

# Travel-Behavior-Based Inference and Forecasting Methods in Metro Systems

Zhanhong Cheng



Department of Civil Engineering  
McGill University  
Montreal, Quebec, Canada

May 5, 2022

A thesis submitted to McGill University in partial fulfillment  
of the requirements of the degree of Doctor of Philosophy

©Zhanhong Cheng, 2022

# Abstract

The metro is indispensable for the urban transportation system. As the world enters the era of informatization and digitalization, data generated from smart card fare collection systems (smart card data) have played an important role in the planning and operation of metro systems. A large body of research uses smart card data to understand passenger travel behavior in metro systems; it has been found that individual mobility in metro systems is highly regular with interpretable patterns. Besides, smart card data have also been extensively used in assisting the operation and control of metro systems, such as inferring trip origins/destinations and forecasting passenger demand. However, a lack in existing research is the connection between the above two aspects—how to use the unique travel behavior characteristics of metro passengers to establish better data-driven applications. To fill this gap, this thesis aims to develop travel-behavior-based inference and forecasting models in metro systems.

The three contributions of this thesis, enclosed in three scientific papers, are (1) trip destination inference, (2) real-time boarding demand forecasting, and (3) real-time origin-destination (OD) matrices forecasting. All models developed in the thesis are tested by real-world smart card data from Guangzhou, China. First, this thesis develops a probabilistic topic model to infer trip destination from tap-in only smart card system. The probabilistic topic model is learned from passengers' historical travel behavior and can predict the most likely destination of a trip given the origin and the departure time. Complementing existing trip-chain-based destination inference methods, the proposed model is particularly useful for isolated trips where conventional methods fail. Besides destination inference, latent topics learned by the probabilistic model can be used to analyze passengers' travel behavior patterns. Second, this thesis aims to incorporate travel behavior regularity into passenger boarding demand/flow forecasting. Utilizing the strong regularity rooted in individuals' travel behavior, a new concept named "returning flow" is proposed to capture the generative mechanism of boarding flow. The returning flow is highly correlated to the boarding flow and can be used as a covariate in a time series model

to improve the boarding flow forecasting. Extensive experiments show the effectiveness of using the travel behavioral feature boarding flow forecasting. The Last part of this thesis addresses the real-time OD matrices forecasting problem in metro systems. Using the low-rank property of OD data, the forecasting is formulated into a low-rank vector autoregression (VAR) problem and is solved by dynamic mode decomposition (DMD). Next, a forgetting ratio is introduced to exponentially reduce the weights for historical data. Moreover, an online update algorithm is developed to update the model efficiently without storing historical data or retraining. Experiments show the proposed model significantly outperforms baseline models in forecasting both OD matrices and boarding flow.

In summary, this thesis uses travel behavioral characteristics to improve inference and forecasting models in metro systems. The proposed models and solutions are beneficial to the intelligent operation of metro systems. The three tasks of destination inference, boarding flow forecasting, and OD matrices forecasting correspond to individual-level, station-level, and-network level applications, respectively. By these three levels, this thesis demonstrates the considerable potential of using travel behavior in various metro applications.

# Résumé

Le métro est indispensable dans plusieurs systèmes de transport urbain. Alors que le monde est entré dans l'ère de l'informatisation et de la numérisation, les données générées par les systèmes de perception par cartes à puce peuvent jouer un rôle important dans la planification et l'exploitation des systèmes de métro. De nombreux travaux de recherche utilisent les données des cartes à puce pour comprendre les comportements de déplacement des passagers dans les réseaux de métros ; on y constate que la mobilité individuelle dans les métros est souvent très régulière, suivant des patrons interprétables. En outre, les données des cartes à puce peuvent également être utilisées pour aider à l'exploitation et au contrôle des systèmes de métro, via par exemple la déduction des origines/destinations des déplacements des passagers et la prévision de l'achalandage. Cependant, une lacune de la recherche existante concerne le lien entre ces deux aspects — comment utiliser les caractéristiques de comportement des déplacements des passagers du métro pour établir de meilleurs outils de planification basés sur ces données. Afin d'y apporter une contribution, cette thèse vise à développer des modèles d'inférence et de prévision basés sur les comportements de déplacement des passagers dans les systèmes de métro.

Les trois contributions spécifiques, chacune intégrée dans cette thèse sous la forme d'un article scientifique, sont (1) l'inférence de la destination du voyage, (2) la prévision de l'achalandage en temps réel et (3) la prévision des matrices origine-destination (OD) en temps réel. Tous les modèles développés dans la thèse sont testés sur des données de carte à puce du système de métro de Guangzhou, en Chine. Tout d'abord, cette thèse développe un modèle de sujet probabiliste (modèle thématique) pour déduire la destination des voyages dans un système de carte à puce à entrée unique (transaction à l'embarquement seulement). Le modèle de sujet probabiliste se base sur le comportement historique des passagers et peut prédire la destination la plus probable d'un voyage compte tenu de son origine et de son heure de départ. En complément des méthodes existantes d'inférence de destination basées sur la séquence de transaction, le modèle proposé est particulièrement utile pour les trajets isolés où les méthodes conventionnelles échouent. Outre l'inférence

de destination, les sujets latents appris par le modèle probabiliste peuvent être utilisés pour analyser les modèles comportementaux des passagers. Deuxièmement, cette thèse propose d'intégrer la régularité des comportements de déplacement dans les prévisions d'achalandage (demande/flot d'embarquement des passagers). Utilisant la forte régularité enracinée dans le comportement de voyage des individus, un nouveau concept nommé "flot de retour" est proposé pour capturer le mécanisme générateur du flot d'embarquement. Le flot de retour est fortement corrélé au flot d'embarquement et peut être utilisé comme covariable dans un modèle de série chronologique pour améliorer la prévision du flot d'embarquement. Des expériences approfondies montrent l'efficacité de l'utilisation de cette prévision du flot d'embarquement des caractéristiques de comportement de voyage. La dernière partie de cette thèse aborde le problème de prévision des matrices origine-destination (OD) en temps réel dans les systèmes de métro. En utilisant la propriété de bas rang des données OD, la prévision est formulée dans un problème d'autorégression vectorielle de bas rang (VAR) et résolue par décomposition en mode dynamique (DMD). Ensuite, un taux d'oubli est introduit pour réduire de façon exponentielle les poids des données historiques. De plus, un algorithme de mise à jour en ligne est développé pour mettre à jour le modèle efficacement sans stocker de données historiques ni se recycler. Les expériences montrent que le modèle proposé surpasse considérablement les modèles de base pour la prévision à la fois des matrices OD et du flot d'embarquement.

En résumé, cette thèse utilise les caractéristiques comportementales des déplacements pour améliorer les modèles d'inférence et de prévision dans les systèmes de métro. Les modèles et solutions proposés sont bénéfiques pour le fonctionnement intelligent des systèmes de métro. Les trois contributions d'inférence de destination, de prévision des flots d'embarquement et de prévision des matrices OD correspondent respectivement aux applications au niveau individuel, au niveau de la station et au niveau du réseau. Par ces trois niveaux, cette thèse démontre le potentiel considérable de l'utilisation du comportement de déplacement dans diverses applications du métro.

# Acknowledgements

Days flew fast, and graduation was approaching in a blink of an eye. Looking back years later, I will definitely miss the busy and fulfilling time at McGill. Thanks to everyone who accompanies me and supports me.

Foremost, I want to express my sincerest gratitude to my supervisors, Prof. Lijun Sun and Prof. Martin Trépanier. They gave me a lot of valuable suggestions during the research, which greatly improved the quality of this thesis and cultivated my academic ability. Thank them for still having regular online meetings with me during the pandemic. Although I was a little nervous before each meeting because I wanted to make more progress to discuss, I always felt motivated and inspired after every meeting. It would be difficult to complete my thesis without their guidance, support, and encouragement. Special thanks to Prof. Lijun Sun, who has also helped me so much in my life and career.

Next, I would like to thank my thesis committee members, including Prof. Luis Miranda-Moreno, Prof. Luc Chouinard, Prof. Wenyi Xia, and Prof. Oded Cats, for their help and valuable comments. Particularly, thanks to my thesis examiners, Prof. Luis Miranda-Moreno and Prof. Oded Cats, their comments greatly improved the quality of this thesis.

Deep acknowledge to exo, Mitacs, and McGill University, who provide financial support for this Ph.D. project; this work would not have been possible without their financial support. Thank you, Hamzeh Alizadeh and Mohsen Nazem, for your help during my internship at exo. I learned a lot from the experience at exo, which is a rare opportunity to experience the industrial environment. Great thanks to Guangzhou metro for providing the smart card data used in the thesis.

I am in a wonderful group. Members of the group are amicable, humorous, and knowledgeable. Being able to study with them is my greatest fortune. I want to thank all my colleagues and friends at McGill, including Jiawei Wang, Xudong Wang, Yuankai Wu, Mengying Lei, Jingbo Tian, Fuqiang Liu, Xiaoxu Chen, Chengyuan Zhang, Zhihao Zheng, Wenshuo Wang, Dingyi Zhuang, Kangli Zhu, Zhenyuan Ma, Mojdeh Sharafi, Seongjin

Choi, Wooseok Do, and others, for their help and accompanies.

I express my deepest gratitude to my parents, who gave me life and unconditional love; this thesis is dedicated to you. I also want to deeply thank my two sisters for their care and support since I was a child. Lastly, thanks to my girlfriend Fan Wu, your support and love are my biggest treasure; the rest of my life is dedicated to you.

# Contribution of Authors

This is an article-based thesis and the main contents from Chapter 3 to Chapter 5 are three journal articles. Details of the three journal publications are listed below.

- **Cheng, Z.**, Trépanier, M., Sun, L., 2021. Probabilistic model for destination inference and travel pattern mining from smart card data. *Transportation* 48(4), 2035-2053.
- **Cheng, Z.**, Trépanier, M., Sun, L., 2021. Incorporating travel behavior regularity into passenger flow forecasting. *Transportation Research Part C: Emerging Technologies* 128, 103200.
- **Cheng, Z.**, Trépanier, M. and Sun, L., 2022. Real-time forecasting of metro origin-destination matrices with high-order weighted dynamic mode decomposition. *Transportation Science* (in press).

I declare that I am the first author and the only student author of the three articles. My contributions to the three publications include designing models and experiments, implementing and testing models, and writing manuscripts. The other two authors Prof. Lijun Sun and Prof. Martin Trépanier are my supervisors. They provide guidance, comments, and editorial revision to the articles. Besides the above three articles, the rest part of the thesis is completed by me. Thanks to my supervisors for helping me proofread the thesis.

# Contents

<b>Abstract</b>	i
<b>Résumé</b>	iii
<b>Acknowledgments</b>	v
<b>Contribution of Authors</b>	vii
<b>Contents</b>	viii
<b>List of Figures</b>	xi
<b>List of Tables</b>	xiv
<b>1 Introduction</b>	<b>1</b>
1.1 Background	1
1.2 Research Scope and Objectives	3
1.2.1 Travel behavior features	3
1.2.2 Smart card data applications	4
1.2.3 Objectives	5
1.3 Thesis Contributions	6
1.4 Thesis Organization	7
<b>2 Literature Review</b>	<b>9</b>
2.1 Travel Behavior Patterns in Smart Card Data	9
2.1.1 Measuring travel behavior patterns	10
2.1.2 Applications of travel behavior patterns	11
2.2 Forecasting Using Smart Card Data	13
2.2.1 Individual mobility prediction	13
2.2.2 Station passenger flow forecasting	15
2.2.3 Origin destination matrices forecasting	17
2.3 Summary of existing research	18

<b>3</b>	<b>Probabilistic Model for Trip Destination Inference</b>	<b>20</b>
3.1	Abstract . . . . .	21
3.2	Introduction . . . . .	21
3.3	Literature Review . . . . .	23
3.3.1	Destination inference in smart card data . . . . .	23
3.3.2	Transit pattern mining . . . . .	24
3.4	Methodology: Topic Model for Destination Inference . . . . .	25
3.4.1	Model formulation . . . . .	25
3.4.2	Model inference . . . . .	27
3.4.3	Destination inference and station-to-rank transformation . . . . .	28
3.5	Case study . . . . .	29
3.5.1	Guangzhou Metro data . . . . .	30
3.5.2	Hyperparameters . . . . .	31
3.5.3	Benchmark models . . . . .	32
3.5.4	Scenario 1: using ground truth training set . . . . .	33
3.5.5	Scenario 2: using estimated training set . . . . .	35
3.5.6	Interpreting latent topics . . . . .	36
3.5.7	Passenger clustering . . . . .	39
3.6	Conclusions and discussion . . . . .	41
<b>4</b>	<b>Using Travel Behavior for Boarding Flow Forecasting</b>	<b>43</b>
4.1	Abstract . . . . .	44
4.2	Introduction . . . . .	44
4.3	Methodology . . . . .	48
4.3.1	Problem description . . . . .	48
4.3.2	Returning flow . . . . .	48
4.3.3	Return probability parallelogram (RPP) . . . . .	50
4.4	Experiments . . . . .	54
4.4.1	ARIMA model . . . . .	54
4.4.2	Model selection and evaluation . . . . .	56
4.4.3	Data . . . . .	56
4.4.4	One-step-ahead forecasting . . . . .	58
4.4.5	Multi-step-ahead forecasting . . . . .	60
4.4.6	Forecasting under special events . . . . .	63
4.4.7	Experiments in other models . . . . .	66
4.5	Conclusions and Discussion . . . . .	68

<b>5</b>	<b>Origin-Destination Matrices Forecasting with Dynamic Mode Decomposition</b>	<b>71</b>
5.1	Abstract . . . . .	72
5.2	Introduction . . . . .	72
5.3	Related Work . . . . .	74
5.4	Problem Description . . . . .	76
5.5	Dynamic Mode Decomposition . . . . .	77
5.6	High-order Weighted Dynamic Mode Decomposition . . . . .	79
5.6.1	Model specification . . . . .	79
5.6.2	Model estimation . . . . .	81
5.6.3	Online update . . . . .	83
5.6.4	Connections with other DMD models . . . . .	86
5.7	Experiments . . . . .	86
5.7.1	Data and experimental settings . . . . .	86
5.7.2	Hyperparameters . . . . .	88
5.7.3	Benchmark models . . . . .	89
5.7.4	Forecast result . . . . .	90
5.7.5	Effect of the low-rank assumption . . . . .	94
5.7.6	Effect of the online update . . . . .	96
5.8	Conclusions and Discussion . . . . .	98
<b>6</b>	<b>Final Conclusion &amp; Future Work</b>	<b>99</b>
6.1	Summary of Results . . . . .	99
6.2	Discussion . . . . .	101
6.2.1	Practical implications . . . . .	101
6.2.2	Limitations . . . . .	101
6.2.3	Forecasting in transportation . . . . .	102
6.3	Future Research . . . . .	103
	<b>Bibliography</b>	<b>106</b>

# List of Figures

1.1	Summary of scope and objectives . . . . .	5
3.1	Plate notation for the graphical model. . . . .	27
3.2	The probability of visiting different stations. (a) For passengers that have been observed to visit 5, 10, 20, and 30 different stations, the rank of the stations (in the order of the visit frequency) and the visiting probabilities in a log-log scale. The insert figure shows that the four groups of $p(r)$ can be well approximated by $p(r) \sim r^{-\eta}$ , when applying $\eta = 3.57N^{-0.38}$ . (b) The histogram for the number of different stations visited by each passenger, in the 10000 passengers. . . . .	31
3.3	Bottom left: the 2D histogram of the number of passengers, gridded by the number of trips of each passenger in the training set and the entropy of their trips. Bottom right: the destination inference accuracy of two models for passengers at different entropy levels and the histogram of entropy. Top: the destination inference accuracy of two models for passengers with different numbers of trips in the training set and the histogram of the number of trips in the training set. (Accuracies are shown by means.) . . . . .	34
3.4	Topic-word distributions. (a) The departure time distributions of the four time topics. (b) The origin distributions of the four origin topics. (c) The destination distributions of the four destination topics. . . . .	37
3.5	The latent topic distribution of a passenger. . . . .	38

3.6	Top: the dendrogram of the hierarchical clustering on 500 passengers. Bottom: the feature matrix for the clustering; each column in the matrix is a vectorized latent topic distribution of one passenger. The tick labels on y-axis represent the semantic meanings of latent topics: such as, R1R2 means the origin topic that peaks at ranked 1 <sup>st</sup> station and the destination topic that peaks at ranked 2 <sup>nd</sup> station. The four temporal topics are separated by the minor ticks between every two major ticks and ordered from early morning topic to night topic. . . . .	40
4.1	Illustration of two passenger groups (G1/2) and the boarding demand forecasting problem at station $s$ . . . . .	49
4.2	The composition of the boarding flow in Guangzhou metro. Based on the smart card data from July 21 to 28, 2017. . . . .	50
4.3	The density histogram of the return time interval ( $\tau_{\text{boarding}} - \tau_{\text{alighting}}$ ) of two example stations. . . . .	52
4.4	The return probability parallelogram (RPP), the alighting flow $m_t^s$ , and the returning flow $r_t^s$ for two representative stations: (a) A typical station in commercial areas. (b) A typical station in residential areas. . . . .	53
4.5	The boarding flow and the estimated returning flow. As marked by red arrows, the returning flow reflects the irregular increases/drops of the boarding flow: (a) Tiyu Xilu station, (b) Luoxi station, (c) Changshou Lu station, and (d) Huijiang station. The green curves represent demand time series $y_t^s$ and the orange curves correspond to the estimated returning flow time series $\hat{r}_t^s$ . . . . .	57
4.6	Top: the dendrogram for the hierarchical clustering based on RPP. Bottom: the test set SMAPE differences between M2 and M0; green and negative values means using $\hat{r}_{t+1}^s$ improves the boarding flow forecast in the test set. . . . .	60
4.7	The cluster centroids. (a), (b), and (c) correspond to the cluster centroids of cluster 1, 2, and 3, respectively. . . . .	61
4.8	The alighting, boarding, and the estimated returning flow of Luogang station under an event (August 30). . . . .	63
4.9	The RPP for passenger flow induced by events in Luogang station. Probabilities are set to zero for time windows without event. . . . .	65
4.10	The boarding and the returning flow (using Eq. (4.12)) of Luogang station. . . . .	65

5.1	The singular values of a ten-day-length $Y_{t-1}$ collected from Guangzhou metro smart card system. . . . .	78
5.2	Model framework for HW-DMD. Model input $X$ contains $hn$ rows for lagged OD snapshots and $2s$ rows for lagged boarding snapshots. Columns in $Y$ and $\hat{Y}$ are, respectively, real and forecasted snapshots for OD flow. Model coefficients are estimated by weighted historical data ( $X_t^w$ and $Y_t^w$ ) and updated daily whenever new data come. . . . .	81
5.3	The histogram of $o_{i,j}$ in an OD snapshot of a morning peak in Guangzhou. . . . .	88
5.4	The effect of $\rho$ to the forecast OD RMSE in the validation set of the Guangzhou data. . . . .	89
5.5	The RMSE of OD flow forecasting at different times and different OD pairs. (a) and (c) shows the RMSE of OD matrix forecasting at every 30-minute interval, along with the total OD flow in the network. Using $2^i$ as boundaries, we divide OD pairs into groups according to their average half-hour OD flow; the forecast RMSE at each group and the number of OD pairs of each group are shown in (b) and (d). . . . .	93
5.6	The real and forecasted time series of four selected OD pairs of Guangzhou metro. (a) is the busiest OD pair in the Guangzhou metro data set. (a) to (d) are in a flow decreasing order. . . . .	94
5.7	Comparing OD flow with its low-rank approximation in two Guangzhou metro OD pairs. (a) and (b) corresponds to the (a) and (d) in Figure 5.6, respectively. . . . .	95
5.8	The evolvement of forecast RMSE under different HW-DMD update methods (Guangzhou data set). Each marker represents the RMSE of forecasted OD flow during a day. Numbered dates in the horizontal axis are Mondays; weekends are skipped. . . . .	97

# List of Tables

2.1	Literature of short-term metro passenger flow forecasting. . . . .	14
3.1	The destination inference accuracy of scenario 1. . . . .	33
3.2	The destination inference accuracy and coverage of scenario 2. . . . .	36
4.1	The one-step boarding flow forecasting of four stations. . . . .	58
4.2	Paired t-test p-values for absolute forecast errors. . . . .	59
4.3	The multi-step boarding flow forecasting of four stations by time series cross-validation (30-min resolution). . . . .	62
4.4	The multi-step boarding flow forecasting of four stations by time series cross-validation (15-min resolution). . . . .	63
4.5	The boarding flow forecasting under special events for Luogang stations. . .	66
4.6	The one-step boarding flow forecasting of four stations by SVR . . . . .	67
4.7	Paired t-test p-values for SVR absolute forecast errors. . . . .	67
4.8	The one-step boarding flow forecasting of four stations by MLP . . . . .	67
4.9	Paired t-test p-values for MLP absolute forecast errors. . . . .	68
4.10	The number of significant stations in the paired t-test between M2 and M0 (0.05 significance level). . . . .	68
5.1	Models' performance for OD flow forecasting. . . . .	91
5.2	Models' performance for boarding flow forecasting. . . . .	92
5.3	The difference between original data and the low-rank approximation. . . .	95

# Chapter 1

## Introduction

### 1.1 Background

As an efficient, high-capacity, and green transportation mode, the metro is playing an ever-important role in shaping future sustainable transportation. By the end of 2017, there were metros in 178 cities and 56 countries, carrying on average 168 million passengers per day ([Union Internationale des Transports Publics \(UITP\), 2018](#)). These numbers are still growing rapidly, especially in Asia-Pacific, Latin America, the Middle East, and North Africa. The rapid growth of metro ridership and infrastructure brings new challenges to metro operations, which calls for better management strategies and a better understanding of the system. In the meanwhile, new technologies, concepts, and uncertainties (e.g., autonomous driving, mobility as a service, COVID-19 pandemic) constantly change our vision for the future of transportation.

Smart card systems have been widely adopted as automatic fare collection systems for metros and general public transit. Aside from fare collection, smart card systems also record refined trip-level information, such as trip start/end locations and times. As a result, data in metro systems are becoming much more accessible than ever before. Smart card data has been extensively used in counting passenger flow, evaluating system performance, planning, and operation. With the advances of intelligent transportation systems (ITS), it is now possible to collect and use data from other devices, such as Wi-Fi, Bluetooth, and video camera. Meanwhile, transit agencies and users desire more timely and accurate access to data (e.g., real-time passenger demand, vehicle location, congestion, and accident information) to support their decision-making and trip planning. The ways of collecting and analyzing data are becoming more and more automated and intelligent. In such an era of rapid technological advancement, utilizing the massive data for better metro services is

of great importance.

The keys to successful data-driven transit service include three progressive levels (Ceder, 2016): (1) collecting and understanding data, (2) planning and decision-making, and (3) operations and control. Located in the first level, understanding passengers' travel behavior in metro systems is a fundamental task—this is not only an engineering problem but also a scientific problem. Typical research for metro travel behavior includes analyzing the regularity (Sun et al., 2013), the variability (Morency et al., 2007), or the evolution (Briand et al., 2017) of travel behavior. These studies have greatly enhanced our understanding of travelers' mobility patterns. For instance, we now know individuals' metro trips are often highly regular; a significant portion of passengers repeat similar trips as they did in the past (Ma et al., 2013). In terms of trip variability, the rank-frequency distribution of an individual's visited locations approximately follows a power law (Hasan et al., 2013). However, despite that our knowledge of metro travel behavior has grown significantly, the road to transforming the knowledge into concrete applications to improve metro services (i.e., the second and the third levels) is still long. In other words, the current practice in the second and third levels does not fully exploit the knowledge we obtained from the first level.

There is still a lack of recognition for the role of travel behavior in supporting metro planning and operation. Taking real-time passenger demand forecasting—a crucial task in metro operation—as an example. Most existing models (such as statistical time series models (Ding et al., 2017), machine learning models (Chen et al., 2011), and deep learning models (Liu et al., 2019b)) overlook the fact that passenger flow consists of trips from each individual traveler with strong regularity rooted in their travel behavior. A commuter's work trip in the morning can help predict their home trip in the evening, while this travel behavior cannot be explicitly encoded in current forecasting models. Current forecasting methods don't "understand" how the ridership is generated, and little research evaluates the impact of individuals' travel behavior on the aggregated ridership. Broadly speaking, the function and the potential of travel behavior patterns in metro operation and planning are not well recognized.

Given the weak connection between the "understanding of travel behavior" and the "application of data", it is essential to explore whether and to what extent we can use the travel behavior characteristics to improve data-driven applications in a metro system. Therefore, a major motivation of this thesis is to build a link between the two isolated parts: using travel-behavior characteristics to improve existing and develop new and better data-driven applications in metro systems. The two specific applications studied in this

thesis are: (1) trip destination inference for tap-in only smart card systems and (2) real-time passenger demand forecasting. For the first application, collecting passenger's origin and destination (OD) has always been a critical step in the long-term planning of metros. For the second application, real-time passenger demand forecasting is a foundation of short-term operation and monitoring (e.g., temporary increasing supply to satiate a sudden demand surge, informing passengers of real-time metro congestion information).

## 1.2 Research Scope and Objectives

Smart card data are no doubt still the largest and the most detailed data available in metro systems. All models and methods developed in this thesis are tested on the anonymous smart card data from Guangzhou or Hangzhou, China. The scope of travel behavior and data-driven applications in smart card data is too general; this thesis particularly focuses on the behaviors and applications related to "mobility", i.e., passengers' movement characterized by time and locations. The basic unit of mobility in this thesis is a trip of a passenger. More refined travel behavior and phenomena, like the route choice behavior within a trip, are not in the scope of the thesis.

### 1.2.1 Travel behavior features

Passengers' travel behavior characteristics have been extensively studied using smart card data (refer to the review in Section 2.1). I extract four main travel behavioral features of metro systems as follows. These travel behavioral features are used to assist the inference and forecasting models developed in this thesis.

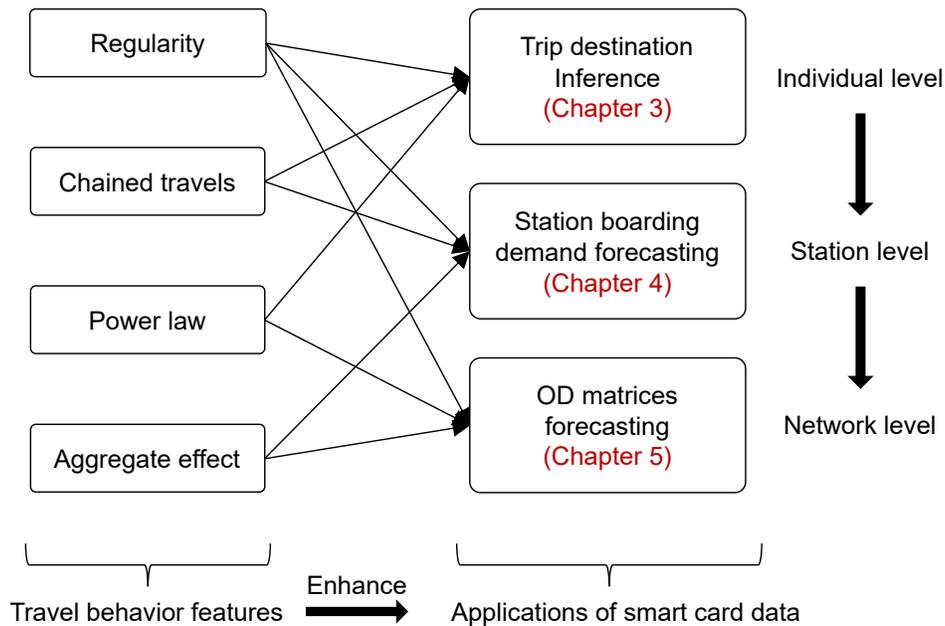
- **Feature 1: Regularity.** Research has shown that passengers traveling in metro systems are highly regular (e.g., Sun et al., 2013; Goulet-Langlois et al., 2017). A passenger usually repetitively visits similar locations at a similar time of the day. For example, a student travels every day from home to school by metro at around 8:00 a.m. In the meanwhile, human behavior comes with randomness, the regularity is in a statistical sense.
- **Feature 2: Chained travels.** In metro systems, a passenger's boarding station is usually identical to the alighting station of their previous trip, especially when the interval between two trips is short. This property has been widely used in literature (e.g., Barry et al., 2002; Trépanier et al., 2007; Zhao et al., 2007) for trip destination inference.

- **Feature 3: Power law.** A power law relationship between two quantities means the value of a quantity is inversely proportional to the power of the other quantity (Newman, 2005). A power law depicting a rank-frequency relationship is also named Zipf's law. Many quantities in metro systems appear to follow a power law. For example, the frequency with which stations are visited by a passenger (see Figure 3.2 (a)), the flow distribution between different origin-destination (OD) pairs (see Figure 5.3).
- **Feature 4: Aggregate effect.** Many macroscopic quantities in metro systems are the aggregate effect of individual travelers. For instance, a metro system's passenger flow/demand consists of the movements of all individual travelers; the boarding demand at a metro station is the summation of the OD demand from this station to all other stations.

## 1.2.2 Smart card data applications

Metro smart card data have a wide range of applications (Pelletier et al., 2011). Certainly, some applications are related to passengers' travel behavior while others are not. This thesis addresses two specific types of applications: trip destination inference and short-term passenger demand forecasting.

- **Application 1: Trip destination inference.** Many smart card systems only require passengers to tap their smart card at the origin station. As a result, trip destination—a critical piece of information in transportation planning—is not recorded in these tap-in-only smart card systems. Traditional destination inference methods (based on the chained travel feature) cannot properly estimate the destinations of isolated trips. Therefore, this thesis aims to develop a probabilistic model based on travel regularity for trip destination inference. The probabilistic model can improve the destination inference accuracy for isolated trips.
- **Application 2: Passenger demand forecasting.** Short-term (from minutes to days) passenger demand forecasting is crucial for real-time metro operation and monitoring. This thesis focuses on the forecasting of boarding passenger demand (1) at metro stations and (2) in OD matrices. Particularly, the travel behavior characteristics listed in Section 1.2.1 will be used to address the challenges of multi-step forecasting, forecasting under special events, and OD matrices forecasting.



