Bayesian Inference of Time-varying Origin-Destination Matrices from Boarding/Alighting Counts for Transit Services

Xiaoxu Chen, Zhanhong Cheng, and Lijun Sun*

Department of Civil Engineering, McGill University, Montreal, Canada
xiaoxu.chen@mail.mcgill.ca, zhanhong.cheng@mcgill.ca, lijun.sun@mcgill.ca
* Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 8, 2024

Keywords: origin-destination matrix estimation, Bayesian inference, Markov chain Monte Carlo, matrix factorization, Gaussian processes, transit services

1 INTRODUCTION

The origin-destination (OD) matrix for a bus route captures passenger flows from one stop to another, serving as a comprehensive representation of passenger demand. These matrices can be defined either in an aggregated manner (e.g., overall demand during morning rush hours) or in detail for each bus journey. An accurate estimation of the OD matrices and a good understanding of how these matrices evolve over time are critical for transit agencies in making planning and operational decisions, such as route design (Ahern *et al.*, 2022), service scheduling (Martínez *et al.*, 2014), timetabling (Sun *et al.*, 2014), and fleet allocation (Gkiotsalitis *et al.*, 2019).

This study aims to infer time-varying OD matrices from boarding/alighting counts with high temporal resolution. We extend the work of Hazelton (2010) and develop a temporal Bayesian model for inferring transit OD matrices at the individual bus level. To model the discrete count data, we assume that the number of alighting passengers at subsequent bus stops, given a boarding stop, follows a multinomial distribution. To better characterize the temporal patterns in passenger demand, we assume that the parameters (i.e., assignment probabilities) of the multinomial distribution vary smoothly over time, thus allowing for building a time-varying model using counts observed from a limited number of bus journeys. We introduce a latent variable matrix and use it to parameterize the time-varying multinomial distributions through the softmax transformation. In addition, we propose using matrix factorization to parameterize the latent matrix as the product of a mapping factor matrix and a temporal factor matrix, which substantially reduces the number of parameters. To encode a temporally smooth structure in the matrix, we impose Gaussian process priors on the columns of the temporal factor matrix, which consequently ensure that the assignment probabilities vary smoothly over time. For model inference, we follow Hazelton (2010) and also develop a two-stage algorithm based on MCMC. In the first stage, we sample latent OD matrices conditional on parameters using the Metropolis-Hastings sampling algorithm with the proposal distribution proposed by Hazelton (2010), which efficiently bypasses the need to enumerate the large number of feasible OD matrices that align with observed boarding and alighting counts for each bus trip. In the second stage, we sample model parameters conditional on latent OD matrices obtained from the first stage. The key challenge in this step is to efficiently sample latent Gaussian processes with non-Gaussian likelihood, where the posterior no longer has an analytical formulation. To address this issue, we use elliptical slice sampling (ESS), an efficient algorithm developed by Murray *et al.* (2010), to sample the temporal factor matrix. We evaluate our proposed model using real-world APC data and true OD matrices from three bus routes in an anonymous city. We compare the performance of the proposed temporal model to a non-temporal variant, and the results show that the temporal Bayesian model outperforms the non-temporal variant, confirming the importance and value of developing a time-varying model. In addition, we also compare our model with the widely used IPF method, and the results show that our model can achieve superior performance in deterministic estimation.

2 PROBLEM DEFINITION

Consider a bus route comprising S stops at which passengers can board and alight. Let u_i and v_i denote the numbers of boarding passengers and alighting passengers at stop i, respectively, for $i = 1, 2, \ldots, S$. Such boarding/alighting counts are available from the APC systems. In general, we will see neither alighting passengers at stop 1 nor boarding passengers at stop S, so we can fix $v_1 = u_S = 0$. For a bus journey (i.e., a trip from stop 1 to stop S), we denote by $y_{i,j}$ the number of passengers who board at stop i and alight at stop j, which cannot be observed directly. Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_S)^\top$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_S)^\top$ be the vectors of boarding and alighting counts at the stops, respectively; we then denote by $\boldsymbol{x} = \left(\boldsymbol{u}^\top, \boldsymbol{v}^\top\right)^\top = (u_1, u_2, \ldots, u_S, v_1, v_2, \ldots, v_S)^\top$ as the aggregation of observed counts for a bus trip.

