Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://amstat.tandfonline.com/loi/utch20

Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand

Xiaochen Xian, Honghan Ye, Xin Wang & Kaibo Liu

To cite this article: Xiaochen Xian, Honghan Ye, Xin Wang & Kaibo Liu (2020): Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand, Technometrics, DOI: 10.1080/00401706.2019.1704887

To link to this article: https://doi.org/10.1080/00401706.2019.1704887

TECHNOMETRICS
2020, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/00401706.2019.1704887

Taylor & Francis
Taylor & Francis Group

Spatiotemporal Modeling and Real-Time Prediction of Origin-Destination Traffic Demand

Xiaochen Xian[a], Honghan Ye[b], Xin Wang[b,c], and Kaibo Liu[b]

[a]Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL; [b]Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI; [c]Grainger Institute for Engineering, University of Wisconsin-Madison, Madison, WI

ABSTRACT

Traffic demand prediction has been a crucial problem for the planning, scheduling, and optimization in transportation management. The prediction of traffic demand counts for origin-destination (OD) pairs has been considered challenging due to the high variability and complicated spatiotemporal correlations in the data. Though several articles have considered estimating traffic flows from counts observed at specific locations, existing traffic prediction models seldom dealt with spatiotemporal demand count data of certain OD pairs, or they failed to effectively consider domain knowledge of the traffic network to enhance the prediction accuracy of traffic demand. To tackle the aforementioned challenges, we formulate and propose a multivariate Poisson log-normal model with specific parameterization tailored to the traffic demand problem, which captures the spatiotemporal correlations of the traffic demand across different routes and epochs, and automatically clusters the routes based on the demand correlations. The model is further estimated using an expectation-maximization algorithm and applied for predicting future demand counts at the subsequent epochs. The estimation and prediction procedures incorporate Markov chain Monte Carlo sampling to overcome the computational challenges. Simulations as well as a real application on a New York yellow taxi data are performed to demonstrate the applicability and effectiveness of the proposed method. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received October 2018
Accepted November 2019

KEYWORDS
MCMC sampling;
Origin-destination pairs;
Real-time prediction;
Spatiotemporal correlation;
Traffic demand

1. Introduction

In the past decades, intelligent transportation system (ITS) has brought advanced technology that enables a data-rich environment and many opportunities for traffic research. Collected from a large number of auxiliary instruments (e.g., cameras, inductive-loop detectors, Global Positioning System (GPS)-based receivers, and microwave detectors), a massive quantity of data can be acquired to provide useful information and generate new functions and services (Zhang et al. 2011). With massive traffic data readily available, it offers unprecedented opportunities for traffic prediction, which is considered as one of the most prevalent issues facing ITS (Li et al. 2015) and provides crucial inputs for the planning, scheduling, and optimization in transportation management.

In this article, we are interested in the online prediction of the origin-destination (OD) demand count in traffic networks, which represents the number of trips between certain combinations of an origin and a destination. Indeed, the study of OD demand prediction based on count data has a growing impact on many traffic control and management policies (Ashok 1996; Ashok and Ben-Akiva 2002; Li 2005; Hazelton 2008; Shao et al. 2014). For example, dynamic OD demand prediction is critical in planning for the charging services of the electrical vehicles (EVs; Zhang, Kang, and Kwon 2017). As an important type of new-emerging clean energy vehicles, a well-designed charging facility network is necessary to extend the vehicle range and popularize the use of EVs. In particular, the dynamic demand between nodes of the traffic network plays a key role in determining the availability of the charging facilities, planning the multi-period charging schedules, and meeting the customer needs at the maximum extent (Brandstatter, Kahr, and Leitner 2017; Zhang, Kang, and Kwon 2017). In another example, OD taxi demand prediction has been shown very important to help dynamically allocate resources to meet travel demand and to reduce empty taxis on streets which waste energy and worsen the traffic congestion (Yao et al. 2018). In addition, the online OD demand prediction is especially desired in the car-sharing system for making optimal dispatch decisions (Agatz et al. 2011).

There are numerous studies concerning traffic prediction, which can be categorized in the following two types. The first category considers predicting some traffic-related values for a specific location or independently for multiple locations, by applying temporal prediction techniques such as Kalman filter (Okutani and Stephanedes 1984), regression (Liu et al. 2017), time series (Ishak and Al-Deek 2002; Moreira-Matias et al. 2013), functional data analysis (Zhou and Mahmassani 2007; Chen and Müller 2014; Wagner-Muns et al. 2018), or incorporating new data sources (Ashok 1996; Carrese et al. 2017). The second category predicts traffic-related quantities at numerous

CONTACT Kaibo Liu *kliu8@wisc.edu* Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/TECH*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/TECH*.

locations while considering the spatial interactions among the traffic network for modeling and prediction. Chen et al. (2017) provided a traffic forecasting method that introduced a belief-desire-intention model into agent-based simulations, which can simulate traffic activities under different scenarios considering human psychological characteristics. Though it regards the traffic network as a whole system and considers the spatial interactions, this method is not data-driven and requires detailed model inputs which are hard to obtain in practice. Data-driven prediction methods, on the contrary, find the spatiotemporal correlations using the historical demand data without explicitly detailed knowledge of the entire system. In the literature of data-driven traffic prediction methods, Sun, Zhang, and Yu (2006) modeled traffic flows among adjacent road links in a transportation network as a Bayesian network and considered the joint probability distribution for spatially adjacent nodes. Deng et al. (2016) proposed a latent space model that applies matrix decomposition on road networks to capture the spatial correlation within the system. Tong et al. (2017) adopted a simple unified linear regression method with regularizations that smooths the prediction differences for locations and time points with similar features. Yao et al. (2018) proposed a deep learning framework to model the complex nonlinear spatial and temporal relations among the traffic network. These studies, however, do not take into consideration of the stochastic nature of traffic demands, and thus have to require frequent model redesigns (Tong et al. 2017). Moreover, these methods focus on either flow data or demand associated with the origins, instead of OD demand count data. Currently, there is still a literature gap in collectively modeling and predicting OD demand count data considering both the spatial and temporal correlations.

The challenges for modeling and predicting OD demand count lie in several aspects. First, most existing traffic prediction methods work with only traffic flow data at specific locations (Ashok and Ben-Akiva 2002; Ishak and Al-Deek 2002; Moreira-Matias et al. 2013; Deng et al. 2016); however, these methods ignore the natural characteristics of the demand that are count data and correlate in both spatial and temporal domains (Wemegah and Zhu 2017), and thus they are not suitable for modeling OD demand counts. Though methods dealing with spatiotemporal correlations among count data are developed in different application contexts, such as public health and ecology (Achcar et al. 2011; Conn et al. 2015; Flaxman et al. 2015), these methods are still inapplicable for modeling OD pair traffic demand as they do not consider the unique directionality of the traffic data, thus failing to appropriately model the spatiotemporal correlation structure in the traffic network. Figure 1 shows a toy example of a small traffic network with four OD pairs (i.e., $a-b$, $a-c$, $d-e$, $e-d$). It can be observed that the demand at origin a is only associated with and influenced by destinations b and c. Therefore, unlike the conventional spatiotemporal count data, the OD demand is directional and associated with the features of both the origin and the destination.

Second, existing literature that deals with count data models the traffic demand count data separately for different OD pairs without considering spatial correlations and domain knowledge (Bera and Rao 2011; Perrakis et al. 2012). Some traffic network features, for example, region characteristics of origins

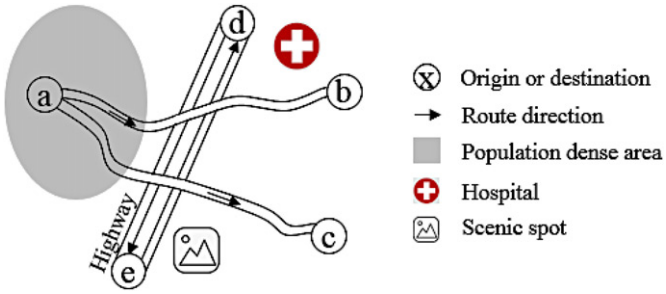

Figure 1. An illustrative example of traffic network.

or destinations, maximum speed limit on a route, and travel distance, are demonstrated to play important roles in trip generation and distribution analysis (Fotheringham and O'Kelly 1989; de Grange, Ibeas, and González 2011), and are easy to obtain from the traffic network offline. Following the example in Figure 1, although the presence of some traffic network information may relate to traffic demands, it is not clear how this information can be effectively used. Third, collectively modeling the spatiotemporal correlations in a traffic network involves estimating a large number of correlation parameters, which may lead to computational issues and affect the estimation accuracy. For example, to predict the demand of the traffic network in Figure 1 at 8 time epochs, it may involve up to hundreds of parameters in the covariance matrix. Therefore, to provide accurate estimations and avoid overfitting, it is crucial to incorporate physical knowledge of the traffic network and consider the sparsity of the correlation structure.

The objective of this study is to appropriately model the stochastic OD traffic demand counts considering the spatiotemporal correlations between different routes and epochs, while incorporating physical knowledge of the traffic network in the estimation. The estimation results are expected to enhance the prediction accuracy and robustness of the online traffic demand prediction for future epochs. In particular, we investigate a multivariate Poisson log-normal model with a block-diagonal covariance matrix and incorporate domain knowledge of the traffic network features to account for spatial correlations. In this way, we can fully explore the complicated spatiotemporal correlation structure of the traffic network demand and automatically cluster the routes with high correlations, without introducing a large number of parameters that impact the estimation accuracy. Besides transportation systems, the proposed method can be easily extended to other network applications with count data through little modifications, such as communication systems, supply chain management, smart grid, or even three-dimensional networks (Wang, Liu, and Zhang 2019). The remainder of this article is organized as follows. Section 2 provides a detailed mathematical formulation of the OD traffic demand problem. Section 3 introduces the main idea of the proposed method for the offline parameter estimation and online prediction. Section 4 conducts simulation experiments and Section 5 performs a real case study on a New York yellow taxi dataset to thoroughly evaluate and compare the proposed method with existing benchmarks. Section 6 concludes the article and discusses future studies.