**Figure 1.1:** Summary of scope and objectives

A large body of literature has proposed solutions to the above two applications (see Section 2.2 and Section 3.3.1) given their great importance. The major purpose is not to claim methods developed in this thesis are better than those in the literature; comparisons are always conditional. The value of this thesis is proposing a new way of thinking—from a travel behavioral perspective—to the above inference and forecasting tasks. Integrating travel behavior to inference and forecasting achieves certain effects and advantages that other methods do not have.

### 1.2.3 Objectives

The overall objective of the thesis is to develop better inference and forecasting methods for metro planning and control with travel behavior characteristics. The two major focuses are utilizing the domain knowledge from travel behavior and achieving good performance in inference and forecasting problems. All models developed in the thesis utilize certain types of travel behavioral characteristics listed in Section 1.2.1. The links between travel behavioral features and their corresponding applications are shown in Figure 1.1.

As shown in Figure 1.1, Chapter 3 to Chapter 5 develop different applications using certain types of travel behavior features. The research targets in these applications lift

from an individual level to a station level, and finally to a network level. I summarize the objectives of each chapter as follows:

- **Objective 1: Develop a probabilistic model for trip destination inference.** Passengers' trip destination is important information for the planning and evaluation of metro systems. However, the trip destination is not available in tap-in-only smart card systems. Although most trips' destinations can be estimated by the origin of the immediate next trip, there lacks an effective method to infer the destinations of isolated unlinked trips. Therefore, Chapter 3 aims to develop a probabilistic model to infer the destinations of unlinked trips by learning travel patterns from historical trips. The trip destination inference is an individual-level application.
- **Objective 2: Use travel behavior regularity to improve passenger flow forecasting.** Forecasting the short-term passenger demand at each metro station has always been an important and popular topic. Previous studies mainly model passenger flow as time series by aggregating individual trips and then forecasting based on the values in the past several steps. However, this approach overlooks the fact that passenger flow consists of trips from each individual with predictable travel behavior. Chapter 4 thus attempts to incorporate travel behavior regularity into passenger flow forecasting for improved performance.
- **Objective 3: Develop a forecasting model for high-dimensional OD matrices.** The ridership among OD pairs contains more refined details than the station-level passenger demand. However, forecasting OD ridership is notoriously difficult due to the high-dimensional, sparse, noisy, and skewed nature of OD matrices. In response to these challenges, Chapter 5 focuses on developing an applicable and accurate OD matrices forecasting model. Using the proposed model as a bottom-up approach for station-level boarding flow forecasting is also a meaningful attempt.

### 1.3 Thesis Contributions

The detailed contributions of each model/application are provided individually in each chapter. The following are the high-level summary of the contributions of this thesis:

- **Contribution 1:** This thesis bridges travel behavior patterns and certain data-driven applications in metro systems; results show using the travel behavior features can

significantly enhance the performance of inference and forecasting models for the individual, station, and network levels applications in smart card data. Compared with previous research that implicitly/sporadically uses travel behavior characteristics in their models, the contribution of this thesis is providing a holistic and systematic view of the critical role of travel behavior in the planning and operation of metro systems. This new perspective has broad application scenarios and creates a new future research direction.

- **Contribution 2:** This thesis proposes several inference and forecasting models for smart card data with improved accuracy. Specifically, the probabilistic topic model in Chapter 3 improves the inference accuracy of unlinked trips' destinations. The travel-behavior-based boarding flow forecasting method in Chapter 4 enhances the accuracy of multi-step forecasting and the forecasting under special events, and a paired t-test shows the improvement is significant. The forecasting model proposed in Chapter 5 achieves a good forecasting performance in both OD matrices and boarding flow forecasting. With these better models, this thesis helps enhance the service and intelligence of metro systems.
- **Contribution 3:** This thesis addresses the scalability and model maintenance issue in a practical forecasting model. Chapter 5 proposes to use dynamic mode decomposition (DMD) to address the scalability issue in OD matrices forecasting. Moreover, a tailored online update algorithm is developed to update the coefficients of a forecasting model efficiently without storing historical data or retraining, which maintains the model performance over the long term at low costs.

## 1.4 Thesis Organization

This is a manuscript-based thesis with six chapters, where Chapter 3 to Chapter 5 are based on articles that were either published or accepted by peer-reviewed journals. The follows are the chapter-level organization of the thesis:

- **Chapter 1** introduces the background, motivation, objectives, and contributions of this thesis. The links between each chapter are also explained.
- **Chapter 2** gives an overall literature review of smart card data from three aspects: (1) travel behavior patterns, (2) data inference and imputation, and (3) passenger demand forecasting.

- **Chapter 3** presents a probabilistic topic model to infer the destinations of unlinked trips for tap-in-only smart card systems. This Chapter also uses the probabilistic model for travel behavior pattern analysis and passenger clustering.
- **Chapter 4** proposes a travel-behavior-based feature—returning flow. This chapter shows using the returning flow in a forecasting model can significantly improve the passenger flow forecasting under various scenarios (one-step, multi-step, under special event).
- **Chapter 5** formulates real-time OD matrices forecasting into a high-order vector autoregression problem and proposes a dynamic mode decomposition (DMD) approach to solve the problem.
- **Chapter 6** summarizes the thesis with final conclusions and future directions.

# Chapter 2

## Literature Review

The core of this thesis is incorporating travel behavior characteristics into forecasting models in metro smart card data. Accordingly, the literature review consists of two parts: Section 2.1 reviews research on modeling individual travel behavior patterns by smart card data, and Section 2.2 summarizes three levels of forecasting (individual mobility, station passenger demand, and OD matrices) in metro systems. The purpose of this chapter is not to gather all the related articles, but to summarize the methodological development in this field through comparison and classification of representative studies. Most of the literature in this chapter is from the past two decades, after the widespread adoption of smart card automatic fare collection systems in public transit. Readers are referred to a great review by [Pelletier et al. \(2011\)](#) for the history and early-stage application of smart card systems in public transit. This chapter focuses on summarizing the overall research development and trend; more detailed reviews for destination inference and OD matrices forecasting are supplemented in Section 3.3 and Section 5.3, respectively.

### 2.1 Travel Behavior Patterns in Smart Card Data

Measuring individuals' travel regularity and discovering common mobility patterns from the population are critical for various applications. Travel survey is the traditional approach to study travel behavior (e.g., [Kitamura, 1990](#); [Axhausen et al., 2002](#)). In recent years, smart card data have largely replaced survey data in studying the travel behavior of public transit users. Although it only contains limited types of information (e.g., origin, destination, time, and card type), smart card data enables recording long-term and large-scale passenger mobility data, bringing new methods and applications in travel behavioral studies. Section 2.1.1 and Section 2.1.2 summarize research on measuring travel behavior

patterns and corresponding applications using transit smart card data, respectively.

### 2.1.1 Measuring travel behavior patterns

Passengers' travel in metro systems is highly regular (Sun et al., 2013; Hasan et al., 2013). Although there is not a unified definition for travel behavior patterns in metro systems, a large number of methods have been developed to extract, measure, and understand meaningful facts/routines encoded in smart card transactions. These methods are categorized into the following two types.

- **Using handcrafted features:** In an early work, Morency et al. (2007) designed indicators, such as the number and the frequency of stations used for boarding, to measure passengers' spatial and temporal variability. A k-means algorithm was then conducted to cluster passengers. Ma et al. (2013) defined spatial travel pattern as "transit rider repeatedly visits the same or adjacent places on a multi-day basis" and temporal travel pattern as "transit rider repeatedly starts (and/or finishes) his/her daily trip during the same time period". Density-based spatial clustering of applications with noise (DBSCAN) algorithm was used to identify these repeated patterns. A similar definition was also used in a few follow-up studies (Bhaskar et al., 2014; Kieu et al., 2015; Ma et al., 2017). Zhao et al. (2017) defined spatial pattern and temporal patterns by designing statistics on how often a passenger repeatedly takes metro trips in certain stations and times. Ghaemi et al. (2017) proposed a projection method to project smart card transactions to a polar axis, in which the spatiotemporal closeness can be conveniently measured Euclidean distance, bringing benefits to pattern visualization and clustering. Goulet-Langlois et al. (2017) proposed to use entropy rate to measure individual travel regularity. The aforementioned handcrafted approaches are very flexible for specific analyses and are intuitively straightforward. However, using handcrafted features often requires proficient expertise and the result can be subjective. Besides, this approach is usually limited to discovering superficial phenomena.
- **Using learned latent patterns:** Recent research started using algorithms to automatically extract latent representations (also called embedding) from smart card data to understand travel behavior patterns. This approach learns shared latent features from a population rather than independent individuals, discovering intrinsic patterns that cannot be defined by a human expert. Matrix/tensor factorization is a primary method for latent pattern discovery. For example, Goulet-Langlois et al. (2016) used

principal component analysis (PCA) to extract eigen-patterns from passengers' multi-week activity sequences. [Du et al. \(2019\)](#) developed a coupled factorization on a (*origin*  $\times$  *transfer*  $\times$  *destination*) tensor and a (*origin*  $\times$  *destination*) matrix to extract mobility patterns from transit data. Developing probabilistic models with latent variables is another approach for latent pattern discovery. For instance, [Mohamed et al. \(2014\)](#) used a mixture of unigrams to capture passengers' weekly travel profiles. [Sun and Axhausen \(2016\)](#) proposed a probabilistic tensor factorization to extract mobility patterns from smart card data. [Briand et al. \(2016\)](#) developed a mixed Gaussian model to extract latent features to mining passengers' temporal travel patterns. [Zhao et al. \(2020b\)](#) extended Latent Dirichlet Allocation (LDA) to discover spatiotemporal patterns from smart card data. Chapter 3 of this thesis is also a variant of LDA.

Beyond the scope of smart card data, similar methods have been developed to investigate latent mobility patterns in other human mobility data sources, such as the cell phone data ([Farrahi and Gatica-Perez, 2009](#)), location-based social media data ([Hasan and Ukkusuri, 2014](#)), and license plate recognition data ([Sun et al., 2021](#)).

## 2.1.2 Applications of travel behavior patterns

Understanding passengers' travel behavior itself is an important "application". The incentive of many studies is simply answering certain interesting questions, such as how often users travel ([Morency et al., 2007](#))? How loyal are users ([Trépanier et al., 2012](#))? Is there any common law for users' travel ([Hasan et al., 2013](#))? Essentially, nearly all articles reviewed in this charter, to some extent, enhance the understanding of the metro system and users' travel behavior. Besides enhancing understanding, applications of travel behavior patterns in smart card data are categorized by the following five aspects. There are intersections among different applications because the following aspects share similarities.

- **Passenger clustering:** Having represented each passenger's travel behavior by either handcrafted or learned features, an immediate application is passenger clustering (or segmentation), allowing transit agencies to understand the composition of different types of users. In fact, most models reviewed in Section 2.1.1 can be used to provide features for passenger clustering. Commonly used cluster algorithm includes the K-means ([Morency et al., 2007](#); [Ma et al., 2013](#)) and hierarchical clustering ([Ghaemi et al., 2017](#); [He et al., 2020](#)). An extra clustering step is sometimes unnecessary when using latent class models (e.g., [Mohamed et al., 2014, 2016](#); [Sun and Axhausen, 2016](#);

Briand et al., 2016), because the learned latent patterns can be regarded as cluster centroids and each passenger's travel behavior is characterized by a mixture of clusters.

- **Analyze pattern evolution:** Many researchers investigated how travel patterns change over time by long-term data, where a travel pattern extraction and a clustering step are often combined with. For example, Trépanier et al. (2012) used a hazard model to analyze the influential factors to the survival (keep using their smart cards) of different types of users. Using the mixed Gaussian model developed in (Briand et al., 2016), Briand et al. (2017) analyzed the pattern changes by a five-year smart card data collected in Gatineau, Canada. Similar research includes (Deschaintres et al., 2019; Viallard et al., 2019). Zhao et al. (2018a) developed a Bayesian approach to detect pattern changes in individual travel behavior. (Ma et al., 2020) evaluated the behavioral response of different groups of passengers to promotion in public transport. Gao et al. (2022) evaluated passengers' travel behavior before and after a transit service adjustment. It is worth mentioning a recent study by Mützel and Scheiner (2021) investigated changes in metro mobility patterns under the impact of COVID-19.
- **Data imputation:** Smart card data lack information like trip purpose, passengers' social demographic, or even trip destinations. To impute trip purpose from smart card data, Han and Sohn (2016) proposed to use continuous hidden Markov model (CHMM) to derive clusters of activities and transition probabilities between clusters; each cluster was then attached to an activity pattern (e.g., home, work). Mo et al. (2021) used input-output hidden Markov model (IOHMM) to infer and interpret activities patterns. The spatiotemporal topic model developed by Zhao et al. (2020b) also reveals meaningful latent activities. Alsgar et al. (2018) combined travel behavior patterns with land use and survey data to impute trip purpose from smart card data. Besides, there is a large body of research that uses trip-chain continuity property to infer trip destination from smart card data (e.g., Barry et al., 2002; Trépanier et al., 2007); a detailed review on this topic is organized in Section 3.3.1.
- **Individual mobility prediction:** Individual mobility prediction in smart card data refers to predicting the origin, destination, and time of a passenger's next metro trip. The idea is to forecast individuals' future travel based on their historical travel patterns, such as in the articles by Zhao et al. (2018c); Mo et al. (2021). A detailed review on this topic is organized in Section 2.2.1.

- **Anomaly detection:** Abnormal travel behavior can be identified after understanding normal travel patterns. [Du et al. \(2018\)](#) defined several criteria as abnormal travel behavior, and proposed an unsupervised anomaly detection model and a supervised classification model to identify pickpocket suspects from smart card transactions. Similarly, [Xue et al. \(2020\)](#) treated frequent entry and exit from a station as a suspected travel behavior. [Zhang et al. \(2021c\)](#) identified invalid fare machine association (e.g., wrongly associate a fare machine of station A to station B in Automatic Fare Collection (AFC) database) by detecting abnormal transaction behavior.

## 2.2 Forecasting Using Smart Card Data

This section reviews short-term (from minutes to days) forecasting models in metro smart card data. Unlike long-term planning, the purposes of short-term forecasting are fine-grained operation, control, monitoring, and providing real-time services. The literature is organized by three levels of forecasting: (1) individual mobility prediction, (2) station-level passenger demand forecasting, and (3) network-level OD matrices forecasting. Individual mobility prediction centers on learning individual behavioral patterns; station-level passenger demand forecasting models focus more on capturing spatiotemporal correlations; OD matrices forecasting methods have to handle the high-dimensionality challenge. References reviewed in this section are summarized in [Table 2.1](#), from which we can see an increasing trend of research items year by year.

### 2.2.1 Individual mobility prediction

Predicting the location and time of an individual's future movement has received great attention because of its critical role in trip recommendation and demand forecasting. A fundamental assumption of individual mobility prediction is that an individual follows the same travel behavior pattern in the past and future. Sequential models, like Markov models and recurrent neural networks (RNN), are commonly used in individual mobility prediction. For example, [Hasan et al. \(2013\)](#) proposed a simple model that chooses among home, work, and others (with fixed probabilities) as an individual's next trip destination, and the stay duration at each location is modeled by a hazard function. Because the frequency of visiting different metro stations of an individual follows a power law, [Hasan et al. \(2013\)](#) showed even this simple model makes reasonable forecasting. [Zhao et al. \(2018c\)](#) used Bayesian  $n$ -gram model to simultaneously predict the origin, destination, and

**Table 2.1:** Literature of short-term metro passenger flow forecasting.

Author year	Forecast objective(s)	Method	Remarks
(Wei and Chen, 2012)	Boarding	EDM, MLP	N/A
(Hasan et al., 2013)	Next destination, duration	Heuristic, hazard function	Individual mobility
(Sun et al., 2015)	Boarding	Wavelet analysis, SVM	N/A
(Ni et al., 2016)	Boarding	Linear regression, SARIMA	Events
(Toqué et al., 2016)	OD matrices	LSTM	N/A
(Ding et al., 2017)	Boarding	ARIMA, GARCH	N/A
(Li et al., 2017)	Alighting	Multi-scale RBF network	Events
(Toqué et al., 2017)	Boarding	LSTM, random forest	N/A
(Ren and Xie, 2017)	OD	Tensor decomposition	Dimensionality reduction
(Dai et al., 2018)	Boarding, alighting	KNN, Adaboost,	N/A
(Wang et al., 2018)	Boarding	Linear regression	N/A
(Tang et al., 2018)	Boarding	SVM, ARIMA, linear regression	A comparison
(Ma et al., 2018)	Boarding	CNN, Bi-LSTM	Spatiotemporal correlation
(Noursalehi et al., 2018)	Boarding	Dynamic Factor Models	Event
(Zhao et al., 2018c)	Next origin, destination, time	$n$ -gram model	N/A
(Guo et al., 2019)	Boarding, alighting	SVM, LSTM	N/A
(Han et al., 2019)	Boarding, alighting	GCN	Spatiotemporal correlation
(Liu et al., 2019b)	Boarding, alighting	LSTM	N/A
(Hao et al., 2019)	Alighting	Bi-LSTM, Seq2Seq	Multi-step forecasting
(Wang et al., 2019a)	Boarding	SARIMA	Events
(Zhang et al., 2019b)	OD	LSTM	Skewed and sparse OD
(Chen et al., 2020a)	Boarding, alighting	ARIMA, GARCH	Events
(Sha et al., 2020)	Boarding	LSTM, GRU, MLP	N/A
(Chen et al., 2020b)	Boarding, alighting	GCN, GRU, Seq2Seq	Spatiotemporal correlation
(Zhao et al., 2020a)	Boarding, alighting	LSTM, Holt-Winters	N/A
(Gong et al., 2020)	OD	Matrix factorization	Dimensionality reduction
(Shen et al., 2021)	OD	Gravity Model + deep learning	N/A
(Yang et al., 2020)	Boarding	Attention mechanism	Spatiotemporal correlation
(Zhao et al., 2021)	Boarding	LSTM, RF, gradient boosting	Events
(Cheng et al., 2021a)	Boarding	SARIMA	Behavioral feature, events
(Wang et al., 2021)	Boarding	Hypergraph GCN	Spatiotemporal correlation
(Ye et al., 2021)	OD	GCN, Transformer	Spatiotemporal correlation
(Noursalehi et al., 2021)	OD	Wavelet transform + CNN	Spatiotemporal correlation
(Zhang et al., 2021b)	OD	CNN	Spatiotemporal correlation
(Cheng et al., 2022)	OD, boarding	Dynamic mode decomposition	Dimensionality reduction Online update
(Zúñiga et al., 2021)	OD	MLP	N/A
(Mo et al., 2021)	Next origin, destination, time	IOHMM	Individual mobility
(Wu et al., 2021)	Next origin, destination, time	Neural temporal point processes	Individual mobility
(Xue et al., 2022)	Boarding	CNN, MLP	Events

time of the next trip. [Mo et al. \(2021\)](#) developed an IOHMM model to predict individuals' next trip; an advantage of this model is that trip purposes can be interpreted by meaningful latent activity patterns. [Wu et al. \(2021\)](#) proposed a deep learning approach that uses attentive marked temporal point processes (AMTPP) for individual mobility; results show AMTPP significantly improves the forecasting of the trip start time. Besides using smart card data, individual mobility prediction has also been widely studied in GPS and cell phone data (e.g., [Gambs et al., 2012](#); [Feng et al., 2018](#)).

## 2.2.2 Station passenger flow forecasting

The passenger flow in this section refers to the number of boarding (or alighting) passengers at a metro station per unit time (usually from minutes to hours for short-term forecasting). Passenger flow is an aggregated quantity of individuals' travel. In contrast to modeling users' behavior regularity in individual mobility prediction, passenger flow forecasting is often formulated into time series models, and the core is capturing the temporal (also spatial when multivariate time series) correlation/dependencies in the time series. In general, short-term metro passenger flow forecasting methods can be categorized into statistical time series models, (shallow) machine learning models, and deep learning models.

- **Statistical time series models:** Statistical time series forecasting assumes time series follows certain structures (such as a stochastic process), representative models include exponential smoothing model, autoregressive integrated moving average (ARIMA). The book by [Box et al. \(2015\)](#) is a great reference for this class of models. Statistical time series models have been widely used in general traffic forecasting problems (e.g., [Williams, 2001](#); [Williams and Hoel, 2003](#)) and are often jointly used with other models. For example, [Ni et al. \(2016\)](#) applied a linear regression model to capture the correlations between passenger flow and social media data; the regression was embedded into a seasonal ARIMA (SARIMA) to better forecast the passenger flow under special events. Combined with ARIMA, much research applied generalized autoregressive conditional heteroskedasticity (GARCH) model to forecast the passenger flow under special events ([Ding et al., 2017](#); [Chen et al., 2020a](#)); results show GARCH gives more accurate forecasting and confidence interval. [Noursalehi et al. \(2018\)](#) used a dynamic factor model for station passenger flow forecasting considering the impact of special events. Based on travel behavior, [Cheng et al. \(2021a\)](#) (Chapter 4 of this thesis) proposed a concept called "returning flow" and

used it as an external covariate in a SARIMA model; results show the behavior-based variable can significantly improve the forecasting. In addition to being used alone, models like ARIMA and exponential smoothing were used as a component in more sophisticated models to improve the forecasting (Sun et al., 2020; Zhao et al., 2020a).

- **Machine learning models:** Machine learning-based passenger flow forecasting uses non-linear/non-parametric regression functions to capture the complex spatiotemporal correlations. This class of models usually focuses more on minimizing regression errors but does not impose assumptions on the process/dynamic of time series. Examples include multilayer perceptron (MLP) (Wei and Chen, 2012; Li et al., 2017), k-nearest neighbors (KNN) (Dai et al., 2018), support vector machine (SVM) (Sun et al., 2015; Tang et al., 2018; Guo et al., 2019), random forest (Toqué et al., 2017; Dai et al., 2018), and adaptive boosting (AdaBoost) (Dai et al., 2018). The forecasting performance is usually largely determined by the quality of input features. For instance, Wei and Chen (2012) decomposed passenger flow into intrinsic mode functions (IMF) by empirical mode decomposition (EMD). The meaningful IMFs were extracted as inputs for an MLP. Sun et al. (2015) applied wavelet analysis to passenger flow and used a support vector machine (SVM) model to forecast the resulting signature; the final results were reconstructed to passenger flow. Tang et al. (2018) established a spatial feature based on OD matrix and used the feature as an input for an SVM model. Li et al. (2017) used the boarding flows of 18 transfer stations as the input of multi-scale radial basis function (RBF) networks to forecast the alighting flow of a crucial station. Wang et al. (2018) used the flow of loyal travelers as a variable for a linear regression model to forecast passenger flow. Zhao et al. (2021) applied a synthetic minority oversampling technique (SMOTE) to alleviate the imbalanced data sample to improve the forecasting under special events.
- **Deep learning models:** Recent years have seen a surge of research using deep learning for metro passenger flow forecasting. Compared with traditional machine learning methods, deep learning does not require manually designed features, rather extracting features automatically through deep artificial neural networks. Moreover, Traditionally models are usually station-specific, while deep learning models can take the passenger flow of multiple stations as a vector and forecast their ridership in a batch. Long short-term memory (LSTM) networks gated recurrent unit (GRU) networks are two primary RNN structures used for passenger flow forecasting (e.g., Toqué et al., 2017; Sha et al., 2020). These structures are special recurrent neural net-

works (RNN) that can avoid the gradient vanishing or exploding gradient problems. [Ma et al. \(2018\)](#) combined bi-directional LSTM (Bi-LSTM) with convolutional neural networks (CNN) to learn spatiotemporal correlations of ridership at different stations. Results show the deep learning framework outperforms traditional forecasting methods. [Liu et al. \(2019b\)](#) used three LSTM modules to learn the weekly, daily, and recent trends, respectively. The three trends were then fused together to produce the final forecast. [Hao et al. \(2019\)](#) applied attention to a sequence to sequence (Seq2Seq) LSTM network to perform a multi-step forecast of alighting flow. Some recent research in deep learning (e.g., [Vaswani et al., 2017](#)) found that an attention mechanism or CNN in some cases can be as good as or even outperform RNN in modeling sequential data. Therefore, [Yang et al. \(2020\)](#) developed an attention-based neural network to forecast metro passenger flow. Besides temporal correlations, researchers have also developed innovative methods to capture spatial correlations among stations ([Ma et al., 2018](#); [Han et al., 2019](#); [Wang et al., 2021](#)), where CNN and Graph convolutional neural networks (GCN) are extensively used for learning spatial correlations.

In addition to learning spatiotemporal correlation, using/designing appropriate covariates to incorporate the influence of exogenous factors is also critical for forecasting. In this perspective, much research has shown that using social media data ([Ni et al., 2016](#); [Xue et al., 2022](#)), weather information ([Tang et al., 2018](#)), event information ([Zhao et al., 2021](#)), time of a day ([Liu et al., 2019b](#)), metro timetable ([Liu et al., 2019b](#)), metro network structure ([Wang et al., 2021](#)), and returning flow from previous trips ([Cheng et al., 2021a](#)) can be helpful for metro passenger flow forecasting.

### 2.2.3 Origin destination matrices forecasting

Forecasting the short-term ridership among a metro system's origin-destination pairs (OD matrix) has become a new research hotspot. This is because OD matrices contain richer information and have wider application than station-level ridership. For example, we can obtain station-level passenger flow through aggregating ridership in OD pairs that have the same origin station, and we can also assign an OD demand to a network to further forecast the congestion of trains/lines. However, OD matrices forecasting is highly challenging due to the high-dimensional and skewed nature of OD matrices. Specifically, the number of OD pairs is the square of the number of stations, and the demand among OD pairs is power-law distributed (highly skewed). Most station-level forecasting methods

do not scale well to OD matrices forecasting. Significant effort has been made to tackle the high-dimensionality issue; two major types of approaches are (1) using matrix/tensor factorization for dimensionality reduction (Ren and Xie, 2017; Dai et al., 2018; Gong et al., 2020; Cheng et al., 2022), and (2) using CNN or GCN to replace the fully connected layer in deep neural networks to reduce the parameters space (Shen et al., 2021; Chen et al., 2020b; Noursalehi et al., 2021; Zhang et al., 2021b; Ye et al., 2021). Section 5.3 of this thesis gives a more detailed review of these OD matrices forecasting methods.

## 2.3 Summary of existing research

This chapter presents a literature review on metro smart card data from two perspectives: (1) travel behavior pattern mining and (2) forecasting using smart card data. As categorized in Section 2.1.1, designing handcrafted features and learning latent patterns from data are two types of methods to measure travel behavior patterns. Handcrafted features are more flexible for specific purposes, but how to choose such a feature can be subjective and usually requires proficient expertise. On the other hand, learning latent patterns from data can discover intrinsic mobility patterns (through techniques like matrix factorization and topic models) that human experts cannot define; this approach is more demanding in data quality and quantity and, therefore, less flexible. In Section 2.1.2, we present various applications that are closely related to travel behavior. However, most existing passenger demand forecasting models reviewed in Section 2.2 are time series models without utilizing unique travel behavior characteristics in metro systems. Therefore, one of the aims of this thesis is to enhance passenger demand forecasting with travel behavior properties.

For passenger demand forecasting, we can find an increasing trend in the complexity of models. However, a problem occurs that many studies appear to be too homogeneous: it is not difficult to use a new model for passenger demand forecasting, but a new model is not equal to making contributions to this problem. See Section 6.2.3 for a discussion about what is good research on transportation forecasting. Besides, despite prosperous forecasting models, implementations of these models in real-world metro systems are rare. The difficulties in implementation are twofold. Technologically, the data quality of many smart card systems is insufficient, and the data transmission and process speed do not satisfy the requirement of real-time forecasting. Methodologically, passenger demand forecasting, as an “upstream” model, should be combined with a “downstream” control model to maximize its utility. Future research on passenger demand forecasting should

focus more on its practical implications.

## Chapter 3

# Probabilistic Model for Trip Destination Inference

---

This chapter is an article published in *Transportation*:

- Cheng, Z., Trépanier, M., Sun, L., 2021. Probabilistic model for destination inference and travel pattern mining from smart card data. *Transportation* 48(4), 2035-2053.

This chapter corresponds to the travel-behavior-based inference method of this thesis. The probabilistic model learns latent travel behavior patterns from individuals' historical trips and can be used to infer the probabilities of unknown destinations given the trip's origin and departure time.

---

## 3.1 Abstract

Inferring trip destination in smart card data with only tap-in control is an important application. Most existing methods estimate the trip destination based on the continuity of trip chains, while the destinations of isolated/unlinked trips cannot be properly handled. We address this problem with a probabilistic topic model. A three-dimensional Latent Dirichlet Allocation (LDA) model is developed to extract latent topics of departure time, origin, and destination among the population; each passenger's travel behavior is characterized by a latent topic distribution defined on a three-dimensional simplex. Given the origin station and departure time, the most likely destination can be obtained by statistical inference. Furthermore, we propose to represent stations by their rank of visiting frequency, which transforms divergent spatial patterns into similar behavioral regularities. The proposed destination estimation framework is tested on Guangzhou Metro smart card data, in which the ground-truth is available. Compared with benchmark models, the topic model not only shows increased accuracy but also captures essential latent patterns in passengers' travel behavior. The proposed topic model can be used to infer the destination of unlinked trips, analyze travel patterns, and passenger clustering.

## 3.2 Introduction

Origin and Destination (OD) Matrix is an essential input for transit planning and operation. Most transit agencies have been relying on travel surveys to collect representative OD information. However, conducting such a survey with a reasonable scale is not only costly but also time-consuming. With the recent advances of intelligent transportation systems, researchers and practitioners have started taking advantage of the transit operation data and smart card data for better planning and operation practices (Pelletier et al., 2011).

