McGill

Bayesian Inference of Time-Varying Transit OD Matrices from Boarding and Alighting Counts

Xiaoxu Chen

Department of Civil Engineering
McGill University
October 11, 2024

Xiaoxu Chen

Zhanhong Cheng

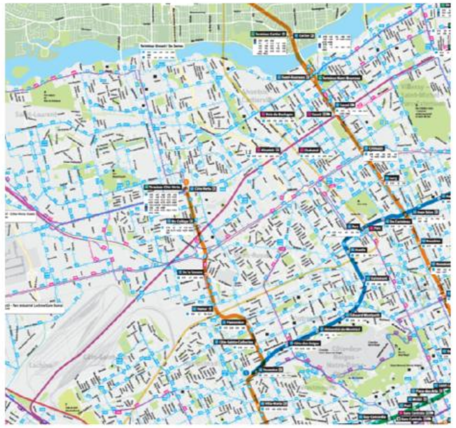
Lijun Sun

Motivation

- What is OD matrix?
  - An OD matrix represents the number of passengers traveling between a pair of stops along a bus route
    - ✓ Aggregated manner (e.g., overall demand in morning peak hours)
    - ✓ **Individual bus-journey level** (**time-varying**)

	D1	D2	D3	D4	D5
O1	0	2	3	1	1
O2	0	0	4	2	1
O3	0	0	0	2	1
O4	0	0	0	0	2
O5	0	0	0	0	0

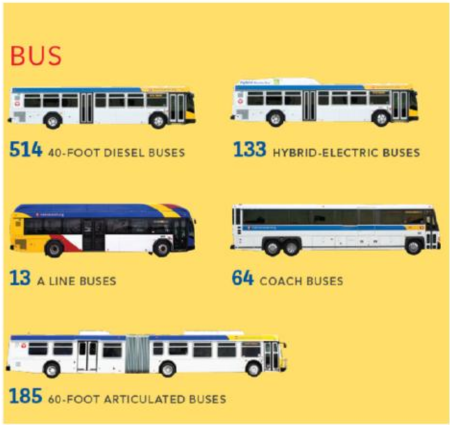
- Why do we need OD matrix?
  - OD matrices are important input for transit service planning, management, and operations.



Bus route design

Depart A	Arrive B
6:00	6:30
6:15	6:45
6:30	7:00
6:45	7:15
7:00	7:30
7:15	7:45
7:30	8:00
7:45	8:15
8:00	8:30
8:15	8:45
8:30	9:00
8:45	9:15
9:00	9:30

Scheduling



Fleet allocation

Advanced data collection techniques (AFC and APC) to obtain OD matrices

- **Automatic fare collection (AFC) system**
  - For systems that require both tapping-in and tapping-off
    - Trip information is generated (time, origin, destination, bus_id)
    - Only available in a few cities around the world (e.g., Singapore)
  - For systems that require only tapping-in
    - Partial trip information (time, origin, ~~destination~~, bus_id)
    - Destination inference models (e.g., with trip chain assumptions)

- **Automatic passenger counting (APC) system**
  - Europe, North America
  - Aggregated boarding/alighting counts
  - No individual trip information
  - The focus of this talk

Bayesian concept

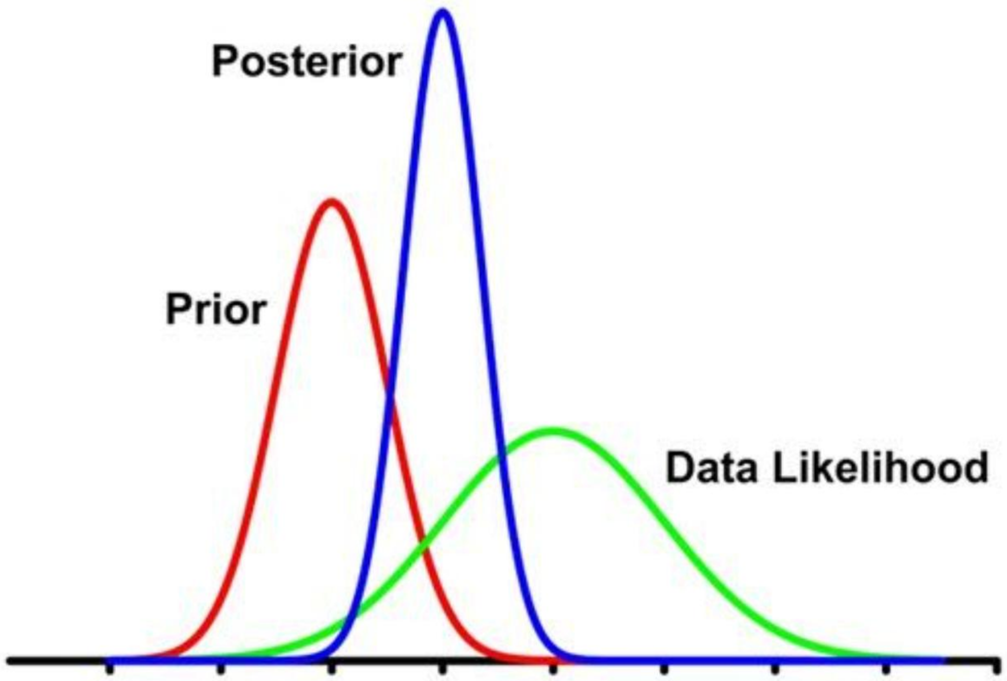
Bayes' theorem: combine prior knowledge and observations

Likelihood Prior

$$p(\boldsymbol{\theta} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathcal{D})}$$

