大學教育資助委員會 University Grants Committee

Lingnan University 嶺南大學 香港 Hong Kong School of Data Science

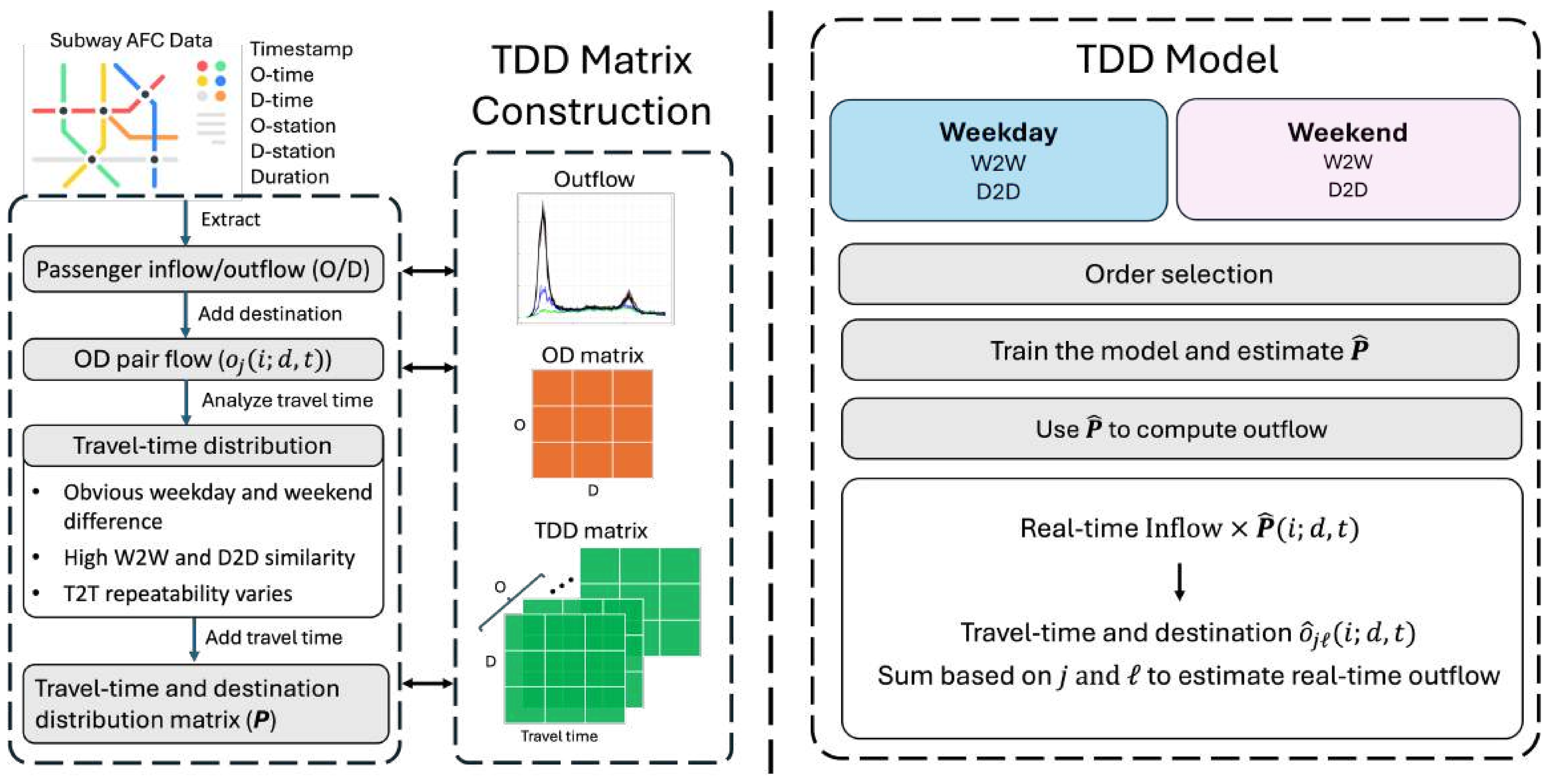
A Novel Passenger Travel-time and Destination Distribution Model for Day-ahead Forecasting

Funding: 深圳市科技创新委员会 **PI: S. Joe Qin**

Introduction

- We propose a travel-time and destination distribution (TDD) matrix, which computes the proportion of passengers' mobility.
- By quantifying the similarity of current time slot TDD with slot-ahead, day-ahead, and week-ahead, in-depth underlying mechanisms of passengers' periodic travel patterns is unvieled, revealing predictable and regular mobility behavior.
- TDD model to predict passengers' TDD one-day ahead.
- By combining estimated TDD with real-time passenger data, we can accurately estimate the passenger flows and travel durations.