Smart card systems are initially designed for the purpose of Automatic Fare Collection (AFC). When the system has both tap-in and tap-out controls (e.g., using a distance-based transit fare scheme), the full itinerary (boarding time/station and alighting time/station) of each trip can be registered. However, most smart card systems across the world adopt a single fare scheme with only tap-in validation, and the alighting information (time/station) is essentially unknown. Inferring the alighting stations is a crucial problem in obtaining the OD matrix from these smart card systems.

Trip destination estimation in smart card data has always been a hot issue. Barry et al. (2002) proposed two assumptions to address this issue: (1) the alighting station

of a trip is very likely to be the boarding station of the immediate next trip; (2) the last alighting station of a day is usually the first boarding station of the same day. This type of “rule-based” model soon became the workhorse algorithm for smart card destination estimation. Depending on the data, current algorithms can obtain around 60% to 85% trips’ destinations; these trips are often called linked trips in the literature, and the rest un-inferred trips are referred to as unlinked trips. Without the information from consecutive trips, the destination estimation of unlinked trips is more challenging. Existing methods address this problem by seeking similar trips in the passenger’s historical trips; we refer to them as individual-history-based models. Such as [He and Trépanier \(2015\)](#) used the spatial and temporal kernel density probability of passengers’ trips and get an additional 10% estimation for unlinked trips.

The prediction of unlinked trips is challenging without the help of the trip-chain continuity information. The solution lies in the regularity of human mobility. As explained by [González et al. \(2008\)](#) and [Song et al. \(2010\)](#), human movement follows certain regularity and is highly predictable. However, there still lacks an appropriate framework to infer the missing destination using mobility regularity. To address this issue, this paper attempts to build an integrated model that estimates the missing destinations drawing on the common mobility patterns among the population. We establish a probabilistic topic model for smart card data by making an analogy with the Latent Dirichlet Allocation (LDA) model ([Blei et al., 2003](#)). We assume transit trips among the population can be summarized in a few latent topics over departure time, origin, and destination. Every passenger is characterized by a latent topic distribution and the whole population share the topic-word distributions for departure time, origin and destination. To share more information among different passengers, we represent each station by each passenger’s rank of visiting frequency, as against to directly using the station ID. A case study is performed on Guangzhou Metro data, where the tap-out data is used as the ground truth to test different models. Results show our topic model has improved accuracy compared with individual-history-based models. We further demonstrate passengers’ latent topic distribution is a useful feature for passenger clustering, commuter identification, and travel pattern mining.

The remainder of the paper is organized as follows. Section 3.3 briefly reviews the current research on smart card data destination inference and transit pattern mining. Section 3.4 elaborates the topic model for transit trips, the Gibbs sampling for model inference, and the destination inference in ranked stations. The case study on Guangzhou Metro will be shown in Section 3.5, where the destination inference will be compared with individual-history-based models; the model interpretation and the passenger clustering

will be demonstrated in Section 3.5. Finally, conclusions and discussions are summarized in Section 3.6.

## 3.3 Literature Review

### 3.3.1 Destination inference in smart card data

Destination inference is an important problem in smart card data. Existing methods primarily take advantage of the continuity of trip chains, and infer the destinations based on assumptions or rules. In a very first study, [Barry et al. \(2002\)](#) proposed that the destination of a trip can be inferred by the origin of the immediate next trip, and they assumed the last destination of a day is often the first origin in the same day. Since then, many refined models have been proposed based on similar assumptions. [Trépanier et al. \(2007\)](#) imposed a distance constraint between consecutive trips, and they further assumed the last destination of a day can also be inferred by the first origin in the next day. [Munizaga and Palma \(2012\)](#) proposed to use generalized time instead of distance in destination inference. Further, [Sánchez-Martínez \(2017\)](#) constructed a generalized disutility minimization objective to determine the paths and transfers between the origin and destination. Research based on similar rule-based methodology has become the mainstream, and more research can be found in [Zhao et al. \(2007\)](#), [Wang et al. \(2011\)](#), [Gordon et al. \(2013\)](#), and [\(Nunes et al., 2016\)](#). Depending on the data, the rule-based method can accomplish around 60% to 85% of the destinations; trips of which the destinations can be inferred by the rule-based model are often called linked trips.

For the O-D of unlinked trips, whose destination cannot be inferred by rule-based models, one treatment is to scale the O-D of linked trips by some methods (e.g., [Munizaga and Palma, 2012](#); [Gordon et al., 2018](#)). This approach assumes the destination distribution of unlinked trips at each origin is the same as the linked trips, which is unverified. On the other hand, the destinations of unlinked trips can be estimated by similar historical trips (individual-history-based model), similar to supervised learning with labeled data. For example, [Trépanier et al. \(2007\)](#) defined a similar trip as a trip on the same route with similar departure time in the previous several days. [He and Trépanier \(2015\)](#) used spatial and temporal kernel density probability estimated by historical trips to infer the destination of unlinked trips. [Zhang et al. \(2015\)](#) conducted an interesting study, where a collaborative space alignment framework was presented to reconstruct smart card trips. Recent studies attempted to use (deep) neural networks to infer trip destinations ([Jung and](#)

[Sohn, 2017](#); [Assemi et al., 2020](#)). These studies were based on smart card systems with full information and extensive features (e.g., time and location, land-use features of stations). Experiments showed promising results, while a large number of labeled destinations are essentially unavailable for a real tap-in-only system.

To summarize, existing research has developed various algorithms based on the trip continuity feature to estimate the destination of linked trips. The destination estimation of unlinked trips relies on similar historical trips. This paper provides a whole new approach to infer the destination of unlinked trips by a topic model. The proposed model is not only a prediction model but also a generative model that captures individuals' behavioral patterns.

### 3.3.2 Transit pattern mining

There has been a large body of literature on passengers' travel behavior patterns. The travel patterns are usually characterized by certain features. A series of analyses (such as commuter identification, passenger clustering, and pattern evolving) can then be conducted using these features. Next, we briefly review related literature based on how these features are obtained.

In many studies, the features for travel patterns are designed based on domain knowledge. For example, [Morency et al. \(2007\)](#) defined two indicators to measure passengers' spatial and temporal variability. Then a k-mean algorithm was conducted to cluster passengers. In another research, [Ma et al. \(2013\)](#) designed four features based on how often did a passenger repeatedly visits the same or adjacent places on a multi-day basis; these features can be used to identify regular passengers. A similar approach is also applied in [Ma et al. \(2017\)](#). [Mohamed et al. \(2014\)](#) established a temporal profile by passengers' travel time on a weekly basis to analyze the travel patterns. [He et al. \(2020\)](#) directly used time series on transit smart card activities' data as features, and used the distance between time series for passenger clustering.

On the other hand, the travel patterns can also be represented by latent features that are learned from data; the topic model developed in this paper also falls in this category. For example, [Goulet-Langlois et al. \(2016\)](#) used principal component analysis (PCA) to extract eigen-patterns from passengers' multi-week activity sequences. [Briand et al. \(2016\)](#) applied a mixed Gaussian model to extract latent features to mining passengers' temporal travel patterns. Based on the same method, [Briand et al. \(2017\)](#) further analyzed the year-to-year pattern changes in a public transportation system. [Zhao et al. \(2018b\)](#) applied a topic model to discover latent activity patterns from smart card data, which is very relevant to

our research. (Zhao et al., 2018c) and this paper both extend the LDA for travel behavior mining. The main difference is that we organize the latent features in a three-dimensional manner, which captures the interaction of spatial and temporal topics.

Besides the public transportation domain, topic models have been widely applied for mobility mining. For example, Hasan and Ukkusuri (2014) classified individuals' activity patterns by applying LDA to geo-location data collected from Twitter. Sun and Axhausen (2016) applied a probabilistic tensor factorization to smart card transactions to understand urban mobility patterns. Fan et al. (2016) applied LDA to mobile phone call data, and further developed a Hidden Markov Model for complete missing mobility data. Sun et al. (2021) developed a two-dimensional LDA on license plate recognition data, where the spatial and temporal topics are modeled separately, and their interactions are characterized in a two-dimensional simplex. This research applies the same methodology as Sun et al. (2021) and extends it to smart card data with three-dimensional features (origin, destination, and time).

### 3.4 Methodology: Topic Model for Destination Inference

This section details the probabilistic topic model for trip destination inference in smart card data. The objective is to infer the unknown trip destination in a tap-in-only system. A large portion of the destinations of linked trips could be inferred by rule-based models (Barry et al., 2002; Trépanier et al., 2007; Munizaga and Palma, 2012); some trip surveys could also provide a sample of complete trip information (MTL Trajet, 2019). We can train the proposed topic model by those trips with complete (ground truth or inferred) itineraries. Next, the destinations of unlinked trips could be inferred by the trained topic model.

#### 3.4.1 Model formulation

A smart card trip could be characterized by a three-element tuple  $(w^t, w^o, w^d)$  representing the departure time, origin, and destination, where  $w^t$  is assumed to be a discrete variable in one-hour intervals. Then, all the historical trips of a passenger  $u$  can be represented as  $\mathbf{w}_u = \{(w_i^t, w_i^o, w_i^d) : i = 1, \dots, N_u; w_i^t \in \{1, \dots, T\}; w_i^o, w_i^d \in \{1, \dots, S\}\}$ ; where  $N_u$  is the total number of trips for passenger  $u$ ,  $T$  is the number of possible departure hours, and  $S$  is the number of boarding/alighting locations.

The LDA model in NLP assumes there are several topics (e.g., sport, and cooking) among the corpus, the probability for each word's occurrence varies from topic to topic

(e.g., the probability for the word “basketball” occurs in the sports topic is higher than which in the cooking topic). A document is characterized by a mixture of topics, which explains the probability of each word’s occurrence in the document. By making an analogy to the LDA model, we treat each trip  $(w^t, w^o, w^d)$  as a word and  $\mathbf{w}_u$  as a document (a bag of words). Thus, all the trips belonging to a passenger compose a document with each trip being regarded as a word, each passenger’s trips are characterized by a mixture of latent topics.

The traditional LDA cannot directly model metro trips, because of the three interdependent attributes of a trip (i.e. time, origin and destination). A common solution is to combine different attributes into one dimension with the vocabulary size of  $T \times S \times S$ , such as in [Hasan and Ukkusuri \(2014\)](#) and [Fan et al. \(2016\)](#). The main drawback of this approach is that it considerably increases the vocabulary size, while the new combined words are sparse with many unobserved/unlikely trips. Moreover, the interdependency between original attributes is lost (e.g., two trips with the same origin and destination but different times can become unrelated words). To address this problem, we use an innovative method that expands the latent topics into a three-dimensional tensor, similar to the probabilistic tensor factorization as in [Sun and Axhausen \(2016\)](#) and [Sun et al. \(2021\)](#). By increasing the dimension of latent topics, we have three types of topic-word distributions, which avoid the large vocabulary set and capture inter-dependencies of different types of words in the latent space.

The latent topic is organized as a three-dimensional tensor  $\mathcal{Z} \in \mathbb{R}^{J \times K \times L}$ , where  $J$ ,  $K$ , and  $L$  are the number of latent topics of time, origin, and destination, respectively. The element  $z_{j,k,l}$  of tensor  $\mathcal{Z}$  corresponds to the  $j^{\text{th}}$  temporal topic  $z_j^t$ , the  $k^{\text{th}}$  origin topic  $z_k^o$ , and the  $l^{\text{th}}$  destination topic  $z_l^d$ . Each passenger’s trips are characterized by a Multinomial distribution over latent topics  $\mathcal{Z}$  (the topic distribution), parameterized by  $\theta_u$ . Given a latent topic  $z_{j,k,l}$ , the topic-word distributions for departure time, origin, and destination are Multinomial distributions parameterized by  $\varphi_{z^t}$ ,  $\psi_{z^o}$ , and  $\omega_{z^d}$ , respectively. The overall picture of the model can be clearly depicted by a graphical model shown in [Figure 3.1](#); where  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\eta$  are parameters for Dirichlet priors;  $U$  is the number of passengers. We describe the generative process in [Figure 3.1](#) follows:

- Draw topic distribution for each passenger  $\theta_u \sim \text{Dirichlet}_{J \times K \times L}(\boldsymbol{\alpha})$ .
- Draw topic-time distribution for each time topic  $\varphi_j \sim \text{Dirichlet}_J(\boldsymbol{\beta})$ .
- Draw origin distribution for each origin topic  $\psi_k \sim \text{Dirichlet}_K(\boldsymbol{\gamma})$ .
- Draw destination distribution for each destination topic  $\omega_l \sim \text{Dirichlet}_L(\boldsymbol{\eta})$ .

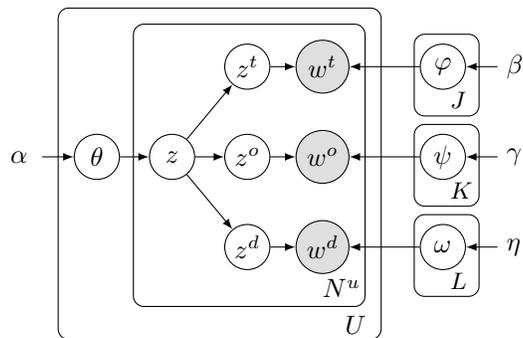


Figure 3.1: Plate notation for the graphical model.

- For each passenger  $u$ , for each trip record:
  - Draw latent topic  $z \sim \text{Multinomial}(\theta_u)$ .
  - Obtain  $z^o$ ,  $z^d$ , and  $z^t$  by  $z$ .
  - Draw  $w^t \sim \text{Multinomial}(\varphi_{z^t})$ .
  - Draw  $w^o \sim \text{Multinomial}(\psi_{z^o})$ .
  - Draw  $w^d \sim \text{Multinomial}(\omega_{z^d})$ .

We apply Multinomial distribution to departure time by discretizing time into one-hour intervals. This is a reasonable simplification and has been widely used in literature (Hasan and Ukkusuri, 2014; Sun and Axhausen, 2016; Sun et al., 2021). Continuous distributions, such as Normal and Log-Normal distributions (Zhao et al., 2018b), are more refined in time representation, but they are also more computational costly and to some extent restrictive in the shape of the distribution. Considering one-hour resolution is normally enough to distinguish different travel/activity patterns, this paper uses the discrete representation of time.

### 3.4.2 Model inference

The model inference involves estimating the parameters for latent topic distribution of each passenger and the topic-word distribution of each topic. In the generative process, each trip is generated from a latent topic  $z$ , which is unobserved. We use a collapsed Gibbs sampling algorithm Griffiths and Steyvers (2004) to iteratively sample the topic for each

trip by the conditional probability shown in Eq. (3.1):

$$P(z_i^t = j, z_i^o = k, z_i^d = l | w_i^t = t, w_i^o = o, w_i^d = d, \mathbf{z}_{-i}^t, \mathbf{z}_{-i}^o, \mathbf{z}_{-i}^d, \mathbf{w}_{-i}^t, \mathbf{w}_{-i}^o, \mathbf{w}_{-i}^d) \propto \frac{N_{z^t=j}^{w^t=t} + \beta}{N_{z^t=j} + T\beta} \times \frac{N_{z^o=k}^{w^o=o} + \gamma}{N_{z^o=k} + S\gamma} \times \frac{N_{z^d=l}^{w^d=d} + \eta}{N_{z^d=l} + S\eta} \times \frac{N_{z^t=j, z^o=k, z^d=l}^u + \alpha}{N^u + JKL\alpha}. \quad (3.1)$$

Where  $\mathbf{w}_{-i}^{(\cdot)}$  and  $\mathbf{z}_{-i}^{(\cdot)}$  are trip attributes and latent topics for all other trips except trip  $i$ ;  $N_{(\cdot)}^{(\cdot)}$  denotes the number of trips that satisfy the condition listed in the subscript and the superscript. Note that the current trip  $i$  is excluded when counting  $N$ .

The sampling procedure will converge after sufficient iterations, and by then we can estimate the parameters in topic distributions and topic-word distributions by Eq. (3.2):

$$\begin{aligned} \varphi_{t,j} &= \frac{N_{z^t=j}^{w^t=t} + \beta}{N_{z^t=j} + T\beta'} \\ \psi_{o,k} &= \frac{N_{z^o=k}^{w^o=o} + \gamma}{N_{z^o=k} + S\gamma'} \\ \omega_{d,l} &= \frac{N_{z^d=l}^{w^d=d} + \eta}{N_{z^d=l} + S\eta'} \\ \theta_{u,j,k,l} &= \frac{N_{z^t=j, z^o=k, z^d=l}^u + \alpha}{N^u + JKL\alpha}. \end{aligned} \quad (3.2)$$

### 3.4.3 Destination inference and station-to-rank transformation

Having estimated all the parameters in the model, we can infer the missing destination for a trip with only the origin and the departure time observed. According to the Bayes' theorem, the probability for passenger  $u$  alighting at a location  $d$  given the departure time  $t$  and the boarding location  $o$  takes the form:

$$\begin{aligned} P(w^d = d | w^t = t, w^o = o; u) &\propto P(w^t = t, w^o = o, w^d = d; u) \\ &= \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L P(w^t = t | z_j^t) P(w^o = o | z_k^o) P(w^d = d | z_l^d) P(z_j^t, z_k^o, z_l^d; u). \end{aligned} \quad (3.3)$$

Next, the most likely destination of a trip is the one that takes the highest probability in Eq. (3.3).

By now we have shown the complete topic model for destination inference, but there is

a final impediment that prevents the model from giving a good destination estimation—the giant heterogeneity among passengers’ spatial patterns. In essence, the topics of an LDA model are learned from the word co-occurrences across different documents. However, the origin-destination set is generally diverse from person to person (few word co-occurrences), which means a very large number of spatial topics are required to capture the spatial heterogeneity of the entire population. The large latent space not only fails to extract representative patterns among individuals but also increases the number of unknown parameters.

To address this problem, we do not use unique IDs for stations; instead, we label locations by the frequency-rank in each passengers’ historical trips. Studies have shown the frequency of individuals’ historical locations follows Zipf’s law (González et al., 2008), indicating most of the trips of a passenger are between several frequently visited locations. Therefore, the first several ranks can well characterize a person’s travel behavior. Specifically, denote  $r_u^i$  to be the rank (by the order of visiting frequency) of station  $s_i$  in all the historical origins of passenger  $u$ . We transform each passenger’s visited locations into the rank representation and store a mapping function  $M_u(r_u^i) \rightarrow s_i$  to restore real stations. By doing this, the diverse spatial patterns are essentially transformed into similar behavioral regularities (e.g., travel from the most visited station to the second most visited station). The same-ranked location for different passengers’ does not correspond to the same real stations, but represents a similar degree of importance of these stations to these passengers. We build the topic model and infer the destination in the ranked reference; the estimation for the real destination is then retrieved by the mapping function  $M_u$ .

### 3.5 Case study

We use the Guangzhou Metro smart card data—a tap-in and tap-out system—to examine the proposed topic model. As the topic model requires a portion of complete itineraries (training set) to learn passengers’ travel patterns, we will investigate our model under two scenarios:

1. using a ground truth training set;
2. using an estimated training set.

In scenario 1, we randomly select 70% trips and preserve the real destinations, as a training set; the destination inference is tested on the rest 30% data. Scenario 2 is a more realistic case where the ground truth destinations are unknown. We train the model with the

destinations inferred by the rule-based model as in [Trépanier et al. \(2007\)](#) and demonstrate our model’s performance under the “noisy” training set. The case study part is organized as follows: we will first introduce the data set, hyperparameters, and baseline models; next test the destination inference accuracy, interpret the latent patterns; finally present an application of passenger clustering.

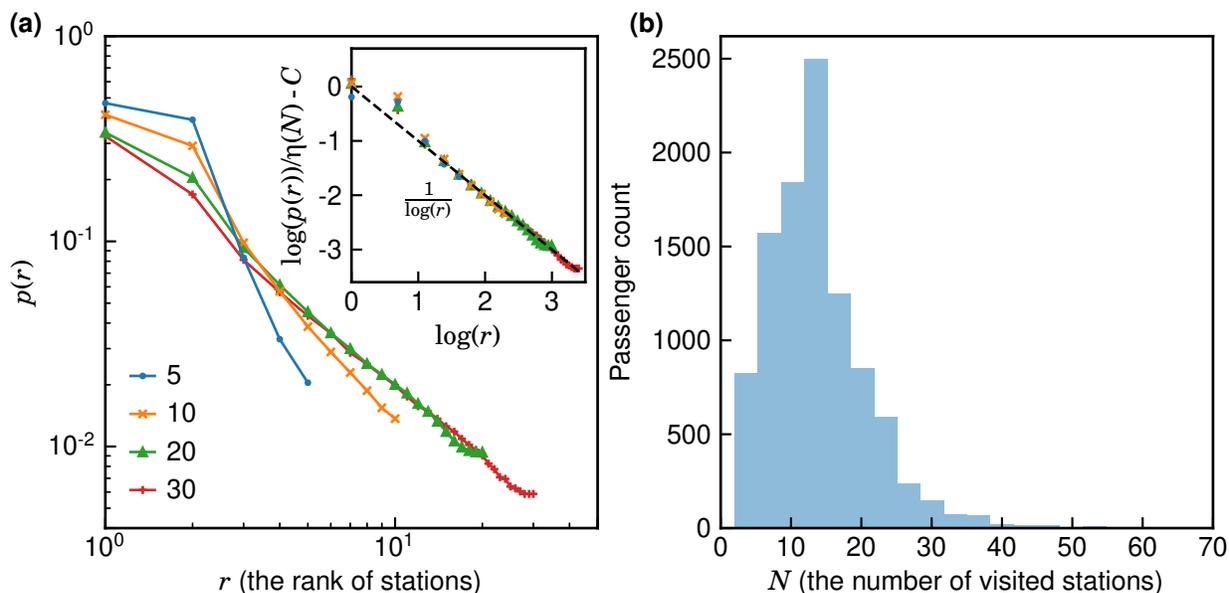
### 3.5.1 Guangzhou Metro data

Guangzhou Metro is one of the busiest metro systems in the world. As of August 2019, Guangzhou Metro has 14 operating lines with a total length of 478 kilometers. It is the third-largest metro system in China, after Beijing and Shanghai. The average daily ridership exceeds 8.6 million, taking over 50% of the ridership in the public transportation system of Guangzhou city ([Guangzhou Metro, 2019](#)). Except for line 9, 14, APM, and THZ1, our data covers the rest 11 lines of Guangzhou Metro with 159 stations from July 1<sup>st</sup> to September 30<sup>th</sup>, 2017. The metro operates 19 hours per day from 5:00 to 24:00. Therefore, the vocabulary size for time is 19, for origin and destination is 159.

Guangzhou Metro is a tap-in and tap-out system with both origin and destination registered, we can compare the estimated destination with the real destination to test the inference accuracy. There are single pass, day pass, Yang Cheng Tong and Lingnan pass (including various subclasses for students, elderly and disabled people), and digital tickets on smartphone apps. Around 1/3 trips are accomplished by single or day pass, and the destinations of these short-term users are barely estimable because of the lack of information. We only focus on the passengers with a minimum of 20 observations in the three months; later we will discuss the effect of the number of observations on the estimation accuracy. We showcase our model in 10000 randomly selected passengers. The total number of selected trips is 667,033, which means on average each person took 67 trips in the three months.

As discussed in [Section 3.4.3](#), instead of station IDs, we train our model by each passenger’s rank of stations. A preliminary analysis of the data shows that the first several ranks can capture most of the trips. [Figure 3.2 \(a\)](#) shows the rank  $r$  of stations and the visiting probabilities  $p(r)$  in a log-log plot. It can be found that the visiting probability drops significantly as the rank gets large. Further, the relation can be approximated by the Zipf’s law  $p(r) \sim r^{-\eta}$  with the exponent term  $\eta$  relates to the number of visited stations  $N$  ([Hasan et al., 2013](#)). When applying  $\eta = 3.57N^{-0.38}$ ,  $p(r)$  and  $r$  can be approximated by a single distribution, shown in the inserted figure. Similar to [Hasan et al. \(2013\)](#), the rank 2 station deviates from this relation, showing higher visiting probability. This indicates that

there is a bi-central mobility pattern in metro usage compared to the common Zipf’s law (González et al., 2008).



**Figure 3.2:** The probability of visiting different stations. (a) For passengers that have been observed to visit 5, 10, 20, and 30 different stations, the rank of the stations (in the order of the visit frequency) and the visiting probabilities in a log-log scale. The insert figure shows that the four groups of  $p(r)$  can be well approximated by  $p(r) \sim r^{-\eta}$ , when applying  $\eta = 3.57N^{-0.38}$ . (b) The histogram for the number of different stations visited by each passenger, in the 10000 passengers.

Figure 3.2 (b) shows the histogram for the number of different stations visited by each passenger. We can find most passengers visited between 5 to 20 different metro stations in the three-month period; the number of people who visited more than 20 stations tails off. Therefore, we cut off the frequency-rank at 20, marking all stations ranked larger than 20 as 20. By doing this, each passenger’s spatial vocabulary size is aligned at 20. Because the possibility of choosing cut stations is very low, as long as the cut-off point is not too small, the choice of cut-off point has little effect to the performance of our model. Representing stations by rank significantly decreases the number of latent topics needed on the spatial dimension.

### 3.5.2 Hyperparameters

There are two types of hyperparameters in our models—the number of latent topics and Dirichlet priors. In literature, the number of latent topics is often determined by perplexity, which measures the average likelihood of the test data set (Blei et al., 2003; Hasan and

Ukkusuri, 2014). In our context, we use the destination inference accuracy in the test set to select the number of topics. We perform a grid search over  $J = [3, 4, 5]$  and  $K, L = [2, 3, 4, 5]$  and select the best configuration by the minimal destination inference error, and we prefer a smaller model when the errors are close. Based on the result, we choose  $J = 4$  and  $K = L = 4$  for scenario 1,  $J = 4$  and  $K = L = 3$  for scenario 2; more topics do not contribute to the inference accuracy. Note that the number of spatial topics for scenario 2 is less than scenario 1. This is because the training set of scenario 2 is estimated from the rule-based models, and the noisy training set prevents the model from learning more patterns.

There are four Dirichlet priors in our model. These hyperparameters affect the smoothness of the Multinomial distribution; a larger value will increase the smoothness. Besides, we found hyperparameters (within a range) have little effect to destination inference accuracy, which is more relevant to the peak rather than the smoothness of distributions. We adopt the typical value in NLP and set  $\beta = \gamma = \eta = 0.1$  (Griffiths and Steyvers, 2004). The hyperparameter  $\alpha$  affects the smoothness of individuals' topic distribution. Considering it is rare for an individual to possess a wide range of travel patterns; we apply a small value  $\alpha = 5/(J \times K \times L)$  to learn a relatively sparse topic distribution that captures individual's specific character. Note a typical setting for  $\alpha$  in NLP is  $50/(\text{number of topics})$  (Griffiths and Steyvers, 2004).

### 3.5.3 Benchmark models

We compare our topic model with five benchmark models. For the first four benchmark models, we predict the destination by the most visited destination in a passenger's historical similar trips. The four kinds of "similar trips" are defined as follows:

- (SO) Trip with the same origin.
- (ST) Trip with the same departure time (one-hour interval).
- (SOT\_O) Trip with the same origin and departure time; if no such trip, use SO.
- (SOT\_T) Trip with the same origin and departure time; if no such trip, use ST.

We adopt the method proposed by He and Trépanier (2015) as the fifth benchmark model, where the destinations of unlinked trips are predicted by the multiplication of spatial and temporal kernel density at potential destinations. This method was developed for bus systems with all potential destinations on the same bus line as the origin. Because

the potential destinations of metro systems could be on different lines, we extend the potential destinations with historical destinations that have the same origin as the current trip, and replace the spatial kernel density by the visiting frequency. We choose 1 hour as an appropriate bandwidth for the temporal kernel density estimation after comparing different alternatives. We refer to this model as the “kernel-based” method in the following text. When any of the above benchmark models fails, the destination is predicted by the most visited destination of the corresponding passenger.

### 3.5.4 Scenario 1: using ground truth training set

In scenario 1, we randomly select 70% of the trips as the training set, where the ground truth destinations are known. Table 3.1 shows the destination inference accuracy of different models in both training and test sets. As the Gibbs sampling depends on the initial value, the accuracies of topic models are means of 50 runs and the standard deviations are shown in parentheses. It can be found that our rank-based topic outperforms the best benchmark models (SOT\_O) around 2% in the test set. On the other hand, the no-rank topic model—directly uses station ID in the model—has the worst accuracy, even under a very large number of topics. Our station-to-rank preprocessing greatly improves the inference accuracy and reduces the latent parameters. As expected, the accuracy of the training set is slightly higher than the test set.

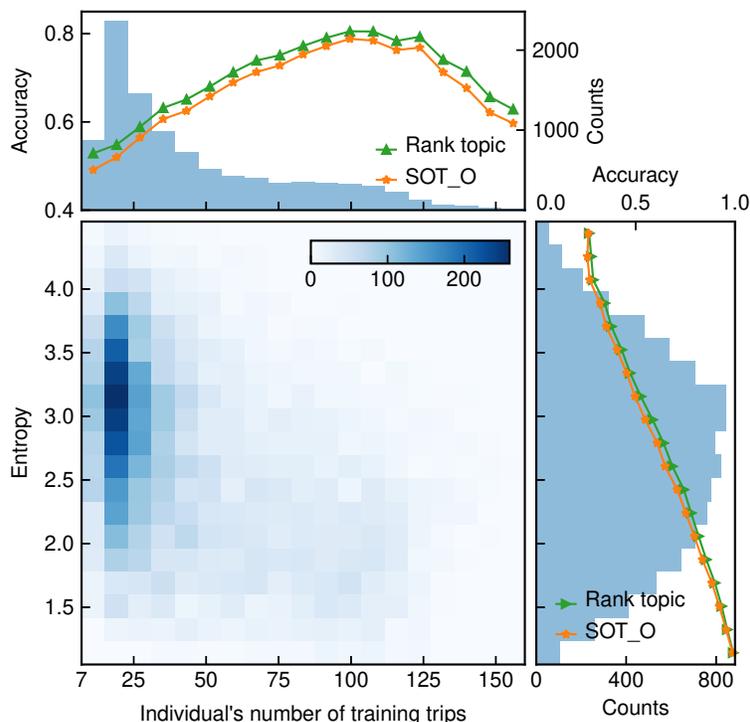
**Table 3.1:** The destination inference accuracy of scenario 1.