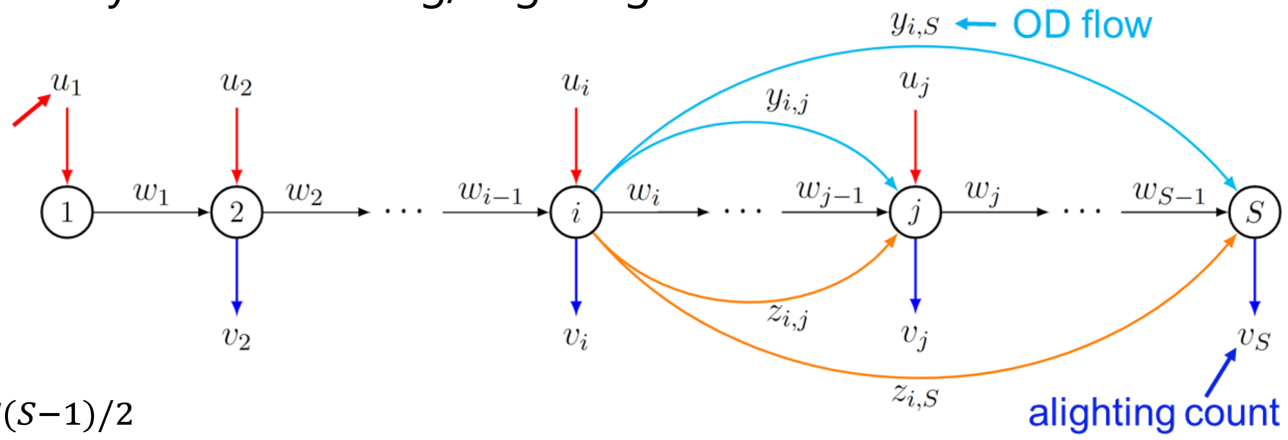
Posterior

$$p(\boldsymbol{y}^* \mid \mathcal{D}) = \int p(\boldsymbol{y}^* \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D})\, d\boldsymbol{\theta}$$



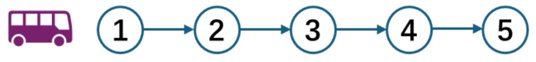
Research question

- Estimate OD matrix for each bus journey from boarding/alighting counts
- For each bus:
  - OD matrix $\boldsymbol{Y}$
  - Boarding vector $\boldsymbol{u} \in N^{S-1}$
  - Alighting vector $\boldsymbol{v} \in NS^{-1}$



- Denote by $\boldsymbol{y}$ the OD vector $\boldsymbol{y} \in N^{S(S-1)/2}$
- We have $\boldsymbol{x} = \binom{\boldsymbol{u}}{\boldsymbol{v}} = \boldsymbol{A}\boldsymbol{y}$
- Question: estimate the distribution $p(\boldsymbol{y})$
- This is a typical **inverse problem**:
  - Not identifiable when $S \geq 4$
  - There are a large # of feasible solutions

$$\underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 1&1&1&1&1&0&0&0&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&1&1&1&0&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1 \\ 1&0&0&0&0&0&0&0&0&0&0&0&0&0&0 \\ 0&1&0&0&0&1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&1&0&0&1&0&0&0&0&0 \\ 0&0&0&1&0&0&0&1&0&0&1&0&1&0&0 \\ 0&0&0&0&1&0&0&0&1&0&0&1&0&1&1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} y_{1,2} \\ y_{1,3} \\ y_{1,4} \\ y_{1,5} \\ y_{1,6} \\ y_{2,3} \\ y_{2,4} \\ y_{2,5} \\ y_{2,6} \\ y_{3,4} \\ y_{3,5} \\ y_{3,6} \\ y_{4,5} \\ y_{4,6} \\ y_{5,6} \end{bmatrix}}_{\boldsymbol{y}}$$



	D1	D2	D3	D4	D5	B_i
O1	0	?	?	?	?	7
O2	0	0	?	?	?	7
O3	0	0	0	?	?	3
O4	0	0	0	0	?	2
O5	0	0	0	0	0	0
A_j	0	2	7	5	5	

Industry practice

- Iterative Proportional Fitting (IPF)
- Morning/afternoon rush hours, off-peak hours (total boarding and alighting)
- “Onboard survey”: asking people their **origin-destination** and **purpose** for the trip
- Use to obtained matrix for that bus as a “seed” matrix, scaling the “seed” matrix to match the data
- Limitation: “biased” seed, and results should **NOT** be deterministic

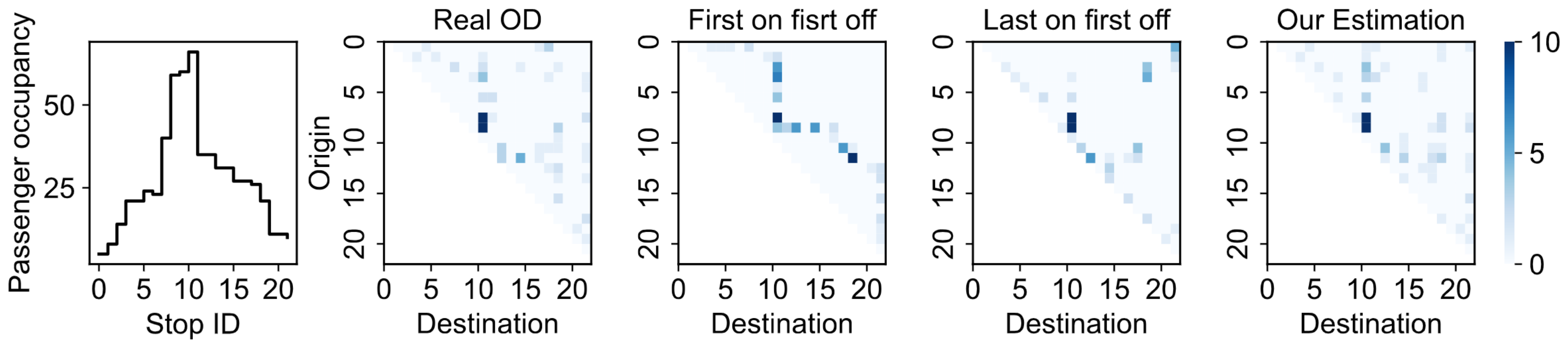


Figure 1 Illustration of the uncertainty of OD solution.