Research Outline



Constructing TDD Matrix

Passenger flow: The inflow, denoted as $O(i;d,t)$; the ouflow, denoted as $D(i;d,t)$.

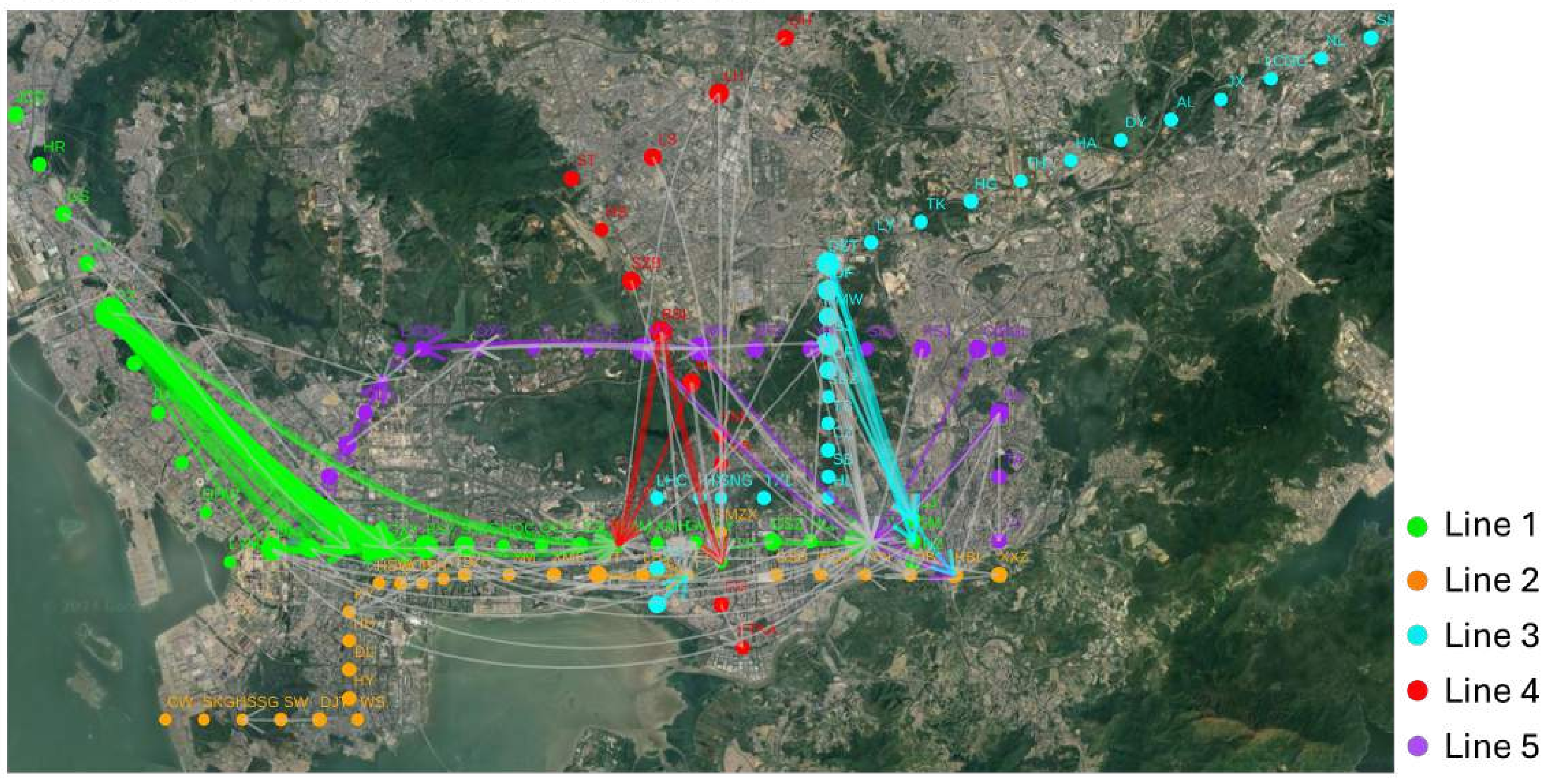
OD matrix: denoted with $o_j(i;d,t)$.

