Uncovering Unmet Demand in Bike-sharing Systems Based on Bayesian Gaussian Decomposition of Time-varying OD Tensor

Xinghang Zhu[1], Xiaoxu Chen[1], Luis Miranda-Moreno[1], Lijun Sun[1,*]
[1]Department of Civil Engineering, McGill University, Canada (*: Corresponding Author)

Extended abstract submitted for presentation at the 12th Triennial Symposium on Transportation Analysis conference (TRISTAN XII)
June 22-27, 2025, Okinawa, Japan

February 15, 2025

Keywords: Bike-sharing system, Unmet demand, Bayesian Inference, Gaussian Process, Markov chain Monte Carlo

1 Introduction

The global rise in carbon emissions remains a pressing environmental issue, driven significantly by the transportation sector, which is now the third-largest source of greenhouse gas (GHG) emissions (Sun *et al.*, 2021). Bike-sharing systems (BSS), have emerged as a key sustainable transportation mode to mitigate urban congestion and emissions. Currently, BSS are experiencing rapid growth. In 2023, total Bike-sharing trips in North America increased by 20%, surpassing the pre-pandemic peak recorded in 2019 (NACTO, 2023).

However, BSS often struggle with capacity imbalances, leaving substantial unmet demand. Users frequently encounter shortages of available bikes and docks, creating "unmet demand". While observed data can record the origin-destination (OD) of each bike-sharing trip, it fails to capture the number of users who fail to pick up or drop off bikes, resulting in decreased utilization rates of BSS. Consequently, existing research focusing on BSS optimization and prediction relies solely on observed demand data and overlooks unmet demand. This limitation leads to inaccurate OD demand estimates and suboptimal operations. Thus, identifying unmet demand is crucial for the sustainable growth of bike-sharing systems.

Considering the difficulty of directly determining unmet demand, our study innovatively develops a new model to estimate the actual demand pattern, by mining the mathematical distribution of observed data. Then, unmet demand can be identified by comparing the discrepancies between actual and observed demand. Therefore, the primary objective of our research is to estimate the new indicators: the trip occurrence probabilities, $(\lambda_{i,j}^m, j = 1, 2, ..., S)$, which can reflect the true demand. We conceptualize the event: "During the m^{th} hour, the user arriving at Origin Station i chooses the Destination Station j from S Stations" as an experiment following a multinomial distribution. In this context, the trip occurrence probabilities, $\lambda_{i,j}^m$, serve as fixed parameters of the multinomial distribution, reflecting the fundamental demand of users, which remains unaffected by equipment shortages. However, $\lambda_{i,j}^m$ is a high-dimensional tensor with time-varying patterns, making it almost impossible to solve using conventional statistical methods. To extract this parameter, the proposed innovative model integrates a Bayesian Gaussian decomposition framework and employs Markov Chain Monte Carlo (MCMC) methods—specifically, Gibbs Sampling and Elliptical Slice Sampling—for parameter estimation, which will be introduced in Section 2.

2 Methodology: Bayesian Gaussian Decomposition

2.1 Problem Definition: Multinomial Distribution Framework

Assuming there are S bike-sharing stations and M time units, the tensor representing the observed origin-destination (OD) trips, denoted as $\mathbf{Y}$, has dimensions $S \times S \times M$. Given a specific origin station i, the choice of destination stations follows a multinomial probabilistic pattern characterized by a set of probabilities. Specifically, let $\boldsymbol{\lambda}_i^m = (\lambda_{i,1}^m, \lambda_{i,2}^m, \ldots, \lambda_{i,S}^m)$ represent the probability distribution over the S possible destination stations during the m^{th} time unit. The number of trips $\mathbf{y_i^m}$ taken from the origin i to each destination j, where $j = 1, 2, \ldots, S$, can then be described by a multinomial distribution:

$$\mathbf{y}_i^m \sim \text{Multinomial}(n_i^m, \boldsymbol{\lambda_i^m}), \quad P(\mathbf{y}_i^m \mid \boldsymbol{\lambda}_i^m) = n_i^m! \prod_{j=1}^{S} \frac{\left(\lambda_{i,j}^m\right)^{y_{i,j}^m}}{y_{i,j}^m!} \tag{1}$$

where n_i^m denotes the total trips generating from Origin Stations i during the m^{th} time unit. Now, we shift the crux of the problem from predicting the observed $\mathbf{Y}$ to estimating the trip occurrence probabilities $\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} \in \mathbb{R}^{S \times S \times M}$.