Solution for a single bus

- Current literature: Hazelton (2010)
- Assuming each entry (count) follows a Poisson distribution
  - With demand "rate" $\boldsymbol{\lambda}^n$ for bus n
  - The likelihood is

$$L(\boldsymbol{\lambda}^n) = p(\boldsymbol{x}^n \mid \boldsymbol{\lambda}^n) = \sum_{\boldsymbol{y}^n} p(\boldsymbol{x}^n \mid \boldsymbol{y}^n, \boldsymbol{\lambda}^n)\, p(\boldsymbol{y}^n \mid \boldsymbol{\lambda}^n)$$
$$= \sum_{\boldsymbol{y}^n \in \mathcal{H}(\boldsymbol{x}^n)} p(\boldsymbol{y}^n \mid \boldsymbol{\lambda}^n)$$

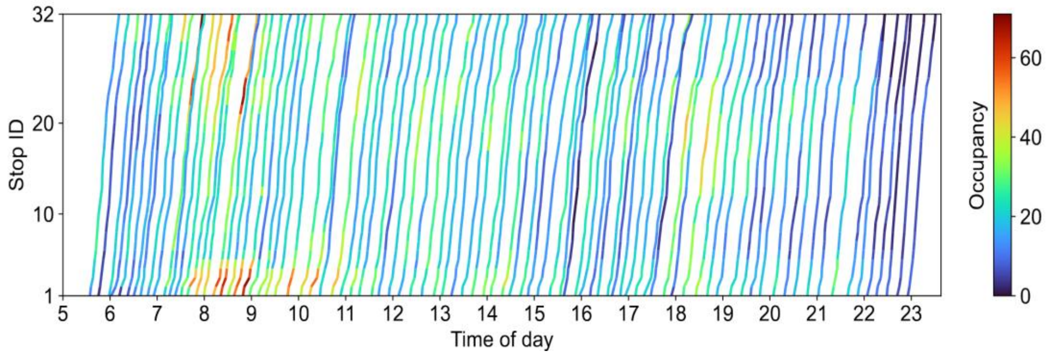
i \ j	1	2	3	4	5
1		1.5	2.0	1.8	5.1
2			4.9	1.5	2.5
3				0.9	5.2
4					2.4
5					

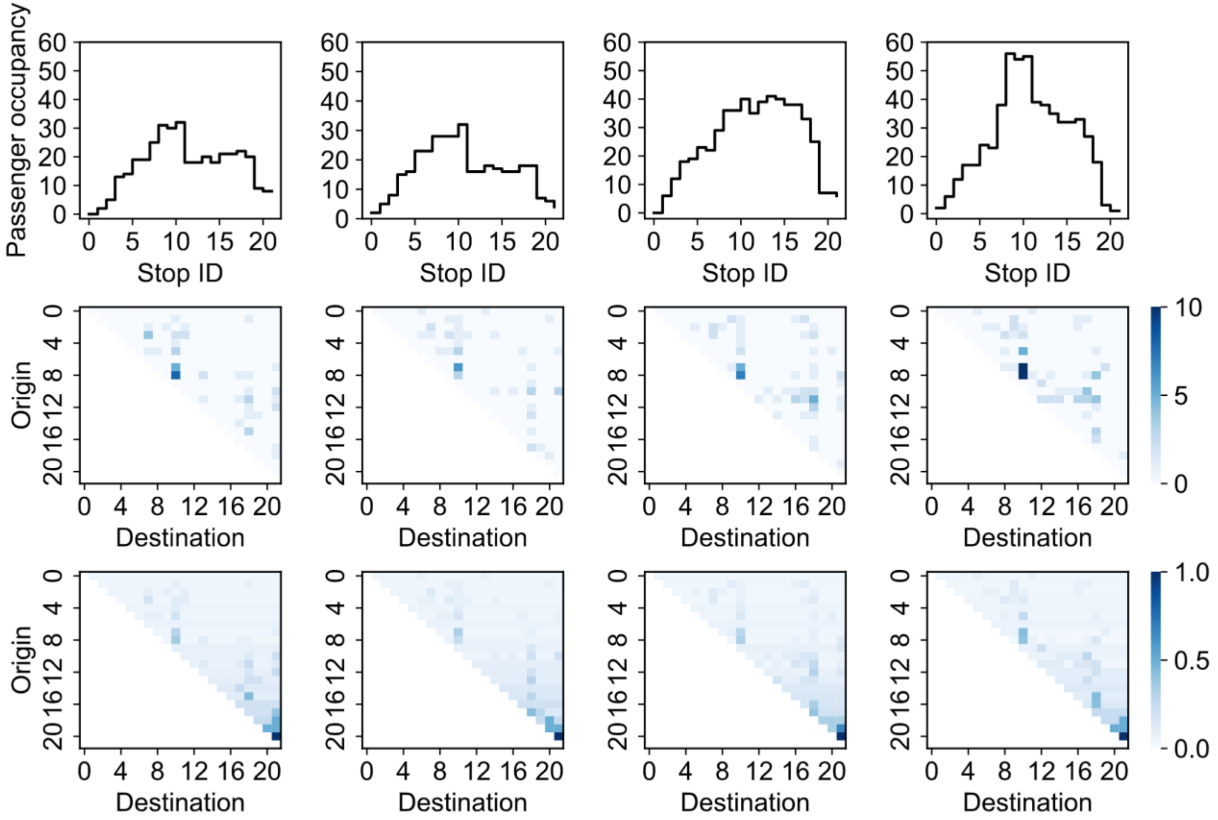
- Computing the likelihood: Enumerating all potential solutions
- Assuming all buses share the same "rate" parameter (static model)

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

For multiple buses

- For multiple buses in a day, can we assume $\boldsymbol{\lambda}^n$ remain the same?





