

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

Xiaoxu Chen

Postdoctoral Fellow,
Department of Logistics and Operations Management,
HEC Montréal

BIRS Workshop, July 1, 2026

Collaborators



Xiaoxu Chen

Logistics and Operations Management,
HEC Montréal



Zhanhong Cheng

Intelligent Transportation Systems,
Zhejiang University



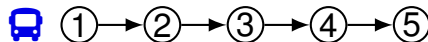
Lijun Sun

Department of Civil Engineering,
McGill University

Chen, X., Cheng, Z., & Sun, L. (2025). Bayesian inference of time-varying origin-destination matrices from boarding and alighting counts for transit services. *Transportation Research Part B: Methodological*, 199, 103278.

OD matrix for a bus route

- A bus route has a sequence of **stops**.
- An OD matrix shows passenger flows between each pair of bus stops.
- Two types of OD matrices for bus route:
 - **Individual bus-journey** level
 - **Aggregated** level (e.g., overall demand in morning peak hours)
- OD matrices are important **input** for bus operations management.



	D_1	D_2	D_3	D_4	D_5
O_1	-	2	3	1	1
O_2	-	-	4	2	1
O_3	-	-	-	2	1
O_4	-	-	-	-	2
O_5	-	-	-	-	-

Transit data sources for OD information

Traditional on-board survey data

- **Full trip** data: time, origin, destination, bus ID
- **OD matrices**
- Time-consuming and labor-intensive



Transit data sources for OD information

AFC data: Tap-in + Tap-off

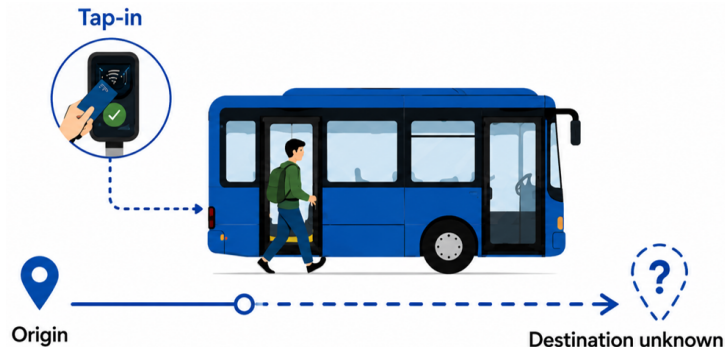
- **Full trip** data: time, origin, destination, bus ID
- Available in some systems (e.g., Singapore, Beijing)
- **Destination known**



Transit data sources for OD information

AFC data: Tap-in only

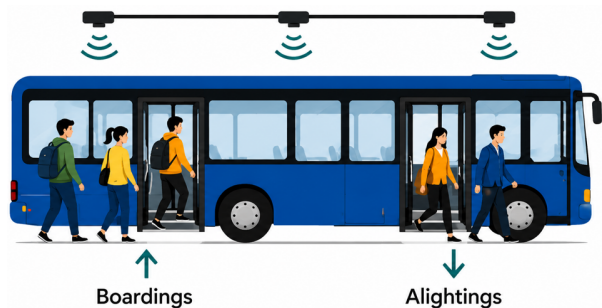
- **Partial trip** data: time, origin, bus ID
- Destination inference models (e.g., trip-chain assumptions)
- **Destination unknown**



Transit data sources for OD information

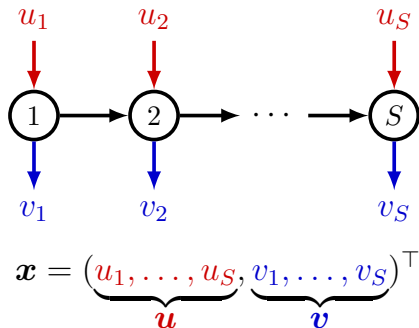
APC data

- **Aggregated boarding and alighting counts**
- No individual trip information
- Common in Europe and North America
- **Aggregated counts only**

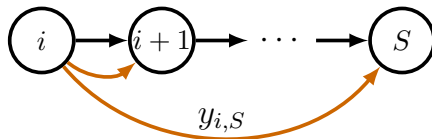


Research problem

- Observation for each bus.



- Unknown OD flows for each bus.



