/(1+\phi_{t_0}^2))Y_{t_0(s_{ij}-1)} \\ +Y_{t_0(s_{ij}+1)}, \sigma_{t_0}^2/(1+\phi_{t_0}^2)) \end{aligned} \tag{9}$$

Where the dimensional distributions are determined by the mean function $\mu(s,t)$, and the covariance matrices $cov(Y(s,t), Y((s,t)\prime)$ for all spatial points $s \in D$. Since we are using the GP to build the HBM, we define the hierarchy of the GP in Equation 10:

$$Y_{l(s_{ij};t)} = f(x_l)_{(s_{ij};t)} + \epsilon_{l(s_{ij};t)} \tag{10}$$

The long time unit is denoted by L, where $l = 1, ..., L$ and the short time unit is denoted by Tl, where $t = 1, ..., Tl$.

Our dataset is represented by $Y_{l(s_{ij};t)}$, while $\epsilon_{l(s_{ij};t)}$ is a random error that we assume to be independently normally-distributed that follows $\epsilon_i \sim N(0, \sigma^2)$, and we can breakdown Equation 2 as follows:

$$\eta_{(s_{ij};t_0)} = (\eta_{(s_{11};t_1)}, \ldots, \eta_{(s_m;t_l)})^{\top\prime} \tag{11}$$

$$\mu_{(s_{ij};t_0)} = (\mu_{(s_{11};t_1)}, \ldots, \mu_{(s_m;t_l)})^{\top\prime} \tag{12}$$

The mean $\mu_{(s_{ij};t_0)}$ can be represented by $\chi\beta$, where β represents the vector of regression coefficients, and χ represents the matrix of the covariates between time and space. Thus, Equation 2 can be written as:

$$f(x_l)_{(s_{ij};t)} = \chi_{l(si;t)}\beta + \eta_{(s_{ij};t)} \tag{13}$$

Different covariance matrices show significant positive results on the prediction outcomes where estimating the correlation for the space-time effect on a specific observed value is a major step in fitting the model. In the following, we briefly describe these four covariance matrices.

Covariance Matrices: In GP, the spatial correlation parameter is calculated by applying one of the four covariance matrices: exponential, Gaussian, spherical and Matérn. We refer to the covariance function by S_η which includes three parameters: ϕ, ν and the distance between two spatial points s_i and s_j that will be calculated as $\parallel s_i - s_j \parallel$.

$$S_\eta = \phi + \nu + coordinates\ (s_i - s_j) \tag{14}$$

When the distance between s_i and s_j increases, their correlation level decays, and we refer to this by the parameter alpha where it dominates the rate of the correlation of s_i and s_j locations; ν is the smoothness parameter that softens the fitted curve of the model. The spTimer package uses *exponential* as the default covariance matrix. The decay of the correlation function is calculated as:

$$Cov_E(s_i, s_j; \phi) = exp - (2\sqrt{\nu} \parallel s_i - s_j \parallel \phi) \tag{15}$$

Where ϕ and $\nu > 0$. Similarly, in the Gaussian covariance matrix, the square of the exponential covariance matrix is calculated as:

$$Cov_G(s_i, s_j; \phi) = exp - (2\sqrt{\nu} \parallel s_i - s_j \parallel \phi)^2 \tag{16}$$

The spherical covariance matrix takes in consideration the range "distance" over pairs of spatial points. The covariance vanishes when the distance between s_i and s_j is zero [17].

$$\begin{aligned} Cov_S(s_i, s_j; \phi) = 1 - 1.5 \times (2\sqrt{\nu} \parallel s_i - s_j \parallel \phi) \\ +0.5(2\sqrt{\nu} \parallel s_i - s_j \parallel \phi)^3 \end{aligned} \tag{17}$$

The Matérn covariance matrix includes the modified Bessel functions of the second kind that is sometimes called the Basset functions, and it is given by Equation 18:

$$Cov_M(s_i, s_j; \phi, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (2\sqrt{\nu} \parallel s_i - s_j \parallel \phi)^{\nu} K_{\nu}(2\sqrt{\nu} \parallel s_i - s_j \parallel \phi) \quad (18)$$

Gibbs Sampler: The Gibbs sampler is an MCMC algorithm that generates a sequence of observations (samples) from a specific multivariate distribution of the hierarchical model parameters.