Alighting probabilities

- Both demand and **supply** are time-varying
- $\boldsymbol{\lambda}^n$ vary substantially even for consecutive buses
- Anything in common among these buses?
- **Smooth alighting probability**

A Bayesian model

- Important **prior knowledge**
  - We expect alighting probability to vary, but it should vary **smoothly** over time
  - In other words, the passengers that take two consecutive vehicles (veh i and veh $i+1$) at the same stop should have similar destination profiles

- **Our solution**
  - Instead of modeling $\boldsymbol{\lambda}^n$, we model the probability $p_i^n(d=j)$
  - Each row of OD matrix represents a multinomial distribution (a simplex: $\sum_j p_i^n(d=j) \; = 1$)
  - Replace the Poisson likelihood with Multinomial likelihood

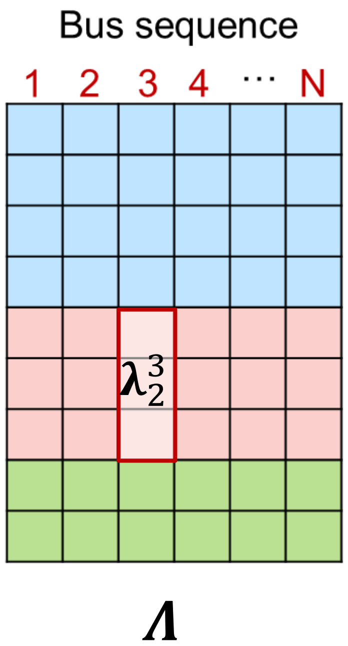
i \ j	1	2	3	4	5
1		0.1	0.2	0.1	0.7
2			0.3	0.2	0.5
3				0.4	0.6
4					1.0
5					

A Bayesian model

- **Parameterization—Time-varying multinomial probability**

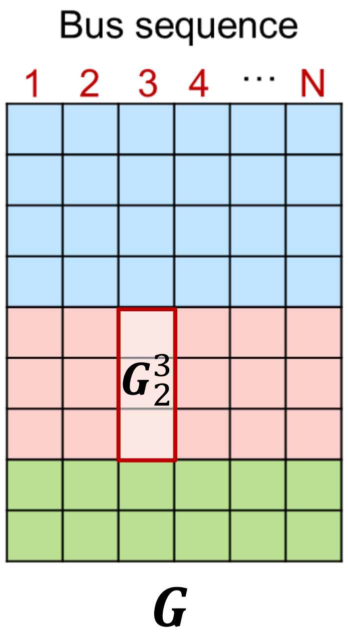
i \ j	1	2	3	4	5
1		0.1	0.2	0.1	0.7
2			0.3	0.2	0.5
3				0.4	0.6
4					1.0
5					

Vectorization →



$$\boldsymbol{\lambda}_i^n = \operatorname{Softmax}\left(\rho \boldsymbol{G}_i^n\right)$$

$$\begin{pmatrix} \frac{\exp\left(\rho G_{i,i+1}^n\right)}{1+\sum_{j=i+1}^{S-1} \exp\left(\rho G_{i,j}^n\right)} \\ \frac{\exp\left(\rho G_{i,i+2}^n\right)}{1+\sum_{j=i+1}^{S-1} \exp\left(\rho G_{i,j}^n\right)} \\ \vdots \\ \frac{\exp\left(\rho G_{i,S-1}^n\right)}{1+\sum_{j=i+1}^{S-1} \exp\left(\rho G_{i,j}^n\right)} \\ \frac{1}{1+\sum_{j=i+1}^{S-1} \exp\left(\rho G_{i,j}^n\right)} \end{pmatrix}$$

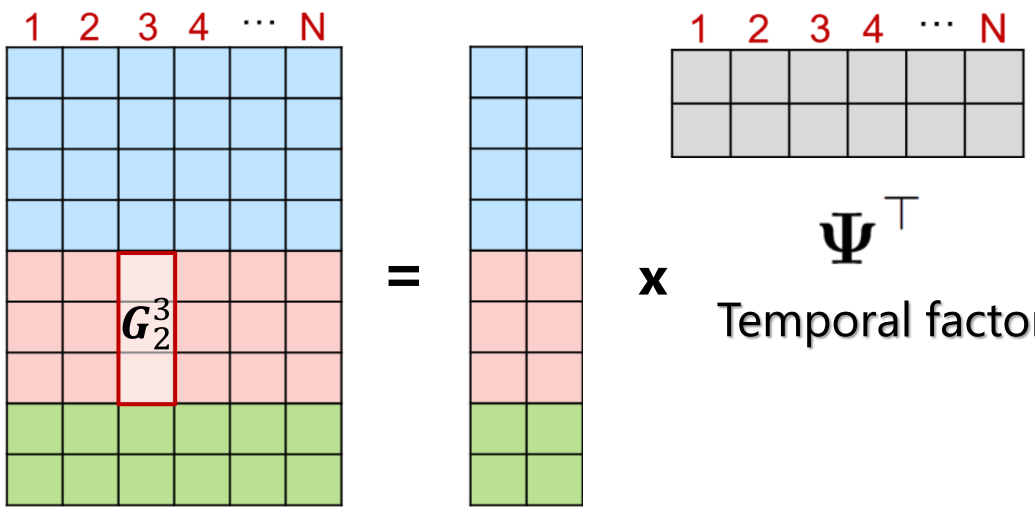


Two problems:

- (1) Numerous parameters: $(S-1)(S-2)/2 \times N$
- (2) The columns of **G** should vary smoothly from 1 to N