Method	Accuracy (test set)	Accuracy (training set)
SO	67.38%	--
ST	63.49%	--
SOT_O	67.75%	--
SOT_T	64.90%	--
Kernel-based	67.15%	--
<b>Rank topic<sup>a</sup></b>	69.78% (0.14%)	73.24% (0.14%)
No-rank topic <sup>b</sup>	31.15% (0.19%)	35.30% (0.18%)

<sup>a</sup> The number of topics  $J = 4$  and  $K = L = 4$ .

<sup>b</sup> The number of topics  $J = 5$  and  $K = 10, L = 100$ .

As the topic model requires some historical trips for training, we want to evaluate the effect of an individual’s number of training trips on the prediction accuracy. Besides, it is also interesting to investigate the relationship between destination inference and individuals’ travel regularity. There are many metrics to measure individual’s travel regularity, such as entropy (Scheiner, 2014), actual entropy (Song et al., 2010), and entropy



**Figure 3.3:** Bottom left: the 2D histogram of the number of passengers, gridded by the number of trips of each passenger in the training set and the entropy of their trips. Bottom right: the destination inference accuracy of two models for passengers at different entropy levels and the histogram of entropy. Top: the destination inference accuracy of two models for passengers with different numbers of trips in the training set and the histogram of the number of trips in the training set. (Accuracies are shown by means.)

rate (Goulet-Langlois et al., 2017). Entropy measures the randomness of a probability distribution. In metro trips, the entropy of passenger  $u$  is defined as

$$E_u = - \sum_{i=1}^{N_u} p_u(i) \log_2 p_u(i),$$

where  $p_u(i)$  is the historical probability that location  $i$  was visited,  $N_u$  is the total number of visited stations. The larger the entropy value is, the more random the distribution is, and the harder the prediction task becomes. Unlike the actual entropy and the entropy rate, the order of the trips does not affect the entropy. As the LDA is a bag-of-words model regardless of the order of words, we use entropy to reflect an individual's travel regularity and evaluate its relation to the accuracy of destination inference.

Figure 3.3 illustrates the destination inference accuracy under different numbers of training trips and entropy levels. Overall, the number of training trips concentrates on

the small end, with most passengers having 10 to 25 training trips. On the contrary, the entropy distribution is centered in the middle level with a decreasing trend in the high and low levels. From the bottom right of Figure 3.3, it is conspicuous that the prediction accuracy steadily increases with the decrease of the entropy, this is because more regular travelers are easier to predict. With this in mind, it is not hard to understand the relation between the number of training trips and prediction accuracy. The group with around 110 training trips has the highest prediction accuracy, this is because this group has the lowest entropy level (see bottom left of Figure 3.3). It is not hard to conclude that the changes in the prediction accuracy are mainly caused by the entropy rather than the number of training trips. The most predictable people are those that have around 110 training trips (around  $110/0.7=157$  trips with test set) in the three months; this number indicates that these people are very likely to be regular commuters. The SOT\_O and the rank-based topic model follow the same trend under different numbers of training trips and entropy levels, but our topic model always has higher accuracy.

### 3.5.5 Scenario 2: using estimated training set

Scenario 2 imitates the real-world tap-in only system, where the ground truth destinations are unknown. We first use rule-based models to infer the destinations of all linked trips as a training set, and then train our topic model using the estimated training set. The rule-based model that we applied is similar to [Trépanier et al. \(2007\)](#):

- Rule 1: predict the destination as the origin of the next trip on the same day.
- Rule 2: predict the last destination of a day as the first origin of the same day.
- Rule 3: predict the last destination of a day as the first origin of the next day.

The next rule will be only applied when the previous rule is not applicable to a trip. Note that any two metro stations can be connected by transfers; therefore, we do not need to verify whether the origin of the next trip is in the vicinity of the first Metro line, which is different to the bus network in [Trépanier et al. \(2007\)](#).

The accuracy and the coverage of the three rules are shown in Table 3.2. The assumptions of these rules have been indirectly verified by cordon count data ([Barry et al., 2002](#)) and survey data ([Barry et al., 2002](#); [Munizaga et al., 2014](#)), only a few studies examined these workhorse assumptions by ground-truth destinations ([Alsger et al., 2016](#)). We can tell from Table 3.2 that Rule 1 using the consecutive trips could reach 86% accuracy. Although destinations inferred by Rule 2 and Rule 3 are less reliable, they are indispensable parts for

the training set, because they represent the other side of passengers’ travel patterns (e.g., returning home at night). The three rules together handle 85.26% of the trips.

**Table 3.2:** The destination inference accuracy and coverage of scenario 2.

	Coverage	Cumulative coverage	Method	Accuracy (test set)	Accuracy (training set)
Linked trips	44.44%	44.44%	Rule 1	86.33%	--
	35.49%	79.93%	Rule 2	76.80%	--
	5.34%	85.26%	Rule 3	60.50%	--
Unlinked trips	14.74%	100.00%	SO	49.63%	--
			ST	43.02%	--
			SOT_O	48.93%	--
			SOT_T	44.19%	--
			Kernel-based	50.51%	--
			<b>Rank topic<sup>a</sup></b>	51.43% (0.14%)	66.48%(0.16%)
			No-rank topic <sup>b</sup>	31.14% (0.20%)	35.48%(0.18%)

<sup>a</sup> The number of topics  $J = 4$  and  $K = L = 3$ .

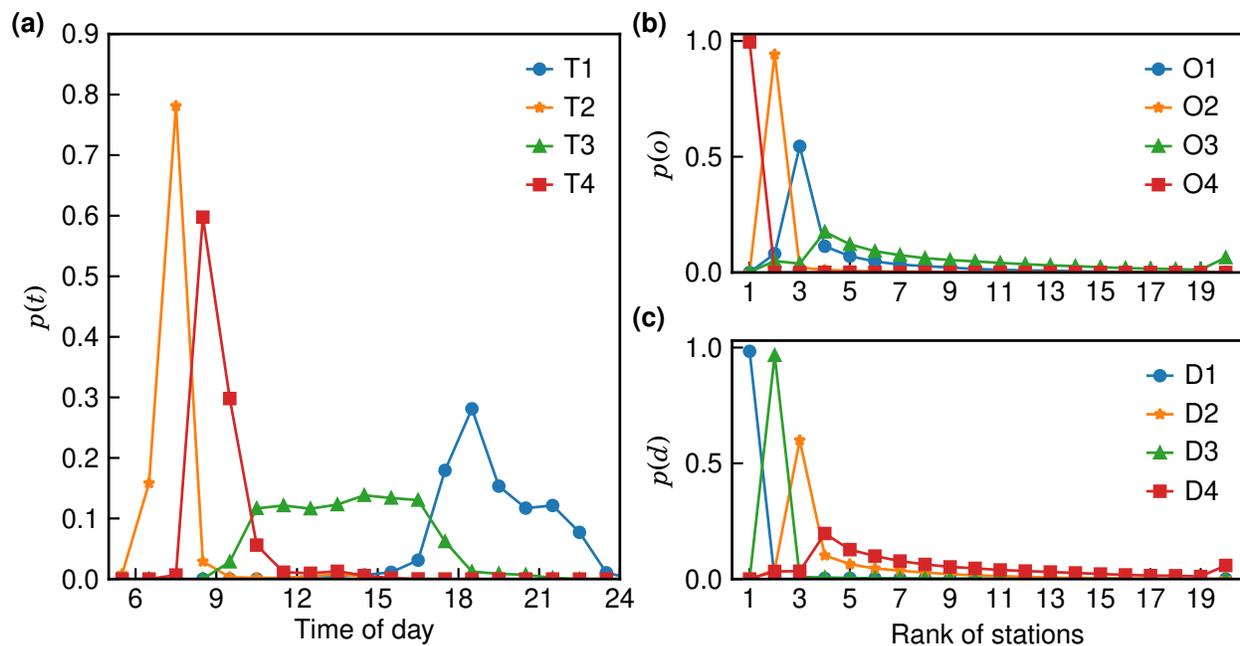
<sup>b</sup> The number of topics  $J = 5$  and  $K = 10, L = 100$ .

We then infer the destinations of unlinked trips by our rank-based topic model. Note that for scenario 2, we only use the origins for the ranking. Because the real destinations are unknown and the frequency of destinations is roughly the same as the origins if a passenger uses the smart card to and from. The destination inference results of the topic models and four benchmark models are shown in Table 3.2, and the standard deviations are shown in parentheses. It can be found that the best benchmark model is the kernel-based method with 50.51% accuracy, and our rank-based topic model performs slightly better than the kernel-based method with around 51.43% accuracy in the test set. It is noteworthy that the accuracy of the training set is significantly higher than the test set, despite they are both trained by the noisy data. This is because the training set and the test set are not randomly partitioned; there are some differences in the distributions of linked trips and unlinked trips. Besides the lack of ground truth, this difference further impacts the accuracy of scenario 2.

### 3.5.6 Interpreting latent topics

In the proposed topic model, each topic is characterized by a distribution over time (T), origin (O), or destination (D). By looking at these topic-word distributions, we can endow semantic meanings to latent topics. Therefore, we illustrate the topic-word distribution

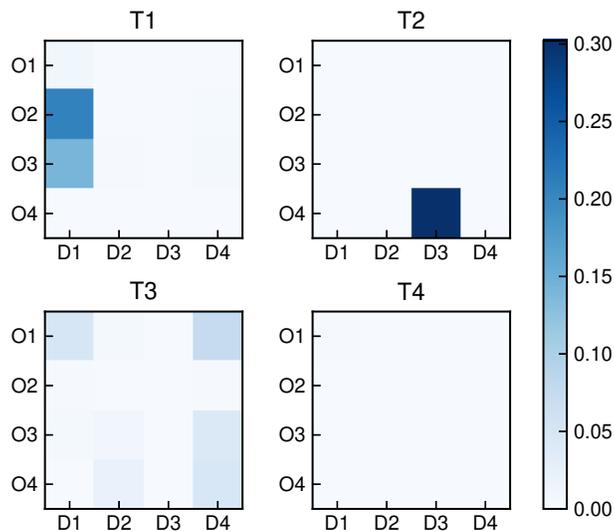
of scenario 1 by Figure 3.4. For time topics in Figure 3.4 (a), we can find topic T2 and T4 have very high probabilities of traveling in the morning, and could be interpreted as early and late morning peaks topics, respectively. Contrarily, topic T1 indicates trips in the night and T3 takes the rest of the day. For spatial topics shown in Figure 3.4 (b) and (c), it can be found that O4 and D1 take an almost 1 probability for the ranked 1<sup>st</sup> station, representing boarding and alighting at the most visited station, respectively. Meanwhile, O2 and D3 represent boarding and alighting at the second most visited station; O1 and D2 represent boarding and alighting at the third most visited station. For O3 and D4, the probabilities peak at the ranked 4<sup>th</sup> station and then gradually tail off. Moreover, we found the topic-word distribution is quite stable across different runs. Although the order of topics could switch, the shapes of the topic-word distributions maintain unchanged. This suggests the model is insensitive to initial values and the latent topics are good representations for travel patterns.



**Figure 3.4:** Topic-word distributions. (a) The departure time distributions of the four time topics. (b) The origin distributions of the four origin topics. (c) The destination distributions of the four destination topics.

It is worth mentioning that although the probability of alighting at a station after rank 4 is not zero, it is impossible to predict the destination of a trip as a station ranked after 4 by Eq. (3.3). Because Eq. (3.3) always predicts the destination as the most likely one, which is always the most likely destination (peak) in a particular latent destination topic. This

limitation causes the accuracy of ranked 4<sup>th</sup> destination being compromised by stations after rank 4; Luckily, these trips are sparse and with high randomness, the first three destinations make up the majority.



**Figure 3.5:** The latent topic distribution of a passenger.

After model training, each passenger is assigned with a distribution over topics, representing to what extent the passenger belongs to each topic. This topic distribution is a high-level summary of a passenger’s travel pattern. For example, Figure 3.5 shows the latent topic distribution of a passenger. Each matrix represents the probabilities over origin and destination topics under a time topic. Although we don’t know the exact mapping relation between the rank of a station and its real function (e.g., home/work), we can easily understand these travel patterns by common sense. It is conspicuous that there are two latent topics with significantly higher probability, indicating a possible commuting pattern. The most significant latent topic is [T2, O4, D3]; according to the semantic meaning shown in Figure 3.4, [T2, O4, D3] represents this passenger frequently departure from the most visited station to the second most visited station in the early morning, indicating a possible home-work behavior. Similarly, the second significant topic [T1, O2, D1] represents traveling from the second most visited station to the most visited location in the night, which could be the work-home trip. Besides, [T1, O3, D1] also has a high probability, which could be backing home from the third most visited location (such as a shop) at night. Other noticeable topics are mostly in T3 and have relatively low probabilities; these could be recreational activities. Further, we can find this passenger often uses metro in the early morning (T2), night (T1), and afternoon (T3), but seldom uses it between 8:00 AM to 10:00

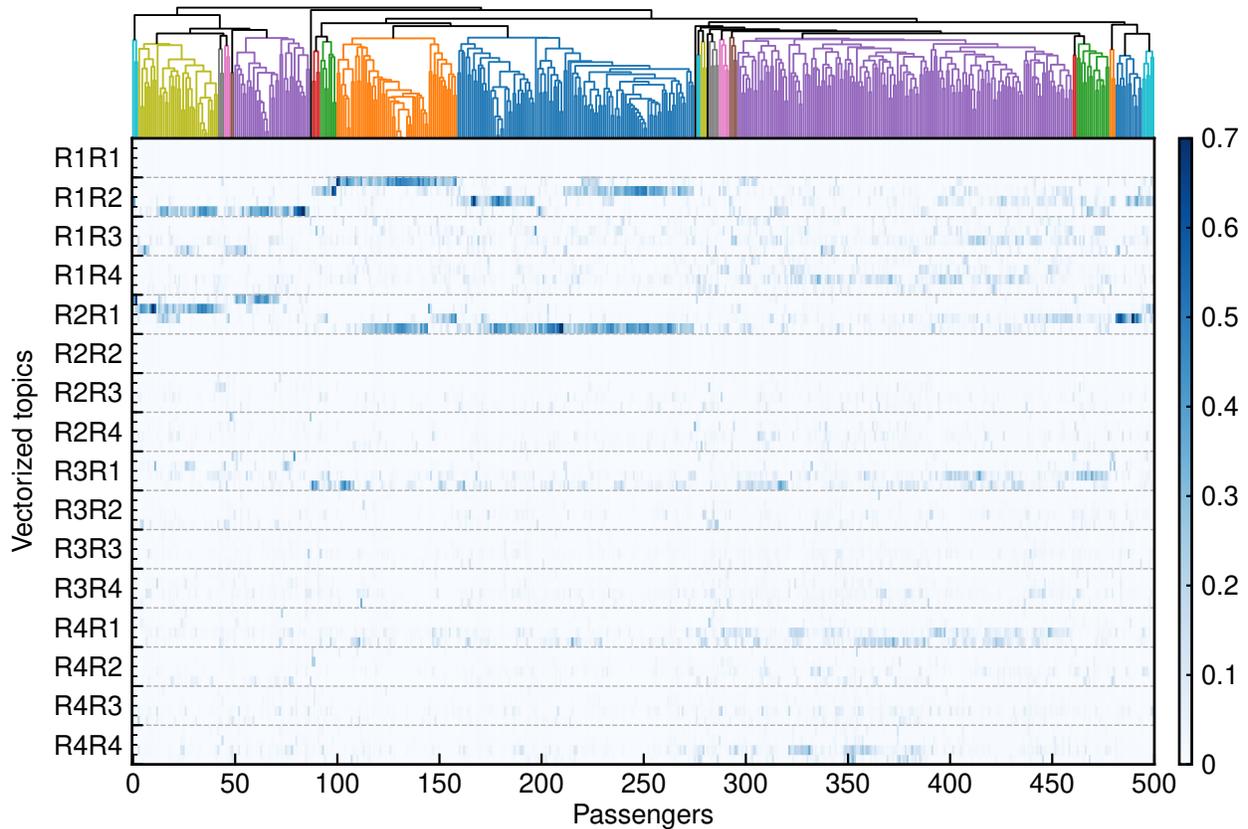
AM (T4).

### 3.5.7 Passenger clustering

Passenger clustering is important for personalized service, improving demand models, and various applications. The latent topic distribution is an excellent feature for passenger clustering. There has been a large body of research that uses smart card data for passenger clustering and travel pattern mining. Most existing methods capture either spatial or temporal features. [Ma et al. \(2013\)](#) clustered passengers based on spatial and temporal features, but the two kinds of features are independently defined and then combined. Our latent topic distribution jointly captures the spatial and temporal patterns in a compact manner, which provides a useful approach for investigating people's travel behaviors.

The feature used for clustering is passenger' latent topic distribution. Jensen-Shannon divergence (JSD) is a metric of measuring the similarity between two probability distributions; we apply the square root of the JSD as the distance between two latent topic distributions. Next, we select 500 passengers and apply hierarchical clustering to illustrate common travel behaviors among the population. Hierarchical clustering is a useful way to visualize the structure of the clustering component. It is also useful in providing the centroid and the number of clusters for faster clustering methods, such as K-means.

The hierarchical clustering of 500 passengers by their latent topic distribution is shown in [Figure 3.6](#). Noticeably, passengers on the left half of the figure (From 0 to around 275) show distinct two travel directions: one from the rank 1 station to the rank 2 station and the other from the rank 2 to the rank 1 station, indicating a commuting pattern. More specifically, the temporal topic within R1R2 and R2R1 are different in each cluster, showing these passengers regularly leave from a place at a certain time and then come back at another time, the time at which passengers leave and back distinguishes different clusters. On the other hand, passengers on the right half of [Figure 3.6](#) (around 275 to 500) do not have an as significant commuting pattern as those on the left part, and therefore correspond to non-commuters. The latent topic distributions of non-commuters show more diverse interactions between different topics, especially between rank 1 stations and others. It is also interesting to find that the early morning topic (in the top minor tick between two major ticks, corresponding to T2 in [Figure 3.4](#)) mostly belongs to the commuters; most metro trips from non-commuters are in the late morning, afternoon, and night. The distinct pattern between commuters and non-commuters validates the returners and explorers dichotomy in human mobility ([Pappalardo et al., 2015](#)). By utilizing the proportion of people under different clusters, a potential application of our topic model is to generate



**Figure 3.6:** Top: the dendrogram of the hierarchical clustering on 500 passengers. Bottom: the feature matrix for the clustering; each column in the matrix is a vectorized latent topic distribution of one passenger. The tick labels on y-axis represent the semantic meanings of latent topics: such as, R1R2 means the origin topic that peaks at ranked 1<sup>st</sup> station and the destination topic that peaks at ranked 2<sup>nd</sup> station. The four temporal topics are separated by the minor ticks between every two major ticks and ordered from early morning topic to night topic.

synthetic itineraries of the population for transit simulation.

### 3.6 Conclusions and discussion

This paper uses a probabilistic topic model for smart card data destination estimation and travel pattern mining. We establish a three-dimensional LDA model that captures the time, origin, and destination attributes in smart card trips. Moreover, we introduce a station-to-rank preprocessing that reduces the spatial divergence among passengers to discover more compact latent topics. The case study of Guangzhou Metro shows our model outperforms individual-history-based model by around 2% more accurate, in both scenarios with ground-truth or estimated training set. As a probabilistic model, the destination estimation accuracy is more related to an individual's travel regularity than the number of trips in the training set. Other than a prediction model, the proposed topic model is also a generative model that explains the probability of a trip by the individual's latent topics (i.e. the probability of traveling from rank  $o$  station to rank  $d$  station under time topic  $t$ ) and can be used for travel pattern analysis, and passenger clustering, and trip generation.

For the spatial topics, we introduce a station-to-rank transformation that enhances word co-occurrences among passengers and greatly improves the inference accuracy. The limitation of rank representation is the loss of spatial information. As shown in Figure 3.4 (b) (c), each spatial topic actually corresponds to one rank, rather than a mixture of words. Therefore, the topic model cannot be used for spatial clustering as to the vocabulary clustering in natural language processing. Indeed, related research (Hasan and Ukkusuri, 2014; Zhao et al., 2018b) primarily focused on passengers' pattern rather than the spatial similarity. How to derive spatial/region similarity from individuals' transit itineraries is an interesting direction, such as (Du et al., 2019). Besides, representing stations by labels (home/work/shop) could further improve the model interpretability; incorporating geographical and land use features could be promising future research. Finally, the effect of our model in the denser bus network is also worth exploring.

There is also improvement space for temporal topics. Firstly, distinguishing weekdays and weekends could be helpful to the prediction. Secondly, how to transform the topic model to a time-varying version is an interesting direction, such as Fan et al. (2016). A problem with our smart card data is that it is a coarse sample for individuals' life trajectory. Even in our sampled 10000 "frequent" users, around 50% of passengers use metro no more than four times a week; it is hard to utilize connections between neighboring trips under

such big travel intervals. Finally, similar to [Yin et al. \(2017\)](#), we can include extraneous variables to improve prediction accuracy. Out of the context of smart card data, it is promising to extend our model for a more general mobility prediction, such as next trip prediction ([Zhao et al., 2018c](#)).

## Chapter 4

# Using Travel Behavior for Boarding Flow Forecasting

---

This chapter is an article published in *Transportation Research Part C: Emerging Technologies*:

- Cheng, Z., Trépanier, M., Sun, L., 2021. Incorporating travel behavior regularity into passenger flow forecasting. *Transportation Research Part C: Emerging Technologies* 128, 103200.

The appendix of the original publication is added to Section 4.4.7 of this Chapter for fluent writing. This chapter corresponds to the travel-behavior-based forecasting method of this thesis. This chapter and the previous Chapter 3 both utilize the regularity and chained trips in passengers' travel behavior. Chapter 3 is an application on disaggregated individuals' trips while this chapter raises to an aggregated level about station passenger demand. Both chapters show the important role of travel behavior in metro operation and management.

---

## 4.1 Abstract

Accurate forecasting of passenger flow (i.e., ridership) is critical to the operation of urban metro systems. Previous studies mainly model passenger flow as time series by aggregating individual trips and then perform forecasting based on the values in the past several steps. However, this approach essentially overlooks the fact that passenger flow consists of trips from each individual traveler. For example, a traveler’s work trip in the morning can help predict his/her home trip in the evening, while this causal structure cannot be explicitly encoded in standard time series models. In this paper, we propose a new forecasting framework for boarding flow by incorporating the generative mechanism into standard time series models and leveraging the strong regularity rooted in travel behavior. In doing so, we introduce returning flow from previous alighting trips as a new covariate, which captures the causal structure and long-range dependencies in passenger flow data based on travel behavior. We develop the return probability parallelogram (RPP) to summarize the causal relationships and estimate the return flow. The proposed framework is evaluated using real-world passenger flow data, and the results confirm that the returning flow—a single covariate—can substantially and consistently improve various forecasting tasks, including one-step-ahead forecasting, multi-step-ahead forecasting, and forecasting under special events. And the proposed method is more effective for business-type stations with most passengers coming and returning within the same day. This study can be extended to other modes of transport, and it also sheds new light on general demand time series forecasting problems, in which causal structure and long-range dependencies are generated by the user behavior.

## 4.2 Introduction

Recent years have witnessed the rapid development of metro systems and the continued growth of metro ridership worldwide ([Union Internationale des Transports Publics \(UITP\), 2018](#)). As an efficient and high-capacity transportation mode, the metro is playing an ever-important role in shaping future sustainable transportation. Given the growing importance of metro systems, it is critical to have a good understanding of passenger demand patterns to support service operation. A key task is to make accurate and real-time forecasting of passenger demand/ridership, which plays a vital role in a wide range of applications, including service scheduling, crowd management, and disruption response, to name but a few.

Short-term passenger flow forecasting typically focuses on forecasting the passenger flow in the next few minutes to several hours, and has been extensively studied in public transportation research. Most existing studies formulate passenger flow data as time series and follow similar methods as those applied in traffic flow forecasting. For example, statistical time series models have been widely applied to ridership forecasting problems, including autoregressive integrated moving average (ARIMA) (Williams and Hoel, 2003; Ding et al., 2017; Chen et al., 2020a), exponential smoothing (Tan et al., 2009), and state-space/Kalman filter (Stathopoulos and Karlaftis, 2003; Jiao et al., 2016). Most of these classical time series models are linear by nature; to better characterize the non-linearity in time series data, non-linear versions or ensemble extensions of these models have also been studied (e.g., Jiao et al., 2016; Carrese et al., 2017). Recent research starts regarding the forecasting a supervised machine learning problem. On this track, some representative supervised learning models have been applied, such as support vector machine (SVM) (Chen et al., 2011; Sun et al., 2015), artificial neural network (ANN) (Vlahogianni et al., 2005; Tsai et al., 2009; Li et al., 2017), random forest (Toqué et al., 2017), and recurrent neural network (RNN)/long short-term memory (LSTM) as emerging deep learning approaches (Hao et al., 2019; Liu et al., 2019b). The aforementioned research mainly focuses on modeling a univariate time series for a single metro station. However, the metro system is a network in which stations exhibit strong spatial and temporal correlations/dependencies. To extend the univariate analysis to network-wide passenger flow forecasting, some state-of-the-art models have been proposed to better characterize the complex spatiotemporal patterns and dynamics. For example, Gong et al. (2020) proposed matrix factorization models to estimate passenger flow data for each origin-destination (OD) pair; Li et al. (2020) introduced local smoothness prior based on auxiliary information (e.g., flow correlation, network typology, and POI composition) into tensor completion models to forecast passenger flow; Chen et al. (2020b) developed graph convolutional network (GCN) models to capture the complex spatiotemporal dependencies in a metro network. These new machine learning-based models have shown superior performance over traditional time series models, and they are more effective in capturing the complex patterns by incorporating domain knowledge and external features such as weather, event, time of day, and day of week.

In all the studies mentioned above, passenger flow data is generally modeled as an aggregated count time series obtained by counting the number of unique card IDs in smart card transactions. Despite the simplicity and effectiveness of these models, we would argue that the most important characteristic of passenger flow is overlooked due

to the aggregation: passenger flow consists of the movement of individuals with strong regularity rooted in their travel behavior. For instance, if a passenger alights at a metro station for work in the morning, he/she will probably depart at the same station when he/she goes home in the evening. If he/she does not travel in the morning, it becomes less likely we will observe a corresponding return trip. This example clearly shows that past trips should be utilized to predict future demand, and individual travel behavior actually can result in causal structure and long-range dependencies in passenger flow time series data. Some recent studies have shown that travel behavior plays a substantial role in dynamic traffic assignment (Cantelmo and Viti, 2019) and online demand estimation (Cantelmo et al., 2020). This effect is particularly true for metro systems where passengers' travel patterns are highly regular (Sun et al., 2013; Goulet-Langlois et al., 2017; Zhao et al., 2018c). Therefore, when developing a passenger flow forecasting model, it is essential to integrate this type of behavior-driven and long-range dependencies in addition to the local input (e.g., the past  $n$  steps in the time series).

The goal of this study is to explore the potential of incorporating an additional travel behavior component into the forecasting of passenger flow time series. Specifically, we propose a new scheme to forecast boarding/incoming passenger demand at a station by integrating historical alighting time series at the same station. We define returning passengers as those who finish their first trip at station  $s$  and also start their second trip at the same station. In other words, returning passengers refer to the individuals who stay at station  $s$  to perform an activity (e.g., home and work). In general, these return trips are not random and often exhibit strong regularity due to the activities performed. This motivates us to forecast the incoming/boarding demand from these "returning passengers" using the information on their previous trips. To achieve this, we introduce a new concept of return probability parallelogram (RPP) to better estimate returning flow, and we find that the estimated returning flow highly correlates with the overall boarding demand in a real-world data set. To further quantify the benefits of incorporating this returning flow measure, we evaluate the proposed models for one-step-ahead forecasting, multi-step-ahead forecasting, and forecasting under special events. Our results show that incorporating returning flow as an additional variable will consistently improve the accuracy of forecasting.

The idea of leveraging trip-level information has been introduced and examined in some recent studies, which predict the alighting flow of a station using the recent boarding flow from other related stations (see e.g., Li et al., 2017; Hao et al., 2019; Liu et al., 2019b). However, the large number of boarding-alighting station pairs makes it difficult to learn

an informative model at a trip level, and eventually these studies develop deep neural networks to learn the correlation from the aggregated count data in a purely data-driven approach. Our model, instead, uses the alighting of “this trip” to predict the boarding of the “next trip”, where the alighting and the boarding stations are usually the same (Barry et al., 2002; Trépanier et al., 2007). We examine this idea on a boarding flow forecasting application, which is more important to service operation and planning. The “returning flow” proposed in this paper is solely based on the intrinsic travel regularity of travelers, and it does not require external information/knowledge. Our work is closely related to Zhao et al. (2018c), which proposes a probabilistic model to predict the next trip for an individual based on his/her trip history. However, instead of predicting individual trips, our primary goal is to forecast the overall passenger flow to support the decision-making in service operation. In doing so, we estimate the returning flow in an aggregated approach; therefore, the framework does not require individual-based data sets that are confidential and sensitive for privacy reasons. The main contribution of this work is summarized as follows.

- We define returning flow to characterize the causal structure and long-range dependencies in passenger flow data, which are essentially overlooked in previous time series-based studies.
- We integrate returning flow as an additional covariate into standard time series models, and the proposed behavior-integrated model shows consistently improved performance in our case studies based on a real-world data set.
- Our model also provides a new approach to forecast passenger flows under special events.

To the best of our knowledge, this is the first research that incorporates a travel behavior component into the longstanding passenger flow forecasting problem. The remainder of the paper is organized as follows. Section 4.3 introduces the concept of returning flow and return probability parallelogram as the tool to integrate travel behavior regularity into the passenger flow forecasting framework. In Section 4.4, we develop case studies based on real-world smart card data and demonstrate the effectiveness of the proposed models in different scenarios. Finally, Section 4.5 concludes our research and discusses future work.