Algorithm 1 Gibbs Sampler

initialize $\mathrm{y}^{(0)} \sim q(y)$
for *iteration m= 1,2,3,...* **do**
  $\mathrm{y}_1^m \sim p(Y_1 = y_1 | Y_2 = y_2^{m-1}), Y_3 = y_3^{m-1}$
  $y_2^m \sim p(Y_2 = y_2 | Y_1 = y_1^m), Y_3 = y_3^{m-1}$
  $\vdots$
  $y_D^m \sim p(Y_D = y_D | Y_1 = y_1^m), Y_D = y_{D-1}^{m-1}$
  end

Gibbs sampling allows us to examine each variable and calculate its conditional distribution. For example, assume the random variables $Y_1, Y_2, \ldots, Y_n$. The value for each random variable $y_1, y_2, \ldots, y_n$ is initialized from the prior distribution. In each iteration m, the sampler produces the samples of $y_1, y_2, \ldots, y_n$ as follows:

$$y_1^m \sim p(Y_1 = y_1 | Y_2 = y_2^{m-1}), Y_3 = y_3^{m-1} \quad (19)$$

$$y_2^m \sim p(Y_2 = y_2 | Y_1 = y_1^m), Y_3 = y_3^{m-1} \quad (20)$$

$$y_3^m \sim p(Y_3 = y_3 | Y_1 = y_1^m), Y_2 = y_2^m \quad (21)$$

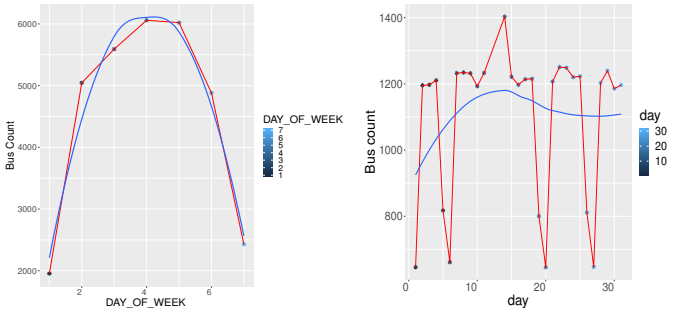
The Gibbs sampler stops when all generated sampling values have the same distribution size. Algorithm 1 provides the procedure that the Gibbs sampler uses to generate the samples. Samples are generated by examining each random variable one at a time and obtaining samples from the conditional distributions of each variable. A sequence of pairs of random variables is generated like: $(Y_1, y_1), (Y_2, y_2), (Y_3, y_3)$.

IV. Spatiotemporal GP Prediction

A. Study Area and Data Preprocessing

The data we use in this study is collected from the Chicago Transit Authority (CTA). Public transit plays an essential role in the development of large cities and are considered an economical way of transportation. However, this mode of transportation is often involved in traffic congestion [18]. Understanding the public transit traffic data will help improving public transport services and generally enhances road traffic management.

The study area includes 29 sensors distributed mostly in downtown Chicago. The data includes bus count, speed, and



(a) Bus count in the weekdays. (b) Bus count in a month.

Figure 1: The bus traffic data exploration plots.

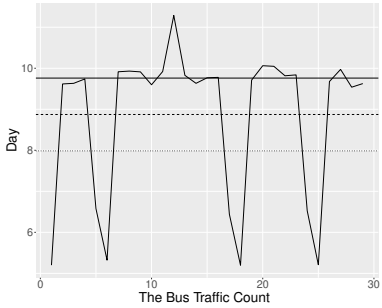


Figure 2: The average volume of the bus traffic count per day.

the number of sensor readings at a 10-minutes time interval from August 2018 until December 2019.

A quick exploration of the data reveals that the bus count increases during the weekdays and in a month as shown in Figure 1 (a) and (b). Figure 2 provides more visual analysis of the bus count data and presents the aggregate number of buses for each day. It shows three horizontal lines where the dash line refers to the mean, solid line represents 10% higher than mean, and dotted line indicates 10% less than the mean. The reason we are showing this visual analysis is to illustrate the variations in traffic patterns over days, which does not exhibit normal distribution. Unlike most of the other spatiotemporal analysis techniques, HBM can efficiently deal with data that does not have a normal distribution [19].