Our study aims to infer the OD matrix $\boldsymbol{Y} = (y_{i,j})_{S\times S}$. For a bus route/service, it is clear that passengers can only travel to downstream stops. Thus, we fix $y_{i,j} = 0$ for all cases where $i \geq j$, and focus exclusively on the upper-triangular part of $\boldsymbol{Y}$. Following Hazelton (2010), we denote by $\boldsymbol{y}_i$ the number of passengers traveling from the i-th stop to the subsequent stops along the bus route. For instance, $\boldsymbol{y}_1 = (y_{1,2}, y_{1,3}, \ldots, y_{1,S})^\top$ denotes the passenger counts from the initial stop to all subsequent stops along the route. This definition continues to $\boldsymbol{y}_{S-1} = (y_{S-1,S})$, which represents the number of passengers traveling from the second-to-last stop to the last stop. Next, we stack these passenger counts into a single OD vector $\boldsymbol{y} = \left(\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_{S-1}^\top\right)^\top \in R^{S(S-1)/2}$. Although $\boldsymbol{y}$ is not directly observable, its relationship with the observed boarding and alighting counts can be expressed as follows $\boldsymbol{x} = \mathbf{A}\boldsymbol{y}$, where $\mathbf{A}$ is a $2S \times M$ binary routing matrix.

To account for multiple bus journeys, we extend the notation to include a bus index n, which represents the n-th bus journey, with $\boldsymbol{x}^n = \left(\boldsymbol{u}^{n\top}, \boldsymbol{v}^{n\top}\right)^\top$ and $\boldsymbol{y}^n = \left(\boldsymbol{y}_1^{n\top}, \ldots, \boldsymbol{y}_{S-1}^{n\top}\right)^\top$. To effectively model the dynamic/time-varying nature of OD matrices/vectors, our model incorporates temporal information. Specifically, we denote by t^n the departure time at the initial stop for the n-th bus trip/journey. For a total of N bus journeys over a studied period, we define $\mathcal{X} = \{\boldsymbol{x}^n \mid n = 1, 2, \ldots, N\}$ as the set of observed boarding and alighting counts and $\boldsymbol{t} = \left(t^1, t^2, \ldots, t^N\right)^\top$ as the vector of observed departure time. The primary objective is to estimate the set of OD vectors, denoted as $\mathcal{Y} = \{\boldsymbol{y}^n \mid n = 1, 2, \ldots, N\}$, using observed data set $\mathcal{X}$ and $\boldsymbol{t}$. This problem is challenging because the number of unknown quantities (OD vector) is much larger than the number of observations (boarding and alighting counts) in the linear system, resulting in a challenging statistical linear inverse problem (Vardi, 1996, Hazelton, 2010). We denote by $\mathcal{H}(\boldsymbol{x}^n) = \{\boldsymbol{y}^n \mid \boldsymbol{x}^n = \mathbf{A}\boldsymbol{y}^n\}$ the solution space that encompasses all feasible OD vectors consistent with the observation $\boldsymbol{x}^n$. In general, the solution space could be very large even for a route with a modest number of stops.

3 METHODOLOGY

Let $\lambda_{i,j}^n$ be the probability that a passenger boarding at stop i will alight at stop j during the n-th bus trip. Furthermore, let $\boldsymbol{\lambda}_i^n = \left(\lambda_{i,i+1}^n, \ldots, \lambda_{i,S}^n\right)^\top$ be the alighting probabilities of downstream stops for a passenger boarding at stop i, and the sum of these probabilities is one, i.e.,

$\sum_{j=i+1}^{S} \lambda_{i,j}^{n} = 1$. Next, let $\boldsymbol{\lambda}^{n} = \left(\boldsymbol{\lambda}_{1}^{n\top}, \ldots, \boldsymbol{\lambda}_{S-1}^{n}{}^{\top}\right)^{\top}$ denote probabilities for all the corresponding OD entries of the n-th bus trip. Assuming that passengers make decisions independently, $\boldsymbol{y}_{i}^{n}$ follows a multinomial distribution $\boldsymbol{y}_{i}^{n} \sim \text{Multinomial}\left(u_{i}^{n}, \boldsymbol{\lambda}_{i}^{n}\right)$. Specifically, it can be represented as $p\left(\boldsymbol{y}_{i}^{n} \mid u_{i}^{n}, \boldsymbol{\lambda}_{i}^{n}\right) = u_{i}^{n}! \prod_{j=i+1}^{S} \frac{\lambda_{i,j}^{n}{}^{y_{i,j}^{n}}}{y_{i,j}^{n}!}$, and the likelihood of observing $\boldsymbol{x}^{n}$ becomes

$$L\left(\boldsymbol{\lambda}^{n}\right) = p\left(\boldsymbol{x}^{n} \mid \boldsymbol{\lambda}^{n}\right) = \sum_{\boldsymbol{y}^{n} \in \mathcal{H}(\boldsymbol{x}^{n})} \prod_{i=1}^{S-1} u_{i}^{n}! \prod_{j=i+1}^{S} \frac{\lambda_{i,j}^{n}{}^{y_{i,j}^{n}}}{y_{i,j}^{n}!}. \tag{1}$$