## 4.3 Methodology

In this section, we introduce returning flow and the return probability parallelogram as two fundamental building blocks in the behavior-based boarding flow forecasting framework. The proposed forecasting models are constructed by integrating returning flow as a new feature/covariate into traditional time series forecasting models. We start with a brief description of the passenger flow forecasting problem.

### 4.3.1 Problem description

Suppose that in a metro system we have access to all smart card transactions, i.e., we know the anonymous ID of passengers, the time and the locations/stations of both boarding (tapping-in) and alighting (tapping-out) for each trip. In this case, a station  $s$  will generate two passenger flow time series: the alighting/arriving flow for passengers with station  $s$  as their destination, and the boarding/incoming flow for passengers who start their trips from station  $s$ . We denote by  $y_t^s$  and  $m_t^s$  the boarding flow and the alighting flow at station  $s$  in time interval  $t$ , respectively.

We focus on the case of forecasting the boarding flow  $y_t^s$ . Given some recent observations  $y_1^s, \dots, y_{t-1}^s, y_t^s$ , our goal is to predict the values of  $y_{t+1}^s, y_{t+2}^s, \dots, y_{t+L}^s$  in the next  $L$  time steps/intervals. This is a standard time series analysis problem on which traditional statistical models such as ARIMA can be applied. In this paper, we aim to achieve better forecasting results over the traditional models by integrating additional information of the behavioral regularity of passengers associated with alighting flow  $m_t^s$ .

### 4.3.2 Returning flow

We begin our model by introducing the concept of the “returning flow”. To facilitate model development, we divide all the passengers associated with stations  $s$  (both boarding and alighting) into two groups (see Figure 4.1):

**(G1)** Passengers who alight at station  $s$ ;

**(G2)** Passengers who board at station  $s$  without a previous trip alighting at  $s$ .

With this definition, we can model the total boarding flow  $y_t^s$  by combining the boarding flow in the two groups. The passengers in G1 can be further separated into two subgroups given if they have their next trip originating from station  $s$  within a certain time window. We define the subgroup with a following trip as G1A and the other as G1B. Thus, G1A

Group		$t-h$	$\dots$	$t-2$	$t-1$	$t$	$t+1$	$t+2$	$\dots$	$t+L$
G1	A			○	→	●				
	B				○	→	●			
	sum ○ in G1	$m_{t-h}^s$	$\dots$	$m_{t-2}^s$	$m_{t-1}^s$	$m_t^s$				
G2						●				
	sum ● in G1+G2	$y_{t-h}^s$	$\dots$	$y_{t-2}^s$	$y_{t-1}^s$	$y_t^s$	$\hat{y}_{t+1}^s$			

○ represents alighting, ● represents boarding,  
 → represents observed trip chain, --→ represents trip chain to be predicted.

**Figure 4.1:** Illustration of two passenger groups (G1/2) and the boarding demand forecasting problem at station  $s$ .

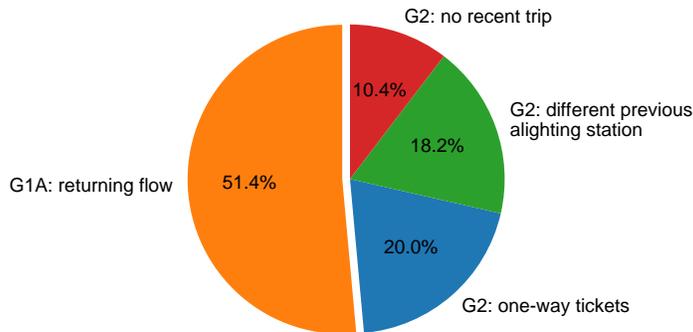
actually consists of those passengers who conduct certain activities (e.g., home/work) around station  $s$ . We define “returning flow” at time  $t$  as the number of people in G1 who will finish their activities and start their return trips at time  $t$  by station  $s$ , denoted by  $r_t^s$ . In fact, these chained trips (with departing station being identical to the alighting station of a previous trip) make up a substantial proportion of all trips. As shown in Figure 4.2, for the Guangzhou metro in our case study, the returning flow accounts for over 50% of all boarding demand. We thus hypothesize that having the “returning flow” as an additional variable will enhance the forecasting of  $y_{t+1}^s$ . We refer to the forecasting model with  $r_{t+1}^s$  as a covariate as M2:

$$\text{M2} : \hat{y}_{t+1}^s = f(y_{1:t}^s, r_{t+1}^s). \quad (4.1)$$

It should be noted that we do not have access to  $r_{t+1}^s$  (i.e., those dashed arrows in Figure 4.1), as the returning flow in G1 is only observed up to time  $t$  (i.e., those solid arrows in Figure 4.1). Therefore, in practice, we need to first estimate  $\hat{r}_{t+1}^s$  and then use it as a proxy for  $r_{t+1}^s$  in M2. On the other hand, a possible alternative is to use  $r_t^s$ —which we have access in real-time—instead of  $\hat{r}_{t+1}^s$  as the covariate. We define this alternative model as M1 and use it as a baseline model:

$$\text{M1} : \hat{y}_{t+1}^s = f(y_{1:t}^s, r_t^s). \quad (4.2)$$

We also consider a standard time series model without any additional variables as a



**Figure 4.2:** The composition of the boarding flow in Guangzhou metro. Based on the smart card data from July 21 to 28, 2017.

baseline (M0):

$$M0 : \hat{y}_{t+1}^s = f(y_{1:t}^s). \quad (4.3)$$

Note the returning flow (G1A) in this paper does not cover the G2 part of the boarding flow (Figure 4.2). Because it is hard to forecast the one-way tickets or standalone trips in G2 by data solely from a metro system. However, a metro ride can be one trip in an activity chain or even one mode in a multi-modal trip; it is possible to infer the G2 part if we have complete trip chain information supported by other data sources, which will be greatly helpful for the boarding flow forecasting. Like in most cases, we have no complete trip/activity chain information. We thus establish the returning flow concept upon consecutive metro trips with the first destination and the next origin overlapped. This is a specific subset of activity chains. Luckily, we can obtain a very accurate estimation for the future returning flow (G1A) and it already takes a substantial part of the total boarding flow; using the returning flow as a covariate is still beneficial for the boarding flow forecasting.

### 4.3.3 Return probability parallelogram (RPP)

In this subsection, we propose a method to estimate the returning flow. We only consider the returning flow within a time window  $H$  when we define G1A. For the current and past returning flow,  $r_t^s$  can be readily obtained by

$$r_t^s = \sum_{t_a=t-H}^{t-1} r_{t_a,t} \quad (4.4)$$

where  $r_{t_a,t}$  is the number of passengers that come (alight) at station  $s$  at  $t_a$  and return (board) at  $t$ . Using Eq. (4.4),  $r_t^s$  can be obtained in real-time and used in M1.

However, M2 requires a returning flow in the future that cannot be accessed by Eq. (4.4). Therefore, we propose a method to estimate  $\hat{r}_{t+1}^s$  based on the returning flow generalization mechanism. Our fundamental assumption is that there exists a universal distribution  $p^s(\tau_{\text{boarding}} | \tau_{\text{alighting}})$  characterizing the conditional probability that a passenger in G1 who alights at time  $\tau_{\text{alighting}}$  will start his/her returning trip at time  $\tau_{\text{boarding}}$ . Note that we define  $p^s$  on the whole group G1, so the subgroup G1B is also modeled in this distribution. For the passengers who alight at time  $t_a$  we have:

$$\sum_{t=t_a+1}^{t_a+H} p^s(\tau_{\text{boarding}} = t | \tau_{\text{alighting}} = t_a) + p^s(\tau_{\text{boarding}} = \text{NA} | \tau_{\text{alighting}} = t_a) = 1, \quad (4.5)$$

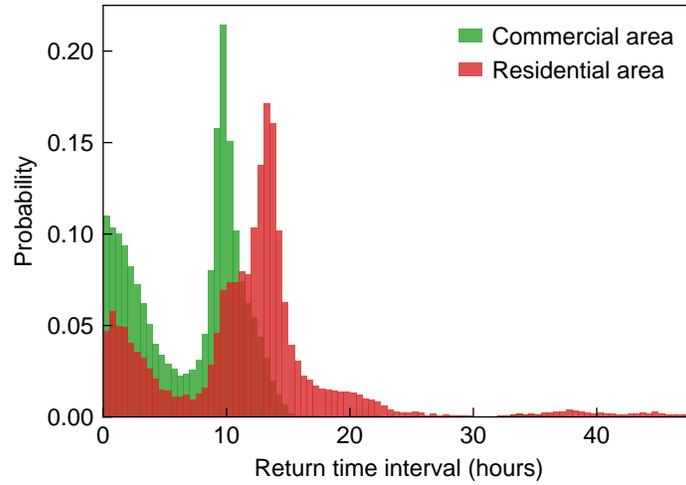
in which the term  $p^s(\tau_{\text{boarding}} = \text{NA} | \tau_{\text{alighting}} = t_a)$  represents the conditional probability of an arriving passenger does not return within the time window  $H$  (i.e., subgroup G1B).

If the conditional distribution  $p^s(\tau_{\text{boarding}} | \tau_{\text{alighting}})$  is available for all  $\tau_{\text{alighting}}$ , we can estimate the expectation of the returning flow  $r_{t+1}^s$  at time  $t+1$  by:

$$\hat{r}_{t+1}^s = \sum_{h=1}^H m_{t-h+1}^s p^s(\tau_{\text{boarding}} = t+1 | \tau_{\text{alighting}} = t-h+1). \quad (4.6)$$

It is important to note that the estimation of  $\hat{r}_{t+1}^s$  using Eq. (4.6) is very different from predicting  $\hat{r}_{t+1}^s$  using a time series model based on past observations. This is because a simple time series model such as ARIMA cannot characterize the unique generative mechanisms (e.g., come-and-return) and the corresponding long-range dependencies/causal structure provided by these mechanisms in the passenger flow data.

The time window length  $H$  is an additional parameter to be determined before applying Eq. (4.6). To choose an appropriate  $H$ , we quantify the inter-trip time/duration ( $\tau_{\text{boarding}} - \tau_{\text{alighting}}$ ) for all those passengers with both the alighting trip and the next boarding trip at the same station. We conduct this analysis on the Guangzhou metro data set. Figure 4.3 shows the distribution of the inter-trip time of two representative stations in a commercial area and a residential area, respectively. The distribution is obtained by aggregating all alighting records on a typical Monday, and we track the returning flow within 48 hours after the alighting. The return time intervals in both stations are characterized by a bi-modal pattern. The first peak (less than 3 hours) corresponds to certain short-duration activities (e.g., dining and shopping). The longer peaks largely



**Figure 4.3:** The density histogram of the return time interval ( $\tau_{\text{boarding}} - \tau_{\text{alighting}}$ ) of two example stations.

correspond to “work” activities (9-12 hours) in the commercial area and “home” activities (10-16 hours) in the residential area, respectively. Relatively few activities take around 6 hours, and thus the return time intervals in both stations exhibit a “U” shape pattern. More importantly, as we can see, almost of the return trips start within a 24-hour window after finishing previous trips. Therefore, for simplicity, we only take the alighting flow within the past 24 hours into account when estimating  $\hat{r}_{t+1}^s$ .

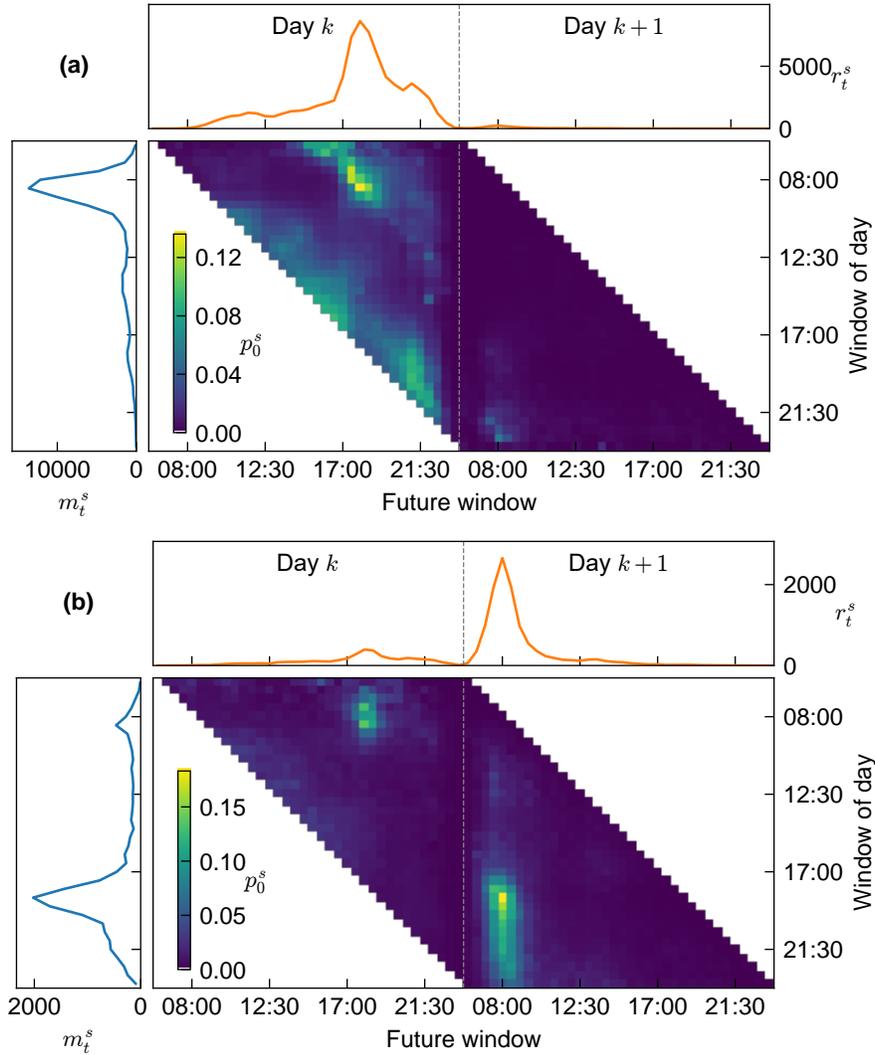
Having determined  $H$ , the next step is to obtain a good estimate of the conditional probability distribution  $p^s(\tau_{\text{boarding}} \mid \tau_{\text{alighting}})$ . However, the current formulation involves a set of conditional probabilities for each value of  $\tau_{\text{alighting}}$ , making it difficult to estimate. For simplicity, we assume that the conditional distributions are universal across different days:

$$\begin{aligned} p^s(\tau_{\text{boarding}} = t_b \mid \tau_{\text{alighting}} = t_a) &= p_0^s(\text{future window}(t_b) \mid \text{window of day}(t_a)) \\ &= p_0^s(\text{window of day}(t_a) + t_b - t_a \mid \text{window of day}(t_a)), \end{aligned} \quad (4.7)$$

where we refer to the reduced distribution  $p_0^s$  as the return probability parallelogram (RPP). As the new conditional distribution  $p_0^s$  is defined given the time of day of  $t_a$ , we can estimate it using historical trip data of passengers in group G1A (i.e., the solid arrows in Figure 4.1). Denote  $r_{t_a, t_b}^s$  to be the number of passengers that come (alight) at station  $s$  at  $t_a$

and return (board) at  $t_b$ . For a time window  $w$  of day,  $p_0^s$  can be estimated by

$$p_0^s(w + h|w) = \frac{\sum_{\text{window of day}(t_a)=w} r_{t_a, t_b}^s}{\sum_{\text{window of day}(t_a)=w} m_{t_a}^s} \quad (h = 1, 2, \dots, H). \quad (4.8)$$



**Figure 4.4:** The return probability parallelogram (RPP), the alighting flow  $m_t^s$ , and the returning flow  $r_t^s$  for two representative stations: (a) A typical station in commercial areas. (b) A typical station in residential areas.

We use Figure 4.4 to illustrate the idea of RPP. Panel (a) and (b) show two sets of conditional distributions for a commercial area and a residential area, respectively, in Guangzhou metro. The resolution for time slot is set to half an hour, and the range is

from 6:00 to 24:00 (operation time of the metro system). Note that in this parallelogram representation we concatenate the 24:00 of day  $k$  and the 6:00 of day  $k + 1$  on the horizontal axis. There are two blank triangles in this diagram: the one on the left corresponds to the  $t_b \leq t_a$ , where the distribution is not defined; the one on the right corresponds to the conditional probability with  $t_b > t_a + H$  ( $H = 48$  for 24 hours), which is also ignored for simplicity. It should be noted that in RPP the sum of each row is less than 1, as it does not include the passengers in G1B (with no returning trips, i.e.,  $\tau_{\text{boarding}} = \text{NA}$ ). With this formulation, we can replace  $p^s (\tau_{\text{boarding}} = t + 1 \mid \tau_{\text{alighting}} = t - h + 1)$  in Eq. (4.6) by the corresponding conditional probability in RPP.

As shown in Figure 4.4, it is obvious that different stations exhibit different RPP patterns. For example, for the commercial station in Figure 4.4(a), most trips arrive (alight) in the morning and return in the evening on the same day, which essentially captures work activities. It is very rare to see returning trips on the next day. As for the station in a residential area in Figure 4.4(b), on the contrary, we can see that the distribution mainly characterizes home activities, where alighting flow generally peaks in the evening and the returning flow concentrates in the morning of the next day. The RPP representation demonstrated in Figure 4.4 further suggests that the unique come-and-return dynamics for a station should be considered in passenger flow forecasting applications.

## 4.4 Experiments

In this section, we conduct numerical experiments to evaluate the effectiveness of the proposed behavior-integrated models. We choose the standard SARIMA model as the core model for time series forecasting (M0). On top of this model, we create two regression with SARIMA error models—M1 and M2—by simply incorporating the observed  $r_t^s$  and the estimated  $\hat{r}_{t+1}^s$  as additional covariates, respectively. We evaluate the performance of these models in three scenarios: (1) one-step-ahead forecasting, (2) multi-step-ahead forecasting, and (3) forecasting under special events. Besides, we also test using Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP) as M0 in Section 4.4.7 and observe consistent results with the SARIMA.

### 4.4.1 ARIMA model

We choose seasonal ARIMA as the main baseline model—M0. ARIMA is a well-established time series forecasting model which has been widely used in traffic/passenger flow

forecasting (Williams and Hoel, 2003; Ding et al., 2017; Chen et al., 2020a). Considering the strong periodicity from day to day, we apply Seasonal ARIMA (SARIMA) to model passenger flow. Here we give a brief introduction of the SARIMA model, and we refer readers to the book by Hyndman and Athanasopoulos (2018) for a comprehensive review of time series models. A SARIMA model is usually denoted by  $ARIMA(p, d, q)(P, D, Q)[m]$ , where  $p$ ,  $d$ , and  $q$  represent the order of autoregressive, differencing, and moving-average;  $P$ ,  $D$ , and  $Q$  are the order of autoregressive, differencing, and moving-average for the seasonal part; and  $m$  is the number of period in each season. For a time series  $y_1, \dots, y_T$ , the  $ARIMA(p, d, q)(P, D, Q)[m]$  model (M0) takes the form

$$\Phi(B) (1 - B^m)^D \phi(B)(1 - B)^d y_t = \theta(B)\Theta(B)e_t, \quad (4.9)$$

where  $B$  is the backshift notation defined by

$$\begin{aligned} B^a y_t &= y_{t-a}, \\ \Phi(B) &= \left(1 - \Phi_1 B^m - \dots - \Phi_P B^{P \times m}\right), \\ \phi(B) &= \left(1 - \phi_1 B - \dots - \phi_p B^p\right), \\ \theta(B) &= \left(1 + \theta_1 B + \dots + \theta_q B^q\right), \\ \Theta(B) &= \left(1 + \Theta_1 B^m + \dots + \Theta_Q B^{Q \times m}\right); \end{aligned}$$

$\Phi_i$ ,  $\Theta_i$ ,  $\phi_i$ , and  $\theta_i$  are ARIMA coefficients to be estimated;  $e_t$  is an error assumed to follow a white noise process (i.e., zero mean and iid).

When incorporating the returning flow  $r_1, \dots, r_T$ , as a covariate, the forecasting model (M2) becomes

$$\begin{aligned} y_t &= \beta r_t + \eta_t, \\ \Phi(B) (1 - B^m)^D \phi(B)(1 - B)^d \eta_t &= \theta(B)\Theta(B)e_t. \end{aligned} \quad (4.10)$$

Where  $\beta$  is the regression coefficient,  $\eta_t$  is a regression error term that follows the ARIMA procedure. Note the regression coefficient and the ARIMA coefficients are estimated in one step, rather than estimated separately. All ARIMA models in this study are estimated using the **forecast** package for R (Hyndman et al., 2020).

### 4.4.2 Model selection and evaluation

We apply the same order of SARIMA to the demand time series for all the stations. The seasonal frequency is set to  $m = 36$  (i.e., daily, from 6:00 to 24:00, as we use half an hour as the temporal resolution). For most stations, after a seasonal differencing, the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) indicates no further differencing is required to make the time series stationary, we thus set  $D = 1$  and  $d = 0$ . We search over possible models and finally select ARIMA(2, 0, 1)(1, 1, 0)[36] as the baseline M0, which is shown to be appropriate for most stations. Indeed, we can achieve better forecasting results by designing station-specific models with different orders. However, as our goal is to evaluate the effect of using the returning flow as a covariate, we still select a universal model for all stations for simplicity.

We use the root mean square error (RMSE) and the symmetric mean absolute percentage error (SMAPE) to evaluate model accuracy:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t^s - \hat{y}_t^s)^2},$$

$$\text{SMAPE} = \frac{2}{N} \sum_{t=1}^N \frac{|y_t^s - \hat{y}_t^s|}{|y_t^s| + |\hat{y}_t^s|} \times 100(\%),$$

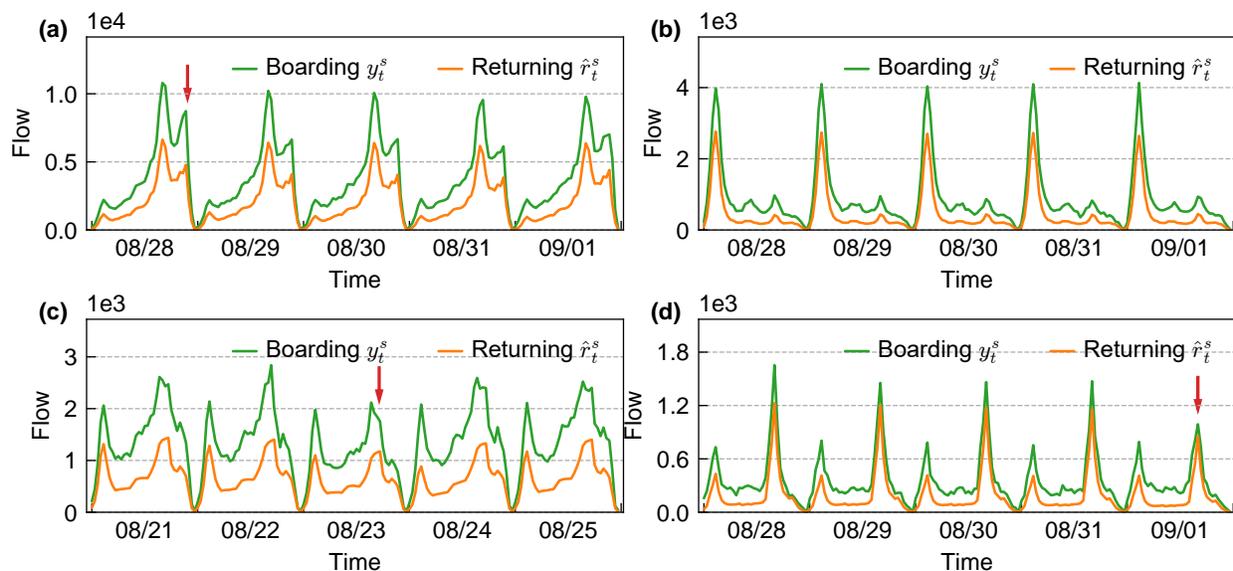
where  $y_t^s$  and  $\hat{y}_t^s$  are the real boarding flow and the predicted boarding flow, respectively. In addition to RMSE and SMAPE, we also use the Akaike information criterion (AIC) (Akaike, 1998) to measure the trade-off between the goodness of fit and the complexity of a model. A smaller AIC suggests a better model.

### 4.4.3 Data

We use the passenger flow data retrieved from Guangzhou metro in China as a case study. The smart card data set covers 159 stations from July 24 to September 8 in 2017. Note that the data on weekends are not included in our analysis (i.e., we concatenate Friday with the next Monday), as the RPP has different patterns on weekends. We divide the whole data set into three parts:

- (D1) July 24 to August 4 (two weeks): estimate the RPP  $p_0^s$  for a station  $s$  (for M2 only);
- (D2) August 7 to August 25 (three weeks): estimate model parameters for all the three SARIMA models (training set);

(D3) August 28 to September 8 (two weeks): evaluate model performance (test set).



**Figure 4.5:** The boarding flow and the estimated returning flow. As marked by red arrows, the returning flow reflects the irregular increases/drops of the boarding flow: (a) Tiyu Xilu station, (b) Luoxi station, (c) Changshou Lu station, and (d) Huijiang station. The green curves represent demand time series  $y_t^s$  and the orange curves correspond to the estimated returning flow time series  $\hat{r}_t^s$ .

After estimating RPP from D1, we compute  $\hat{r}_{t+1}^s$  on data sets D2 and D3 following Eq. (4.6). Before estimating the SARIMA models, we first empirically examine the relationship between the time series of returning flow  $\hat{r}_t^s$  and the time series of incoming demand  $y_t^s$ . Figure 4.5 shows the demand time series  $y_t^s$  and the estimated returning flow time series  $\hat{r}_t^s$  on data set D2/D3 for four representative stations. Station (a) and (d) are commercial areas, where the boarding flows concentrate in the afternoon and evening. Station (b) is a residential area that has an extremely high morning peak. Station (c) has high boarding flow in both the morning and afternoon peaks. We can see that the estimated returning flow  $\hat{r}_t^s$  matches  $y_t^s$  for all types of stations very well. Notably, the returning flow  $\hat{r}_t^s$  makes up a large proportion of the total boarding demand, and it can correctly characterize the temporal dynamics in  $y_t^s$ . More importantly, as marked by the red arrows in panels (a), (c) and (d), the returning flow can even reproduce some irregular increases/drops (i.e., anomalies) of the boarding flow, which are very difficult to capture using conventional time series models with  $y_t^s$  alone.

#### 4.4.4 One-step-ahead forecasting

We use data set D2 to estimate model parameters and apply the model to D3 for evaluation. Table 4.1 shows the results of one-step-ahead forecasting for the four stations in Figure 4.5. Compared with M0, M2 consistently reduces the RMSE and SMAPE of both training and test sets of the four stations by incorporating  $\hat{r}_{t+1}$ . Meanwhile, M2 is also superior with a larger log-likelihood and a lower AIC. However, with the observed  $r_t^s$  as input, M1 performs almost the same with M0. This might be due to the fact that  $r_t^s$  correlates highly with  $y_t^s$  (the observation at the last step), since the returning flow covers a considerable proportion of the overall boarding flow. Thus, the amount of additional information brought by this term is rather marginal. While on the contrary,  $\hat{r}_{t+1}^s$  estimated externally by combining RPP and the alighting time series  $m_t^s$  actually encodes the generative mechanisms and long-range dependencies, and thus M2 produces much better forecasting results.

**Table 4.1:** The one-step boarding flow forecasting of four stations.

Stations	Model	RMSE (train)	RMSE (test)	SMAPE (train)	SMAPE (test)	Log-likelihood	AIC
(a) Tiyu Xilu	M0	398.09	363.06	11.89%	11.63%	-3747.19	7504.37
	M1	396.95	362.00	12.36%	12.22%	-3745.76	7503.52
	M2	<b>372.94</b>	<b>319.60</b>	<b>10.59%</b>	<b>9.29%</b>	<b>-3713.9</b>	<b>7439.79</b>
(b) Luoxi	M0	64.06	71.61	9.51%	9.93%	-2830.25	5670.49
	M1	64.06	71.61	9.51%	9.93%	-2830.25	5672.49
	M2	<b>63.92</b>	<b>71.37</b>	<b>9.48%</b>	<b>9.89%</b>	<b>-2829.25</b>	<b>5670.48</b>
(c) Changshou Lu	M0	93.78	94.34	10.36%	11.88%	-3018.66	6047.31
	M1	93.64	95.16	10.31%	11.99%	-3017.84	6047.67
	M2	<b>90.02</b>	<b>92.74</b>	<b>9.39%</b>	<b>10.83%</b>	<b>-2998.07</b>	<b>6008.15</b>
(d) Huijiang	M0	37.32	47.37	14.05%	14.27%	-2557.94	5125.88
	M1	37.29	47.29	14.05%	14.25%	-2557.61	5127.21
	M2	<b>36.67</b>	<b>38.61</b>	<b>13.49%</b>	<b>13.91%</b>	<b>-2549.47</b>	<b>5110.93</b>

To further evaluate whether the improvement of M2 is statistically significant, we apply paired t-test to compare the absolute forecast errors on the test set D3. For each station, denote the forecast error of model M to be a random variable  $\varepsilon_M = \hat{y} - y$ . When comparing M2 and M0, the null hypothesis  $H_0 : \mu(|\varepsilon_{M2}| - |\varepsilon_{M0}|) = 0$  means no significant difference between the absolute forecast error of M2 and M0. We use the lower-tailed alternative hypothesis  $H_a : \mu(|\varepsilon_{M2}| - |\varepsilon_{M0}|) < 0$ , which means the absolute forecast error of M2 is smaller than M0. We also compare M2 with M1 in the same way. Based on the p-values in Table 4.2, we reject  $H_0$  for stations (a)(c)(d). Therefore, M2 indeed improves the forecast

for stations (a)(c)(d). Although M2 also reduces the RMSE and SMAPE for station (b), Table 4.2 shows the improvement is not significant at the 0.05 level.