2.2 Model Structure: Tensor Decomposition with Latent Gaussian process

Since λ_i^m reflects the endogenous mode of human behavior in utilizing the bike-sharing system, general sampling methods are inadequate for capturing its temporal dynamics. To ensure the two properties of λ_i^m, namely being a probability vector and exhibiting time-varying nature, we combine natural Softmax parameterization with a Gaussian Process to estimate λ_i^m. Thus, we can summarize the research structure as shown in Fig 1.

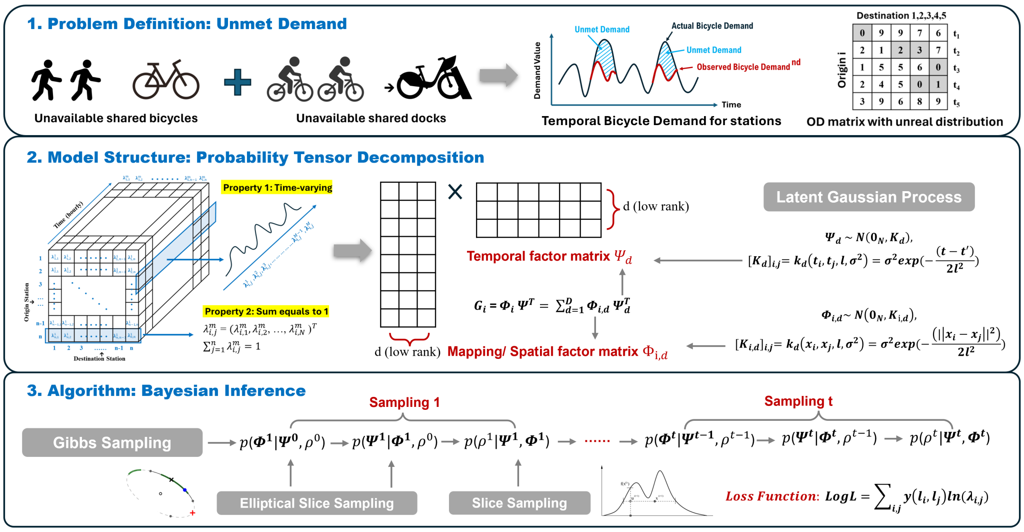

Figure 1 – *Research Structure*

Softmax parameterization can effectively convert the output elements into a probability vector that respects the constraint that all elements sum to one. The function is as follows:

$$\boldsymbol{\lambda}_i^m = \mathrm{Softmax}(\rho \mathbf{G}_i^m) = \begin{pmatrix} \frac{\exp(\rho G_{i,1}^m - \max(\rho \mathbf{G}_i^m))}{\sum_{j=1}^S \exp(\rho G_{i,j}^m)} \\ \vdots \\ \frac{\exp(\rho G_{i,S}^m - \max(\rho \mathbf{G}_i^m))}{1+\sum_{j=1}^S \exp(\rho G_{i,j}^n)} \end{pmatrix}, \tag{2}$$

where $\mathbf{G}_i^m = \left(G_{i,1}^m, G_{i,2}^m, \ldots, G_{i,S}^m\right)^\top \in \mathbb{R}^S$, and $\rho > 0$ is the temperature parameter, which is used to control the smoothness of the output probabilities.

To capture the time-varying nature of these distributions after applying the softmax parameterization, we introduce a Gaussian Process to model the temporal dynamics and spatial correlations. The temporal evolution of $\mathbf{G}_i$ is represented as $\mathbf{G}_i^m \mid \mathbf{G}_i^{m-1} \sim \mathcal{N}(\mathbf{G}_i^{m-1}, \sigma^2 \mathbf{I})$. However, the high dimensionality of $\mathbf{G} \in \mathbb{R}^{S \times S \times M}$ poses significant challenges for finding the optimal solution, although computing the likelihood of the multinomial distribution remains feasible. To address this, we reduce the dimensionality of $\mathbf{G}$ by transforming it from a tensor into a matrix:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_S \end{bmatrix} = \begin{bmatrix} \Phi_1 \\ \vdots \\ \Phi_S \end{bmatrix} \times \Psi^\top = \Phi \Psi^\top, where\ \mathbf{G}_i = \Phi_i \Psi^\top = \sum_{d=1}^D \phi_{i,d} \psi_d^\top, \tag{3}$$

where Φ denotes the mapping factor matrix and Ψ denotes the temporal factor matrix, $\Phi_i \in \mathbb{R}^{S \times D}$ and $\Psi_d \in \mathbb{R}^{N \times D}$. The low-rank assumption implies that $D \ll M$ and $D \ll S \times S$, thereby significantly reducing the number of parameters through the factorization of $\mathbf{G}$.