2. Model Formulation

In this section, we provide detailed model formulations and assumptions.

Let N_{ijt} denote the observed traffic demand (i.e., the count of vehicles) for route j ($j = 1, \ldots, J$) on day i ($i = 1, \ldots, I$), at epoch t ($t = 1, \ldots, T$). Please note that the routes are directed as shown in Figure 1. Based on the natural characteristics of the demand counts, it is reasonable to model each observation N_{ijt} with a Poisson log-linear model (Perrakis et al. 2012; Xian, Li, and Liu 2018) such that

$$N_{ijt} \sim \text{Poisson}\left(\lambda_{ijt}\right),$$
$$u_{ijt} = \log \lambda_{ijt}.$$

Here λ_{ijt} is the intensity of the Poisson process, and u_{ijt} is the log transformation of the intensity. It should be noted that the demand count can be regarded as Poisson distributed since the occurrence of two vehicles can be considered independent in a large system. To characterize the spatiotemporal correlations across different routes and time points, we model u_{ijt} as a mixed-effect Gaussian process based on K basis functions $B_k(t)$ that

$$u_{ijt} = \mu_{jt} + \sum_{k=1}^{K} \gamma_{jk} B_k(t) + Z_{ijt}. \tag{1}$$

Here $\boldsymbol{\mu} = \left[\mu_{11}, \mu_{12}, \ldots, \mu_{JT}\right]'$ is the fixed effect coefficient that models the common characteristics of the whole traffic network, and $\boldsymbol{\gamma}_k = \left[\gamma_{1k}, \gamma_{2k}, \ldots, \gamma_{Jk}\right]$ is the random effect coefficient with prior distribution $\boldsymbol{\gamma}_k \sim N(0, \mathbf{R}_{\theta_y})$ that characterizes the uniqueness of different routes. Here $\mathbf{R}_{\theta_y}$ is the correlation matrix which takes into consideration of the traffic network information, where $\left[\mathbf{R}_{\theta_y}\right]_{j_1,j_2} = \sigma_{j_1,j_2} \exp\left\{-\theta_y \left|\boldsymbol{y}_{j_1} - \boldsymbol{y}_{j_2}\right|^2\right\}$. In this expression, $\boldsymbol{y}_j$ denotes the unique features of route j, such as information about the origin and destination, the maximum speed limit on a route, and the travel distance. The squared distance $\left|\boldsymbol{y}_{j_1} - \boldsymbol{y}_{j_2}\right|^2$ between features $\boldsymbol{y}_{j_1}$ and $\boldsymbol{y}_{j_2}$ is defined as $\sum_l \left|y_{j_1,l} - y_{j_2,l}\right|^2$, where the distance $\left|y_{j_1,l} - y_{j_2,l}\right|$ for each feature l is standardized to have unit variance. According to the expression of $\mathbf{R}_{\theta_y}$, the correlations between routes j_1 and j_2 are higher if they are associated with more similar features $\boldsymbol{y}_{j_1}$ and $\boldsymbol{y}_{j_2}$; θ_y is a nonnegative tuning parameter for the effect of traffic network features; and σ_{j_1,j_2} is a binary coefficient indicating whether the demands of route j_1 and route j_2 correlate or not. Here, the coefficient $\boldsymbol{\sigma} = \left[\sigma_{j_1,j_2}\right]_{J\times J}$ accounts for the sparsity of the correlation matrix $\mathbf{R}_{\theta_y}$ and zeros out the elements if no strong correlations are observed, and thus it helps to obtain an accurate parameter estimation.

The term Z_{ijt} in model (1) is the random error that follows a Gaussian distribution which has the covariance structure

$$\text{cov}\left(Z_{ij_1t_1}, Z_{ij_2t_2}\right) = \sigma_{j_1,j_2} \exp\left\{-\theta_y \left|\boldsymbol{y}_{j_1} - \boldsymbol{y}_{j_2}\right|^2\right\} \cdot \tau^2 \exp\{-\theta_t \left|t_1 - t_2\right|\}. \tag{2}$$

It should be noted that the covariance between $Z_{ij_1t_1}$ and $Z_{ij_2t_2}$ depends on both the features of routes j_1 and j_2, and the time points t_1 and t_2, which we refer to as the spatial and temporal covariance structures, respectively. Specifically, the spatial structure of Z_{ijt} remains the same as that in $\mathbf{R}_{\theta_y}$. In Equation (2), τ^2 is a constant parameter which is the variance for Z_{ijt}, and θ_t is a nonnegative tuning parameter for the effect of time lag.

Based on the above model and the covariance structure, we can derive the distribution of the log-transformed intensities. Denote the log-transformed intensity of the OD traffic demand on day i as $\boldsymbol{u}_i = \left(u_{i11}, u_{i12}, \ldots, u_{iJT}\right)'$. From the above model formulation, we can further derive that conditioning on parameters $\left(\boldsymbol{\mu}, \theta_y, \theta_t, \sigma, \boldsymbol{\tau}^2\right)$, u_i follows normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\mu = \left(\mu_{11}, \mu_{12}, \ldots, \mu_{JT}\right)', \quad \boldsymbol{\Sigma} = \mathbf{R}_{\theta_y} \otimes \left[\mathbf{R}_B + \tau^2 \mathbf{R}_{\theta_t}\right]. \tag{3}$$

Here, the symbol $\otimes$ denotes the Kronecker product, $\mathbf{R}_B$ is a fixed $T \times T$ matrix with the (t_1, t_2) element equal to $\sum_{k=1}^{K} B_k(t_1) B_k(t_2)$, and $\mathbf{R}_{\theta_t}$ is a $T \times T$ matrix with the (t_1, t_2) element equal to $\exp\{-\theta_t |t_1 - t_2|\}$. Therefore, the large covariance matrix is parametrized based on only the parameters $\theta_y, \theta_t, \boldsymbol{\sigma}$, and τ^2. This parsimonious model has several advantages, such as high interpretability tailored to the traffic demand count data, increased stability of the estimation results, and reduced computational burden for parameter estimation. Please note that though there may be some weekly or monthly trends in the demand, such trend information can be easily estimated and addressed based on the historical demand data. Without loss of generality, here we assume the trend information has already been removed and we focus on modeling the residuals parts. As a result, $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_I$ can be considered as independently and identically distributed Gaussian random vectors. In addition, it is assumed that here the demand counts are fully observable for specified OD pairs, which typically holds true for certain fleet vehicles such as taxis, last-mile EVs, and shared rides. In these applications, the pick-up and drop-off locations are naturally specified or recorded, where the proposed method can be directly applied.

3. Methodology

In this section, we propose a method to estimate the model parameters introduced in Section 2 based on historical observations, and further online predict the traffic demand for future epochs. The details of the parameter estimation framework are presented in Section 3.1. Then Section 3.2 demonstrates the online prediction scheme based on the estimated parameters.

3.1. Parameter Estimation

Denote the parameter set as $\boldsymbol{\Theta} = \left\{\boldsymbol{\mu}, \theta_y, \theta_t, \boldsymbol{\sigma}, \tau^2\right\}$, which are the parameters to be estimated in model (1). As the observed traffic demand counts $\boldsymbol{N}_i = \left(N_{i11}, N_{i12}, \ldots, N_{iJT}\right)'$ follow a multivariate Poisson-lognormal distribution, the likelihood of the data given parameter $\boldsymbol{\Theta}$ can be derived as

$$p(\boldsymbol{N}|\boldsymbol{\Theta}) = \int_{u} p(\boldsymbol{N}|\boldsymbol{u})\, p(\boldsymbol{u}|\boldsymbol{\Theta})\, d\boldsymbol{u}$$

$$= \int_{u} \prod_{i} \prod_{j} \prod_{t} \text{Poisson}\left(N_{ijt}|u_{ijt}\right) \prod_{i} p\left(\boldsymbol{u}_i|\boldsymbol{\mu},\ \boldsymbol{\Sigma}\right) d\boldsymbol{u}$$

$$\propto \prod_{i} \int_{u_i} \prod_{j} \prod_{t} \exp\left(-\exp\left(u_{ijt}\right)\right) \exp\left(N_{ijt} u_{ijt}\right)$$

$$\times \exp\left(-\frac{1}{2}\left(u_i-\mu\right)^T \boldsymbol{\Sigma}^{-1}\left(u_i-\mu\right)\right) |\boldsymbol{\Sigma}|^{-1/2}\, du_i.$$

For the estimation and prediction, we need to have a positive definite covariance matrix $\boldsymbol{\Sigma}$ as it requires to take the inverse of $\boldsymbol{\Sigma}$ in the likelihood function. To ensure this property, Proposition 1 provides a sufficient condition for $\boldsymbol{\Sigma}$ to be positive definite.

Proposition 1. If $\boldsymbol{\sigma}$ is equivalent to a block diagonal binary matrix after permutations, then the matrix $\boldsymbol{\Sigma}$ in Equation (3) is positive definite. Here, a block diagonal binary matrix $\mathbf{A}$ refers to a square block matrix of which the elements in the main diagonal block square matrices are all one, and the off-diagonal blocks are zero matrices, that is,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & 0 & \cdots & \mathbf{A}_n \end{pmatrix},$$

where $\mathbf{A}_i$ $(i = 1, \ldots, n)$ is a square matrix with all elements equal to 1.

The proof can be found in Appendix A. According to Proposition 1, we will focus on finding a block diagonal $\boldsymbol{\sigma}$ to guarantee the positive definiteness of the covariance matrix $\boldsymbol{\Sigma}$. Please note that estimating the block diagonal matrix $\boldsymbol{\sigma}$ is equivalent to clustering the routes into a list of blocks, where only the routes within a block highly correlate with each other. It can also be regarded as a sparse estimation approach because for the routes not in the same block, their correlation is considered zero if the correlation is not high. Based on the parameterization, the covariance matrix $\boldsymbol{\Sigma}$ is also a block diagonal matrix.