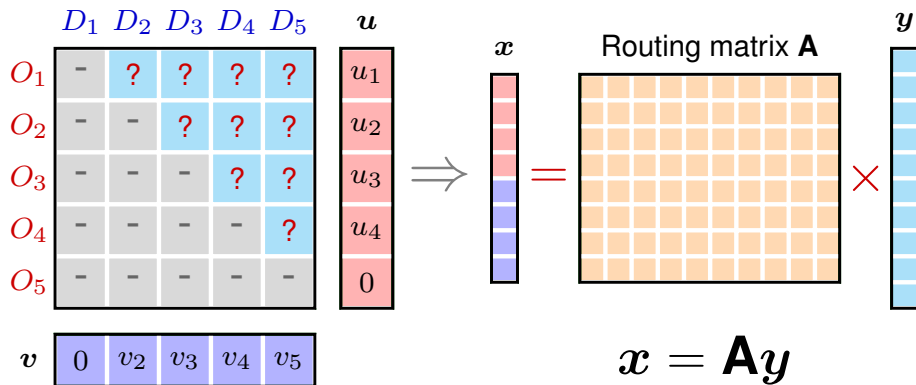
$$\mathbf{y}_i = (y_{i,i+1}, \dots, y_{i,S})^\top$$

Stack all $\mathbf{y}_i$ vectors

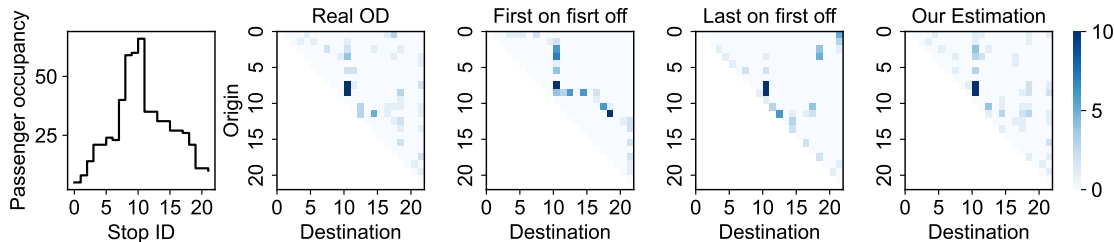
$$\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_{S-1}^\top)^\top$$

Research problem

- Estimating OD matrix is an linear inverse problem.
 - $u_i = \sum_j^S y_{ij}$, $v_j = \sum_i^S y_{ij}$; Not identifiable $S > 4$; Numerous feasible solutions.

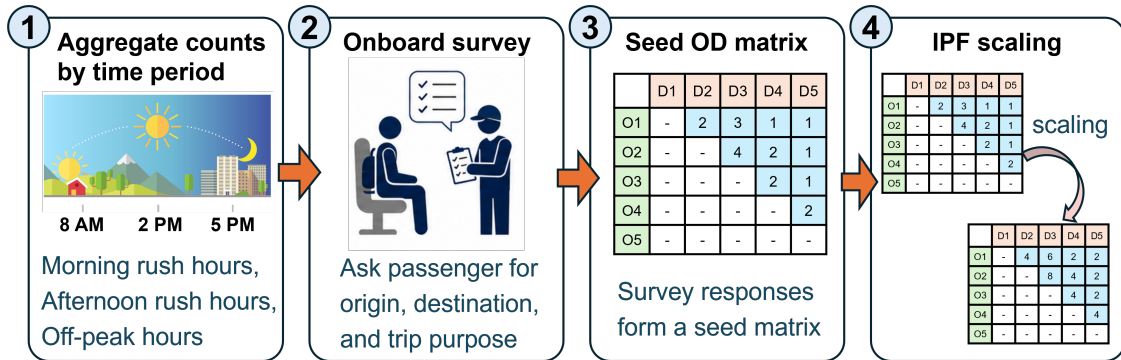


Research problem



- Uncertainty quantification in OD estimation is important.
- Let $\mathcal{X} = \{\mathbf{x}^n \mid n = 1, 2, \dots, N\}$ be the observations for all buses.
- This research focuses on estimate $p(\mathbf{y}^n \mid \mathcal{X})$.

Industry practice: IPF-based OD estimation



- Limitation: labor-intensive, “biased” seed, and deterministic estimation.