To capture the temporal dynamic smoothness, we assume that each column ψ_d of the temporal matrix Ψ is governed by a latent Gaussian process (Hall *et al.*, 2008). We select the squared exponential kernel as the kernel function for each column ψ_d. The SE kernel provides a continuous representation of temporal variations, which does not impose a strict assumption on the temporal dynamics. The functions of Latent Gaussian Process on temporal matrix Ψ are defined as follows:

$$\psi_d \sim \mathcal{N}(\mathbf{0_M}, \mathbf{K_d}), [\mathbf{K_d}]_{i,j} = k_d(t_i, t_j; l, \sigma^2) = \sigma^2 \exp\left(-\frac{(t_i - t_j)^2}{2l^2}\right) \tag{4}$$

where, l and σ^2 are the hyperparameters representing the lengthscale and variance, respectively. Additionally, spatial correlations may be present in bike-sharing systems, closer stations may have similar distribution. Consequently, for each vector column $\phi_{i,d}$, we can also set:

$$\phi_{i,d} \sim \mathcal{N}\left(\mathbf{0_N}, [\mathbf{K_{i,d}}]_{i,j}\right), [\mathbf{K_{i,d}}]_{i,j} = k_d(\mathbf{x}_i, \mathbf{x}_j; l, \sigma^2) = \sigma^2 \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2l^2}\right) \tag{5}$$

where $\|\mathbf{x}_i - \mathbf{x}_j\|$ denotes the Euclidean distance between spatial locations $\mathbf{x}_i$ for bike-sharing station i and $\mathbf{x}_j$ for bike-sharing station j. To ensure the positivity of the temperature parameter ρ, we model it using its log-transformed value. Specifically, we assign a Gaussian prior to $\log(\rho)$ by $\log(\rho) \sim \mathcal{N}(\mu_\rho, \sigma_\rho^2)$, where μ_ρ and σ_ρ^2 are scalar hyperparameters.

2.3 Algorithm: Gibbs sampling and Elliptical Slice sampling

Based on the Bayesian Gaussian Decomposition model, we utilize Bayesian inference to determine the posterior distributions of the parameters λ. The parameters of interest in the posterior distribution are denoted by $\Theta = \{\Phi, \Psi, \rho\}$. The posterior distribution $p(\Theta \mid \mathbf{Y}, t)$ is proportional to the product of the prior distribution and the likelihood function, as expressed by:

$$p(\Theta \mid \mathbf{Y}, \mathbf{t}) \propto p(\Theta \mid \mathbf{t})\, p(\mathbf{Y} \mid \Theta) \propto p(\Theta \mid \mathbf{t}) \prod_{m=1}^{M} p(\mathbf{y}_m \mid \Theta). \tag{6}$$

For complex models, this posterior distribution is often intractable to compute directly. Therefore, we employ Gibbs sampling, which is a specific MCMC method, which simplifies the sampling process by iteratively sampling each parameter from its conditional distribution, given the current values of the other parameters and the observed data (Gelfand, 2000).

Considering the high dimensionality of ψ_d and $\phi_{i,d}$, we select Elliptical Slice Sampling (ESS) as their sampling algorithm, which can handle the complexity of high-dimensional parameter spaces without requiring gradient information (Murray *et al.*, 2010), as shown in Algorithm 1. Given that the temperature factor ρ is a scalar, we utilize Slice Sampling to iteratively generate samples. The likelihood $\mathcal{L}(\boldsymbol{\psi}_d)$ can be calculated as follows.

$$\mathcal{L}(\boldsymbol{\psi}_d) = p(\mathbf{Y} \mid \Phi, \Psi, \rho) = \prod_{n=1}^{N} p(\mathbf{y}_n \mid \lambda_n), \tag{7}$$

Algorithm 1 Elliptical slice sampling for each column $\boldsymbol{\psi}_d$ of factor matrix $\boldsymbol{\Psi}$.