For parameter estimation, we need to obtain the parameter set $\boldsymbol{\Theta}$ by solving the following minimization problem

$$\boldsymbol{\Theta} = \text{argmin}\left(-\log p\left(N|\boldsymbol{\Theta}\right)\right),$$

s.t. $\boldsymbol{\sigma}$ is a symmetric matrix that can be permuted to a block diagonal matrix.

However, it is very challenging to directly minimize this objective function. The closed form solution of Poisson log-normal likelihood has been discussed in the literature, and it is shown that the integration part of the likelihood cannot be simplified directly (Stewart 1994). To tackle this challenge, here we apply data augmentation techniques to work around the direct integral (Chib, Greenberg, and Winkelmann 1998; Ma, Kockelman, and Damien 2008). In statistical analysis, data augmentation techniques regard unobservable variables as unknown parameters to be estimated, which can greatly simplify the original problem (Tanner and Wong 1987; Ye and Tang 2016). Here we treat the unobserved intensity $\boldsymbol{u}$ as an unknown parameter, which is to be estimated simultaneously with the parameter set $\boldsymbol{\Theta}$. Consequently, we can write the log of joint density of the observations $\boldsymbol{N}$ and the intensity $\boldsymbol{u}$ as

$$\log p\left(\boldsymbol{N},\ \boldsymbol{u}|\boldsymbol{\Theta}\right) = \sum_{i} \Bigg(\sum_{j} \sum_{t} \left(-\exp\left(u_{ijt}\right) + N_{ijt} u_{ijt}\right)$$

$$- \frac{1}{2}\left(\boldsymbol{u}_i-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{u}_i-\boldsymbol{\mu}\right) - \frac{1}{2}\log|\boldsymbol{\Sigma}| \Bigg).$$

In this way, we circumvent the intractable integral and simplify the original likelihood function. From now on, we treat $\boldsymbol{u}$ as a latent variable and further employ the Expectation-Maximization (EM) algorithm to obtain the maximum likelihood estimation (MLE) for the parameters, which is elaborated in the following subsections.

3.1.1. The E-Step

The conditional expectation of the log-likelihood function of $\boldsymbol{\Theta}$ with respect to $\boldsymbol{u}$ given $\boldsymbol{N}$, $\boldsymbol{\Theta}^{(s)}$ at iteration s is

$$Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(s)}\right) = \mathbb{E}_{\boldsymbol{u}|N,\ \boldsymbol{\Theta}^{(s)}}\left(\log p\left(\boldsymbol{N},\ \boldsymbol{u}|\boldsymbol{\Theta}^{(s)}\right)\right).$$

Here $\boldsymbol{\Theta}^{(s)}$ represents the current estimates of the parameters at iteration s, and $p\left(\boldsymbol{u}|\boldsymbol{N}, \boldsymbol{\Theta}^{(s)}\right)$ is the conditional distribution of $\boldsymbol{u}$. However, it is very challenging to directly obtain the analytical expression of $Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(s)}\right)$ given the complicated form of the likelihood. To tackle this issue, here we employ a Monte Carlo EM algorithm that incorporates Markov chain Monte Carlo (MCMC) sampling methods (Hung, Joseph, and Melkote 2015) to draw random samples of $\boldsymbol{u}$ to approximate the expected log-likelihood function $Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(s)}\right)$.

In particular, $\boldsymbol{\Theta}^{(s)}$ is a constant at a given iteration s, and thus the conditional probability $p\left(\boldsymbol{u}|\boldsymbol{N}, \boldsymbol{\Theta}^{(s)}\right)$ is proportional to $p\left(\boldsymbol{u},\ \boldsymbol{N}|\boldsymbol{\Theta}^{(s)}\right)$. Therefore, we draw m random samples of $\boldsymbol{u}_i$: $\left\{\boldsymbol{u}_i^{(1)}, \boldsymbol{u}_i^{(2)}, \ \ldots, \ \boldsymbol{u}_i^{(m)}\right\}$ from density $p\left(u,\ N|\boldsymbol{\Theta}^{(s)}\right)$ via the Metropolis–Hastings algorithm. The details of the Metropolis–Hastings algorithm are provided in Appendix B. Then we approximate the expected log-likelihood function with

$$\hat{Q}\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(s)}\right) = \frac{1}{m}\sum_{i=1}^{l}\sum_{i=1}^{m} \log p\left(\boldsymbol{N}, \boldsymbol{u}_i^{(l)}\|\boldsymbol{\Theta}^{(s)}\right).$$

3.1.2. The M-Step

In the M-step, we maximize the expected log-likelihood to obtain the estimation of the parameters for the new iteration. According to the formulation of the E-step, the parameter estimation at iteration $s+1$ is acquired by

$$\boldsymbol{\Theta}^{(s+1)} = \underset{\boldsymbol{\Theta}}{\text{argmax}}\left(\hat{Q}\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(s)}\right)\right)$$

$$= \text{argmin}\left(-\frac{1}{m}\sum_{i=1}^{I}\sum_{l=1}^{m} \log p\left(N, u_i^{(l)}|\boldsymbol{\Theta}^{(s)}\right)\right).$$

The objective function can be further derived as

$$-\frac{1}{m}\sum_{i=1}^{I}\sum_{l=1}^{m} \log p\left(\boldsymbol{N},\ \boldsymbol{u}_i^{(l)}|\boldsymbol{\Theta}^{(s)}\right)$$

$$\propto -\frac{1}{2m}\sum_{l=1}^{m}\left(\sum_{i}\left(-\left(\boldsymbol{u}_i^{(l)}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{u}_i^{(l)}-\boldsymbol{\mu}\right) - \log|\boldsymbol{\Sigma}|\right)\right).$$

Denote $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$, and $\boldsymbol{\Psi} = \Big[\boldsymbol{u}_1^{(1)} - \boldsymbol{\mu}, \boldsymbol{u}_1^{(2)} - \boldsymbol{\mu}, \ \ldots, \ \boldsymbol{u}_1^{(m)} - \boldsymbol{\mu}, \ \boldsymbol{u}_2^{(1)} - \boldsymbol{\mu}, \boldsymbol{u}_2^{(2)} - \boldsymbol{\mu}, \ \ldots, \ \boldsymbol{u}_2^{(m)} - \boldsymbol{\mu}, \ \ldots, \ \boldsymbol{u}_I^{(1)} - \boldsymbol{\mu}, \boldsymbol{u}_I^{(2)} - \boldsymbol{\mu}, \ \ldots, \ \boldsymbol{u}_I^{(m)} - \boldsymbol{\mu}\Big]$. Note that $|\boldsymbol{\Sigma}| = |\boldsymbol{\Omega}|^{-1}$, the objective function is then

$$\frac{1}{2m}\sum_{l=1}^{mI}\left(\boldsymbol{\Psi}_l^T\boldsymbol{\Omega}\boldsymbol{\Psi}_l\right) - \frac{1}{2}I\log\left(|\boldsymbol{\Omega}|\right) = \frac{1}{2m}tr\left(\boldsymbol{\Psi}^T\boldsymbol{\Omega}\boldsymbol{\Psi}\right) - \frac{1}{2}I\log\left(|\boldsymbol{\Omega}|\right).$$

The challenge here lies in the fact that the above objective function is not convex with regard to all the parameters due to the special parameterization rooted in the unique traffic problem. To estimate $\boldsymbol{\Theta} = \left\{\boldsymbol{\mu}, \boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\sigma}, \boldsymbol{\tau}^2\right\}$, we first estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using the MLE. Then, the parameters $\boldsymbol{\sigma}$, $\boldsymbol{\theta}_y, \boldsymbol{\theta}_t$, $\boldsymbol{\tau}^2$ are further estimated based on the estimated covariance matrix $\boldsymbol{\Sigma}$.

1. To estimate $\boldsymbol{\mu}$, it is the same as minimizing

$$\frac{1}{2m}\mathrm{tr}\left(\boldsymbol{\Psi}^T\boldsymbol{\Omega}\boldsymbol{\Psi}\right) = \frac{1}{2m}\sum_{l=1}^{m}\left(\sum_i\left(\boldsymbol{u}_i^{(l)} - \boldsymbol{\mu}\right)^T\boldsymbol{\Omega}\left(\boldsymbol{u}_i^{(l)} - \boldsymbol{\mu}\right)\right).$$

The solution is equivalent to the MLE $\hat{\boldsymbol{\mu}} = \frac{1}{mI}\sum_{l=1}^{m}\sum_i \boldsymbol{u}_i^{(l)}$.

2. To estimate $\boldsymbol{\Sigma}$, we minimize

$$\frac{1}{2m}\mathrm{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Psi}^T\boldsymbol{\Omega}\right) - \frac{1}{2}I\log\left(|\boldsymbol{\Omega}|\right).$$

Then this problem is equivalent to finding the MLE of the normal covariance matrix with scatter matrix $\boldsymbol{\Psi}\boldsymbol{\Psi}^T/m$. In other words, the updated sample covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \frac{\boldsymbol{\Psi}\boldsymbol{\Psi}^T}{mI}.$$

3. Next, we estimate the matrix σ based on the estimated sample covariance matrix $\hat{\boldsymbol{\Sigma}}$. Practically, identifying the nonzero structure of $\boldsymbol{\sigma}$ is equivalent to finding out the clusters of routes such that the demands of routes in the same cluster are strongly correlated.