A Bayesian model

Solutions: Problem (1) — Matrix factorization



$\boldsymbol{\Psi}^\top$

Temporal factor matrix

$\boldsymbol{G}$ $\boldsymbol{\Phi}$

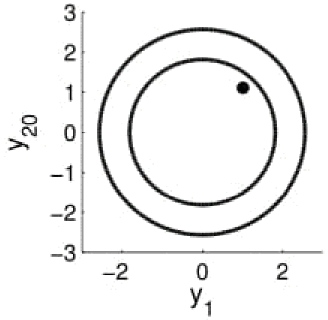
Mapping factor matrix

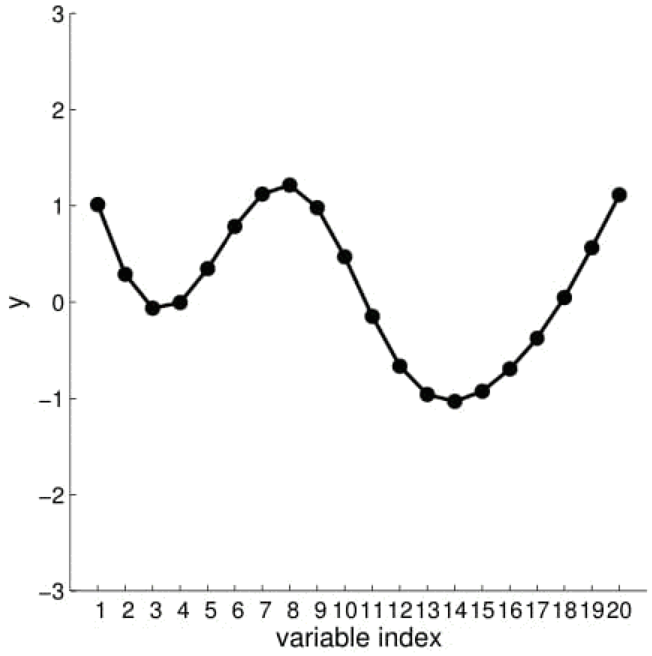
$$\boldsymbol{G}_s = \boldsymbol{\Phi}_s \boldsymbol{\Psi}^\top$$

Problem (2) — Gaussian process

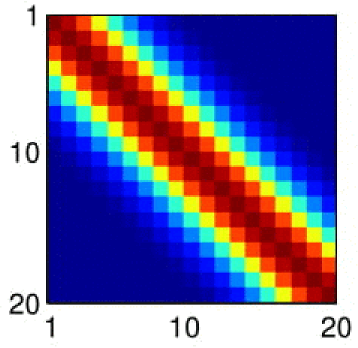
$$\psi_d \sim \mathcal{GP}(\mathbf{0}, \boldsymbol{K}_d), \quad d = 1, \ldots, D,$$

Gaussian process $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$





$\Sigma =$



Bayesian inference

MCMC sampling procedure:

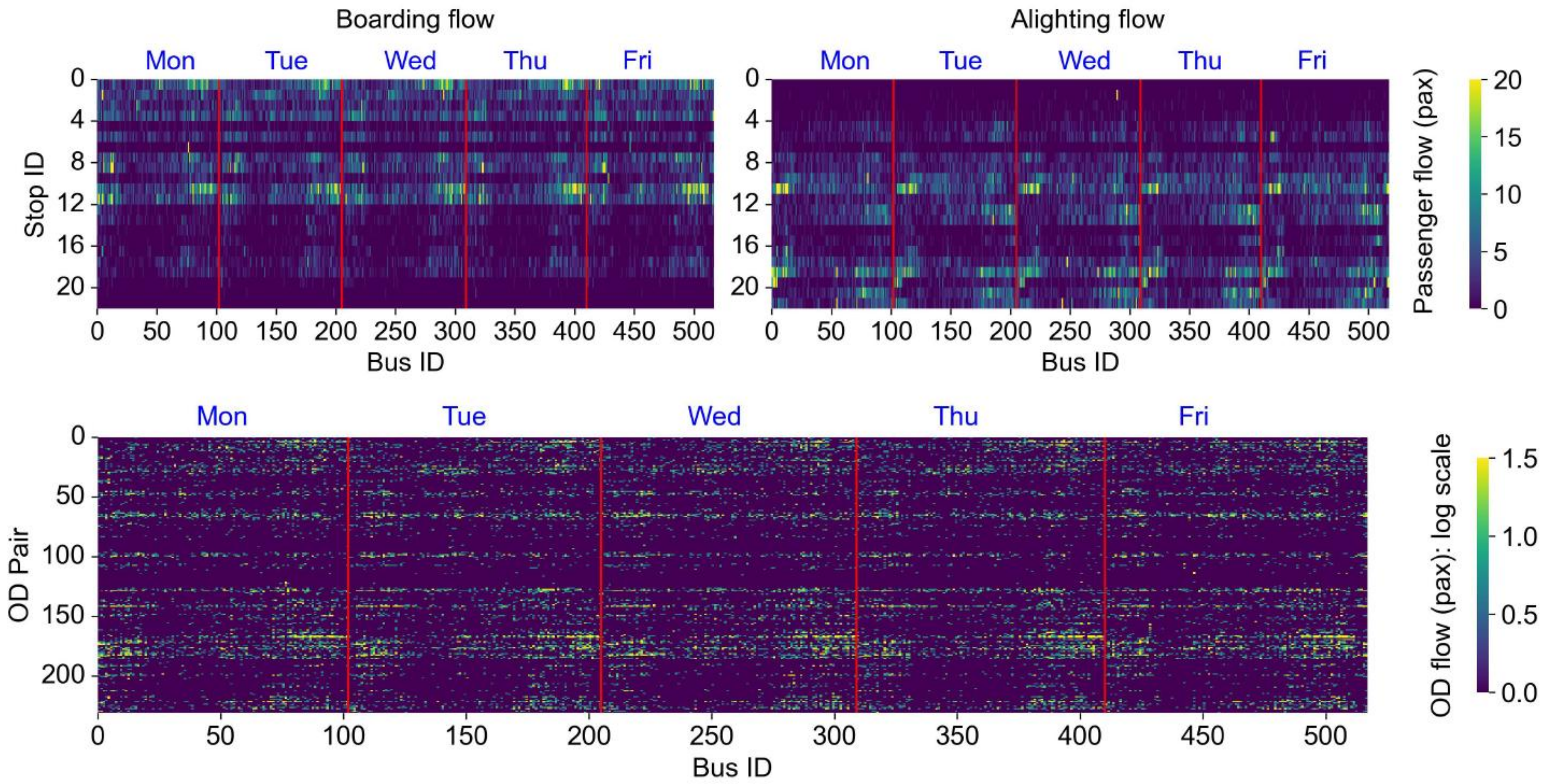
- Sample $\mathcal{Y}$ from $p\left(\mathcal{Y} \mid \Theta, \mathcal{X}\right)$ with Metropolis-Hastings sampling.
- Sample $\boldsymbol{\Phi}$ from $p\left(\boldsymbol{\Phi} \mid \boldsymbol{\Psi}, \rho, \mathcal{Y}\right)$ with Elliptical slice sampling.
- Sample $\boldsymbol{\Psi}$ from $p\left(\boldsymbol{\Psi} \mid \boldsymbol{\Phi}, \rho, \mathcal{Y}\right)$ with Elliptical slice sampling.
- Sample ρ from $p\left(\rho \mid \boldsymbol{\Psi}, \boldsymbol{\Phi}, \mathcal{Y}\right)$ with slice sampling.