Travel-time and Destination: Variable $o_{j\ell}(i;d,t)$ quantifies the number of passengers from inflow $O(i;d,t)$ entering Station i and exiting from Station j during Time slot t on Date d destined with travel time ℓ.

Proportion of Travel-time and Destination: $p_{j\ell}(i;d,t)$ represents the proportion of passengers originating from Station i to Station j during Time slot t on Date d destined with travel time ℓ, i.e., $p_{j\ell}(i;d,t) = \frac{o_{j\ell}(i;d,t)}{O(i;d,t)}$.

$$\mathbf{P}(i;d,t) = \begin{bmatrix} p_{1,0}(i;d,t) & p_{1,1}(i;d,t) & \dots & p_{1,8}(i;d,t) \\ p_{2,0}(i;d,t) & p_{2,1}(i;d,t) & \dots & p_{2,8}(i;d,t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{n,0}(i;d,t) & p_{n,1}(i;d,t) & \dots & p_{n,8}(i;d,t) \end{bmatrix} \in \mathbb{R}^{n\times 9}$$

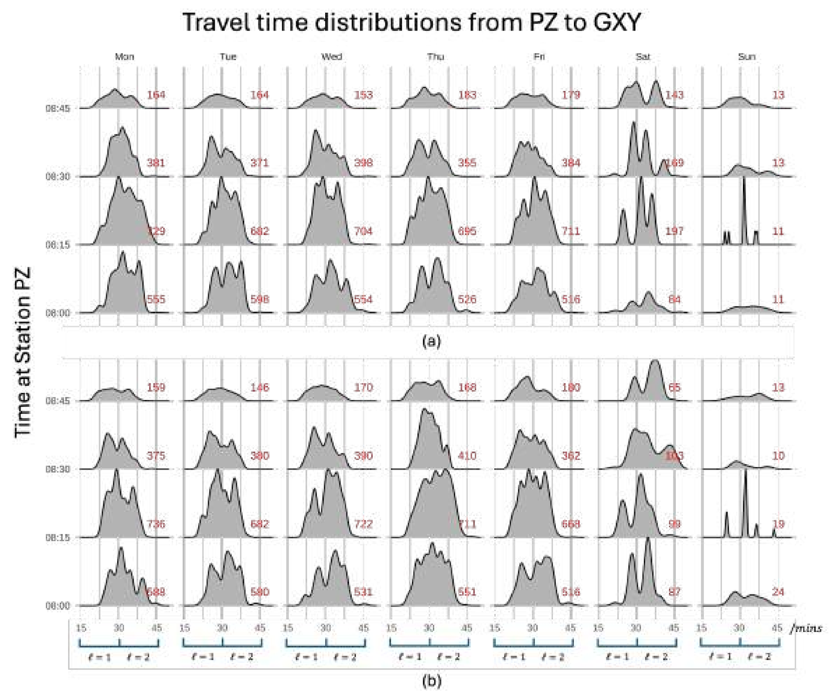
08:00 AM – 08:15 AM, October 14, 2013.



The OD pair from Pingzhou (PZ) to Gaoxinyuan (GXY) attracts the highest number of passengers.

Similarity

- The W2W repeatability is obvious.
- The D2D repeatability is also noticeable, especially for weekdays.
- However, the T2T repeatability varies from one time slot to the next. Rush-hour slots differs significantly from the post-rush-hour slot.