To reduce the computation complexity, we propose an algorithm to determine $\boldsymbol{\sigma}$. The main idea is to first estimate a numeric (continuous-valued) matrix $\boldsymbol{\pi}$ which is a direct approximation of $\boldsymbol{\sigma}$, based on the structure of $\hat{\boldsymbol{\Sigma}}$. Then, the algorithm evaluates the clusters in the rows and columns of $\boldsymbol{\pi}$, and assign the elements of $\boldsymbol{\sigma}$ in the same cluster as 1. The details and property of this algorithm are elaborated in Appendix C.

4. Update the parameters θ_y, θ_t, and τ such that the parametrized matrix $\hat{\boldsymbol{\Sigma}}^{\boldsymbol{\Theta}} = \boldsymbol{R}_{\theta_y} \otimes \left(\boldsymbol{R}_B + \tau^2\boldsymbol{R}_{\theta_t}\right)$ is the nearest to the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$. In particular, we apply the Frobenius norm here which leads to the following objective function

$$\left(\theta_y^{(s+1)}, \theta_t^{(s+1)}, \boldsymbol{\tau}^{(s+1)}\right) = \arg\min_{\theta_y,\theta_t,\tau}\left(\hat{\boldsymbol{\Sigma}}^{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Sigma}}_F^2\right) = \arg\min_{\theta_y,\theta_t,\tau}\left(\sum_{i=1}^{JT}\sum_{j=1}^{JT}\left(\hat{\Sigma}_{i,j}^{\boldsymbol{\Theta}} - \hat{\Sigma}_{i,j}\right)^2\right).$$

This optimization problem can be solved by the first-order optimization method. The detailed derivation of the objective function and the optimization method is discussed in Appendix D.

3.2. Online Prediction of OD Traffic Demand at Future Epochs

In this subsection, we propose a method for online predicting the future OD traffic demand counts based on both the estimated model parameters (as detailed in Section 3.1) and the online observations. For a test day, denote all observations on and before time t as $\tilde{\boldsymbol{N}}_{\cdot,t} = \left\{N_{j,\tau} : \tau \leqslant t, j = 1, 2, \cdots, J\right\}$, a set of future observations at time $t+k$ as $\boldsymbol{N}_{\cdot,\,t+k}$, $k \geqslant 1$, and their corresponding log-transformed intensities as $\tilde{\boldsymbol{u}}_{\cdot,t}$ and $\boldsymbol{u}_{\cdot,\,t+k}$.

First, according to the properties of Gaussian processes, $\left[\tilde{\boldsymbol{u}}_{\cdot,t}, \boldsymbol{u}_{\cdot,\,t+k}\right]$ jointly follows a normal distribution $N\left(\left[\boldsymbol{\mu}_{\cdot,\,t}, \boldsymbol{\mu}_{\cdot,\,t+k}\right]^T, \boldsymbol{\Sigma}^{((1:t,\,t+k),\,(1:t,\,t+k))}\right)$, where the superscript refers to the subsets of the corresponding arrays. Then the conditional distribution of $\boldsymbol{u}_{\cdot,\,t+k}|\tilde{\boldsymbol{u}}_{\cdot,t}$, which is the future intensity conditioned on the current ones, follows a normal distribution with mean $\boldsymbol{\mu}_{\cdot,t+k|t} = \boldsymbol{\mu}_{\cdot,t+k} + \boldsymbol{\Sigma}^{(t+k,1:t)}\boldsymbol{\Sigma}^{(1:t,1:t)^{-1}}\left(\tilde{\boldsymbol{u}}_{\cdot,t} - \boldsymbol{\mu}_{\cdot,t}\right) \triangleq \boldsymbol{\mu}_{\cdot,t+k} + \boldsymbol{L}\left(\tilde{\boldsymbol{u}}_{\cdot,t} - \boldsymbol{\mu}_{\cdot,t}\right)$ and covariance matrix $\boldsymbol{\Sigma}_{t+k|t} = \boldsymbol{\Sigma}^{(t+k,\,t+k)} - \boldsymbol{L}\boldsymbol{\Sigma}^{(1:t,\,t+k)}$, where $\boldsymbol{L} = \boldsymbol{\Sigma}^{(t+k,\,1:t)}\boldsymbol{\Sigma}^{(1:t,\,1:t)-1}$.

By the tower property of expectations and the expectation of lognormal distributions, the conditional expectation of the future traffic demand counts at time $t+k$ can be derived as

$$\begin{aligned}\mathbb{E}\left(\boldsymbol{N}_{\cdot,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) &= \mathbb{E}\left(\mathbb{E}\left(\boldsymbol{N}_{\cdot,t+k}|\boldsymbol{u}_{\cdot,t+k}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\ &= \mathbb{E}\left(\exp\left(\boldsymbol{u}_{\cdot,t+k}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\ &= \mathbb{E}\left(\mathbb{E}\left(\exp\left(\boldsymbol{u}_{\cdot,t+k}\right)|\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\ &= \mathbb{E}\left(\exp\left(\boldsymbol{\mu}_{\cdot,t+k|t} + \frac{1}{2}\mathrm{diag}\left(\boldsymbol{\Sigma}_{t+k|t}\right)\right)\middle|\tilde{\boldsymbol{N}}_{\cdot,t}\right).\end{aligned}$$

Then each element in the above expectation can be further derived as

$$\begin{aligned}&\mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) \qquad (4)\\ =&\mathbb{E}\left(\exp\left(\mu_{j,t+k|t} + \frac{1}{2}\Sigma_{t+k|t}^{jj}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\ =&\exp\left(\mu_{j,t+k} + \frac{1}{2}\Sigma_{t+k|t}^{jj}\right)\mathbb{E}\left(\exp\left(\left(\boldsymbol{L}\left(\tilde{\boldsymbol{u}}_{\cdot,t} - \boldsymbol{\mu}_{\cdot,t}\right)\right)_j\right)\middle|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\ =&\exp\left(\mu_{j,t+k} + \frac{1}{2}\Sigma_{t+k|t}^{jj} - \boldsymbol{L}_j\boldsymbol{\mu}_{\cdot,t}\right)\mathbb{E}\left(\exp\left(\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right).\end{aligned}$$

The last term in Equation (4) can be calculated as

$$\int_u \exp\left(\mathbf{L}_j u\right)\prod_k\prod_t \exp\left(-\exp\left(u_{kt}\right)\right)\exp\left(N_{kt}u_{kt}\right)$$

$$\times \exp\left(-\frac{1}{2}(u-\mu)^T \boldsymbol{\Sigma}^{-1}(u-\mu)\right)|\boldsymbol{\Sigma}|^{-1/2}\,du$$

$$\propto \int_u \exp\left(\left(\mathbf{L}_j + N^T\right)u - 1^T \exp(u)\right.$$

$$\left.-\frac{1}{2}(u-\mu)^T \boldsymbol{\Sigma}^{-1}(u-\mu)\right)|\boldsymbol{\Sigma}|^{-1/2}\,du.$$

Therefore, the last term is equivalent to a likelihood of lognormal distribution, which cannot be obtained analytically. Instead, we again propose to use Monte Carlo approximation to evaluate the quantity. Based on the estimated parameters, we can generate l samples $\left\{\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right\}_{i=1}^{l}$ of the intensity $\tilde{\boldsymbol{u}}_{\cdot,t}$ based on normal distribution $N\left(\boldsymbol{\mu}_{\cdot,t}, \boldsymbol{\Sigma}^{(1:t,\,1:t)}\right)$. Then we can approximate the conditional distribution by

$$\mathbb{E}\left(\exp\left(\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right) = \int \exp\left(\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) d\tilde{\boldsymbol{u}}_{\cdot,t} \quad (5)$$

$$\approx \frac{\sum_{i=1}^{l} \exp\left(\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{N}}_{\cdot,t}|\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right)}{\sum_{i=1}^{l} p\left(\tilde{\boldsymbol{N}}_{\cdot,t}|\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right)}.$$

The Monte Carlo procedure can be done offline before the online prediction, and thus it will not lead to high computational burden in the online prediction process.

Combining Equations (4) and (5), we can predict the traffic demand counts in the future epochs utilizing the current observations at the maximum extent. Besides the mean prediction, we can also derive the variance of the predicted values. First of all, the expected squared demand count of route j at time $t+k$ on a test day given the observed demand counts can be calculated as follows:

$$\begin{aligned}
&\mathbb{E}\left(N_{j,t+k}^2|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\
&= \mathbb{E}\left(\mathbb{E}\left(N_{j,t+k}^2|u_{j,t+k}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\
&= \mathbb{E}\left(\exp\left(u_{j,t+k}\right) + \exp\left(2u_{j,t+k}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\
&= \mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) + \mathbb{E}\left(\exp\left(2\mu_{j,t+k|t} + \Sigma_{t+k|t}^{jj}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\\
&= \mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) + \exp\left(2\mu_{j,t+k} + \Sigma_{t+k|t}^{jj} - 2\boldsymbol{L}_j\boldsymbol{\mu}_{\cdot,t}\right)\\
&\quad\times \mathbb{E}\left(\exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right).
\end{aligned}$$

Similar to Equation (5), the quantity $\mathbb{E}\left(\exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right)$ can be approximated by the Monte Carlo procedure as

$$\mathbb{E}\left(\exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right) = \int \exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) d\tilde{\boldsymbol{u}}_{\cdot,t} \quad (6)$$

$$\approx \frac{\sum_{i=1}^{l} \exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{N}}_{\cdot,t}|\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right)}{\sum_{i=1}^{l} p\left(\tilde{\boldsymbol{N}}_{\cdot,t}|\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right) p\left(\tilde{\boldsymbol{u}}_{\cdot,t}^{(i)}\right)}.$$