Require: Current state $\boldsymbol{\psi}_d$, covariance matrix $\boldsymbol{K}_d$, likelihood function $L(\boldsymbol{\psi}_d)$
Ensure: a new state $\boldsymbol{\psi}_d'$
1: Choose ellipse: $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{K}_d)$
2: Log-likelihood threshold: $\gamma \sim \text{Uniform}[0,1]$, $\log c = \log L(\boldsymbol{\psi}_d) + \log \gamma$
3: Draw an initial sampling range: $\theta \sim \text{Uniform}[0, 2\pi]$, $\theta_{\min} = \theta - 2\pi$, $\theta_{\max} = \theta$
4: $\boldsymbol{\psi}_d' = \boldsymbol{\psi}_d \cos\theta + \boldsymbol{\nu} \sin\theta$
5: **if** $\log L(\boldsymbol{\psi}_d') > \log c$ **then**
6:     **return** $\boldsymbol{\psi}_d'$
7: **else**
8:     Shrink the sampling range and try a new point:
9:     **if** $\theta \leq 0$ **then**
10:         $\theta_{\min} = \theta$
11:     **else**
12:         $\theta_{\max} = \theta$
13:     **end if**
14:     $\theta \sim \text{Uniform}[\theta_{\min}, \theta_{\max}]$
15:     **Go to** Step 4.
16: **end if**

3 Experiments and Results

Based on the Bayesian Gaussian Decomposition model, we utilized the OD trip data from Montreal's BIXI bike-sharing system to conduct experiments. We selected the data in August 2023 and set a bi-hourly time interval. Since BIXI station locations frequently change, we divided the map of Montreal into a grid with cell dimensions of 500 x 500 meters, resulting in 420 grid cells containing BIXI bike-sharing stations ($\mathbf{Y} \in \mathbb{R}^{420\times420\times84}$). According to the pre-experiment, we set $\sigma_0^2, \sigma_p^2 = 1$ and $\mu_p = \ln(0.1)$. Through multiple experiments, we found that a low-rank setting of 6 achieved the maximum log-likelihood and optimal performance, as shown in Fig 2.

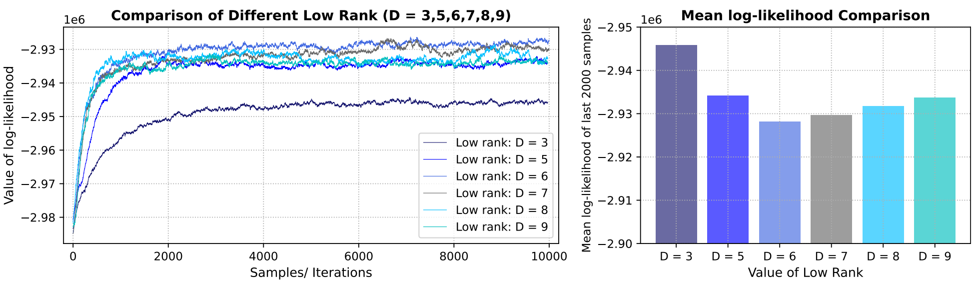

Figure 2 – *Performance comparison of different low rank (Right: Comparison of log-likelihoods during 10000 iterations; Left: Comparison of mean log-likelihood over the last 2000 iterations).*

Figure 3 displays the results for the temporal vector ψ_d from the last 200 samples of the 10,000-iteration sampling process. Each temporal vector demonstrates a clear daily cyclic pattern, despite no periodic constraints in the Gaussian Kernel. This outcome indicates that our model accurately auto-captures the fundamental demand patterns and temporal dynamics across the bike-sharing network.

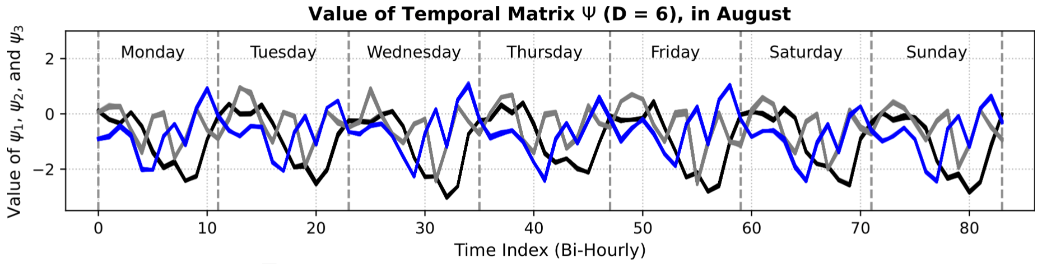

Figure 3 – *Value of Temporal Matrix* Ψ *(*$D = 6$*)*

We randomly selected Cluster 1 as origins to analyze the distribution of λ (Figure 4). The observed OD trip data $\mathbf{Y}$ is highly scattered and sparse, making it difficult to identify clear dynamic patterns or latent demand, suggesting that $\mathbf{Y}$ may not fully reflect the true OD demand. In contrast, our proposed model extracts potential dynamic patterns from $\mathbf{Y}$, offering a clearer representation of demand under ideal conditions. The temporal distribution of λ for trips from fixed origins follows specific patterns influenced by the six time vectors ψ_d.