Approximating posterior distribution of OD vectors:

$$\begin{aligned} p\left(\boldsymbol{y}^{n} \mid \mathcal{X}, \boldsymbol{t}\right) &= \int p\left(\boldsymbol{y}^{n} \mid \boldsymbol{x}^{n}, \Theta\right) p\left(\Theta \mid \mathcal{X}, \boldsymbol{t}\right) d\Theta \\ &\approx \frac{1}{M} \sum_{m=1}^{M} p\left(\boldsymbol{y}^{n} \mid \boldsymbol{x}^{n}, \Theta^{(m)}\right). \end{aligned}$$

Experiments

- AFC data from three bus routes: short (22 stops), medium (40 stops), and long (72 stops)
- Ground truth is available for model evaluation



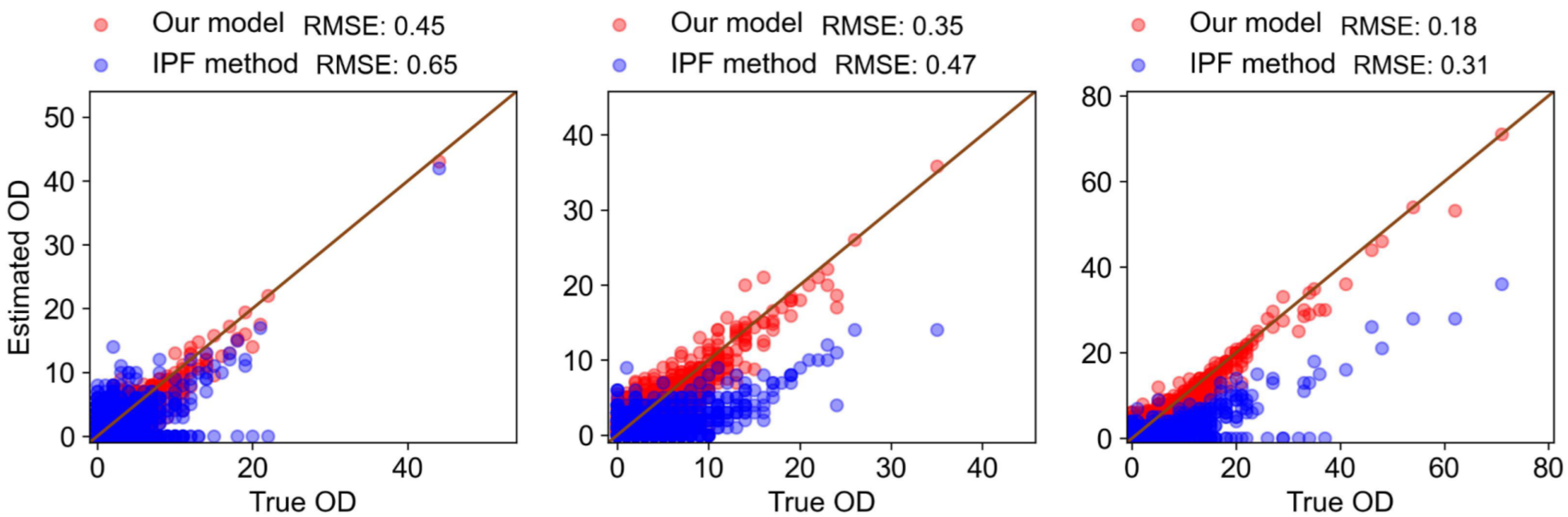
Experiment setting

- Obtain boarding/alighting counts based on the true OD matrices and then apply the model to infer OD matrices based on the counts.
- Compare the performance of our model with the widely used Iterative Proportional Fitting (IPF) method.
- Implement the developed MCMC algorithm and run a total of 100,000 iterations to sample the model parameters.
- Take the first 95,000 iterations as “burn-in” and the last 5000 iterations to approximate the posterior distributions.

Results

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

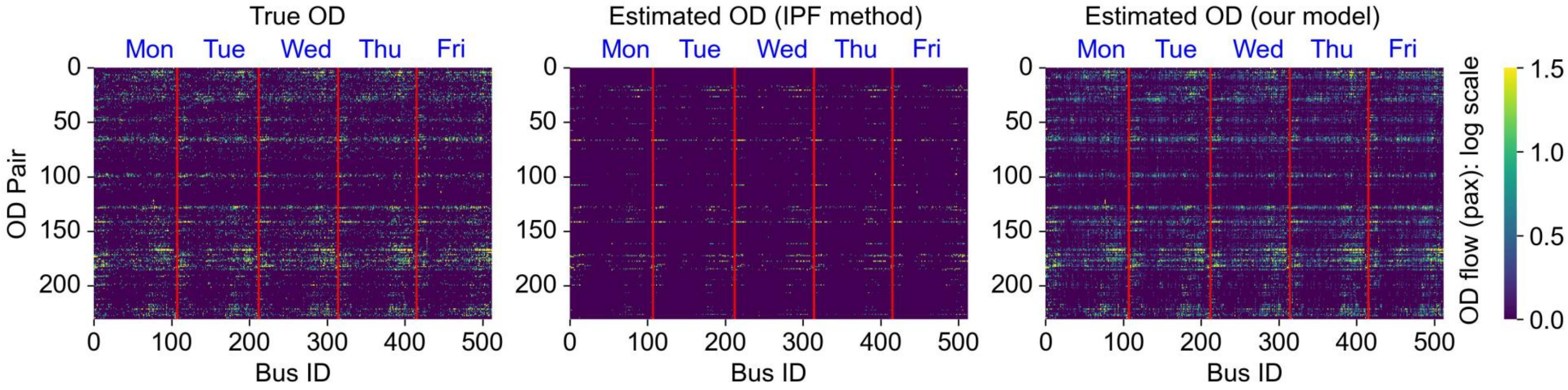
		Static model	Temporal Bayesian model			
			$D=1$	$D=2$	$D=4$	$D=6$
Short route	Mean	-35328.77	-34829.07	-33728.49	-33064.69	-32874.45
	Standard deviation	98.63	86.65	79.14	86.75	85.54
Medium route	Mean	-63599.00	-62915.15	-62141.88	-61652.34	-61539.26
	Standard deviation	165.46	150.24	134.89	156.80	123.00
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