TDD Model

We define the objective as follows, which can be solved as a convex problem to predict the day-ahead TDD matrix.

$$\hat{\mathbf{P}}(i;d,t) = \sum_{a=1}^{n_d} \alpha_a \mathbf{P}(i;d-a,t) + \sum_{b=1}^{n_w} \beta_b \mathbf{P}(i;d-7b,t)$$

$$s.t. \quad \alpha_a \geqslant 0, \beta_b \geqslant 0, \ \forall a \in [1,n_d], \forall b \in [1,n_w]$$

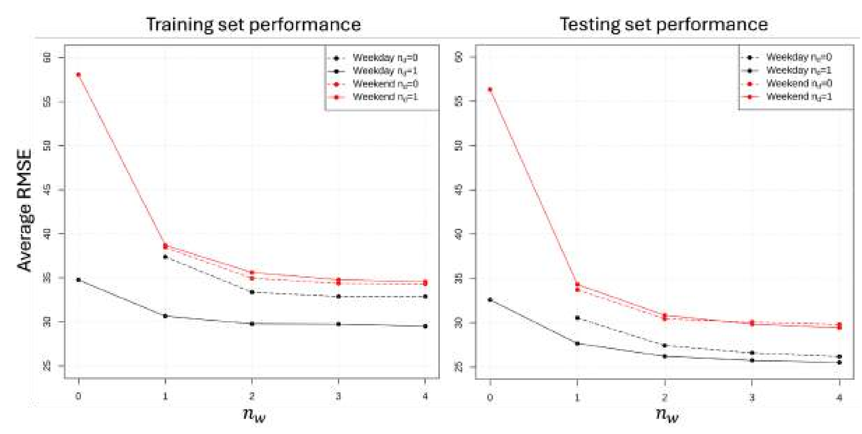
$$\sum_{a=1}^{n_d} \alpha_a + \sum_{b=1}^{n_w} \beta_b = 1$$

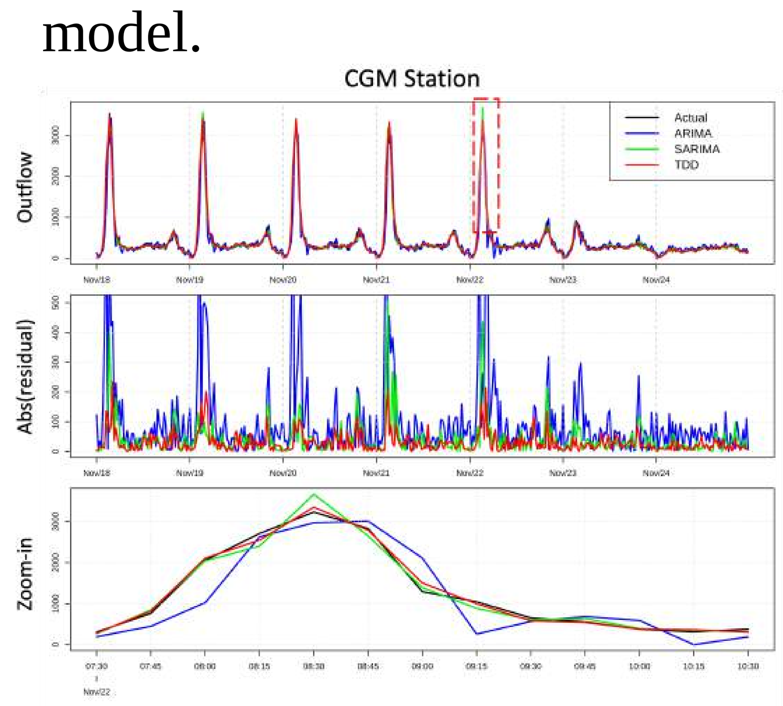
With the predicted $\hat{\mathbf{P}}(i;d,t)$ and real-time inflow $O(i;d,t)$, the passenger outflows at a destined station j can be estimated based on

$$D(j;d,t) = \sum_i \sum_\ell O(i;d,t-\ell) p_{j\ell}(i;d,t-\ell)$$

Results

- Weekdays exhibit improved performance with day-ahead pattern, while weekends are better without it.
- $n_w = 2$ is a good choice for both weekday and weekend model.





- ARIMA family model always have a prediction delay.
- TDD exhibits lower absolute residuals than SARIMA for most of time, especially during peak times, contributing to its overall superior performance

- Out of the 118 stations examined, the TDD model achieves the best test RMSE results for 102 stations.
- SARIMA exhibits an average test RMSE of 25.85 across all stations, while TDD model achieves a mean value of 24.04, showcasing an improvement of nearly 10%.