Then the variance of the predicted values can be calculated as

$$\begin{aligned}
\operatorname{var}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) &= \mathbb{E}\left(N_{j,t+k}^2|\tilde{\boldsymbol{N}}_{\cdot,t}\right) - \left(\mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\right)^2 \quad (7)\\
&= \mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right) - \left(\mathbb{E}\left(N_{j,t+k}|\tilde{\boldsymbol{N}}_{\cdot,t}\right)\right)^2\\
&\quad + \exp\left(2\mu_{j,t+k|t} + \Sigma_{t+k|t}^{jj} - 2\boldsymbol{L}_j\boldsymbol{\mu}_{\cdot,t}\right)\\
&\quad\times \mathbb{E}\left(\exp\left(2\boldsymbol{L}_j\tilde{\boldsymbol{u}}_{\cdot,t}\right)|\tilde{\boldsymbol{N}}_{\cdot,t}\right).
\end{aligned}$$

4. Computational Experiments

In this section, we present the results of numerical experiments on a simulated traffic network. We assume that there are $J = 10$ routes, and traffic network features are randomly generated from normal distribution $N(0, 1)$ for each route. We observe and predict the travel demand counts for each route at an interval of 2 hr, and thus there are $T = 12$ epochs within a day. The simulated data is generated from the lognormal distribution with the following approach. First, the log-transformed intensities $\boldsymbol{u}_i$ on a day i is randomly generated from the normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where the distribution parameters are formulated as in Equation (3). The values of the parameters are set as follows. We take the Fourier basis as the basis function with $K = 8$. For the mean vector, we suppose that all routes share the same mean $\bar{\boldsymbol{\mu}}$ (i.e., $\boldsymbol{\mu} = \mathbf{1}_J \otimes \bar{\boldsymbol{\mu}}$ and $\mathbf{1}_J$ refers to an all-one column vector with length J), which is shown in Figure 2. The other parameters are set as follows: $\theta_y = 0.15, \theta_t = 1$, $\tau = 0.9$, and the binary coefficient matrix is

$$\boldsymbol{\sigma} = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{pmatrix}.$$

This matrix refers to a scenario that the first three routes are closely correlated to each other, the sixth and the seventh routes

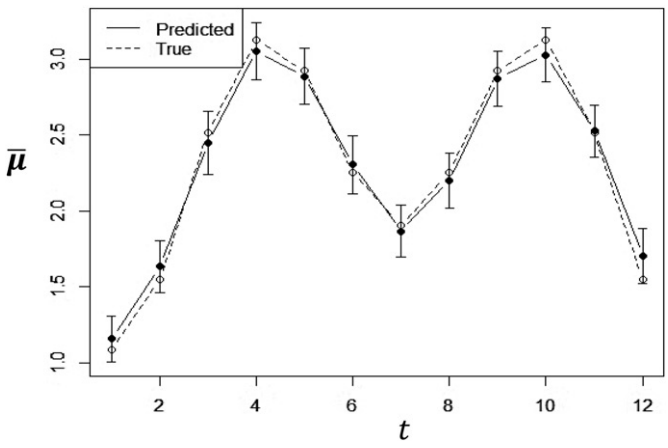

Figure 2. Mean estimation of $\bar{\mu}$ with 90% confidence intervals.

are closely correlated, and so on. Then, the traffic demand counts on day i for the n routes at T epochs are randomly generated from Poisson distributions based on the exponential of $\boldsymbol{u}_i$. In the estimation part, we fix the number of historical observations at $I = 30$ days, and $m = 200$ random samples in the E-step.

4.1. Ideal Case

We first show the parameter estimation results in the ideal case that the traffic network features y is perfectly known and applied in the study. Specifically, we assume that there are five traffic network features randomly generated from the normal distribution $N(0, 1)$, and they are the only features that are related to the traffic demand. Based on the perfect knowledge of this information, the proposed estimation and prediction method is applied to the training data. The reported results are summarized based on 100 simulation runs.

Figure 2 shows the estimation of the mean function $\bar{\boldsymbol{\mu}}$ for a route with 90% confidence interval, from which it can be seen that the estimation is very close to the true values. The averaged estimated value of $\boldsymbol{\sigma}$ is demonstrated in the heat map in Figure 3(b). Comparing to the true value shown in Figure 3(a), there are uncertainties in estimating the off-diagonal nonzero elements in σ. For example, as the estimated values of $\hat{\sigma}_{2,1}$ and $\hat{\sigma}_{3,1}$ are less than one, there is a certain chance that the first route may not be clustered correctly in its true underlying subgroup. A possible reason for this phenomenon is that the randomly generated features of the first route are a little far from routes 2 and 3, and thus the actual correlations between the first route and the other two routes are not strong enough, which leads to high randomness in the estimation process. Table 1 further shows the mean and 90% confidence interval of the estimations for the scalar parameters θ_y, θ_t, and τ. It can be observed that the estimations for all three parameters are close to their true underlying values, and the true values all lie in the 90% confidence interval of the estimation.

Next, the traffic demand counts of $I = 100$ test days are simulated with the same approach as for the training data. We conduct the proposed traffic demand prediction method for the test days, with $l = 500$ random Monte Carlo samples for approximation. Then the prediction performance is compared with three competitive methods that can also be applied for the online prediction of Poisson data: (1) Poisson regression on observations of the same epoch with the traffic network features as the covariates, (2) Poisson regression on observations of the same epoch without covariates, and (3) an intuitive method that predicts the demand using historical means. At an epoch t ($t \geqslant 2$), we apply the above four methods on observations from epochs $1, 2, \ldots, t-1$ to predict the demand count at epoch t. To measure the prediction accuracy, we define the daily prediction error at day i as the L^2-norm of the difference between the predicted demand and the actual demand, that is,

$$e_i = \frac{\hat{\boldsymbol{N}}_i - \boldsymbol{N}_i}{JT}.$$

Figure 4 shows the boxplot of the prediction errors e_i in the 100 test days. For all four methods, the prediction errors exhibit long tails as there may be outliers generated from the Poisson distributions. The proposed method in general achieves the least prediction error, comparing to the other three methods. The other three competitive methods have roughly the same median in the prediction error, but using the historical mean leads to slightly more outliers than the other two methods based on Poisson regression. This is expected as the historical mean method is a naive approach that only averages the historical observations. The methods based on Poisson regression, on the other hand, take the similarity of the routes into consideration (as the traffic demand counts for all routes have the same mean in this simulation study) and thus lead to enhanced prediction

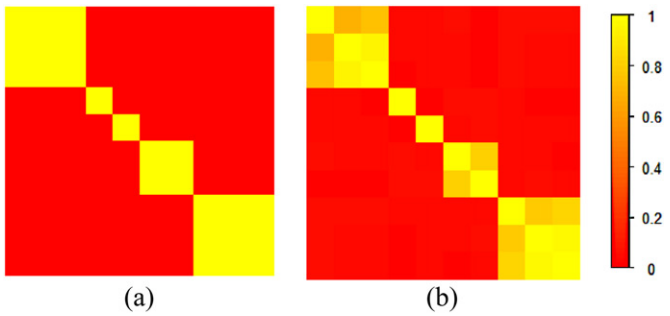

Figure 3. Heat maps of the (a) true and (b) averaged estimated values of σ.

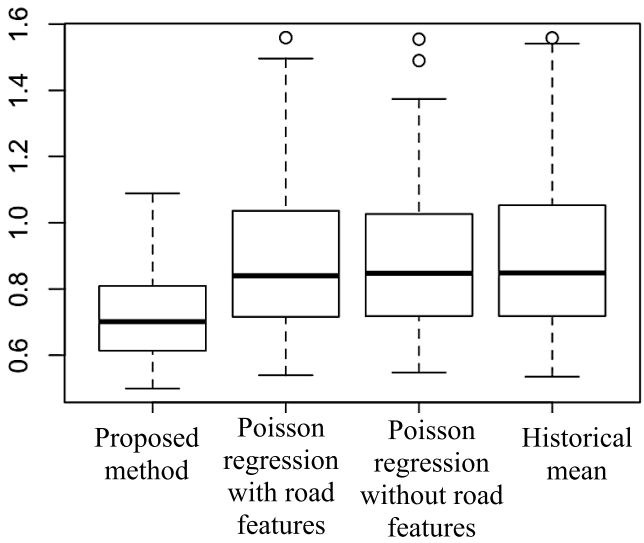

Figure 4. Boxplots of prediction errors for the competitive methods on the test data. The average prediction errors for the four methods are 0.721, 0.894, 0.900, and 0.904, respectively.

Table 1. Parameter estimation results in the ideal case.

	θ_y	θ_t	τ
True values	0.1500	1.0000	0.9000
Estimated values	0.1590 (0.1254, 0.2110)	0.9994 (0.9063, 1.0973)	0.9035 (0.8478, 0.9557)

accuracy. Another interesting observation lies in the comparison between the two Poisson regression methods. By incorporating the traffic network information in the Poisson regression, it even leads to a higher prediction error or more outliers. This observation shows us that incorporating relevant domain knowledge is not always helpful if it is not used appropriately. In this case, the misused model may have led to overfitting by using the traffic network information as the regression covariates. In comparison, the proposed method is more flexible in modeling both the mean and covariance structure of the spatiotemporal patterns and correlations in the traffic demand data. Moreover, the proposed method automatically clusters the routes such that the prediction of traffic demand is based on the sparse spatial structure, which avoids the overfitting.

4.2. Feature Variations

Unlike the ideal case in Section 4.1, we test the performance of the proposed method under the cases when the traffic network information $\boldsymbol{y}$ is not perfectly known. Specifically, we still assume that there are five true underlying traffic network features that have an impact on the traffic demand, and then consider the following two cases. Suppose we collect and apply (i) inadequate features (only three of the five true features are available), and (ii) excessive features (two additional irrelevant features besides the true features are misused). In the two cases, the proposed estimation and prediction method is applied to the training data to test the performance of the proposed method. Other than the traffic network information $\boldsymbol{y}$, the simulation settings in this case are identical to the ones in Section 4.1.