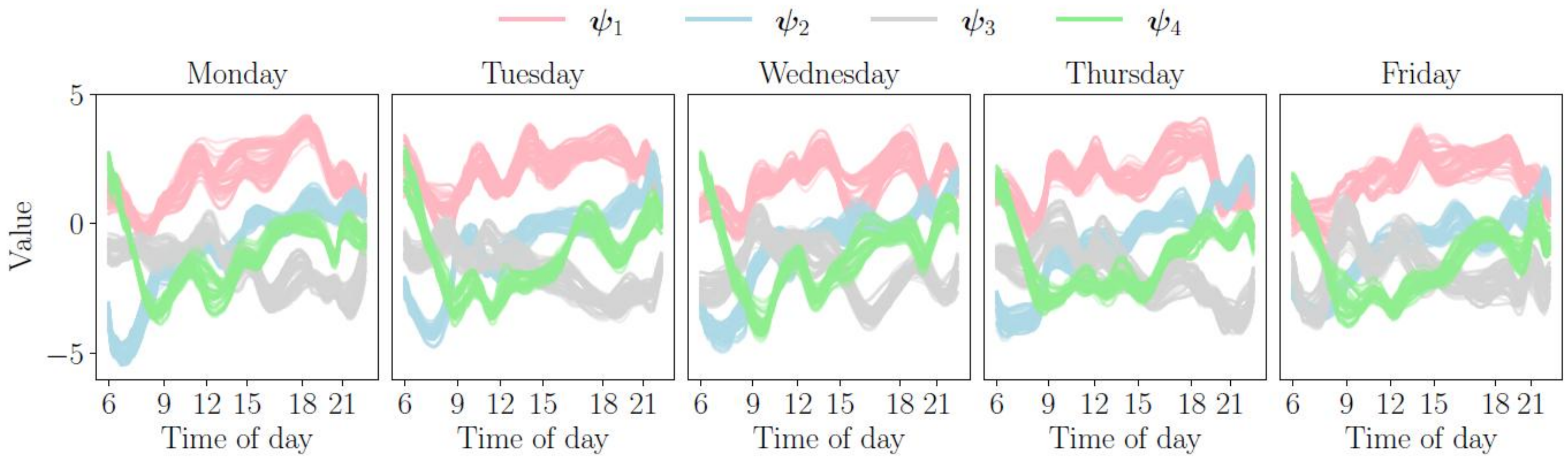
(1) Short route (2) Medium route (3) Long route