**Table 4.2:** Paired t-test p-values for absolute forecast errors.

	(a) Tiyu Xilu	(b) Luoxi	(c) Changshou Lu	(d) Huijiang
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) = 0$	< 0.001*	0.065	0.016*	0.003*
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) < 0$				
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) = 0$	< 0.001*	0.061	0.013*	0.003*
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) < 0$				

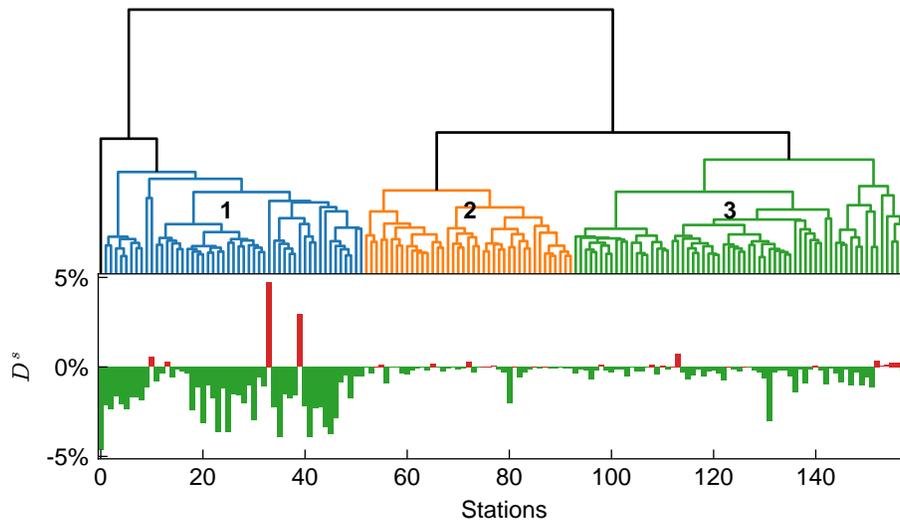
\* Significant at 0.05 level.

The results in Table 4.1 and Table 4.2 indeed show that M2 gives improved accuracy; however, it should be also noted that the improvement varies across different stations. To further explore this variation, we cluster the 159 stations based on their RPPs. In doing so, we transform each RPP into a vector of  $36 \times 36 = 1296$  and perform hierarchical clustering using the Euclidean distance between paired vectors; the distances between clusters are calculated by the Ward's method (Ward Jr, 1963). In the meanwhile, we measure the effect of  $\hat{r}_{t+1}^s$  by the difference in SMAPE:

$$D^s = \text{SMAPE}_{M2}^s - \text{SMAPE}_{M0}^s, \quad (4.11)$$

where  $\text{SMAPE}_{M2}^s$  and  $\text{SMAPE}_{M0}^s$  are SMAPE values of M2 and M0, respectively, on the test data set D3. A negative  $D^s$  means M2 improves the forecasting accuracy.

The dendrogram for the hierarchical clustering is shown in Figure 4.6. We cut the clustering tree at the half-height, which divides the 159 stations into three major clusters (with one station in exception). The cluster centroids (the average RPP for the cluster) for the three clusters are shown in Figure 4.7, where we also show the probabilities of returning on the same day and the next day. From Figure 4.7, we can see cluster 1 corresponds to business-type areas where more passengers return on the same day, such as the Tiyu Xilu station in Figure 4.4 (a). For cluster 2, the probability of returning on the next day is higher than returning on the same day, exhibiting the feature of residential areas, such as the Luoxi station in Figure 4.4 (b). Cluster 3 is a combination of cluster 1 and 2, which has a relatively balanced returning flow in both the current and the next day. The bottom panel of Figure 4.6 shows the  $D^s$  values for all stations following the same order as the clustering result. As can be seen,  $D^s$  values are negative for most stations, confirming the effectiveness of model M2. It should be noted that the effect of  $\hat{r}_{t+1}^s$  is different among clusters. The

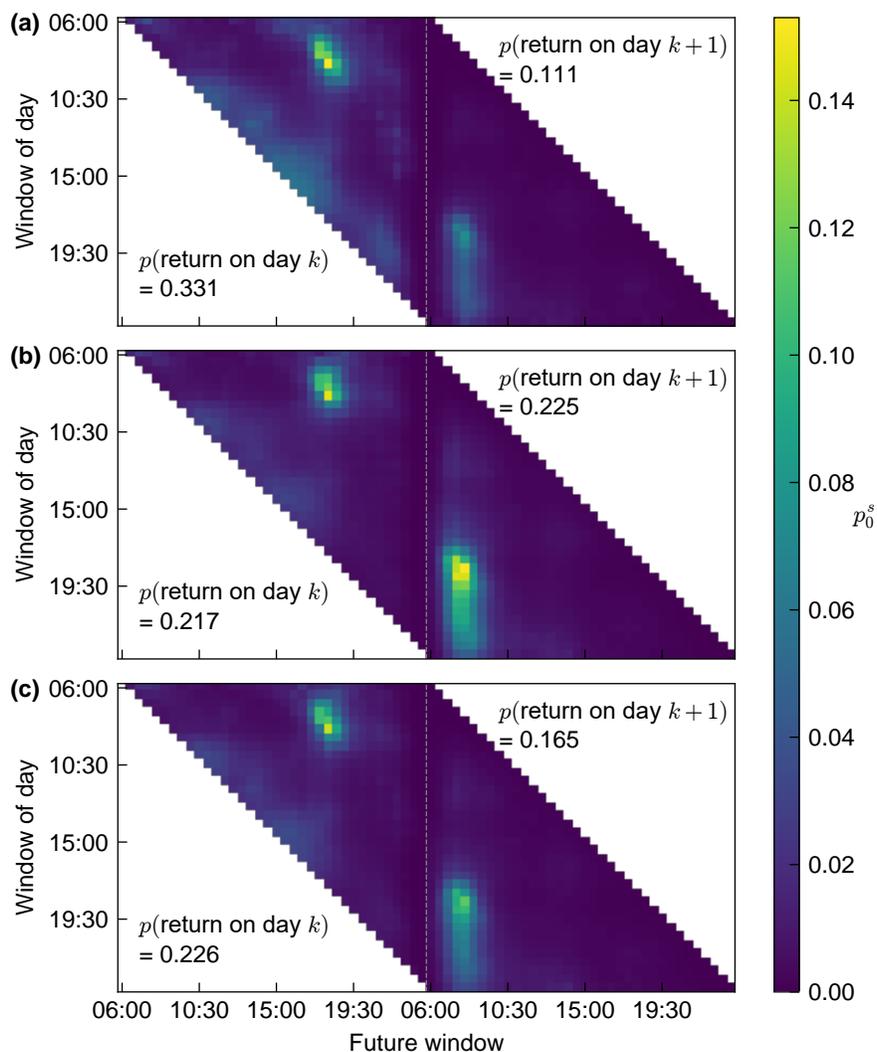


**Figure 4.6:** Top: the dendrogram for the hierarchical clustering based on RPP. Bottom: the test set SMAPE differences between M2 and M0; green and negative values means using  $\hat{r}_{t+1}^s$  improves the boarding flow forecast in the test set.

reduction of the SMAPE is the most profound for cluster 1 (with two exceptions, which we will discuss in detail in Section 4.4.6), while visually the least significant for cluster 2. We use the above paired t-test to check if the improvement of M2 compared with M0 is statistically significant. Using the 0.05 significance level, we find the improvements for 23 out of 51 stations (45.1%) in cluster 1 are significant, 1 out of 41 stations (2.4%) in cluster 2 are significant, and 9 out of 66 stations (13.6%) in cluster 3 are significant. These results suggest the returning flow  $\hat{r}_{t+1}^s$  is more effective for the forecast of business-type stations. The reason could be that the duration for work/shop activities is more fixed than the home activity, so the return flow estimation is more accurate for business-type stations.

#### 4.4.5 Multi-step-ahead forecasting

Even with strong seasonality, multi-step-ahead forecasting is still a challenging task because the errors will accumulate with the rolling forecasting process. A unique advantage of model M2 is that the estimation of returning flow  $\hat{r}_{t+L}^s$  suffers less from this error accumulation problem thanks to the long-range dependencies, and even the alighting flow in many steps ago could still dominate the future returning flow. For example, as shown in Figure 4.4 (a), the alighting flow in the morning (7:00-10:00) plays an important role in determining the returning flow of the evening (17:00-19:00). Therefore, using returning flow as an additional feature in M2 could potentially alleviate the error accumulation



**Figure 4.7:** The cluster centroids. (a), (b), and (c) correspond to the cluster centroids of cluster 1, 2, and 3, respectively.

problem in multi-step-ahead forecasting.

For an  $L$ -step boarding flow forecasting that predicts  $y_{t+L}^s$  by  $y_{1:t}^s$ , M2 requires a series of returning flow  $\hat{r}_{t+1}^s, \dots, \hat{r}_{t+L}^s$  as input. However, in order to estimate  $\hat{r}_{t+L}^s$ , Eq. (4.6) requires the alighting flow series  $m_{t+1:t+L-1}^s$ , which are not available. In this case, we use the average alighting flow at the same window of historical days as the approximation of future alighting flow  $m_{t+1}^s \dots m_{t+L-1}^s$ . This approximation for the future alighting flow should only bring minor errors to the estimation of returning flow in Eq. (4.6), since it only contributes to the last  $L - 1$  components in the summation.

We examine the multi-step-ahead forecasting using a time series cross-validation

method that is known as “evaluation on a rolling forecasting origin” (Hyndman and Athanasopoulos, 2018, Chapter 3). For an  $L$ -step-ahead forecasting, we train a model for each observation in data set D3 using a training set from the first observation in data set D2 to the observation  $L$  steps prior to that observation. The error is only evaluated at the  $L^{\text{th}}$  step, and the overall error is the average error over the test set.

Table 4.3 shows the result of 1, 2, 4, and 6 steps forecasting for M0 and M2. Compared with M0, M2 offers substantially enhanced forecasting in stations (a), (c), and (d), and the errors increase much slower with the growing step  $L$ . Especially, in station (a), the RMSE of M0 increases 252.55 (75.6%) from 1-step forecast to 6-step forecast, the number is only 174.66 (60.4%) for M2. For the residential (cluster 2) station (b), the effect of M2 in multi-step-ahead forecasting is less significant, which validates the different contributions of returning flow among different clusters. Overall, we can see that multi-step-ahead forecasting tasks can benefit substantially from the long-range dependencies encoded in M2 and  $\hat{r}_{t+1}^s$ .

**Table 4.3:** The multi-step boarding flow forecasting of four stations by time series cross-validation (30-min resolution).

Station	Model	30 mins ( $L = 1$ )		1 hour ( $L = 2$ )		2 hours ( $L = 4$ )		3 hours ( $L = 6$ )	
		RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE
(a) Tiyu Xilu	M0	334.06	12.92%	435.16	15.30%	547.35	19.88%	586.61	19.23%
	M2	<b>290.43</b>	<b>9.87%</b>	<b>328.90</b>	<b>12.60%</b>	<b>403.20</b>	<b>10.82%</b>	<b>465.09</b>	<b>11.71%</b>
(b) Luoxi	M0	<b>75.94</b>	10.64%	79.53	11.70%	88.49	12.14%	90.16	12.51%
	M2	76.10	<b>10.60%</b>	<b>79.43</b>	<b>11.65%</b>	<b>88.49</b>	<b>12.08%</b>	<b>89.97</b>	<b>12.38%</b>
(c) Changshou Lu	M0	97.92	10.28%	126.65	14.42%	154.58	16.74%	168.33	15.89%
	M2	<b>84.80</b>	<b>7.85%</b>	<b>103.84</b>	<b>8.84%</b>	<b>129.53</b>	<b>10.37%</b>	<b>153.12</b>	<b>11.58%</b>
(d) Huijiang	M0	50.27	14.50%	54.09	15.52%	56.16	16.22%	56.30	17.33%
	M2	<b>39.76</b>	<b>13.95%</b>	<b>41.13</b>	<b>14.70%</b>	<b>42.76</b>	<b>15.00%</b>	<b>43.52</b>	<b>15.69%</b>

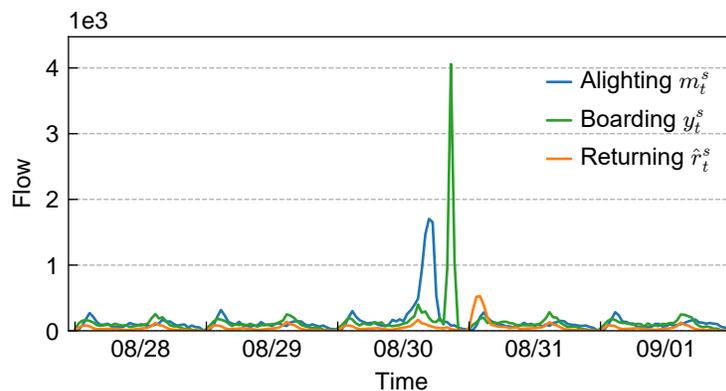
The estimation of the returning trip closely relates to the inter-trip duration. It is thus worth analyzing the effect of time resolution—especially a more refined resolution—to the forecasting performance. We apply a 15-min resolution to further test the impact of the returning flow to the multi-step forecasting. The results are shown in Table 4.4, where the baseline model M0 is ARIMA(2,0,1)(1,1,0)[72]. We can see using the returning flow still greatly alleviates the error accumulation in multi-step forecasting, and the forecasting improvement is the most significant for station (a), (c), and (d), which is consistent with Table 4.3 and Section 4.4.4.

**Table 4.4:** The multi-step boarding flow forecasting of four stations by time series cross-validation (15-min resolution).

Station	Model	15 mins ( $L = 1$ )		30 mins ( $L = 2$ )		1 hour ( $L = 4$ )		1.5 hour ( $L = 6$ )	
		RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE	RMSE	SMAPE
(a) Tiyu Xilu	M0	169.57	11.63%	210.38	13.95%	248.59	16.02%	283.43	18.86%
	M2	<b>163.43</b>	<b>10.35%</b>	<b>190.49</b>	<b>11.44%</b>	<b>197.12</b>	<b>13.38%</b>	<b>206.92</b>	<b>13.24%</b>
(b) Luoxi	M0	<b>41.70</b>	<b>12.45%</b>	<b>44.31</b>	12.79%	46.53	13.61%	<b>48.68</b>	14.29%
	M2	41.72	12.46%	44.33	<b>12.77%</b>	<b>46.48</b>	<b>13.58%</b>	48.71	<b>14.25%</b>
(c) Changshou Lu	M0	52.20	12.29%	60.30	14.26%	72.18	17.44%	79.66	19.55%
	M2	<b>47.79</b>	<b>9.52%</b>	<b>52.76</b>	<b>10.16%</b>	<b>60.42</b>	<b>10.45%</b>	<b>66.06</b>	<b>11.45%</b>
(d) Huijiang	M0	29.22	17.36%	30.56	17.85%	31.90	18.14%	32.42	19.07%
	M2	<b>25.55</b>	<b>17.21%</b>	<b>25.90</b>	<b>17.52%</b>	<b>26.54</b>	<b>17.87%</b>	<b>26.90</b>	<b>18.63%</b>

#### 4.4.6 Forecasting under special events

As shown in Figure 4.6, M2 is less effective only for a few stations. A main reason is that these stations in general have a large variation in RPP from day to day. For these stations,  $\hat{r}_{t+1}^s$  will be less accurate and less informative in supporting the forecasting. Therefore, M2 with  $\hat{r}_{t+1}^s$  estimated by a universal RPP will not benefit as much, if not more, than M0 and M1.

**Figure 4.8:** The alighting, boarding, and the estimated returning flow of Luogang station under an event (August 30).

This is particularly the case for forecasting under special events, since the RPP cannot be well estimated using historical data, and it will involve large variations in nature. For example, as a metro station next to the Guangzhou International Sports Arena, Luogang often experiences surging demand because of large sports events and concerts. Figure 4.8 shows that the come-and-return dynamics under the special event is very different from

normal days. The event on August 30 brought a period of unusually high alighting flow. After the event, the returning passengers caused another peak in the boarding flow. If we adopt a universal RPP estimated by aggregating historical data, we will end up with erroneously distributing the returning flow to the next morning (the morning peak of  $\hat{r}_t^s$  on August 31). To address this problem, we propose to build two separate RPPs for normal and event-induced passengers, respectively. For the passengers alight at time  $t_a$ , the probability of returning at  $t$  becomes the weighted sum of the two parts:

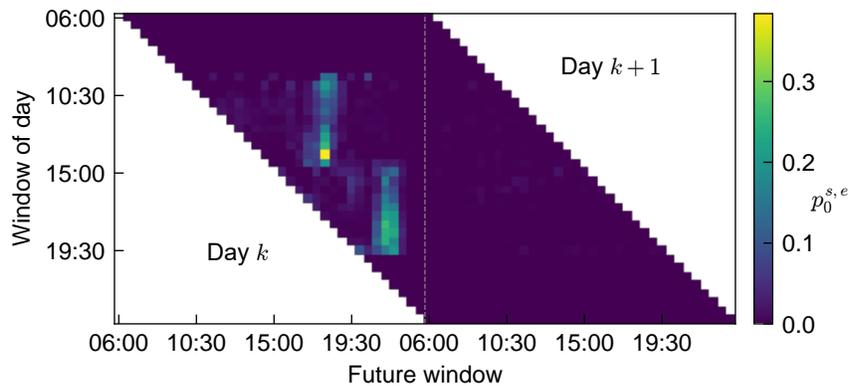
$$p^s (\tau_{\text{boarding}} = t \mid \tau_{\text{alighting}} = t_a) = \frac{m_{t_a}^{s,e}}{m_{t_a}^s} p^{s,e} (\tau_{\text{boarding}} = t \mid \tau_{\text{alighting}} = t_a) + \frac{m_{t_a}^{s,n}}{m_{t_a}^s} p^{s,n} (\tau_{\text{boarding}} = t \mid \tau_{\text{alighting}} = t_a), \quad (4.12)$$

where we use superscript  $e$  and  $n$  to denote variables for event and normal conditions, respectively, and thus the alighting flow is  $m_{t_a}^s = m_{t_a}^{s,e} + m_{t_a}^{s,n}$ . When the event-induced alighting flow  $m_{t_a}^{s,e} = 0$ , Eq. (4.12) reduces to the normal RPP.

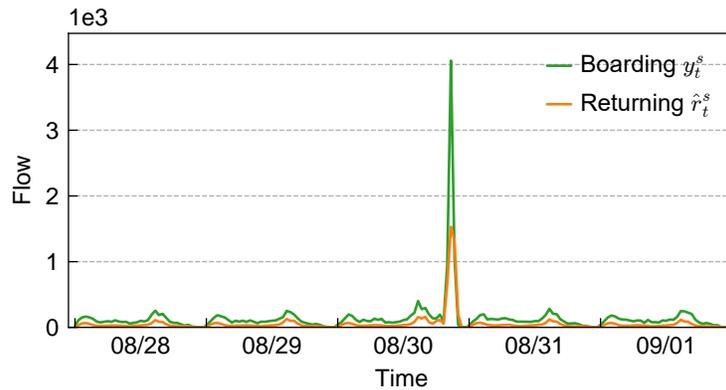
In practice, we have a few approaches to estimate the event-induced alighting flow  $m_{t_a}^{s,e}$  and RPP under events, such as looking into the passenger flow of the specific gate to the event venue or using the time information of an event. When such information is not available, we propose to apply the following method to estimate the RPP under events (assuming all events in station  $s$  follow the same RPP). First, a period with alighting flow larger than a threshold is identified as an event period. For each time window in a day, we use  $Q_3 + 1.5IQR$  as the threshold, where  $Q_3$  is the third quartile and  $IQR$  the interquartile range. Next, the normal RPP can be estimated by the non-event periods. Subtracting the normal alighting (use median) and the normal returning flow from the part identified as event periods, the rest data in event periods are used to estimate the RPP under special events. By separating event and normal flow, we also prevent the normal RPP from being influenced by the event flow.

Using the data from July 1 to August 24, 2017 (weekends included), Figure 4.9 shows the event RPP of Luogang station. It is conspicuous that two types of events exist in this venue—one ends in the afternoon and the other ends in the evening. The returning time for each type of event is more concentrated than the alighting time and is mostly on the same day of the alighting time. Although the RPP is estimated from different events (8 days with significant events), the return probability of these events shows regular and predictive patterns.

Figure 4.10 shows the re-estimated returning flow of Luogang station. Compared with Figure 4.8, we can find the returning flow in Figure 4.10 can now correctly reflect



**Figure 4.9:** The RPP for passenger flow induced by events in Luogang station. Probabilities are set to zero for time windows without event.



**Figure 4.10:** The boarding and the returning flow (using Eq. (4.12)) of Luogang station.

the peak and trend of the boarding flow under the special event. Note that the days in Figure 4.10 are different from the days used for event RPP estimation, which shows the come-and-return dynamic of different events in this station follows similar patterns.

Considering that these events are occasional, we use the non-seasonal ARIMA with the order  $ARIMA(2, 0, 1)$  as our baseline model  $M_0$ . The data separation is the same as Section 4.4.3, except that we use a longer period for the event RPP estimation. We denote by  $M_2'$  the model that uses the “adjusted” returning flow as a covariate. The forecasting results are shown in Table 4.5. To highlight the forecasting performance under events, we only use event periods (when the boarding flow exceeds  $Q_3 + 1.5IQR$ ) to calculate the RMSE and the SMAPE, respectively denoted by  $RMSE_e$  and  $SMAPE_e$ . We can see  $M_2'$  with the “adjusted” returning flow has the best performance under all criteria. The results further show that the returning flow offers substantial improvement even for the

forecasting under special events. Although the simple ARIMA may not be the best model for a time series with apparent “outliers” such as these events, the returning flow could be easily integrated into other models (such as the generalized autoregressive conditional heteroskedasticity, GARCH) for a better prediction.

**Table 4.5:** The boarding flow forecasting under special events for Luogang stations.

Model	RMSE <sub>e</sub> (train)	RMSE <sub>e</sub> (test)	SMAPE <sub>e</sub> (train)	SMAPE <sub>e</sub> (test)	Log-likelihood	AIC
M0	392.64	1003.29	24.55%	46.62%	-2984.11	5978.22
M1	408.15	1043.25	27.19%	45.34%	-2972.46	5956.92
M2	397.14	1018.88	25.95%	46.79%	-2909.38	5830.76
M2'	<b>255.46</b>	<b>535.63</b>	<b>23.16%</b>	<b>36.30%</b>	<b>-2821.29</b>	<b>5654.58</b>

#### 4.4.7 Experiments in other models

To further test if the returning flow can also improve other ridership forecast models, we use two popular machine learning models—Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP)—to repeat the experiment in Section 4.4.4. We rescale data to  $[0, 1]$  by min-max normalization as a preprocessing, and we use the scikit-learn python package to implement these models.

The SVR model is similar to a previous work by Tang et al. (2018), except that we have no external features like the weather. The input features for M0 are the boarding flow at time  $t$  and  $t - 1$ , and 36 dummy variables representing the time of a day. We add  $r_t^s$  to M1 and  $\hat{r}_{t+1}^s$  to M2 as an additional feature. We tune hyperparameters by cross-validation using M0 and select  $C = 0.274$  and  $\varepsilon = 0.016$ ; other hyperparameters are the default setting of the scikit-learn package. The forecast results of the four stations by SVR are shown in Table 4.6 and the significance tests are shown in Table 4.7.

The MLP uses the same features as the SVR. We tune hyperparameters by cross-validation using M0 and select the hidden layer size to be 150 and use the identity activation function; other hyperparameters are the default setting of the scikit-learn package. The forecast results of the four stations by MPL are shown in Table 4.8 and the significance tests are shown in Table 4.9.

In Table 4.6 and Table 4.8, M2 has lower forecast RMSE and SMAPE in the test set than M0 and M1 for stations (a)(c)(d). The hypothesis tests in Table 4.7 and Table 4.9 for these stations also show that the absolute forecast error of M2 is less than M0 and M1. On the other hand, the effect of using the returning flow is not significant for station

**Table 4.6:** The one-step boarding flow forecasting of four stations by SVR

Stations	Model	RMSE (train)	RMSE (test)	SMAPE (train)	SMAPE (test)
(a) Tiyu Xilu	M0	288.00	337.02	11.27%	11.60%
	M1	286.54	337.75	11.59%	11.76%
	M2	<b>280.32</b>	<b>306.79</b>	<b>11.12%</b>	<b>11.45%</b>
(b) Luoxi	M0	<b>64.61</b>	80.66	<b>11.30%</b>	13.41%
	M1	65.81	78.45	11.37%	<b>13.39%</b>
	M2	64.86	<b>78.34</b>	11.47%	13.53%
(c) Changshou Lu	M0	72.80	111.81	<b>6.00%</b>	9.72%
	M1	72.91	109.99	6.13%	9.51%
	M2	<b>65.71</b>	<b>103.83</b>	6.06%	<b>9.33%</b>
(d) Huijiang	M0	31.95	45.37	13.53%	15.75%
	M1	31.98	44.46	13.61%	15.83%
	M2	<b>31.64</b>	<b>42.53</b>	<b>13.46%</b>	<b>15.75%</b>

**Table 4.7:** Paired t-test p-values for SVR absolute forecast errors.

	(a) Tiyu Xilu	(b) Luoxi	(c) Changshou Lu	(d) Huijiang
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) = 0$				
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) < 0$	< 0.001*	0.403	< 0.001*	0.022*
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) = 0$				
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) < 0$	< 0.001*	0.768	< 0.001*	0.011*

\* Significant at 0.05 level.

**Table 4.8:** The one-step boarding flow forecasting of four stations by MLP

Stations	Model	RMSE (train)	RMSE (test)	SMAPE (train)	SMAPE (test)
(a) Tiyu Xilu	M0	312.23	332.87	12.08%	13.52%
	M1	309.10	333.81	11.12%	13.08%
	M2	<b>296.60</b>	<b>307.57</b>	<b>10.68%</b>	<b>12.45%</b>
(b) Luoxi	M0	<b>59.86</b>	<b>74.89</b>	<b>9.92%</b>	<b>11.68%</b>
	M1	71.42	82.27	13.57%	15.56%
	M2	72.66	80.54	14.30%	15.92%
(c) Changshou Lu	M0	74.03	108.54	<b>7.28%</b>	13.37%
	M1	77.38	105.13	9.19%	13.60%
	M2	<b>65.50</b>	<b>95.31</b>	7.41%	<b>12.35%</b>
(d) Huijiang	M0	32.50	43.08	<b>12.62%</b>	15.59%
	M1	34.64	43.85	14.03%	15.59%
	M2	<b>31.39</b>	<b>35.86</b>	14.02%	<b>15.20%</b>

**Table 4.9:** Paired t-test p-values for MLP absolute forecast errors.

	(a) Tiyu Xilu	(b) Luoxi	(c) Changshou Lu	(d) Huijiang
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) = 0$	< 0.001*	0.999	< 0.001*	0.031*
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M0} ) < 0$				
$H_0 : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) = 0$	< 0.001*	0.607	0.003*	< 0.001*
$H_a : \mu ( \varepsilon_{M2}  -  \varepsilon_{M1} ) < 0$				

\* Significant at 0.05 level.

(b)—a residential type station. In summary, the returning flow can improve the boarding flow forecast of SVR and MLP, and the improvement is more significant for business-type stations with more passengers returning on the same day (i.e., station (a)(c)(d)). These results are consistent with the SARIMA model.

Finally, we apply SVR and MLP to all stations and use the paired t-test described in Section 4.4.4 to test if the improvement of M2 is significant compared with M0. Results are shown in Table 4.10, where the clusters are the same as Figure 4.6. The results of different models are consistent, and the proposed returning flow is more effective for business-type stations.

**Table 4.10:** The number of significant stations in the paired t-test between M2 and M0 (0.05 significance level).

	Cluster 1 (business-type)	Cluster 2 (residential-type)	Cluster 3 (combined-type)
Number of stations	51	41	66
SARIMA	23 (45.1%)	1 (2.4%)	9 (13.6%)
SVR	22 (43.1%)	4 (9.8%)	13 (19.7%)
MLP	29 (56.8%)	4 (9.8%)	13 (19.7%)

## 4.5 Conclusions and Discussion

In this paper, we propose a new framework for forecasting passenger flow time series in metro systems. In contrast to some previous studies that capture temporal dynamics in a data-driven way, we try to incorporate the generative mechanisms rooted in travel behavior into modeling passenger flow time series. For that purpose, we introduce returning flow as a new covariate/feature into standard time series models. This returning flow is estimated as the expected returning boarding demand given previous alighting trips; thus, it encodes

the causal structure and long-range dependencies in passenger flow data. We estimate the return probability by aggregating historical data, thereby working around the sensitivity issues and privacy concerns accompanying individual-based data and models as in [Zhao et al. \(2018c\)](#). We examine the proposed framework on a real-world passenger flow data set collected from Guangzhou metro in China. The proposed framework with the returning flow demonstrates superior performance in all three tested scenarios, namely one-step-ahead forecasting, multi-step-ahead forecasting, and forecasting in special events. And we found the returning flow is more useful for the boarding demand forecast of business-type stations, where most returning trips are within the same day. On the contrary, the model does not bring much improvement for residential stations. This result suggests that “home” activity duration demonstrates a higher variance than that of “work” activities. In addition, the experiments in Section 4.4.7 show the returning flow also improves the forecast of machine learning models like SVR and MLP. In fact, the returning flow (as a covariate) and the idea of regularity-based long-range dependency can be used in a diverse range of prediction models (e.g., time series model, machine learning models, deep learning).