Based on our visual analysis, we empirically determined that the minimum distance between sensors should be 0.1 km to capture the spatial correlation. The approach doesn't perform accurately with short-time units "hour" due to implementation limitations in the spTimer package. That's why we opted to use the daily aggregate data. We train the model on 21 locations for the time period starting from 2019-01-01 to 2019-01-29. Testing the model is conducted on the 8 locations that will be defined in the Gibbs sampler.

V. Results and Discussion

We observe that the GP approach achieves an accurate prediction with the bus traffic dataset. A number of performance criteria are used to measure the model accuracy including the mean absolute error(MAE), root mean squared error (RMSE),

Table I: Spatial prediction error.

Prediction error	Matérn	spherical	exponential	Gaussian
MAE	7.6839	8.1898	7.6723	41.8726
RMSE	9.1833	9.7754	9.1444	53.6856
MAPE	37.2477	41.5263	37.1951	201.6881

Table II: Temporal forecasting error.

Forecasting error	Matérn	spherical	exponential	Gaussian
MAE	11.4243	11.4738	11.2184	15.2290
RMSE	13.1387	13.6285	12.8379	19.1728
MAPE	44.5595	45.0353	44.0292	73.0119

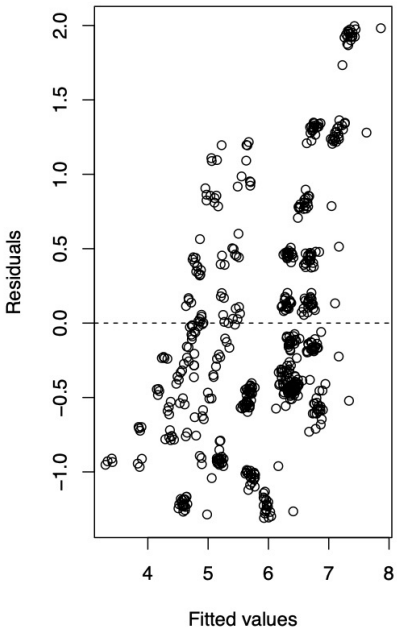


Figure 3: The residuals versus the estimated responses.

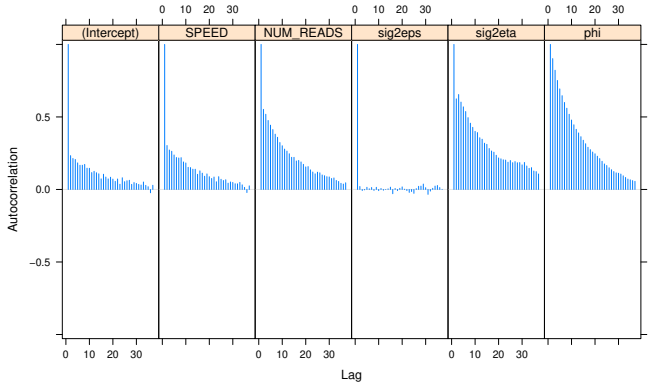


Figure 4: The auto-correlation coefficient estimation using one MCMC chain.

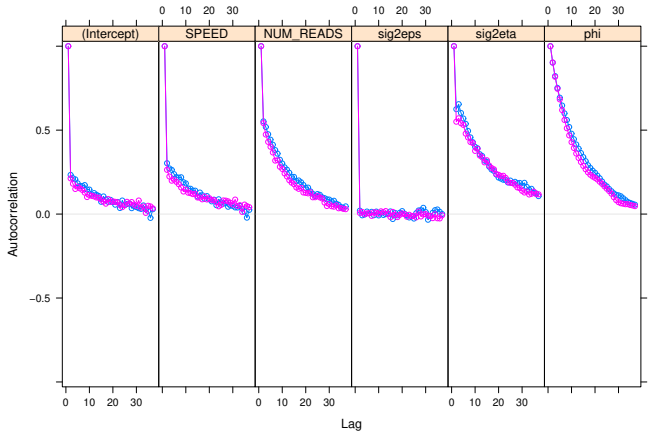


Figure 5: The auto-correlation coefficient estimation using one MCMC chain vs. multiple MCMC chains.