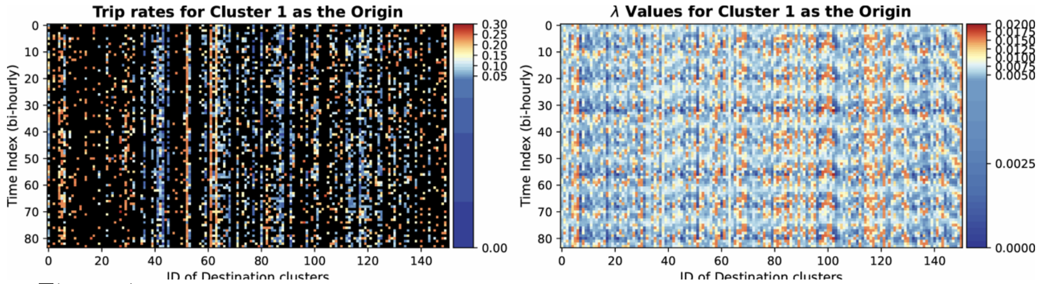

Figure 4 – *Comparison of observed trip rates and estimated trip occurrence probabilities* $\boldsymbol{\lambda}$

Finally, after comparing the actual demand $\boldsymbol{\lambda}$ and observed demand $\boldsymbol{Y}$, we conduct the unmet demand analysis as shown in Figure 5. The results shows that high-demand origin clusters are radially distributed from the city center, indicating a bike shortage despite existing infrastructure. Unmet demand is most prominent in the city center, extending southward, mainly during peak hours (7:00–9:00 AM, 5:00–7:00 PM). These findings validate the model's effectiveness, with future research incorporating urban planning data for deeper insights.

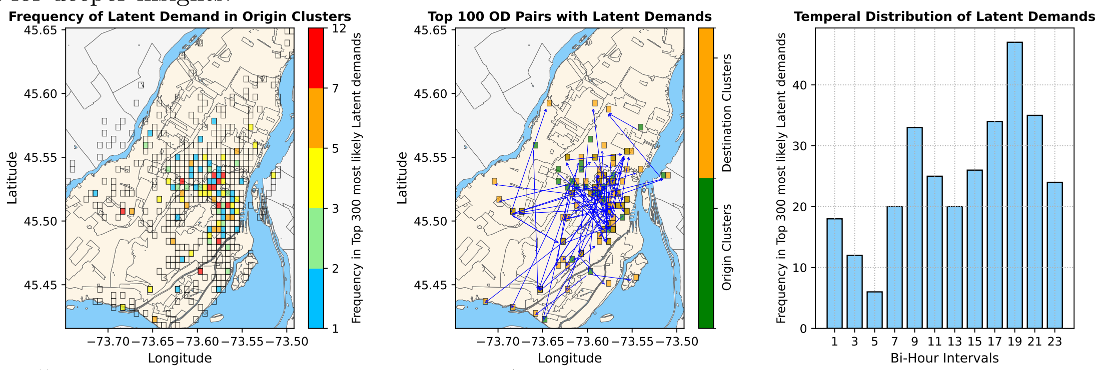

Figure 5 – *Spatial distribution of latent demand (Left: Spatial distribution of origins with the highest unmet demand, Middle: Top 100 OD pairs most likely to exhibit unment demand, Right: Temporal distribution of unmet demand).*

References

Gelfand, Alan E. 2000. Gibbs sampling. *Journal of the American statistical Association*, **95**(452), 1300–1304.

Hall, Peter, Müller, Hans-Georg, & Yao, Fang. 2008. Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **70**(4), 703–723.

Murray, Iain, Adams, Ryan, & MacKay, David. 2010. Elliptical slice sampling. *Pages 541–548 of: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, vol. 9. PMLR.

NACTO. 2023. *National Association of City Transportation Officials: Shared Micromobility in 2023*. Accessed: 2024-07-22.

Sun, Lishan, Zhang, Tong, Liu, Shuli, Wang, Kun, Rogers, Tom, Yao, Liya, & Zhao, Pengfei. 2021. Reducing energy consumption and pollution in the urban transportation sector: A review of policies and regulations in Beijing. *Journal of Cleaner Production*, **285**, 125339.