We focus on the estimation results of parameters relevant to the covariance matrix $\boldsymbol{\Sigma}$ as the variations of traffic network information have a direct influence on $\boldsymbol{\Sigma}$. In particular, the averaged estimated value of $\boldsymbol{\sigma}$ is demonstrated in heat maps in Figure 5. Compared to the ideal case estimation, the results in Figures 5(a) and 5(b) demonstrate similar patterns but are slightly noisier. When using imperfect $\boldsymbol{y}$, the probability of correctly identifying nonzero elements in $\boldsymbol{\sigma}$ generally decreases, which can be explained by the increased uncertainties in the MCMC sampling and thus in determining $\boldsymbol{\sigma}$ in the M step. Table 2 shows the mean and 90% confidence interval of the estimations for the scalar parameters θ_y, θ_t, and τ. It can be observed that when the traffic network features are not perfect, it has the most influence on the estimation of θ_y. This result makes sense as θ_y is the tuning parameter for the effect of the traffic network features. When using inadequate features, the distance between the features of two routes (i.e., $\left|\boldsymbol{y}_{j_1} - \boldsymbol{y}_{j_2}\right|^2$) is less than the underlying truth. Therefore, as compensation, the estimation of θ_y tends to be larger and has a longer upper tail. On the contrary, the estimated value of θ_y is generally less than the true values when excessive features are applied. It should be noted that using excessive features may lead to worse results than using inadequate features, as we can observe that the true value of θ_y lies outside of the 90% confidence interval of the estimation when using excessive features. Irrelevant features may exhibit higher randomness and dominate the feature differences between routes, and thus results in larger estimation bias. There is little evidence that the variation in features causes large estimation bias of θ_t and τ, though it introduces higher variations than the ideal case.

It should be noted that although the estimation may be biased when using imperfect traffic network features, it does not seriously compromise the accuracy of the resulted covariance matrix $\boldsymbol{\Sigma}$ as the estimation of θ_y changes with $\boldsymbol{y}$ to approximate the covariance matrix. Due to the parameterization tailored to the traffic network, the proposed problem formulation regulates the covariance structure and ensures the overall estimation accuracy of $\boldsymbol{\Sigma}$, in spite of using imperfect traffic network features and data variability. Figure 6 shows the prediction errors of the test data when using inadequate and excessive traffic network features. It can be observed that the prediction results in the three cases are very similar, though using imperfect data results

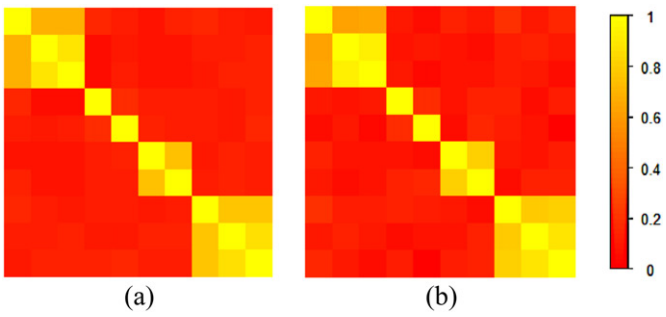

Figure 5. Heat maps of the averaged estimated values of σ when applying (a) inadequate features and (b) excessive features.

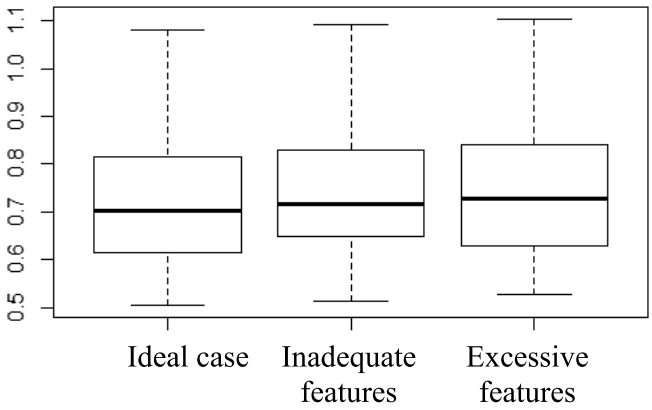

Figure 6. Boxplots of the prediction errors when using imperfect traffic network features.

Table 2. Parameter estimation results with imperfect traffic network features.

	θ_y	θ_t	τ
True values	0.1500	1.0000	0.9000
Inadequate features	0.1668 (0.1366, 0.2575)	1.0148 (0.9221, 1.1732)	0.9074 (0.8546, 0.9573)
Excessive features	0.1072 (0.0861, 0.1417)	0.9470 (0.8951, 1.0167)	0.8958 (0.8339, 0.9490)

in slightly larger medians in the prediction error. There are larger prediction errors when using excessive features, as irrelevant features can be noisy and dominate the feature differences.

4.3. Scalability Analysis

In this section, we study the performance of the proposed method to the size of the traffic network. In particular, we consider traffic networks with $J = 5, 10, 20, 30, 50$ routes and compare the computational time and estimation errors. The results are shown in Figure 7. For different values of J, the black line with round dots represents the estimation time offline, and the blue line with triangular dots represents the prediction time for one day online, where the unit of time is second. It is evident that the proposed method is effective for online prediction though expensive in the estimation procedure. The red line with squared dots represents the estimation error under different values of J, where the estimation error is defined as

$$\widehat{\bar{\boldsymbol{\mu}}} - \bar{\boldsymbol{\mu}} + \hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} + \hat{\theta}_y - \theta_y + \hat{\theta}_t - \theta_t + \hat{\tau} - \tau,$$

and the $\hat{}$ mark denotes the averaged estimated value of a quantity over 10 simulation runs.

5. Application and Results

In this section, we apply the proposed method to a real New York yellow taxi dataset which is collected from June 1st to July 31st in 2017. The data were collected by technology providers authorized under the Taxicab and Livery Passenger Enhancement Programs (TPEP/LPEP) and made publicly available online (NYC Taxi 2017). The dataset records all yellow taxi trips during the aforementioned time period including the pick-up and drop-off dates and times, pick-up and drop-off locations, trip distances, and payment information about the trips. To illustrate the main idea of the proposed method, we focus on the trips between the four busiest zones in Manhattan and investigate the structure of the travel demand counts on these zones as OD pairs. The details of the four taxi zones are shown in Figure 8.

Considering the above 4 taxi zones as origins and destinations, there are in total $n = 16$ routes corresponding to the OD pairs. Denote a route with origin o and destination d as (o, d), the routes can be written as (1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (2, 4), (3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), and (4, 4). For each route, we collect the traffic network information including the origin, destination, zone areas for both the origin and destination, average travel distance, and the number of route selections between the zones (number of major street paths connecting two zones). The zone areas are obtained from the New York City taxi zones data from the NYU Spatial Data Repository. The average travel distance information is directly calculated using the trip distances which are recorded in the data. The number of route selections between the zones is directly observed from the zone maps. These features are shown to play important roles in trip generation and spatial interactions (Fotheringham and O'Kelly 1989; de Grange, Ibeas, and González 2011). Among the features, the origin and destination are considered to be categorical variables, which are encoded using their indices as shown in Figure 8, and discrete metric is applied to calculate the distances between categorical features of two routes. The discrete metric is defined such that the distance between two routes with the same categorical feature is 1 and between two routes with different categorical features is 0. To make sure the categorical features and the numerical features are standardized in the same scale, the distances $|y_{j_1,l} - y_{j_2,l}|$ for each feature l are standardized to have unit variance. For the time scope, we focus on the rush hours between 6:00 p.m. and 8:00 p.m. and observe the trip counts every 15 min ($T = 8$ epochs), as the rush hour traffic demand prediction has been an essential topic in alleviating the traffic burden and responding to sudden rises in traffic demand. We use the data in June 2017

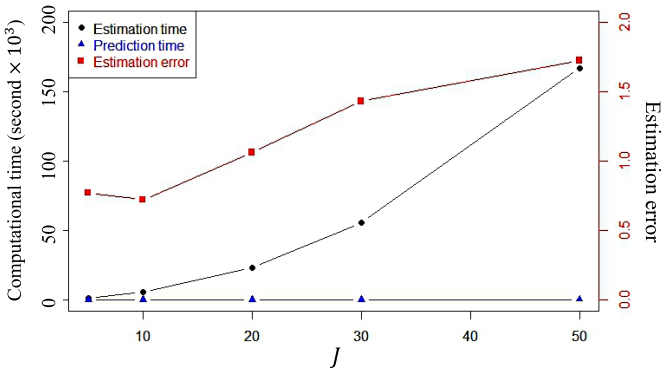

Figure 7. Computational time and estimation error across the number of routes.

Index	Borough	Zones
1	Manhattan	Lincoln Square East
2	Manhattan	Times Square/Theatre District
3	Manhattan	Upper East Side North
4	Manhattan	Upper East Side South

Figure 8. Illustration of the taxi zones in the case study.

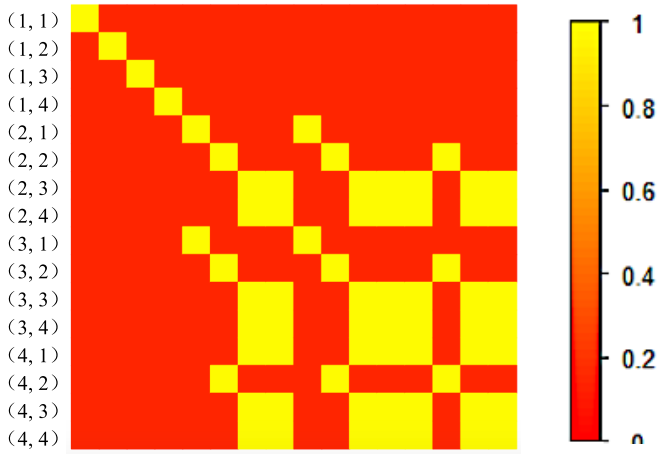

Figure 9. Heat map of the estimated values of $\boldsymbol{\sigma}$.