Hazelton's solution

- Hazelton (2010) assumes each entry (count) follows a Poisson distribution.
- With demand “rate” λ^n for bus n .
- The likelihood is

$$\begin{aligned}\mathcal{L}(\lambda^n) &= p(\mathbf{x}^n \mid \lambda^n) \\ &= \sum_{\mathbf{y}^n} p(\mathbf{x}^n \mid \mathbf{y}^n, \lambda^n) p(\mathbf{y}^n \mid \lambda^n)\end{aligned}$$

- Assuming all buses share the same “rate” parameter λ^n (static model)

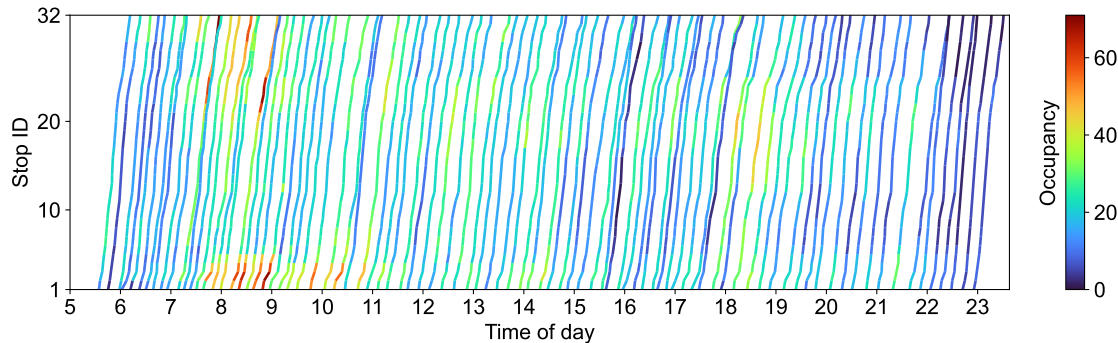
Demand rate λ_{ij}^n

$i \backslash j$	1	2	3	4	5
1	-	2.7	3.5	1.8	1.6
2	-	-	4.1	2.4	1.1
3	-	-	-	2.2	1.1
4	-	-	-	-	2.3
5	-	-	-	-	-

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

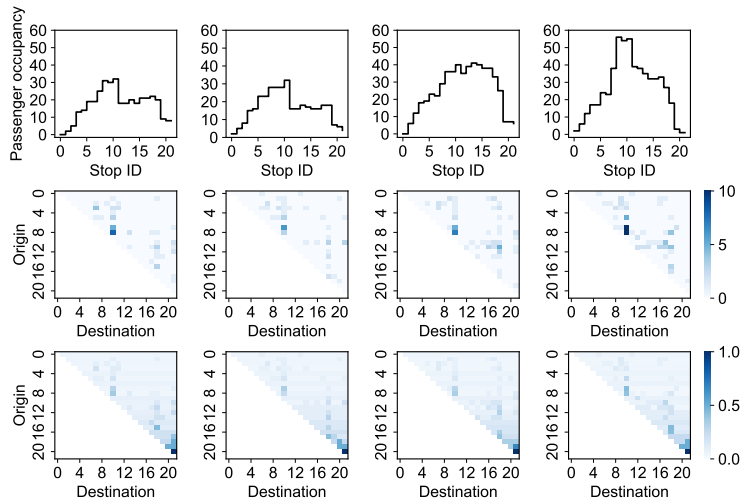
Rethink bus demand and supply

- For multiple buses in a period, can we assume λ^n remain the same?



- Both demand and supply are time-varying.
- λ^n vary substantially even for consecutive buses.

Behavioral perspective



consecutive four buses

OD passenger flows

alighting probabilities
(vary smoothly)

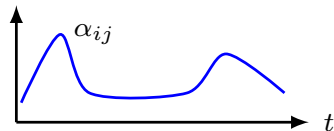
Bayesian model formulation – assumptions

- Alighting probabilities meet $\sum_{j=1}^S \alpha_{ij}^n = 1$.
- Alighting probabilities vary smoothly over time.
- Instead of modeling λ^n , we model α_{ij}^n .
- Assume passengers make decisions independently.
- Given the boarding count u_i^n at stop i , OD flow $\mathbf{y}_i^n \sim \text{Multinomial}(u_i^n, \boldsymbol{\alpha}_i^n)$, i.e.,

$$p(\mathbf{y}_i^n | u_i^n, \boldsymbol{\alpha}_i^n) = u_i^n! \prod_{j=i+1}^S \frac{\alpha_{i,j}^n y_{i,j}^n}{y_{i,j}^n!}$$

Alighting probability α_{ij}^n

$i \backslash j$	1	2	3	4	5
1	-	0.1	0.2	0.1	0.6
2	-	-	0.3	0.2	0.5
3	-	-	-	0.4	0.6
4	-	-	-	-	1.0
5	-	-	-	-	-