For modeling multiple bus journeys in a day, we expect $\boldsymbol{\lambda}_{i}^{n}$ to vary smoothly from one bus to the next (or over time). In this case, we need an effective parameterization that produces time-varying multinomial probabilities. We employ a natural softmax parameterization $\boldsymbol{\lambda}_{i}^{n} = \text{Softmax}\left(\rho \boldsymbol{G}_{i}^{n}\right)$, where $\boldsymbol{G}_{i}^{n} = \left(G_{i,i+1}^{n}, G_{i,i+2}^{n}, \ldots, G_{i,S-1}^{n}\right)^{\top} \in R^{S-i-1}$, and $\rho > 0$ is the temperature parameter, which can help to learn good sharpness/smoothness of the probability distribution. Next, we denote the collection of $\boldsymbol{G}_{i}^{n}$ over N bus journeys by the matrix $\boldsymbol{G}_{i} = \left[\boldsymbol{G}_{i}^{1}, \boldsymbol{G}_{i}^{2}, \ldots, \boldsymbol{G}_{i}^{N}\right]$. Next, we assume $\boldsymbol{G}_{i}$ has a low-rank structure $\boldsymbol{G}_{i} = \boldsymbol{\Phi}_{i} \boldsymbol{\Psi}^{\top} = \sum_{d=1}^{D} \boldsymbol{\phi}_{i,d} \boldsymbol{\psi}_{d}^{\top}$, where $\boldsymbol{\Phi}_{i} \in R^{(S-i-1)\times D}$, $\boldsymbol{\Psi}_{\boldsymbol{d}} \in R^{N\times D}$, and $\boldsymbol{\phi}_{i,d}$ and $\boldsymbol{\psi}_{d}$ are the d-th column of $\boldsymbol{\Phi}_{i}$ and $\boldsymbol{\Psi}$, respectively. To encode temporal smoothness in $\boldsymbol{G}$, we assume that each column $\boldsymbol{\psi}_{d}$ in $\boldsymbol{\Psi}$ is generated from a latent Gaussian process by taking values at bus departure times $\boldsymbol{t}$ with kernel/covariance function $k_{d}\left(t, t'; \boldsymbol{\eta}_{d}\right)$ where $\boldsymbol{\eta}_{d}$ is the vector of kernel hyperparameters. For model inference, we develop a two-stage algorithm based on MCMC, which can be found in the detailed work (Chen *et al.*, 2024).

4 RESULTS

To evaluate our approach, we use high-quality data from three distinct bus routes in a city—a short route with 22 stops, a medium route with 40 stops, and a long route with 72 stops. We apply the proposed model to infer/estimate OD matrices based on the counts and use the real OD matrices to evaluate the performance. We compare the performance of our model with the widely used IPF method (Ben-Akiva *et al.*, 1985).

To demonstrate the importance of integrating temporal dynamics in OD matrix estimation, we compare the performance of the temporal Bayesian model with a non-temporal variant. The non-temporal approach assumes static parameters and is derived from our model with rank $D = 1$ and $\boldsymbol{\Psi} = \mathbf{1}_{N\times 1}$. This ensures that the N journeys share the same alighting probabilities. We evaluate the log-likelihood of true OD matrices given the estimated multinomial parameters. Table 1 presents the log-likelihood of different models for the estimation of OD matrices. For the temporal Bayesian model, we implement four variants with different ranks (1, 2, 4 and 6). First, we compare the static model with the temporal model with $D = 1$. The key difference between these two models is the assumption of $\boldsymbol{\Psi}$—the static model defines $\boldsymbol{\Psi}$ as a column vector of ones, while the temporal model treats $\boldsymbol{\Psi}$ as a random vector generated from a Gaussian process. From the results, we can see that having a temporal component can enhance the quality of the model, confirming the importance of time-varying parameters for OD estimation.