Results

- True and estimated OD vectors for all buses

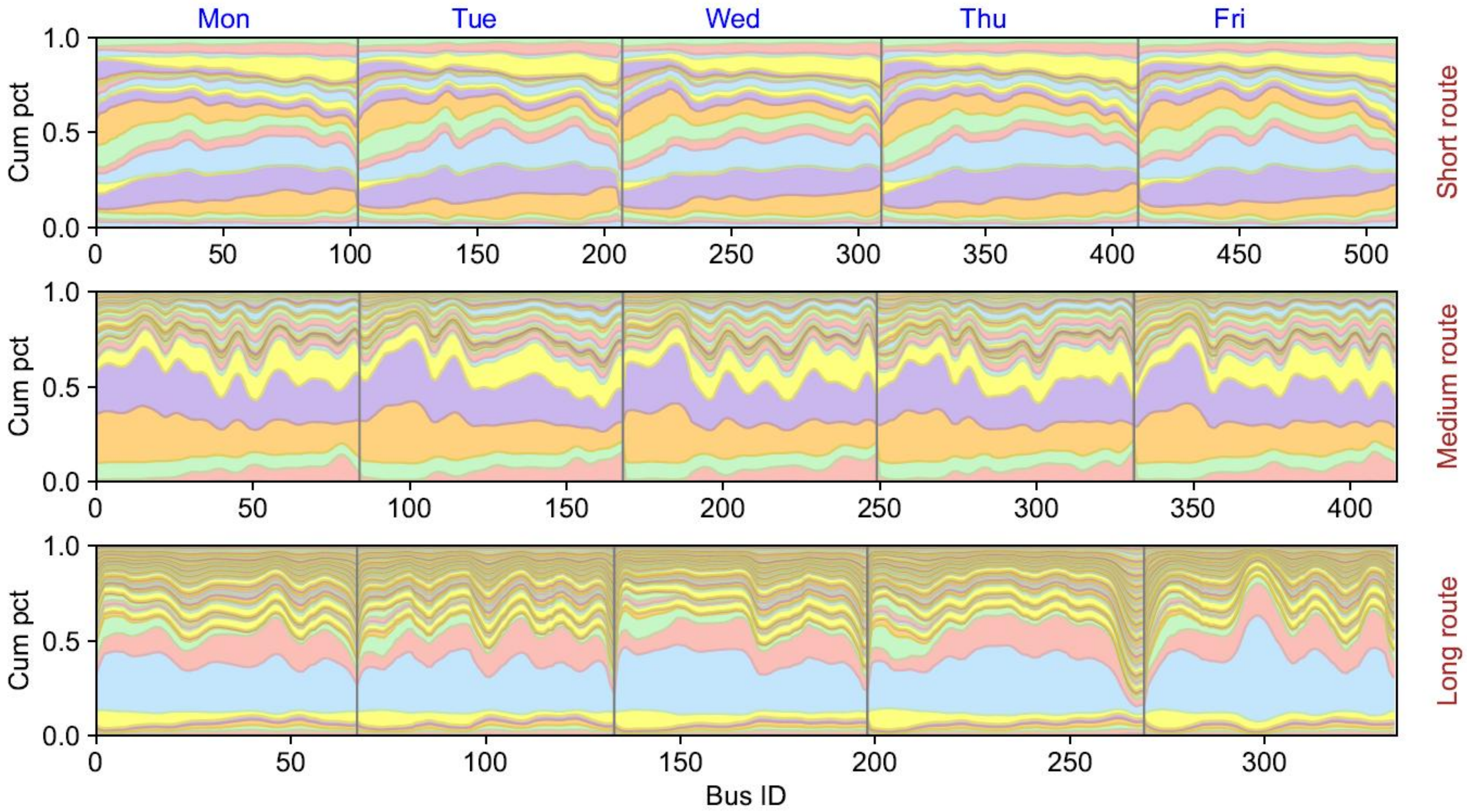


- Posterior samples of Ψ with rank D=4



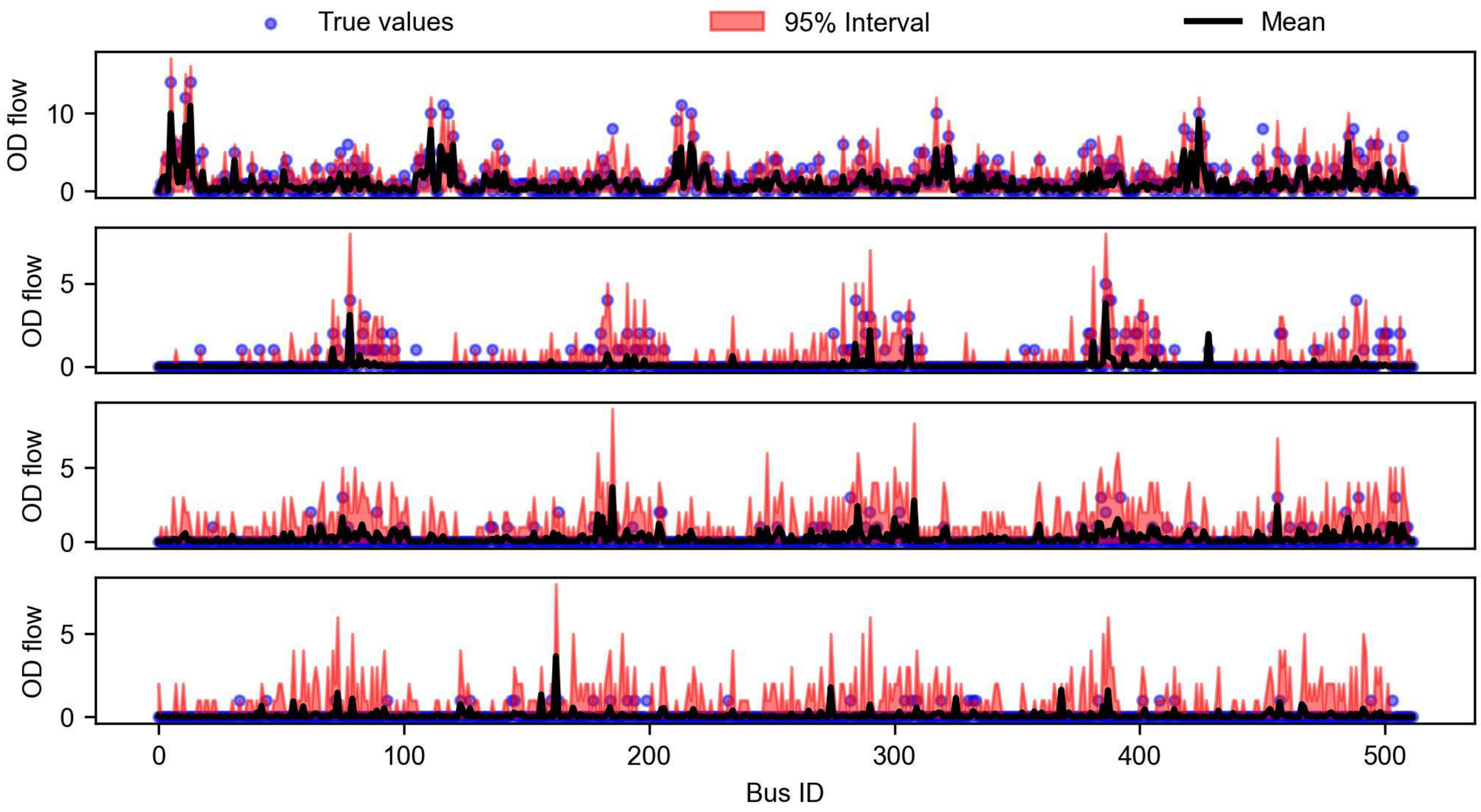
Results

- Temporal patterns of $\boldsymbol{\lambda}_1$ for different bus routes



Results: uncertainty quantification

- 4 selected OD pairs as examples, flow varies substantially across buses



Summary

- Estimate transit OD matrices at the bus-journey level from boarding/alighting counts at bus stops. (It is a challenging **linear inverse problem**!)
- We model the alighting probabilities and assume they are smoothly time-varying.
- In this inverse problem, we use two priors
  - **The "low-rank" prior**
    - Using factor model to reduce the dimensionality in parameters
  - **The "GP" prior**
    - Using the Gaussian process to model the time-varying parameters
- Our proposed model is evaluated in the real-world data and the estimation is good.

Thank You Any Questions?

Xiaoxu Chen
Postdoc researcher
Department of Civil Engineering
McGill University
xiaoxu.chen@mcgill.ca