and mean absolute percentage error(MAPE). We compare the MAE, RMSE, and MAPE outputs of the GP for the spatial prediction with the exponential, Gaussian, spherical and Matérn covariance matrices as shown in Table I. The exponential covariance provides the best performance with the GP in the spatial prediction and is slightly better than the spherical and Matérn. The Gaussian provides the lowest spatial prediction accuracy out of the four covariance matrices.

Table II provides the accuracy error of the temporal forecasting for the different covariance matrices. Interestingly, we observe similarities in the accuracy error between the exponential, spherical and Matérn. The Gaussian provides a better accuracy in the temporal forecasting than it does in the spatial prediction.

In Figure 3, we see the relationship between the residuals and the estimated responses that present the predicted response using the exponential covariance matrix. It appears that residuals roughly form around the zero line. Also, most of the predicted responses fall on the estimated regression line. These points may explain the relatively good correlation between residuals and fits.

Figure 4 presents the correlograms of the auto-correlation coefficient function at different lags where we can see a significant level of correlation between different lags. The ACF plot shows that the current value is constantly determined by the previous values. This means that the daily values are dependent on each other.

We assess the MCMC performance by applying three different diagnostics: Geweke's convergence, Gelman and Rubin, and auto-correlation. Gelman and Rubin's diagnostic requires multiple MCMC chains run in parallel to compare the auto-correlation coefficient of the multiple MCMC chains as shown in Figure 5. The estimated variance of each parameter within the MCMC chain is very small, which indicates that the MCMC chain has converged. The statistical results of the Gelman and Rubin's diagnostic match the results in Figure 5 where the convergence diagnostic is less than 1.1, which also means that the chains are converged.

Figure 6 shows the Gelman and Rubin plot which support the observed results by showing the growth of the Gelman and Rubin scale-reduction factor for each parameter as the number of iterations increases in a chain of 5000 samples.

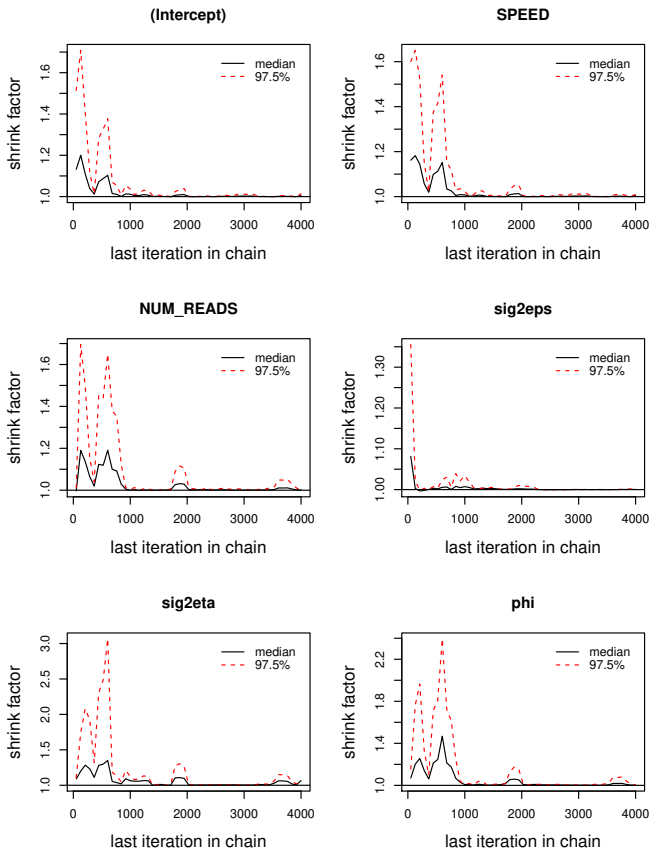


Figure 6: The Gelman and Rubin's diagnostic for a chain of 5000 samples.