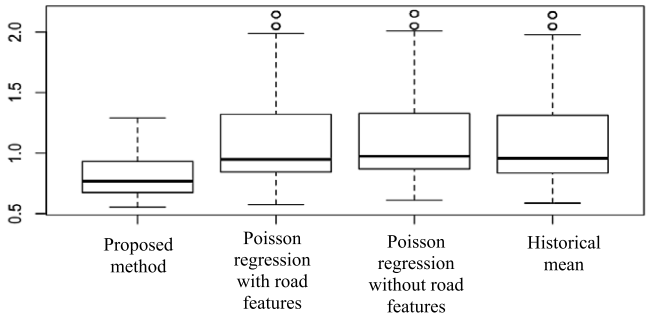

Figure 10. Prediction results for all considered prediction methods.

as the training data and the July data as the test data. There are in total 44,727 and 34,251 taxi trips for the June training data and the July testing data, respectively. The counts are aggregated to 15 min intervals, so in June 2017, this translates to (8 epochs) $\times$ (30 days) $\times$ (16 routes) = 3840 counts, and in July 2019, there are (8 epochs) $\times$ (31 days) $\times$ (16 routes) = 3968 counts. Again, we take the Fourier basis as in the simulation. In the estimation and prediction procedures, the number of Monte Carlo random samples for approximations are fixed at $m = 200$ and $l = 500$. The values of the model parameters in the data are estimated as follows: $\hat{\theta}_y = 0.1472$, $\hat{\theta}_t = 0.1407$, $\hat{\tau} = 0.3577$, and the estimated binary coefficient matrix $\hat{\boldsymbol{\sigma}}$ is shown in Figure 9, which is ordered lexicographically. This result means that the routes are automatically clustered into 7 subgroups by the proposed method. Specifically, the clusters represented by $\hat{\boldsymbol{\sigma}}$ can be rewritten as $\{(1, 1)\}$, $\{(1, 2)\}$, $\{(1, 3)\}$, $\{(1, 4)\}$, $\{(2, 1), (3, 1)\}$, $\{(2, 2), (3, 2), (4, 2)\}$, and $\{(2, 3), (2, 4), (3, 3), (3, 4), (4, 1), (4, 3), (4, 4)\}$. There are some interesting observations from the clusters. First, the clustering results are highly affected by the destination of the routes. Second, zones 3 and 4 as origins or destinations are quite similar in terms of the clustering. In another word, a route that starts from (or ends at) zone 3 to (or from) zone k is very likely to be clustered together with a route that starts from (or ends at) zone 4 to (or from) zone k. This makes sense as they are two parts of the neighborhood Upper East Side and thus may share more similarities in the demand. Third, the routes from origin 1 exhibit very different patterns from the other routes and are thus clustered as individual groups. Therefore, the structure of $\hat{\boldsymbol{\sigma}}$ provides an interesting explanation of the similarity and correlations between the routes, which is essential for gaining a better understanding and ensuring accurate estimation and prediction of the traffic network.

Based on the above parameter estimation, the prediction results are calculated according to the methodology detailed in Section 3.2 and compared with the baseline methods. The boxplots of the daily prediction errors for the test data are shown in Figure 10. It is obvious that the proposed method yields the prediction results with the least errors. The percentage reduction in the mean prediction error of the proposed method over the competing methods is uniformly over 25%, which demonstrates the superiority of our method. Comparing to the other three prediction methods, the proposed method considers the spatiotemporal correlations between the observations in the traffic network and can dynamically adapt to the online observations.

Figure 11 further shows the specific taxi demand prediction results of the routes (4, 3) and (4, 4) for four test days. The solid black line in this figure represents the true dynamic traffic demand counts, where it can be observed that the true taxi demand indeed exhibits high spatial and temporal variation and strong correlations for observations between the routes and across different epochs as we expect. On day 3, for example, the taxi demands for both two routes are far lower than the average and show similar patterns over different epochs. This observation agrees with our assumption of the spatiotemporal correlations hidden in the traffic network. The solid red line in is the predicted demand using the proposed method, and the dashed error bars show the 90% confidence interval of the prediction based on the variance derivation in Equation (7). There are several interesting observations about the predicted values and confidence intervals. First, the prediction results of the proposed method dynamically adapt to the characteristics of the online observations and align with the true observations, as the proposed prediction method exploits the spatiotemporal correlation structures learned from the training data and collectively utilizes the online observations of the routes within the same cluster. Second, for the same epochs, the confidence interval is generally wider for larger predicted values, which is expected according to the property of the Poisson log-normal distribution. For example, comparing the observations on day 1 and day 4 at the same epochs, the confidence intervals for day 4 is narrower than day 1, given that the demand counts observed on day 4 are generally much smaller. Third, the width of the confidence intervals generally grows narrower as time evolves because more and more observations are available. In particular, the predicted value for the first epoch of each route is a fixed prediction based on historical data only (instead of the online observations) and thus has the widest confidence interval. Other than the initial prediction on the first epoch, it can be observed that most of the true observations fall into the 90% confidence interval. The dotted line in Figure 11 is the historical mean, which shows an unsatisfactory prediction result. The Poisson regression based predicted results are omitted in this figure since they are similar to the historical mean.

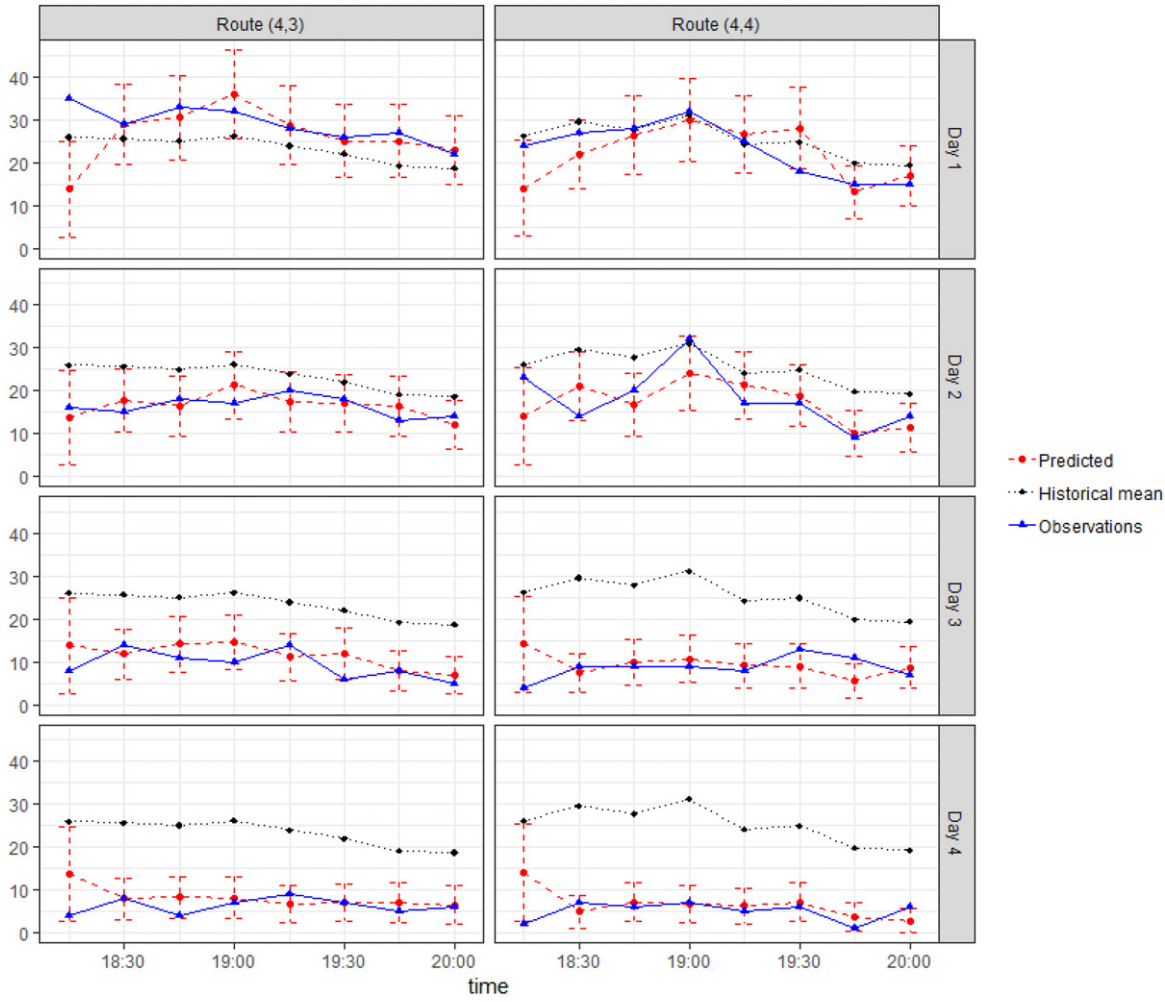

Figure 11. Taxi demand prediction results for routes (4, 3) and (4, 4).

6. Conclusion and Discussion

The modeling and prediction of traffic demand counts in a traffic network have been a challenging task in the literature considering the complex model formulation, the need for effective domain knowledge integration, and the exploitation of acquired knowledge. In this article, we have formulated and proposed a multivariate Poisson log-normal model with a parametrized covariance matrix tailored to the traffic problem formulation. With the specific parameterization, the proposed method can automatically account for the integration of the traffic network domain knowledge and achieve a sparse estimation based on the clustering of the routes. By developing an EM algorithm incorporated with MCMC sampling, we can estimate the parameters of the model with high accuracy, which are then applied for traffic demand count prediction by exploiting the online observations. The results of the simulation studies and a real application to New York yellow taxi data demonstrate that the proposed method achieves a much better estimation and prediction performance than a list of existing baseline methods.