Fig. 1 presents true and estimated OD flows at the journey level derived from the IPF method and our model for the three routes. Because IPF is a deterministic method, for the Bayesian method, we use the posterior mean as the estimated demand for model evaluation. The diagonal line of each plot presents the reference line of perfect estimation where estimated flows would align exactly with the true flows. The method with the dots closer to the reference line has the more accurate estimation. We can observe that the dots obtained from our model are closer to the reference line for all bus routes, indicating that our model outperforms the IPF method. Moreover, we use the root mean square error (RMSE) to compare the performance of

Table 1 – *Log-likelihood of different models for OD matrices estimation.*

		Static model	Temporal Bayesian model			
			$D = 1$	$D = 2$	$D = 4$	$D = 6$
Short route	Mean	-35328.77	-34829.07	-33728.49	-33064.69	-32874.45
	Standard deviation	98.63	86.65	79.14	86.75	85.54
Medium route	Mean	-63599.00	-62915.15	-62141.88	-61652.34	-61539.26
	Standard deviation	165.46	150.24	134.89	156.80	123.00
Long route	Mean	-47449.85	-47078.95	-46179.38	-45722.42	-45722.37
	Standard deviation	158.43	139.10	134.59	134.88	124.12

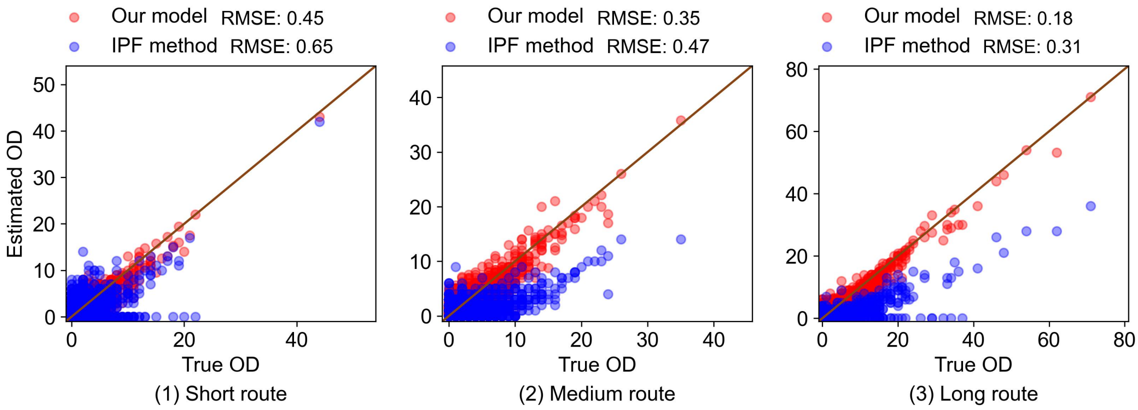

Figure 1 – *True and estimated OD flow of IPF and our proposed model for different routes.*

our Bayesian model and the IPF method. The proposed model gives much smaller RMSE values than those obtained from IPF, which demonstrates that our model outperforms IPF method.

References

Ahern, Zeke, Paz, Alexander, & Corry, Paul. 2022. Approximate multi-objective optimization for integrated bus route design and service frequency setting. *Transportation Research Part B: Methodological*, **155**, 1–25.

Ben-Akiva, Moshe, Macke, Peter P, & Hsu, Poh Ser. 1985. Alternative Methods to Estimate Route-Level Trip Tables and Expand On-Board Surveys. *Transportation Research Record*, **1037**, 1–11.

Chen, Xiaoxu, Cheng, Zhanhong, & Sun, Lijun. 2024. Bayesian Inference of Time-Varying Origin-Destination Matrices from Boarding/Alighting Counts for Transit Services. *arXiv preprint arXiv:2403.04742*.

Gkiotsalitis, Konstantinos, Wu, Zongxiang, & Cats, Oded. 2019. A cost-minimization model for bus fleet allocation featuring the tactical generation of short-turning and interlining options. *Transportation Research Part C: Emerging Technologies*, **98**, 14–36.

Hazelton, Martin L. 2010. Statistical inference for transit system origin-destination matrices. *Technometrics*, **52**(2), 221–230.

Martínez, Héctor, Mauttone, Antonio, & Urquhart, María E. 2014. Frequency optimization in public transportation systems: Formulation and metaheuristic approach. *European Journal of Operational Research*, **236**(1), 27–36.

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*. JMLR Workshop and Conference Proceedings.

Sun, Lijun, Jin, Jian Gang, Lee, Der-Horng, Axhausen, Kay W, & Erath, Alexander. 2014. Demand-driven timetable design for metro services. *Transportation Research Part C: Emerging Technologies*, **46**, 284–299.

Vardi, Yehuda. 1996. Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, **91**(433), 365–377.