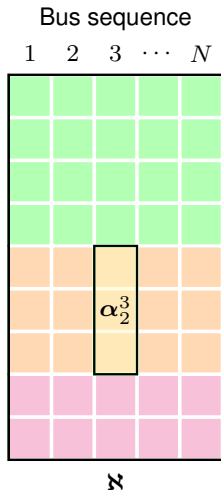


Bayesian model formulation – parameterization

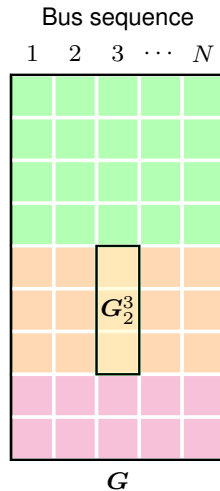
Alighting probability α_{ij}^n

$i \backslash j$	1	2	3	4	5
1	-	0.1	0.2	0.1	0.6
2	-	-	0.3	0.2	0.5
3	-	-	-	0.4	0.6
4	-	-	-	-	1.0
5	-	-	-	-	-

vectorization $\Downarrow$



$\alpha_i^n = \text{Softmax}(\rho G_i^n)$ $\Downarrow$



Challenge 1:

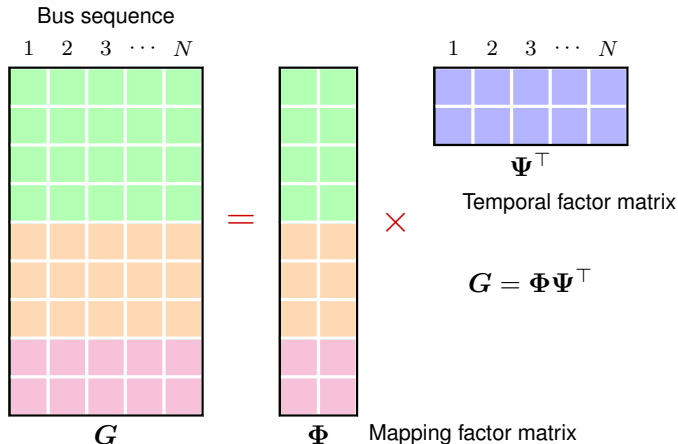
Model parameters:
 $N(S-1)(S-2)/2$

Challenge 2:

The columns of G
 should vary smoothly
 from 1 to N

Bayesian model formulation – parameterization

Solutions: Challenge 1 – Matrix factorization



Challenge 2 – Gaussian process

ψ_d

1 2 3 ... N

$\psi_d \sim \mathcal{GP}(\mathbf{0}_N, \mathbf{K}_d)$

$[\mathbf{K}_d]_{rs} = k_d(t_r, t_s; \boldsymbol{\eta}_d)$

$k(t, t'; l, \sigma^2) = \sigma^2 \exp\left(-\frac{(t-t')^2}{2l^2}\right)$

Model inference: MCMC sampling

MCMC sampling procedure:

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

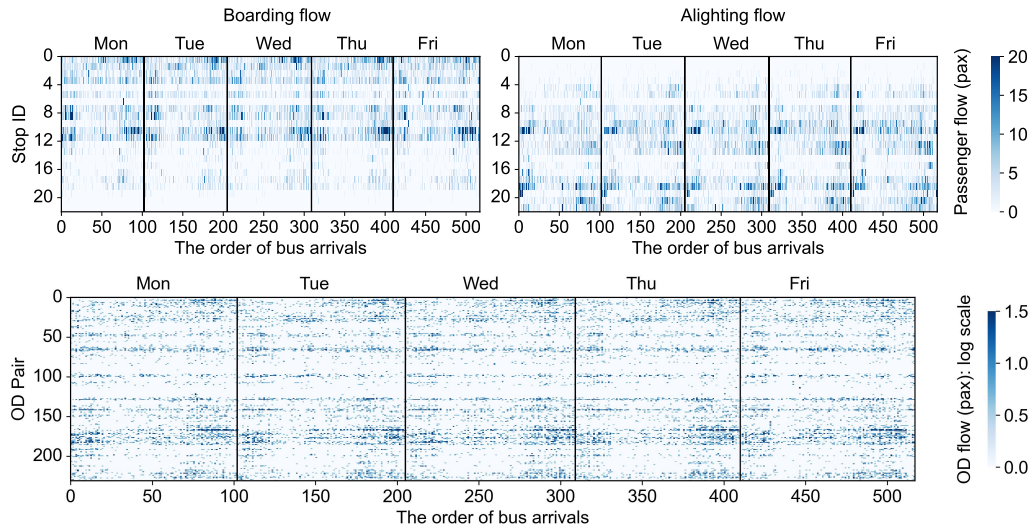
Approximating posterior distribution of OD vectors:

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

Experiments

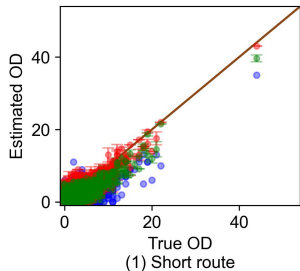
- We use tap-in-tap-out AFC data from three bus routes:
 - short route (22 stops)
 - medium route (40 stops)
 - long route (72 stops)
- Obtain true OD matrices and boarding/alighting counts.
- Apply the model to estimate OD matrices from boarding/alighting counts.
- Compare the performance of our model with IPF method and Poisson model.

Experiments – data from short route

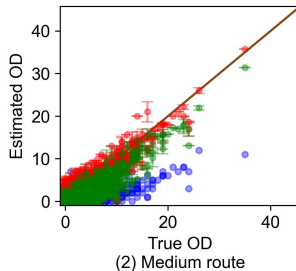


Experiments – performance evaluation

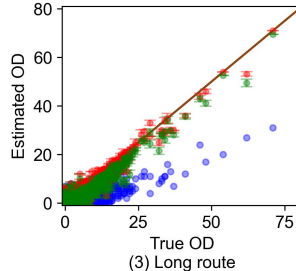
● IPF method RMSE: 0.62
● Our model RMSE: 0.45, CRPS: 0.119
● Poisson-gamma RMSE: 0.47, CRPS: 0.136



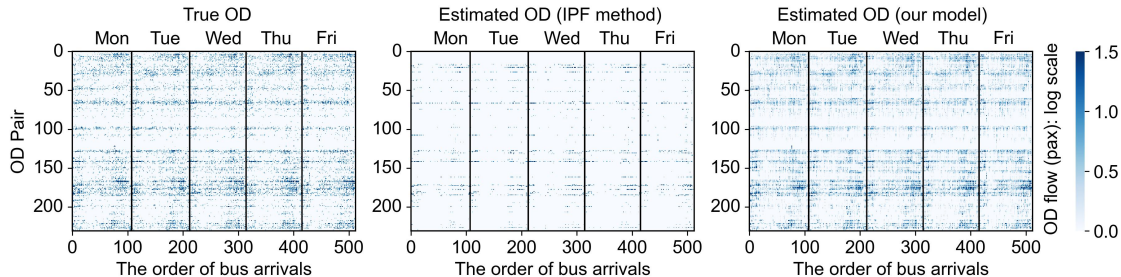
● IPF method RMSE: 0.48
● Our model RMSE: 0.37, CRPS: 0.089
● Poisson-gamma RMSE: 0.40, CRPS: 0.098



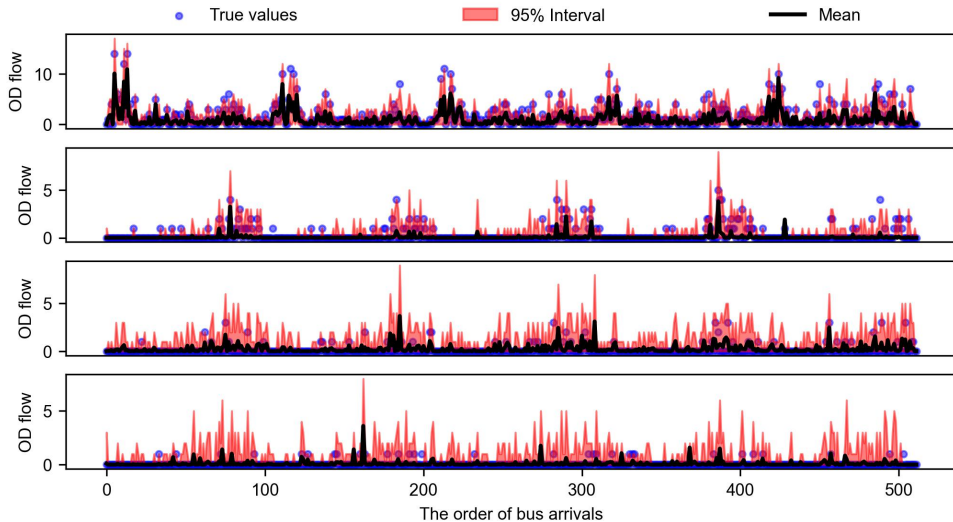
● IPF method RMSE: 0.31
● Our model RMSE: 0.19, CRPS: 0.017
● Poisson-gamma RMSE: 0.21, CRPS: 0.020