There are several directions for future research. First, this study assumes that both the boarding and the alighting time series are available (i.e., both tapping-in and tapping-out are registered by the smart card system), while metro in some cities may have a tapping-in only system. In this case, one should integrate a destination inference model (see e.g., [Barry et al., 2002](#); [Trépanier et al., 2007](#); [Cheng et al., 2021b](#)) into the proposed framework. Estimating the returning probability by other data sources, such as survey, Bluetooth, and call detail records, is also worth exploring. Second, the come-and-return pattern and RPP of a station may change over time. How to detect the pattern changes ([Zhao et al., 2018a](#)) and develop time-varying models should be studied. The current constant RPP is a simplified assumption; further utilizing the returning flow’s auto-correlation is a possible approach to improve the returning flow estimation. There is still space to improve the model by advanced statistical time series and deep learning-based sequence models. Third, the models developed in the current framework are station-specific, while travel behavior regularity is ubiquitous and different stations may share similar patterns. Therefore, the RPP formulation can be generalized using parametric model, approximation, and dimensionality reduction techniques such as principal component analysis and matrix/tensor factorization ([Sun and Axhausen, 2016](#)) to extract common patterns for RPPs across different stations. This can be particularly useful when only limited data are available. Fourth, we should relax the assumption that the returning flow must have the same boarding station as the previous alighting station, because there could be multiple destinations or

multiple transportation modes in one's activity chain (Bowman and Ben-Akiva, 2001). It is possible to improve and extend the forecasting by multi-modal trips since the metro system is often combined with other transportation modes; a route/mode choice model can be integrated into the multi-modal forecasting framework. Sufficient data is a prerequisite for this direction. Lastly, we can extend this framework to other transport modes with non-random and chained travel patterns, such as private vehicles, taxis, and ride-hailing services. This paper also sheds new light on other behavior-driven demand forecasting problems, in which the causal structure and the long-range dependencies play a substantial role. For instance, integrating purchasing behavior into the demand forecasting of retail products.

## Chapter 5

# Origin-Destination Matrices Forecasting with Dynamic Mode Decomposition

---

This chapter is an article accepted by Transportation Science:

- Cheng, Z., Trépanier, M., Sun, L., 2022. Real-time forecasting of metro origin-destination matrices with high-order weighted dynamic mode decomposition. *Transportation Science* (in press) doi: [10.1287/trsc.2022.1128](https://doi.org/10.1287/trsc.2022.1128).

This chapter and the previous Chapter 4 both focus on the short passenger demand forecasting in metro systems. Chapter 4 builds station-specific models while this chapter uses a single model to forecast the OD matrices of the entire metro network. The station-level boarding flow forecasting can also be obtained from the OD matrices forecasting. Thanks to the strong regularity and the power law characteristic of OD matrices, we can use a low-rank model to solve the high-dimensional OD matrices forecasting problem.

---

## 5.1 Abstract

Forecasting short-term ridership of different origin-destination pairs (i.e., OD matrix) is crucial to the real-time operation of a metro system. However, this problem is notoriously difficult due to the large-scale, high-dimensional, noisy, and highly skewed nature of OD matrices. In this paper, we address the short-term OD matrix forecasting problem by estimating a low-rank high-order vector autoregression (VAR) model. We reconstruct this problem as a data-driven reduced-order regression model and estimate it using dynamic mode decomposition (DMD). The VAR coefficients estimated by DMD are the best-fit (in terms of Frobenius norm) linear operator for the rank-reduced full-size data. To address the practical issue that metro OD matrices cannot be observed in real-time, we use the boarding demand to replace the unavailable OD matrices. Moreover, we consider the time-evolving feature of metro systems and improve the forecast by exponentially reducing the weights for historical data. A tailored online update algorithm is then developed for the High-order Weighted DMD model (HW-DMD) to update the model coefficients at a daily level, without storing historical data or retraining. Experiments on data from two large-scale metro systems show that the proposed HW-DMD is robust to noisy and sparse data, and significantly outperforms baseline models in forecasting both OD matrices and boarding flow. The online update algorithm also shows consistent accuracy over a long time, allowing us to maintain an HW-DMD model at much low costs.

## 5.2 Introduction

The metro is a green and efficient travel mode that plays an ever-important role in urban transportation. An accurate real-time ridership/demand forecast is crucial to the efficiency and reliability of metro systems. With the wide application of smart card systems and diverse types of sensors, forecasting real-time metro ridership has become an emerging research question in recent years. Existing research mainly focuses on forecasting the short-term (e.g., 15 or 30 minutes) boarding or alighting ridership at metro stations, such as [Wei and Chen \(2012\)](#), [Sun et al. \(2015\)](#), [Li et al. \(2017\)](#), [Chen et al. \(2020a\)](#), [Liu et al. \(2019b\)](#), and [Zhang et al. \(2020\)](#). In contrast, forecasting the short-term ridership at origin-destination (OD) pairs of a metro system receives little attention. The ridership among all OD pairs of a metro system can be organized into a matrix. For simplicity, an “OD matrix” in this paper refers to such ridership matrix at a certain (short) time interval.

Forecasting metro OD matrices has much broader applications than the station-level

ridership forecast. For example, by assigning OD matrices to a metro network, we can predict and thus regulate the crowdedness of each train. The station-level boarding/alighting flow also can be calculated as the row and column sums of the OD matrix. However, the real-time forecast of metro OD matrices is extremely difficult for the following reasons. (1) The first challenge is the *high dimensionality*. The number of OD pairs of a metro system is the square of the number of stations, often tens of thousands in practice. (2) Short-term OD matrices of a metro system are often *sparse*, and the ridership/flow distribution within an OD matrix is highly *skewed* (see e.g., Figure 5.3). (3) Unlike the boarding or alighting flow, a metro system’s OD matrices cannot be obtained in real-time (*delayed data availability*). Because an OD matrix only becomes available after all the trips belonging to the OD matrix have reached their destinations. Lastly, (4) the complex dynamics of a metro system are *time-evolving*; a well-tuned model may have a short “shelf life” and has expensive retrain/re-tune costs in long-term maintenance. Although a few studies tried to forecast the real-time metro OD matrices by matrix factorization methods (Gong et al., 2018, 2020) or deep learning models (Toqué et al., 2016; Zhang et al., 2021b; Shen et al., 2021), no existing solution overcomes all the four challenges above.

This paper utilizes Dynamic Mode Decomposition (DMD) (Schmid, 2010) – a recent advance in the fluid dynamics community – to address the above challenges in real-time metro OD matrix forecasting problem. DMD is a dimensionality reduction technique that extracts dominating dynamics (modes) from a sequence of high-dimensional vectors. The uniqueness of DMD is that it identifies the best-fit (in terms of Frobenius norm) linear operator that advances a high-dimensional vector sequence forward in time (Tu et al., 2014). We extend the original DMD model by a high-order vector autoregression to incorporate long-term temporal correlations. In dealing with the delayed data availability problem, we replace the latest OD matrices, which are unavailable, with snapshots of boarding flow. We also consider the time-evolving dynamics and introduce a forgetting ratio to reduce the weights of past data exponentially. We name the proposed model High-order Weighted Dynamic Mode Decomposition (HW-DMD). Moreover, we develop a tailored online update algorithm that updates an HW-DMD’s coefficients daily without storing historical data or retraining the model, which greatly reduces the model maintenance costs for long-term implementations. Finally, the proposed model is tested on a Guangzhou metro data set with 159 stations and an Hangzhou metro data set with 80 stations. Our experiments show that HW-DMD can excellently handle the sparse, skewed, and noisy OD data and significantly outperforms baseline models in forecasting both the OD matrices and the boarding flow. The online update algorithm also shows consistent accuracy in

updating an HW-DMD model over a long period. Although the online HW-DMD model is applied to the metro OD matrix forecasting problem, it can be readily applied to general (high-dimensional) traffic flow forecast problems, such as in recent studies about DMD-based traffic flow forecasting (Avila and Mezić, 2020; Yu et al., 2020). We summary main contributions of this paper as follows:

- This paper proposes an HW-DMD model that addresses various difficulties of the real-time metro OD matrix forecasting. Experiments show the forecast of HW-DMD is significantly better than existing models.
- The time-evolving dynamics of a transportation system and the maintenance/update of a forecasting model are often ignored in the literature. This paper considers the time-evolving feature of a metro system by reducing the weights for past data and shows improved performance. An online update algorithm is proposed to reduce the long-term maintenance cost of the HW-DMD model in a time-evolving metro system.
- We propose a DMD-based estimation and online update algorithm for large-scale high-order vector autoregression models with external covariates. The DMD-based estimation produces a best-fit linear operator for rank-reduced full-size data and is particularly useful for the forecast of high-dimensional data with low-rank properties.

The remainder of this paper is organized as follows. We review related work on short-term OD matrix forecasting in Section 5.3. Next, a description of the metro OD matrix forecasting problem is presented in Section 5.4. Section 5.5 briefly introduces the DMD algorithm, which serves as the base for the proposed HW-DMD model. Section 5.6 is the core part of this paper, where the model specification, estimation, and the online update method for HW-DMD are elaborated. In Section 5.7, we conduct numerical experiments on the two metro data sets. Conclusions and future directions are summarized in Section 5.8.

### 5.3 Related Work

In the literature, only a few studies have explored the real-time OD matrix forecasting problem for a “metro” system. Therefore, we extend the range to OD demand forecasting for general road transportation modes, such as the ride-hailing system and the highway tolling system. Note that for a ride-hailing system, the origins and destinations are often defined as zones on a grid.

Matrix/tensor factorization is an effective method to tackle the high-dimensionality problem of OD matrix forecasting. For example, [Ren and Xie \(2017\)](#) applied Canonical Polyadic (CP) decomposition to an *origin*  $\times$  *destination*  $\times$  *vehicle\_type*  $\times$  *time* tensor from highway tolling data. Time series models were then built on the latent temporal matrix to forecast OD matrices. [Dai et al. \(2018\)](#) and [Liu et al. \(2020\)](#) used principal component analysis (PCA) to reduce the dimensionality of OD data and applied several machine learning models to the reduced data for OD flow forecasting. [Gong et al. \(2020\)](#) developed a matrix factorization model to forecast the OD matrices of a metro system. Their work highlights a solution to the delayed data availability problem and various spatial and temporal regularization techniques are introduced to improve the forecast. In summary, the matrix/tensor factorization-based OD matrix forecasting consists of two components: (1) a dimensionality reduction by factorization and (2) a forecasting model applied to the reduced data.

Deep learning is another mainstream method for OD matrix forecasting. In an early study, [Toqué et al. \(2016\)](#) used Long Short-Term Memory (LSTM) networks to forecast the OD matrices of a transit network. They only applied the model to selected high-flow OD pairs because of the high dimensionality and sparsity problems. Convolutional Neural Networks (CNN) and Graph Convolutional Networks (GNN) are two deep learning models that greatly reduce the model size compared with a fully connected neural network. Recently, using CNN/GCN to capture spatial correlations and LSTM to capture temporal correlations started to become the “standard configuration” for deep learning-based OD matrix forecasting. For example, [Chu et al. \(2019\)](#) used multi-scale convolutional LSTM to forecast the real-time taxi OD demand, and [Wang et al. \(2019c, 2020\)](#) used multi-task learning to improve the OD flow forecast of GCN+LSTM networks. A large body of literature focused on better utilizing the spatial/semantic correlations by optimizing the GNN structure or incorporating side information. Such as the local spatial context used by [Liu et al. \(2019a\)](#), the Spatio-Temporal Encoder-Decoder Residual Multi-Graph Convolutional network (ST-ED-RMGC) proposed by [Ke et al. \(2021\)](#), and the Dynamic Node-Edge Attention Network (DNEAT) developed by [Zhang et al. \(2021a\)](#). Some studies combined deep learning models with other models to complement each other. In this direction, [Xiong et al. \(2020\)](#) combined GCN with Kalman filter to forecast the OD matrices of a Turnpike network. [Shen et al. \(2021\)](#) mixed CNN with a Gravity model to forecast OD matrices of a metro system. [Hu et al. \(2020\)](#) considered the travel time between OD pairs as a stochastic variable, and developed a stochastic OD matrix forecasting model based on tensors factorization and GCN. [Noursalehi et al. \(2021\)](#) used discrete wavelet transform

to decompose OD matrices into frequency domain; the outputs were fed into CNN and Convolutional-LSTM networks for forecasting.

The performances of deep learning models are often impaired by the noise in sparse metro OD matrices. To reduce the impact of the noise, Zhang et al. (2019b, 2021b) developed a metric called OD attraction degree (ODAD) to mask insignificant OD pairs. Zhang et al. (2019b) showed that masking near-zero OD pairs improves the forecasting accuracy of an LSTM. Based on ODAD, Zhang et al. (2021b) developed a Channel-wise Attentive Split-CNN (CAS-CNN) model for metro OD matrix forecasting. Another merit of this work is they considered the delayed data availability problem.

In summary, matrix/tensor factorization, CNN, and GCN all aim to reduce model size while maintaining spatial/temporal correlations/dependencies. The HW-DMD model proposed in this paper belongs to the matrix factorization category. Although some ride-hailing systems may not have the delayed data availability problem, most research essentially omitted this problem for simplicity. Particularly, RNN-based deep learning models can barely work without the most recent OD matrices as inputs. In dealing with the delayed data availability problem, existing solutions (Gong et al., 2020; Zhang et al., 2021b; Xiong et al., 2020) used alternative quantities (e.g., boarding ridership, link flow) to compensate for the unavailable OD information. We also adopt this approach in the proposed model.

## 5.4 Problem Description

Many modern metro systems record passengers' entry and exit information using smart cards. We thus know the origin and destination stations, the start and end time for every trip in such a system. Given a fixed time interval (30 minutes in this paper), we denote by  $o_{t,i,j}$  the number of trips that depart from station  $i$  at the  $t$ -th interval to station  $j$ . We call  $o_{t,i,j}$  an *OD flow*. Next, we can describe the number of trips between every OD pair in the system at the  $t$ -th time interval by an *OD matrix*

$$O_t = \begin{bmatrix} o_{t,1,1} & \cdots & o_{t,1,s} \\ \vdots & \ddots & \vdots \\ o_{t,s,1} & \cdots & o_{t,s,s} \end{bmatrix} \in \mathbb{R}^{s \times s},$$

where  $s$  is the number of metro stations. The diagonal elements of a metro OD matrices are always zero. We keep these zero elements because they have a negligible effect on the

forecast. In our model, OD matrices are organized in a vector form

$$\mathbf{f}_t = \text{vec}(O_t) = [o_{t,1,1}, \dots, o_{t,s,1}, o_{t,1,2}, \dots, o_{t,s,2}, \dots, o_{t,1,s}, \dots, o_{t,s,s}]^\top \in \mathbb{R}^n,$$

where  $n = s \times s$  is the number of OD pairs. For convenience, we refer to  $\mathbf{f}_t$  as an *OD snapshot*.

Note that OD snapshots are aggregated by the time when passengers enter the system; the exit time might be in a different time interval. Therefore, the true OD snapshot for interval  $t$  can only be obtained after all those passengers entered at interval  $t$  have reached their destinations; it cannot be observed in real-time (i.e., the delayed data availability). In other words, we often do not have access to  $\mathbf{f}_t$  when forecasting  $\mathbf{f}_{t+1}$ . In contrast, the boarding (entering) flow—another important quantity—is observable in real-time. We denote by  $b_{t,i}$  the number of passengers entering station  $i$  at interval  $t$ . In fact, we have  $b_{t,i} = \sum_j o_{t,i,j}$ . We define a *boarding snapshot* as a vector  $\mathbf{b}_t = [b_{t,1}, b_{t,2}, \dots, b_{t,s}]^\top$ .

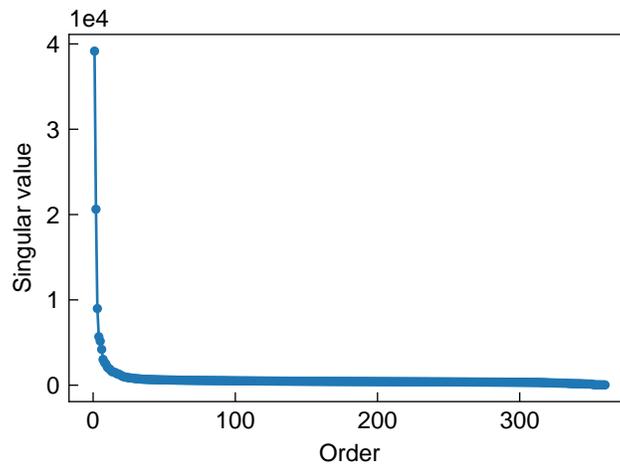
The OD matrices/flow forecasting problem is to forecast future OD snapshots  $\mathbf{f}_{t+1}, \mathbf{f}_{t+2}, \dots, \mathbf{f}_{t+L}$  given a sequence of available historical OD snapshots  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_t$  and boarding snapshots  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_t$ . The reason for using boarding snapshots is to compensate for the delayed data availability problem of recent OD snapshots.

## 5.5 Dynamic Mode Decomposition

Dynamic Mode Decomposition (DMD, [Schmid, 2010](#)) is developed by the fluid dynamic community to extract dynamic features from high-dimensional data. To better illustrate our forecasting model, we briefly introduce DMD in this section.

Consider using a linear dynamical system  $\mathbf{f}_i \approx A\mathbf{f}_{i-1}$  for OD flow forecasting. Similar to many fluid problems,  $n$  is huge for an OD snapshot and even storing  $A \in \mathbb{R}^{n \times n}$  can be prohibitive. Therefore, DMD outputs the (leading) eigenvalues and eigenvectors of  $A$  without calculating the expensive  $A$ . The eigenvectors of  $A$  are referred to as the DMD modes and have clear physical meaning. Each DMD mode is associated with an oscillation frequency and a decay/growth rate determined by its eigenvalue. DMD is also connected to Koopman theory and can model complex non-linear systems by constructing proper measurements ([Rowley et al., 2009](#)). There are many variant algorithms for DMD. We only present the *exact DMD* proposed by [Tu et al. \(2014\)](#), which is closely related to this paper.

We arrange OD snapshots into  $m$ -column matrices  $Y_i = [\mathbf{f}_{i-m+1}, \mathbf{f}_{i-m+2}, \dots, \mathbf{f}_i] \in \mathbb{R}^{n \times m}$ . Typically,  $m \ll n$ . The linear dynamical system follows  $Y_t \approx AY_{t-1}$ . The exact DMD seeks



**Figure 5.1:** The singular values of a ten-day-length  $Y_{t-1}$  collected from Guangzhou metro smart card system.

the leading eigenvalues and eigenvectors of the best-fit linear operator  $A$  by the following procedure.

1. Compute the truncated singular value decomposition (SVD) of  $Y_{t-1} \approx U\Sigma V^T$ , where  $U \in \mathbb{R}^{n \times r}$ ,  $\Sigma \in \mathbb{R}^{r \times r}$ , and  $V \in \mathbb{R}^{m \times r}$  and  $r \ll m$ .
2. Instead of computing the full matrix  $A = Y_t Y_{t-1}^+ \approx Y_t V \Sigma^{-1} U^T$ .<sup>1</sup> We define a reduced matrix  $\tilde{A} = U^T A U \approx U^T Y_t V \Sigma^{-1} \in \mathbb{R}^{r \times r}$ . It can be proved that  $\tilde{A}$  and  $A$  have the same nonzero leading eigenvalues (Tu et al., 2014).
3. Compute the eigenvalue decomposition  $\tilde{A}W = W\Lambda$ . The entries of the diagonal matrix  $\Lambda$  are also the eigenvalues of the full matrix  $A$ .
4. The DMD modes (eigenvectors of  $A$ ) can be obtained by  $\Phi = Y_t V \Sigma^{-1} W$ .

Figure 5.1 shows the singular values of a ten-day-length  $Y_{t-1}$  from the Guangzhou metro system. A few leading singular values explain a significant portion of the variance, confirming the low-rank feature of OD snapshots data. DMD-based model can thus greatly reduce the dimensionality/complexity of such a dynamic system. However, the exact DMD has some limitations for the OD flow forecasting problem. Firstly, the complex temporal correlation of OD flow cannot be well captured by a linear dynamical system. Moreover, using the last OD snapshot is impractical since OD snapshots cannot be observed in real-time. To address these problems, we propose our solution in the next section.

<sup>1</sup> $(\cdot)^+$  denotes the Moore-Penrose inverse of a matrix.

## 5.6 High-order Weighted Dynamic Mode Decomposition

### 5.6.1 Model specification

The forecasting formula of an exact DMD amounts to a high-dimensional vector autoregression of order 1. However, the latest OD snapshots are unknown at the time of forecasting. Therefore, we use the two latest snapshots of the boarding flow as a replacement. We regard OD snapshots of three or more intervals ago as available; because we find more than 96% trips in our data set are completed within one hour (two lags). And we can use a high-order vector autoregression to capture the long-term correlations in OD snapshots. The forecasting model follows

$$\mathbf{f}_i \approx A_{t,1}\mathbf{f}_{i-q_1} + A_{t,2}\mathbf{f}_{i-q_2} + \cdots + A_{t,h}\mathbf{f}_{i-q_h} + A_{t,b1}\mathbf{b}_{i-1} + A_{t,b2}\mathbf{b}_{i-2} \quad \forall i \in \{q_h + 1, q_h + 2, \dots, t\}, \quad (5.1)$$

where time lags for OD snapshots are positive integers satisfying  $3 \leq q_1 < \cdots < q_h < t$ . Note that coefficient matrices  $A_{t,1}, \dots, A_{t,h} \in \mathbb{R}^{n \times n}$  and  $A_{t,b1}, A_{t,b2} \in \mathbb{R}^{n \times s}$  are estimated using the data up to the latest ( $t$ -th) time interval; they are re-estimated when new data become available. This allows model coefficients to be time-varying. We will introduce how to update coefficient matrices using new observations without storing historical data in Section 5.6.3.

To express Eq. (5.1) in a concise matrix form, let  $Y_i = [\mathbf{f}_{i-m+1}, \mathbf{f}_{i-m+2}, \dots, \mathbf{f}_i]$  and  $B_i = [\mathbf{b}_{i-m+1}, \mathbf{b}_{i-m+2}, \dots, \mathbf{b}_i]$ , where  $m = t - q_h$  is the number of target snapshots. Then, Eq. (5.1) is equivalent to

$$Y_t \approx A_{t,1}Y_{t-q_1} + A_{t,2}Y_{t-q_2} + \cdots + A_{t,h}Y_{t-q_h} + A_{t,b1}B_{t-1} + A_{t,b2}B_{t-2} \quad (5.2)$$

$$= [A_{t,1}, A_{t,2}, \dots, A_{t,h}, A_{t,b1}, A_{t,b2}] \begin{bmatrix} Y_{t-q_1} \\ Y_{t-q_2} \\ \vdots \\ Y_{t-q_h} \\ B_{t-1} \\ B_{t-2} \end{bmatrix} \quad (5.3)$$

$$= G_t X_t, \quad (5.4)$$

where  $G_t \in \mathbb{R}^{n \times (hn+2s)}$  and  $X_t \in \mathbb{R}^{(hn+2s) \times m}$  are augmented matrices for coefficients and data, respectively. Note that with this approach, we model forecasting as a regression problem without considering the inter-sequence dependence.

We next introduce a forgetting ratio  $\rho$  ( $0 < \rho \leq 1$ ) that assigns small weights on snapshots to past days. This is because the dynamics of the system may change over time and we prefer to use the most recent dynamics to achieve accurate forecasting. The matrix  $G_t$  can be solved by the following optimization problem

$$\min_{G_t} \sum_{i=1}^m \rho^{\text{day}(m)-\text{day}(i)} \|\mathbf{y}_i - G_t \mathbf{x}_i\|_F^2, \quad (5.5)$$

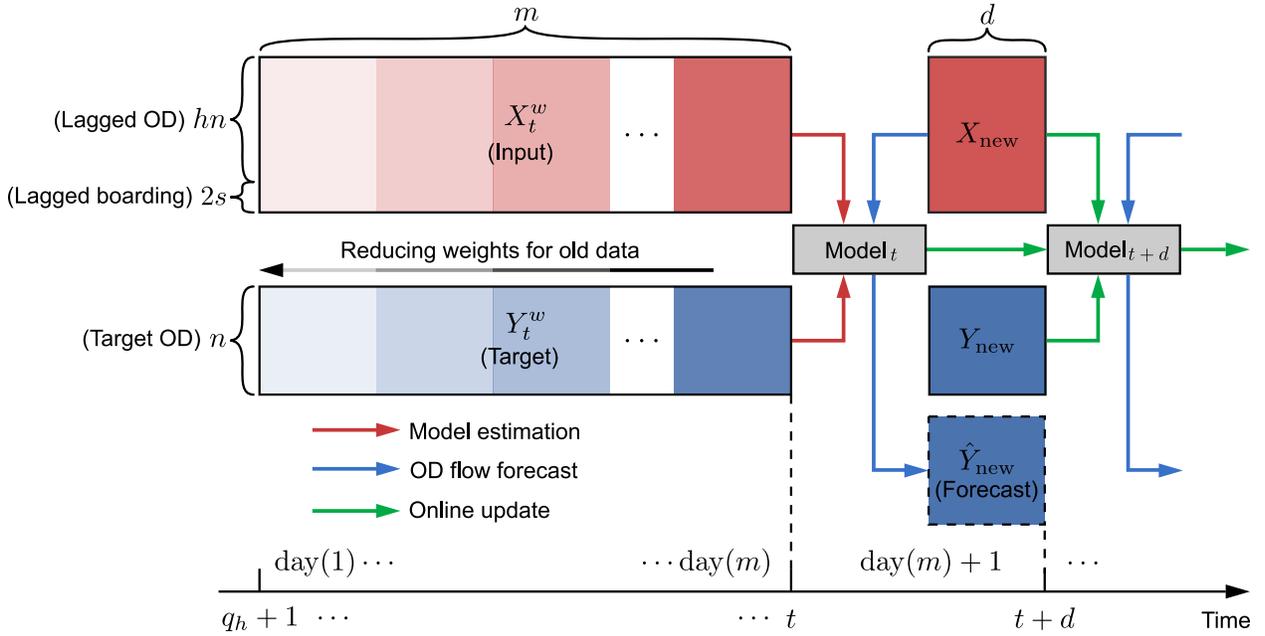
where  $\mathbf{y}_i$  and  $\mathbf{x}_i$  are the  $i$ -th column of  $Y_t$  and  $X_t$ , respectively;  $\text{day}(i)$  represents the day of the snapshot  $\mathbf{y}_i$ . We assign the same weight for snapshots of the same day. The weight  $\rho^{\text{day}(m)-\text{day}(m)} = \rho^0 = 1$  for the latest OD snapshot. For a snapshot in  $j$  days ago, the weight is  $\rho^j$ , which decreases exponentially. This weighting idea is similar to works by [Alfatlawi and Srivastava \(2020\)](#), [Zhang et al. \(2019a\)](#), and [Kwak and Geroliminis \(2020\)](#). For convenience, we define  $\sigma = \sqrt{\rho}$  and the weighted version of  $Y_t$  and  $X_t$  as

$$\begin{aligned} Y_t^w &= [\sigma^{\text{day}(m)-\text{day}(1)} \mathbf{y}_1, \sigma^{\text{day}(m)-\text{day}(2)} \mathbf{y}_2, \dots, \mathbf{y}_m], \\ X_t^w &= [\sigma^{\text{day}(m)-\text{day}(1)} \mathbf{x}_1, \sigma^{\text{day}(m)-\text{day}(2)} \mathbf{x}_2, \dots, \mathbf{x}_m]. \end{aligned}$$

Then, the optimization problem in Eq. (5.5) becomes an ordinary least squares problem

$$\min_{G_t} \|Y_t^w - G_t X_t^w\|_F^2. \quad (5.6)$$

Figure 5.2 summarizes the overall structure of the proposed higher-order weighted DMD (HW-DMD) framework. The underlying forecasting model is a high-order vector autoregression with the boarding flow as extra inputs. A forgetting ratio is introduced to decrease the weights of past data exponentially on a daily basis. In Section 5.6.2, we will introduce a dimensionality reduction technique based on DMD to find a low-rank solution for this large model (w.r.t. number of parameters). Instead of full matrices  $A_{t,(\cdot)}$ , we seek  $\tilde{A}_{t,(\cdot)}$ —much smaller matrices—to capture the system’s dynamic. Finally, an online update method is proposed in Section 5.6.3 to update the model coefficients incrementally without storing historical data. This provides a memory-saving solution that maintains an up-to-date model. Note that the same model framework can be easily extended to incorporate higher-order boarding flow or other external covariates (e.g., days of the week, alighting flow, holidays). For example, we can represent days of the week by one-hot encoding  $\mathbf{w}_i \in \mathbb{R}^{7 \times 1}$  and add an additional regression term  $A_{t,w} \mathbf{w}_i$  to Eq.(5.1) to incorporate the weekly pattern. This paper only presents the model specified in Eq. (5.1) for illustration.



**Figure 5.2:** Model framework for HW-DMD. Model input  $X$  contains  $hn$  rows for lagged OD snapshots and  $2s$  rows for lagged boarding snapshots. Columns in  $Y$  and  $\hat{Y}$  are, respectively, real and forecasted snapshots for OD flow. Model coefficients are estimated by weighted historical data ( $X_t^w$  and  $Y_t^w$ ) and updated daily whenever new data come.

## 5.6.2 Model estimation

We prefer a low-rank approximation of  $G_t$  over a full matrix of the optimal solution of Eq. (5.6). This is because storing the large full matrix is prohibitive, and the optimal solution often leads to overfitting problems, especially for the sparse and noisy OD data. Luckily, we can find a pretty good approximation thanks to the inherent low-rank nature of OD data.