VI. CONCLUSION

This study applies hierarchical Bayesian modeling using the Gaussian process to predict traffic status. The Gaussian process does not apply on data with missing values, however, the normality of the data distribution can be improved by using the transformation log and the square-root using the MCMC algorithm. We use the Gibbs sampler to obtain the samples from the traffic data and use these samples to build the spatial prediction and temporal forecasting. Different covariance matrices are used with the Gibbs sampler including exponential, Gaussian, spherical and Matérn. The results show that the exponential, spherical and Matérn provide a higher accuracy compared to the Gaussian covariance matrix. Results also confirm that HBM can be used effectively in spatiotemporal analysis and yields high prediction accuracy.

REFERENCES

[1] The Economist, "The hidden cost of congestion," Feb 2018. [Online]. Available: https://www.economist.com/graphic-detail/2018/02/28/the-hidden-cost-of-congestion

[2] A. Bull and N. CEPAL, *Traffic Congestion: The Problem and how to Deal with it*. ECLAC, 2003.

[3] A. E. McGlothlin and K. Viele, "Bayesian hierarchical models," *Jama*, vol. 320, no. 22, pp. 2365–2366, 2018.

[4] K. S. Bakar and S. K. Sahu, "sptimer: Spatio-temporal bayesian modelling using R," *Journal of Statistical Software*, vol. 63, no. 15, pp. 1–32, 2015.

[5] S. K. Sahu, A. E. Gelfand, and D. M. Holland, "Spatio-temporal modeling of fine particulate matter," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 11, no. 1, pp. 61–86, 2006.

[6] A. M. Zaslavsky, "Using hierarchical models to attribute sources of variation in consumer assessments of health care," *Statistics in medicine*, vol. 26, no. 8, pp. 1885–1900, 2007.

[7] C. K. Wikle, "Hierarchical models in environmental science," *International Statistical Review*, vol. 71, no. 2, pp. 181–199, 2003.

[8] X. Chen, X. He, C. Xiong, Z. Zhu, and L. Zhang, "A bayesian stochastic kriging optimization model dealing with heteroscedastic simulation noise for freeway traffic management," *Transportation Science*, vol. 53, no. 2, pp. 545–565, 2018.

[9] W.-Z. Lu and Y. Xue, "Prediction of pm10 concentrations at urban traffic intersections using semi-empirical box modelling with instantaneous velocity and acceleration," *Atmospheric Environment*, vol. 43, no. 40, pp. 6336–6342, 2009.

[10] T. Mukhopadhyay, S. Chakraborty, S. Dey, S. Adhikari, and R. Chowdhury, "A critical assessment of kriging model variants for high-fidelity uncertainty quantification in dynamics of composite shells," *Archives of Computational Methods in Engineering*, vol. 24, no. 3, pp. 495–518, 2017.

[11] N. Friedman and D. Koller, "Being bayesian about network structure. a bayesian approach to structure discovery in bayesian networks," *Machine learning*, vol. 50, no. 1-2, pp. 95–125, 2003.

[12] S. N. Wood, *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2017.

[13] G. Brogna, J. Antoni, Q. Leclere, and O. Sauvage, "Engine noise separation through gibbs sampling in a hierarchical bayesian model," *Mechanical Systems and Signal Processing*, vol. 128, pp. 405–428, 2019.

[14] R. Paap, "What are the advantages of mcmc based inference in latent variable models?" *Statistica Neerlandica*, vol. 56, no. 1, pp. 2–22, 2002.

[15] F. Lindgren and H. Rue, "Bayesian spatial modelling with r-inla," *Journal of Statistical Software*, vol. 63, no. 19, pp. 1–25, 2015.

[16] R. Haining, "Statistics for spatio-temporal data by noel cressie and christopher k. wikle. hoboken," 2011.

[17] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[18] D. Q. Nguyen-Phuoc, G. Currie, C. De Gruyter, I. Kim, and W. Young, "Modelling the net traffic congestion impact of bus operations in melbourne," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 1–12, 2018.

[19] J. Parrondo, R. Zurita, J. A. Corrales, and J. Fernandez, "Numerical prediction of the effect of traffic lights on the vehicle noise at urban street intersections," *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3924–3924, 2008.