There are several related topics that worth studying as future works. First, more advanced optimization techniques can be applied in the M-step to learn the covariance matrix, which may increase the estimation accuracy and decrease the computational time. Second, the current method is still challenging when it is applied to very large traffic networks. However, according to expert knowledge, the correlation matrix between routes can be sparse in a very large traffic network. Thus, to increase the scalability of the method, it is worth developing sparse correlation screening techniques that can decompose a large network to several small networks first. Then the proposed method can be applied to each small and closely related network separately and in parallel to reduce the computational time. Another practical issue is that the OD demand might not always be fully observable or easily specified in some contexts (Tebaldi and West 1998). An important future work is thus to estimate and predict the OD demand counts based on partially observed count data. Last but not least, it is important to incorporate advanced learning techniques in the estimation to automatically select the most relevant and informative traffic network features for better prediction results.

Supplementary Materials

"Appendix.pdf" shows the appendices of the article. "Estimation.R" and "Prediction.R" are the code for the proposed method. It also contains all necessary dataset to run the code. All of the supplemental files are contained in a single archive.

Acknowledgments

The authors gratefully acknowledge the editor, associate editor, and anonymous reviewers whose comments helped improve and clarify the manuscript significantly.

Funding

This work was supported in part by the National Science Foundation under grant 1637772.

References

Achcar, J. A., Martinez, E. Z., Souza, A. D. P. D., Tachibana, V. M., and Flores, E. F. (2011), "Use of Poisson Spatiotemporal Regression Models for the Brazilian Amazon Forest: Malaria Count Data," *Revista da Sociedade Brasileira de Medicina Tropical*, 44, 749–754. [2]

Agatz, N. A., Erera, A. L., Savelsbergh, M. W., and Wang, X. (2011), "Dynamic Ride-Sharing: A Simulation Study in Metro Atlanta," *Transportation Research Part B: Methodological*, 45, 1450–1464. [1]

Ashok, K. (1996), "Estimation and Prediction of Time-Dependent Origin-Destination Flows," Doctoral dissertation, Massachusetts Institute of Technology. [1]

Ashok, K., and Ben-Akiva, M. E. (2002), "Estimation and Prediction of Time-Dependent Origin-Destination Flows With a Stochastic Mapping to Path Flows and Link Flows," *Transportation Science*, 36, 184–198. [1,2]

Bera, S., and Rao, K. V. (2011), "Estimation of Origin-Destination Matrix From Traffic Counts: The State of the Art," *European Transport*, 49, 2–23. [2]

Brandstatter, G., Kahr, M., and Leitner, M. (2017), "Determining Optimal Locations for Charging Stations of Electric Car-Sharing Systems Under Stochastic Demand," *Transportation Research Part B: Methodological*, 104, 17–35. [1]

Carrese, S., Cipriani, E., Mannini, L., and Nigro, M. (2017), "Dynamic Demand Estimation and Prediction for Traffic Urban Networks Adopting New Data Sources," *Transportation Research Part C: Emerging Technologies*, 81, 83–98. [1]

Chen, K., and Müller, H. G. (2014), "Modeling Conditional Distributions for Functional Responses, With Application to Traffic Monitoring via GPS-Enabled Mobile Phones," *Technometrics*, 56, 347–358. [1]

Chen, X., Peng, L., Zhang, M., and Li, W. (2017), "A Public Traffic Demand Forecast Method Based on Computational Experiments," *IEEE Transactions on Intelligent Transportation Systems*, 18, 984–995. [2]

Chib, S., Greenberg, E., and Winkelmann, R. (1998), "Posterior Simulation and Bayes Factors in Panel Count Data Models," *Journal of Econometrics*, 86, 33–54. [4]

Conn, P. B., Johnson, D. S., Hoef, J. M. V., Hooten, M. B., London, J. M., and Boveng, P. L. (2015), "Using Spatiotemporal Statistical Models to Estimate Animal Abundance and Infer Ecological Dynamics From Survey Counts," *Ecological Monographs*, 85, 235–252. [2]

de Grange, L., Ibeas, A., and González, F. (2011), "A Hierarchical Gravity Model With Spatial Correlation: Mathematical Formulation and Parameter Estimation," *Networks and Spatial Economics*, 11, 439–465. [2,9]

Deng, D., Shahabi, C., Demiryurek, U., Zhu, L., Yu, R., and Liu, Y. (2016), "Latent Space Model for Road Networks to Predict Time-Varying Traffic," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1525–1534. [2]

Flaxman, S., Wilson, A., Neill, D., Nickisch, H., and Smola, A. (2015), "Fast Kronecker Inference in Gaussian Processes With Non-Gaussian Likelihoods," in *International Conference on Machine Learning*, pp. 607–616. [2]

Fotheringham, A. S., and O'Kelly, M. E. (1989), *Spatial Interaction Models: Formulations and Applications* (Vol. 1), Dordrecht: Kluwer Academic Publishers, p. 989. [2,9]

Hazelton, M. L. (2008), "Statistical Inference for Time Varying Origin–Destination Matrices," *Transportation Research Part B: Methodological*, 42, 542–552. [1]

Hung, Y., Joseph, V. R., and Melkote, S. N. (2015), "Analysis of Computer Experiments With Functional Response," *Technometrics*, 57, 35–44. [4]

Ishak, S., and Al-Deek, H. (2002), "Performance Evaluation of Short-Term Time-Series Traffic Prediction Model," *Journal of Transportation Engineering*, 128, 490–498. [1,2]

Li, B. (2005), "Bayesian Inference for Origin-Destination Matrices of Transport Networks Using the EM Algorithm," *Technometrics*, 47, 399–408. [1]

Li, L., Su, X., Zhang, Y., Lin, Y., and Li, Z. (2015), "Trend Modeling for Traffic Time Series Analysis: An Integrated Study," *IEEE Transactions on Intelligent Transportation Systems*, 16, 3430–3439. [1]

Liu, J., Cui, E., Hu, H., Chen, X., Chen, X. M., and Chen, F. (2017), "Short-Term Forecasting of Emerging On-Demand Ride Services," in *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, IEEE, pp. 489–495. [1]

Ma, J., Kockelman, K. M., and Damien, P. (2008), "A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods," *Accident Analysis & Prevention*, 40, 964–975. [4]

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013), "Predicting Taxi–Passenger Demand Using Streaming Data," *IEEE Transactions on Intelligent Transportation Systems*, 14, 1393–1402. [1,2]

NYC Taxi (2017), available at *http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml*. [9]

Okutani, I., and Stephanedes, Y. J. (1984), "Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory," *Transportation Research Part B: Methodological*, 18, 1–11. [1]

Perrakis, K., Karlis, D., Cools, M., Janssens, D., Vanhoof, K., and Wets, G. (2012), "A Bayesian Approach for Modeling Origin-Destination Matrices," *Transportation Research Part A: Policy and Practice*, 46, 200–212. [2,3]

Shao, H., Lam, W. H., Sumalee, A., Chen, A., and Hazelton, M. L. (2014), "Estimation of Mean and Covariance of Peak Hour Origin-Destination Demands From Day-to-Day Traffic Counts," *Transportation Research Part B: Methodological*, 68, 52–75. [1]

Stewart, J. A. (1994), "The Poisson-Lognormal Model for Bibliometric/Scientometric Distributions," *Information Processing & Management*, 30(2), 239–251. [4]

Sun, S., Zhang, C., and Yu, G. (2006), "A Bayesian Network Approach to Traffic Flow Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 7, 124–132. [2]

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82, 528–540. [4]

Tebaldi, C., and West, M. (1998), "Bayesian Inference on Network Traffic Using Link Count Data" (with discussion), *Journal of the American Statistical Association*, 93, 557–576. [11]

Tong, Y., Chen, Y., Zhou, Z., Chen, L., Wang, J., Yang, Q., Ye, J., and Lv, W. (2017), "The Simpler the Better: A Unified Approach to Predicting Original Taxi Demands Based on Large-Scale Online Platform," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1653–1662. [2]

Wagner-Muns, I. M., Guardiola, I. G., Samaranayke, V. A., and Kayani, W. I. (2018), "A Functional Data Analysis Approach to Traffic Volume Forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 19, 878–888. [1]

Wang, D., Liu, K., and Zhang, X. (2019), "Modeling of a Three-Dimensional Dynamic Thermal Field Under Grid-Based Sensor Networks in Grain Storage," *IISE Transactions*, 51, 531–546. [2]

Wemegah, T. D., and Zhu, S. (2017), "Big Data Challenges in Transportation: A Case Study of Traffic Volume Count From Massive Radio Frequency Identification (RFID) Data," in *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, IEEE, pp. 58–63. [2]

Xian, X., Li, J., and Liu, K. (2018), "Causation-Based Monitoring and Diagnosis for Multivariate Categorical Processes With Ordinal Information," *IEEE Transactions on Automation Science and Engineering*, 16, 886–897. [3]

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., and Ye, J. (2018), "Deep Multi-View Spatial-Temporal Network for Taxi Demand Prediction," arXiv no. 1802.08714. [1,2]

Ye, Z. S., and Tang, L. C. (2016), "Augmenting the Unreturned for Field Data With Information on Returned Failures Only," *Technometrics*, 58, 513–523. [4]

Zhang, A., Kang, J. E., and Kwon, C. (2017), "Incorporating Demand Dynamics in Multi-Period Capacitated Fast-Charging Location Planning for Electric Vehicles," *Transportation Research Part B: Methodological*, 103, 5–29. [1]

Zhang, J., Wang, F. Y., Wang, K., Lin, W. H., Xu, X., and Chen, C. (2011), "Data-Driven Intelligent Transportation Systems: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, 12, 1624–1639. [1]

Zhou, X., and Mahmassani, H. S. (2007), "A Structural State Space Model for Real-Time Traffic Origin-Destination Demand Estimation and Prediction in a Day-to-Day Learning Framework," *Transportation Research Part B: Methodological*, 41, 823–840. [1]