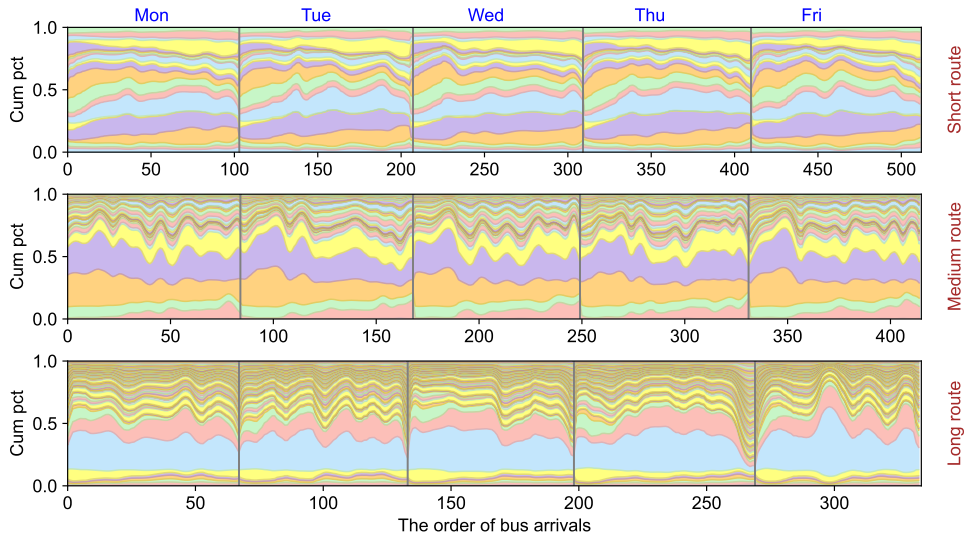
Experiments – OD estimation visualization



Experiments – uncertainty quantification



Experiments – estimated alighting probability (stop 1)



Summary

- Estimate transit OD matrices at the bus-journey level from boarding and 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 matrix factorization to reduce the dimensionality
 - The “GP” prior: Using GP to model the time-varying parameters.
- Our proposed model is evaluated in the real-world data and the estimation is good.

Our method received attentions from industries

Eric Stemmler

Senior Data Analyst
Integrated Expert



giz

Deutsche Gesellschaft
für Internationale
Zusammenarbeit (GIZ) GmbH



Federal Ministry
for Economic Cooperation
and Development



**BASHKIA
TIRANË**

E: Municipality of Tirana, Directorate of
Transportation and Road Traffic
A: Skanderbeg Square, Building No. 2
Tirana 1001, Albania
www.tirana.al

stpg

This position is supported by German development cooperation

David Nelson

Transportation Planner



**CAPITAL REGION
Transportation
Council**

1 Park Place, Suite 101
Albany, NY 12205-2676

Amir H. Ghods, PhD, P. Eng. | CEO

SMATS Traffic Solutions

211-35 Auriga Drive, Nepean, ON K2E 8B7

P: 613-417-0309-x102 | M: 613-406-8765

TF: 888-441-5666 | F: 613-701-4142



exo

Downstream risk-aware decision

Simple example: annual net benefit

Plan A



Expected benefit: **\$12M**



95% likely range: **\$2M to \$20M**



5% bad-case outcome:
service overcrowding and only **\$1M** benefit

Plan B



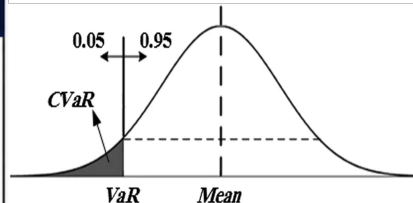
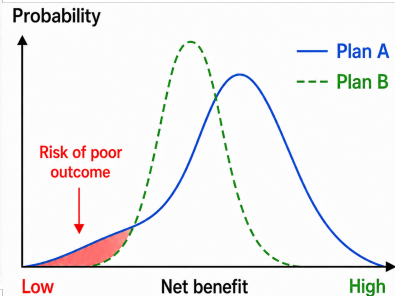
Expected benefit: **\$10M**



95% likely range: **\$7M to \$13M**



5% bad-case outcome:
manageable crowding and only **\$6M** benefit



Thank you!

Questions and discussion welcome!

Xiaoxu Chen `xiaoxu.chen@hec.ca`