Similar to the exact DMD, we first compute the truncated SVD on the weighted augmented data matrix  $X_t^w \approx U_X \Sigma_X V_X^\top$ , where we keep the  $r_X$  ( $r_X \ll m$ ) largest singular values and  $U_X \in \mathbb{R}^{(hn+2s) \times r_X}$ ,  $\Sigma_X \in \mathbb{R}^{r_X \times r_X}$ ,  $V_X \in \mathbb{R}^{m \times r_X}$ . As shown in Figure 5.1, a few leading singular values can well capture the entire data. Therefore, an approximation for coefficient matrices is

$$G_t = Y_t^w X_t^{w+} \approx Y_t^w V_X \Sigma_X^{-1} U_X^\top, \quad (5.7)$$

$$[A_{t,1}, \dots, A_{t,h}, A_{t,b1}, A_{t,b2}] \approx [Y_t^w V_X \Sigma_X^{-1} U_{X,1}^\top, \dots, Y_t^w V_X \Sigma_X^{-1} U_{X,h}^\top, Y_t^w V_X \Sigma_X^{-1} U_{X,b1}^\top, Y_t^w V_X \Sigma_X^{-1} U_{X,b2}^\top], \quad (5.8)$$

where  $U_X^\top = [U_{X,1}^\top, \dots, U_{X,h}^\top, U_{X,b1}^\top, U_{X,b2}^\top]$ ,  $U_{X,1}, \dots, U_{X,h} \in \mathbb{R}^{n \times r_X}$ , and  $U_{X,b1}, U_{X,b2} \in \mathbb{R}^{s \times r_X}$ . This step uses the result from a truncated SVD to replace the original  $X_t^w$ , which reduces the impact of the noise in the data.

The results computed from Eq. (5.7) and Eq. (5.8) are still prohibitive. Therefore, for each column in  $Y_t^w$ , we seek a transformation  $\mathbf{y}_i^w \rightarrow \tilde{\mathbf{y}}_i^w$  such that  $\tilde{\mathbf{y}}_i^w \in \mathbb{R}^{r_Y}$  with  $r_Y \ll n$ . In doing so, we compute another rank- $r_Y$  truncated SVD of the target matrix  $Y_t^w \approx U_Y \Sigma_Y V_Y^\top$ . The columns of  $U_Y$  form an orthonormal basis; thus, the transformation  $\tilde{\mathbf{y}}_i^w = U_Y^\top \mathbf{y}_i^w$  compute the coordinates of  $\mathbf{y}_i^w$  on this basis, which compresses  $y_i^w$  from  $\mathbb{R}^n$  to  $\mathbb{R}^{r_Y}$ . We can project coefficient matrices onto the same basis  $U_Y$  to greatly reduce the dimensionality:

$$\tilde{A}_{t,i} = U_Y^\top A_{t,i} U_Y \approx U_Y^\top Y_t^w V_X \Sigma_X^{-1} U_{X,i}^\top U_Y, \quad \forall i \in \{1, 2, \dots, h\}, \quad (5.9)$$

$$\tilde{A}_{t,bj} = U_Y^\top A_{t,bj} \approx U_Y^\top Y_t^w V_X \Sigma_X^{-1} U_{X,bj}^\top, \quad \forall j \in \{1, 2\}, \quad (5.10)$$

where  $\tilde{A}_{t,i} \in \mathbb{R}^{r_Y \times r_Y}$  and  $\tilde{A}_{t,bj} \in \mathbb{R}^{r_Y \times s}$ . Finally, we can write the model of Eq. (5.2) in the reduced-order subspace

$$\tilde{Y}_t \approx \tilde{A}_{t,1} \tilde{Y}_{t-q_1} + \tilde{A}_{t,2} \tilde{Y}_{t-q_2} + \dots + \tilde{A}_{t,h} \tilde{Y}_{t-q_h} + \tilde{A}_{t,b1} B_{t-1} + \tilde{A}_{t,b1} B_{t-2},$$

where  $\tilde{Y}_i = U_Y^\top Y_i$ . The final forecast of an OD snapshot  $\hat{\mathbf{y}}_i$  can be calculated by transforming back to the original basis by  $\hat{\mathbf{y}}_i = U_Y \tilde{\mathbf{y}}_i$ . With the reduced coefficient matrices  $\tilde{A}_{t,(\cdot)}$  and projection bases  $U_Y$ , we avoid calculating and storing the giant coefficient matrices  $A_{t,(\cdot)}$ .

DMD-based estimation is different from common dimensionality reduction techniques in several ways. For many matrix-factorization-based models and dynamic factor models, a forecasting model is estimated after performing dimensionality reduction (e.g., Ren and Xie, 2017), or latent factors are constructed by keeping the most temporal dynamics (e.g., Forni et al., 2000; Lam et al., 2011; Yu et al., 2016); the forecast ability is designed on the latent (size-reduced) data for these models. In contrast, DMD-based methods first estimate a forecasting model by a least-square fit of rank-reduced full-size data (i.e., Eq. (5.7)), next reduce the dimensionality of the linear operator by projecting to leading SVD modes (i.e., Eq. (5.9)–(5.10)); the resulting linear operator captures the dynamics of the rank-reduced full-sized data. Although the forecast value  $\hat{\mathbf{y}}_i$  by an HW-DMD is restricted on the column space of  $U_Y$ , it is already the best approximation in  $\mathbb{R}^{r_Y}$  (in terms of Frobenius norm (Eckart and Young, 1936)) because the basis is determined by leading singular vectors. Besides, the rank truncation for the data also eases the noise and the overfitting problem. As noted by Schmid (2010), accurate identification of more than the first couple modes can be difficult on noisy data sets without this truncation step.

The major computational cost in parameter estimation of HW-DMD is the SVD part. Current numerical software can solve large-scale SVD very efficiently. Therefore, estimating the HW-DMD model is very fast. We can further derive the eigenvalues and eigenvectors of coefficient matrices  $A_{t,i}$  (Proctor et al., 2016). But this step is not necessary for our task, since they are not used to generate the forecast and there is no clear physical meaning for eigenvectors in a high-order vector autoregression.

### 5.6.3 Online update

A model trained by dated data may not reflect the recent dynamic in a system. Instead of retraining using entire data, we develop an online algorithm that updates HW-DMD day by day with new observations without storing historical data, as shown in Figure 5.2. Similar algorithms for online DMD have been developed by Hemati et al. (2014), Zhang et al. (2019a), and Alfatlawi and Srivastava (2020). We extend the online DMD update algorithm to a high-order weighted version.

To illustrate the update algorithm, we need to reorganize Eq. (5.7)–(5.10). Let  $\tilde{X}_i^w = U_X^\top X_i^w$  and  $\tilde{Y}_i^w = U_Y^\top Y_i^w$  be the projection of data to the coordinates of  $U_X$  and  $U_Y$ , respectively. Using the fact  $(U_X \tilde{X}_t^w)^+ = V_X \Sigma_X^{-1} U_X^\top$ , we can rewrite Eq. (5.7) as

$$\begin{aligned} G_t &\approx Y_t^w (U_X \tilde{X}_t^w)^+ \\ &= Y_t^w \tilde{X}_t^{w\top} \left( \tilde{X}_t^w \tilde{X}_t^{w\top} \right)^+ U_X^\top. \end{aligned}$$

Therefore, Eq. (5.9) and (5.10) becomes

$$\tilde{A}_{t,i} \approx \tilde{Y}_t^w \tilde{X}_t^{w\top} \left( \tilde{X}_t^w \tilde{X}_t^{w\top} \right)^+ U_{X,i}^\top U_Y = P Q_X^+ U_{X,i}^\top U_Y \quad \forall i \in \{1, \dots, h\}, \quad (5.11)$$

$$\tilde{A}_{t,bj} \approx \tilde{Y}_t^w \tilde{X}_t^{w\top} \left( \tilde{X}_t^w \tilde{X}_t^{w\top} \right)^+ U_{X,bj}^\top = P Q_X^+ U_{X,bj}^\top \quad \forall j \in \{1, 2\}, \quad (5.12)$$

where  $P = \tilde{Y}_t^w \tilde{X}_t^{w\top} \in \mathbb{R}^{r_Y \times r_X}$  and  $Q_X = \tilde{X}_t^w \tilde{X}_t^{w\top} \in \mathbb{R}^{r_X \times r_X}$ .

To facilitate the online update, we define an additional matrix  $Q_Y = \tilde{Y}_t^w \tilde{Y}_t^{w\top} \in \mathbb{R}^{r_Y \times r_Y}$ . After the reorganization, model coefficients are represented by three “core” matrices  $P$ ,  $Q_X$ ,  $Q_Y$  and two projection matrices  $U_X$ ,  $U_Y$ . Note these matrices are also time-varying. For simplicity, we omit the  $t$  subscript and regard they are always “up-to-date”. Moreover, there are two important properties for the core matrices.

**Theorem 1.** *Given new observations  $Y_{\text{new}} \in \mathbb{R}^{n \times d}$  and  $X_{\text{new}} \in \mathbb{R}^{(hn+2s) \times d}$  from a new day, where  $d$  is the number of snapshots per day. Under the same projection matrices, the new core matrices*

can be updated by

$$P \leftarrow \rho P + \tilde{Y}_{\text{new}} \tilde{X}_{\text{new}}^\top, \quad (5.13)$$

$$Q_X \leftarrow \rho Q_X + \tilde{X}_{\text{new}} \tilde{X}_{\text{new}}^\top, \quad (5.14)$$

$$Q_Y \leftarrow \rho Q_Y + \tilde{Y}_{\text{new}} \tilde{Y}_{\text{new}}^\top, \quad (5.15)$$

where  $\tilde{X}_{\text{new}} = U_X^\top X_{\text{new}}$  and  $\tilde{Y}_{\text{new}} = U_Y^\top Y_{\text{new}}$ .

*Proof.* Given new observations  $Y_{\text{new}} \in \mathbb{R}^{n \times d}$  and  $X_{\text{new}} \in \mathbb{R}^{(hn+2s) \times d}$  from a new day, Under the same projection matrices, the new core matrix  $P$  can be computed by

$$\begin{aligned} \tilde{Y}_{t+d}^w \tilde{X}_{t+d}^{w\top} &= [\sigma \tilde{Y}_t^w, U_Y^\top Y_{\text{new}}] [\sigma \tilde{X}_t^w, U_X^\top X_{\text{new}}]^\top \\ &= [\sigma \tilde{Y}_t^w, \tilde{Y}_{\text{new}}] [\sigma \tilde{X}_t^w, \tilde{X}_{\text{new}}]^\top \\ &= \sigma^2 \tilde{Y}_t^w \tilde{X}_t^{w\top} + \tilde{Y}_{\text{new}} \tilde{X}_{\text{new}}^\top \\ &= \rho P + \tilde{Y}_{\text{new}} \tilde{X}_{\text{new}}^\top. \end{aligned}$$

Therefore,  $P$  can be updated by  $P \leftarrow \rho P + \tilde{Y}_{\text{new}} \tilde{X}_{\text{new}}^\top$ . Similar proof applies to  $Q_X$  and  $Q_Y$ .  $\square$

**Theorem 2.** Denote by  $\tilde{Y}_t^w = U_Y \tilde{Y}_t^w$  the recovered data from the reduced data. If  $\mathbf{v}_i$  is the  $i$ -th eigenvector of  $Q_Y$ , then  $U_Y \mathbf{v}_i$  is the  $i$ -th left singular vector of  $\tilde{Y}_t^w$ . The same property applies to  $Q_X$  and  $\tilde{X}_t^w = U_X \tilde{X}_t^w$ .

*Proof.* Compute SVD  $\tilde{Y}_t^w = \tilde{U} \tilde{\Sigma} \tilde{V}^\top$ , then

$$\tilde{Y}_t^w \tilde{Y}_t^{w\top} = \tilde{U} \tilde{\Sigma} \tilde{V}^\top \tilde{V} \tilde{\Sigma}^\top \tilde{U}^\top = \tilde{U} (\tilde{\Sigma} \tilde{\Sigma}) \tilde{U}^\top, \quad (5.16)$$

$$\left( \tilde{Y}_t^w \tilde{Y}_t^{w\top} \right) \tilde{U} = \tilde{U} (\tilde{\Sigma} \tilde{\Sigma}) = \tilde{U} \tilde{\Lambda}. \quad (5.17)$$

Therefore, columns of  $\tilde{U}$  are the eigenvectors of  $\tilde{Y}_t^w \tilde{Y}_t^{w\top}$  and the left singular vectors of  $\tilde{Y}_t^w$ . Substitute  $\tilde{Y}_t^w \tilde{Y}_t^{w\top} = U_Y Q_Y U_Y^\top$  to Eq. (5.17), we have

$$\begin{aligned} \left( U_Y Q_Y U_Y^\top \right) \tilde{U} &= \tilde{U} \tilde{\Lambda}, \\ Q_Y \left( U_Y^\top \tilde{U} \right) &= \left( U_Y^\top \tilde{U} \right) \tilde{\Lambda}. \end{aligned}$$

Define  $V = U_Y^\top \tilde{U}$ . Then, each column  $\mathbf{v}_i$  in  $V$  is a eigenvector for  $Q_Y$  and  $U_Y \mathbf{v}_i = U_Y (U_Y^\top \tilde{\mathbf{u}}_i) = \tilde{\mathbf{u}}_i$  is a singular vector of  $\tilde{Y}_t^w$ .  $\square$

Theorem 1 is used to update the core matrices in a memory-saving way. Theorem 2 indicates we can use the eigenvectors of  $Q_Y$  to approximate the left singular vectors of  $Y_t^w$  (because  $Y_t^w \approx \bar{Y}_t^w$ ), which is crucial for updating the projection matrices. Based on these properties, we summarize the online update algorithm in the following three steps.

1. **Expand projection matrices.** Let  $E_Y = Y_{\text{new}} - U_Y U_Y^\top Y_{\text{new}}$  and  $E_X = X_{\text{new}} - U_X U_X^\top X_{\text{new}}$  be the residuals that cannot be represented by the column space of  $U_X$  and  $U_Y$ . To incorporate these residuals, we expand projection matrices by  $U_X \leftarrow [U_X, U_{E_X}]$  and  $U_Y \leftarrow [U_Y, U_{E_Y}]$ , where  $U_{E_X}$  and  $U_{E_Y}$  are the orthonormal bases (obtained by SVD or QR factorization) of  $E_X$  and  $E_Y$ , respectively.
2. **Update core matrices.** To align dimensions, we first pad  $P$ ,  $Q_X$ , and  $Q_Y$  with zeros on the dimensions where  $U_X$  and  $U_Y$  expanded. Then update core matrices by Eq. (5.13)–(5.15).
3. **Compression.** The first two steps incorporate all new information at the cost of expanding dimensions. Next, we compress the model based on Theorem 2. Denote  $V_X$  and  $V_Y$  to be matrices composed by the leading  $r_X$  and  $r_Y$  eigenvectors of  $Q_X$  and  $Q_Y$ , respectively. We can compress projection matrices by  $U_X \leftarrow U_X V_X$ ,  $U_Y \leftarrow U_Y V_Y$  to keep the leading singular vectors of  $\bar{X}_{t+d}^w$  and  $\bar{Y}_{t+d}^w$ . The core matrices can be compressed accordingly by  $Q_X \leftarrow V_X^\top Q_X V_X$ ,  $Q_Y \leftarrow V_Y^\top Q_Y V_Y$ ,  $P \leftarrow V_Y^\top P V_X$ .

Besides the daily update, a more general setting can be updating the model for every  $k$  intervals or only doing the compression step when  $r_X$  or  $r_Y$  exceeds a threshold. This paper adopts the daily update described above because metro systems often have a one-day periodicity. In terms of computational efficiency, the online update algorithm computes the SVD for  $d$ -column data matrices and eigenvalue decomposition of  $Q_X$  and  $Q_Y$ . The computation has a constant cost every day and it is significantly faster than retraining using entire data. In terms of memory efficiency, historical data are not required when updating the model. All we need to store are three “core” matrices and two projection matrices. Regarding the error, the online algorithm does not take into account the previously truncated part. This impact is negligible because the truncated part contains mostly noise, and past data are forgotten exponentially. Our experiments in section 5.7.6 show the online algorithm performs pretty close to or even slightly better than retraining.

### 5.6.4 Connections with other DMD models

The proposed HW-DMD is closely related to Hankel-DMD (Brunton et al., 2017; Arbabi and Mezić, 2017; Avila and Mezić, 2020) and DMD with control (DMDc, Proctor et al., 2016). Hankel-DMD uses Hankel data matrices as input and output to model a non-linear dynamical system by a linear model; its DMD modes approximate to the Koopman modes. There is another model also named Higher Order DMD (HODMD, Le Clainche and Vega, 2017), which requires Hankelizing data in its estimation and is essentially similar to Hankel-DMD. Instead, the proposed HW-DMD uses raw snapshots as the output (the left side of Eq. (5.2)) without using the Hankel structure. This formula is equivalent to a high-order vector autoregression model, which is neater and more suitable in the context of forecasting. Moreover, our model can use non-continuous orders and external variables (e.g., the boarding flow). Essentially, the external variables of our model can be regarded as the control term of a DMDc model.

The three-step online update algorithm for HW-DMD in this paper inherits from the work of Hemati et al. (2014). The original algorithm was developed for the exact DMD introduced in Section 5.5. Besides, the online DMD proposed by (Zhang et al., 2019a) considers the decaying weight of data, but the constant projection matrix in their assumption restricts the update effect. (Alfatlawi and Srivastava, 2020) proposed an online algorithm for weighted DMD using incremental SVD, which is a different technique from our method. Our contribution is extending the algorithm proposed by Hemati et al. (2014) to a high-order weighted version with the consideration of external regression covariates.

## 5.7 Experiments

In this section, we compare the proposed HW-DMD with other forecasting models using real-world data. We begin with an introduction to data and experimental settings. Next, we compare model performances by forecasting the OD matrices and the boarding flow derived from the OD matrices. Finally, we examine the long-term effect of the online HW-DMD update algorithm. The code for experiments is available from <https://github.com/mcgill-smart-transport/high-order-weighted-DMD>.

### 5.7.1 Data and experimental settings

We examine HW-DMD using the metro smart card data from two cities, Guangzhou and Hangzhou. Both data sets record the origin, destination, and entry and exit time of each

metro trip. We focus on the forecast of workdays and connect each Friday to the next Monday. Details of the two data sets are as follows:

- Guangzhou metro data: This data set covers around 301 million trips among 159 metro stations in Guangzhou from July 1st to Sept 30th, 2017. Guangzhou metro operates from 6:00 to 24:00. We use the first twenty weekdays (July 3rd to July 28th) as the training set, the following ten weekdays (July 31st to Aug 11th) as the validation set, and the following ten weekdays (Aug 14th to Aug 25th) as the test set. There are additional one-month data after the test set; we use these data to study the long-term effect of the online HW-DMD update algorithm.
- Hangzhou metro data<sup>2</sup>: This is an open data set that covers 80 effective stations of Hangzhou metro from Jan 1st to Jan 25th, 2019. The operation hours are from 5:30 to 23:30. We use the first ten weekdays (Jan 1 to Jan 14) for training, the following four weekdays (Jan 15 to Jan 18) for validation, and the rest five weekdays (Jan 21 to Jan 25) for testing.

We aggregate OD snapshots by a 30-minute time interval, which means 36 snapshots per day for both cities. Note that a small interval may result in sparse OD matrices; we choose the 30-minute interval to balance the practical requirements. Figure 5.3 shows the distribution of  $o_{i,j}$  from an OD snapshot of a typical morning peak in Guangzhou. The distribution roughly follows a power law, with most OD pairs having small volumes while a few of them are significantly larger. The highly skewed distribution is very difficult to be properly handled by conventional forecasting models.

The performance of a model is quantified using the root-mean-square error (RMSE), the weighted mean absolute percentage error (WMAPE), and the coefficient of determination (denoted as  $R^2$ ):

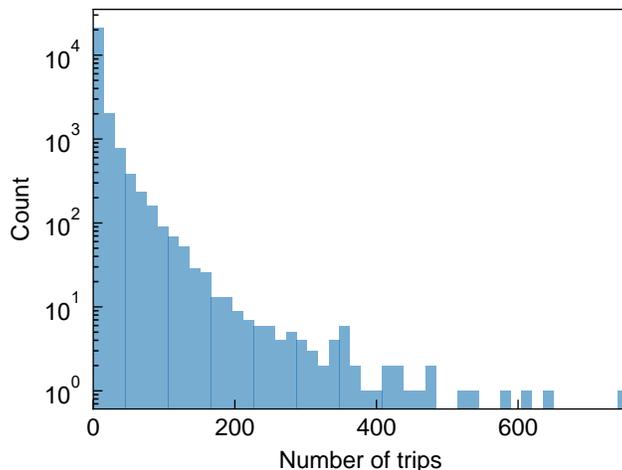
$$\text{RMSE}(\alpha, \hat{\alpha}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\alpha_i - \hat{\alpha}_i)^2},$$

$$\text{WMAPE}(\alpha, \hat{\alpha}) = \frac{\sum_{i=1}^N |\alpha_i - \hat{\alpha}_i|}{\sum_{i=1}^N |\alpha_i|} \times 100\%,$$

$$R^2(\alpha, \hat{\alpha}) = 1 - \frac{\sum_{i=1}^N (\alpha_i - \hat{\alpha}_i)^2}{\sum_{i=1}^N (\alpha_i - \bar{\alpha})^2},$$

where  $\alpha$  and  $\hat{\alpha}$  are, respectively, the real and predicted values;  $\bar{\alpha}$  is the average value of  $\alpha$ ;  $N$  is the total number of elements under different time intervals and locations. The three

<sup>2</sup><https://doi.org/10.5281/zenodo.3145404>



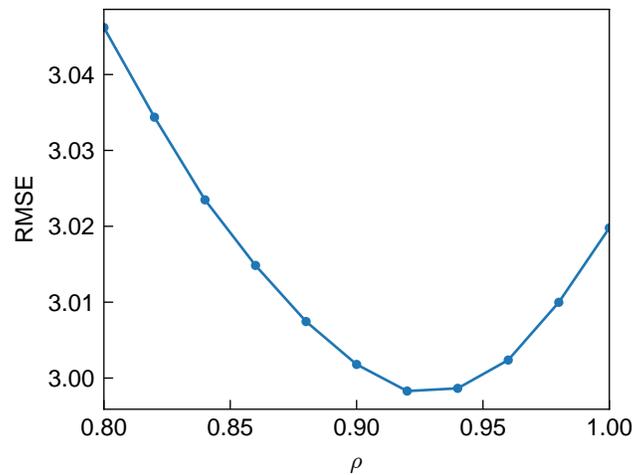
**Figure 5.3:** The histogram of  $o_{ij}$  in an OD snapshot of a morning peak in Guangzhou.

performance metrics are computed for both OD flow  $o$  and boarding flow  $b$  (forecasted by  $\hat{b}_{t,i} = \sum_j \hat{\delta}_{t,i,j}$ ).

### 5.7.2 Hyperparameters

We use the online update algorithm for HW-DMD if not otherwise specified. Hyperparameters for HW-DMD include time lags  $q_1, \dots, q_h$ , the SVD truncation rank  $r_X, r_Y$ , and the forgetting ratio  $\rho$ . These parameters are determined in a sequential order.

We use the Guangzhou data set as an example to elaborate the hyperparameter tuning procedure. We first set  $r_X = r_Y = 100$  and  $\rho = 1$  and select time lags in a greedy manner. For time lags within one day ( $3 \leq q_i \leq 36$ ), we repeatedly add a “currently best” time lag based on the RMSE of the validation set until a new lag brings no improvement or the number of lags reaches ten. This procedure selects  $\{3, 4, 8, 14, 19, 28, 30, 33, 35, 36\}$  as time lags. The considerable high-order time lags in the result indicate long-term auto-correlations of OD time series. For example, the lag 19 roughly equals a typical work duration (9.5 hours), which can be explained as a strong correlation between the departure trips for commuters in the morning and the returning trips in the afternoon (Cheng et al., 2021a). The metro OD flow is also highly regular; the largest several lags (e.g., 33, 35, and 36) capture the one-day periodicity. Next, we determine  $r_X$  and  $r_Y$  by a grid search from 20 to 100 at an interval of 10. The best  $r_Y$  is 50. A larger  $r_X$  than 100 still brings a marginal improvement, but we truncate  $r_X$  at 100 to restrict the model size ( $r_X$  affects the size of  $U_X$  in the online update). Lastly, we set  $\rho$  to be 0.92 based on a line search from 0.8 to 1. As



**Figure 5.4:** The effect of  $\rho$  to the forecast OD RMSE in the validation set of the Guangzhou data.

shown in Figure 5.4, we can see assigning smaller weights for old data indeed improves the forecast. Because  $0.92^8 \approx 0.51$ , using  $\rho = 0.92$  is roughly equivalent to halving the weight every eight days.

The hyperparameter tuning for the Hangzhou data set follows the same procedure. The selected hyperparameters for the Hangzhou data set are time lags={3, 4, 6, 14, 18, 19, 28, 32, 35, 36},  $r_Y = 40$ ,  $r_X = 100$ , and  $\rho = 0.92$ .

### 5.7.3 Benchmark models

We compare HW-DMD with the following benchmark models:

- HA: Historical Average. For the OD flow at a certain period (e.g., 7:00–7:30) of the day, HA uses the average OD flow at that period in the training set as the forecast value.
- TRMF: Temporal Regularized Matrix Factorization (Yu et al., 2016). TRMF is a matrix factorization model that imposes autoregression (AR) processes on each temporal factor. We use time lags  $[1, \dots, 10]$  for the AR processes. We search over  $\{100, 300, 500, 1000, 1500, 2000, 2500, 3000\}$  for the best regularization parameter and search from 30 to 150 with an interval of 10 for the best number of factors.
- ConvLSTM: Convolutional LSTM (Shi et al., 2015). It is a deep recurrent neural network model that forecasts future frames of matrix time series (e.g., videos). Here

we use it to forecast future OD matrices by the most recent ten OD matrices. Following the work by [Zhang et al. \(2021b\)](#), we apply a three-layer LSTM structure with eight, eight, and one filter, respectively, and set the kernel size to be  $3 \times 3$  for all convolutional layers in the model.

- **FNN**: A two-layer Feedforward Neural Network. We use the OD snapshots of 3-10 lags ago and the boarding flow snapshot of 1-2 lags ago as the input features. We perform a grid search over the type of activation functions (linear, sigmoid, and relu) and the number of hidden layers (from 10 to 100 at an interval of 10) for the best model setting.
- **SARIMA**: Seasonal AutoRegressive Integrated Moving Average. We only use SARIMA to forecast the boarding flow since SARIMA only handles one-dimensional time series. We use the order  $ARIMA(2, 0, 1)(1, 1, 0)$  [36] for all the stations and fit 159 separate models. This model configuration is the same as [Cheng et al. \(2021a\)](#) and was tested to be suitable for most metro stations.

Applying TRMF, ConvLSTM, and FNN to the original data (or after a normalization) can hardly obtain a forecast better than HA. This phenomenon was also found by [Gong et al. \(2018, 2020\)](#). This is because the OD data are high-dimensional, sparse, noisy, and highly skewed. To improve the forecast of these models, we apply TRMF, ConvLSTM, and FNN to the residuals after subtracting the HA from the original data. This “mean-removal” processing also weakens the data’s periodicity; therefore, we do not use seasonal lags in these models. Besides, because the standard TRMF and ConvLSTM cannot use the boarding flow as extra inputs, we ignore the delayed data availability problem for these models and assume all historical OD snapshots are available.

#### 5.7.4 Forecast result

We apply trained models to the test set and forecast OD matrices of the next three steps at each time interval. Note OD snapshots of 1-2 lags ago are unknown; they are replaced by previously forecasted OD snapshots when doing multi-step rolling forecasting by HW-DMD/FNN. Next, the boarding flow can be calculated from OD matrices. We compare the forecast accuracy of models in terms of OD flow and boarding flow.

Table 5.1 shows the results of OD flow forecast. We can see HW-DMD with a forgetting ratio  $\rho = 0.92$  outperforms other models in all evaluation metrics. Even the three-step forecast of HW-DMD is better than the one-step forecast of other models. The advantage

of HW-DMD over other models is more significant in the Hangzhou data set. Although TRMF, FNN, and ConvLSTM are trained on the residuals after subtracting the HA from the original data, the improvement of these models compared with HA is limited. In contrast, HW-DMD is directly applied to the original data but provides a significantly better forecast, demonstrating its strong prediction power in handling the sparse, noisy, and high-dimensional OD data. Besides, the performance of the “unweighted” HW-DMD ( $\rho = 1$ ) is slightly behind the weighted version, but still better than other models.

**Table 5.1:** Models’ performance for OD flow forecasting.

Method	Criterion	Guangzhou			Hangzhou		
		One-step	Two-step	Three-step	One-step	Two-step	Three-step
HW-DMD $\rho = 0.92$	RMSE	<b>3.05</b>	<b>3.09</b>	<b>3.11</b>	<b>3.36</b>	<b>3.41</b>	<b>3.44</b>
	WMAPE	<b>29.65%</b>	<b>29.77%</b>	<b>29.79%</b>	<b>31.76%</b>	<b>31.96%</b>	<b>31.84%</b>
	$R^2$	<b>0.957</b>	<b>0.956</b>	<b>0.955</b>	<b>0.934</b>	<b>0.932</b>	<b>0.931</b>
HW-DMD $\rho = 1$	RMSE	3.08	3.12	3.14	3.40	3.45	3.48
	WMAPE	29.71%	29.87%	29.91%	31.94%	32.22%	32.13%
	$R^2$	0.956	0.955	0.954	0.933	0.930	0.929
TRMF	RMSE	3.22	3.24	3.26	3.80	3.89	3.96
	WMAPE	30.61%	30.72%	30.79%	34.02%	34.48%	34.82%
	$R^2$	0.952	0.951	0.951	0.916	0.912	0.908
FNN	RMSE	3.15	3.16	3.18	3.97	4.01	4.05
	WMAPE	30.23%	30.28%	30.32%	33.58%	33.63%	33.65%
	$R^2$	0.954	0.953	0.953	0.908	0.906	0.904
Conv-LSTM	RMSE	3.25	3.26	3.27	4.04	4.06	4.08
	WMAPE	30.11%	30.18%	30.23%	32.96%	32.92%	33.04%
	$R^2$	0.951	0.950	0.950	0.905	0.904	0.903
HA	RMSE	3.43	3.43	3.43	4.34	4.34	4.34
	WMAPE	31.21%	31.21%	31.21%	34.28%	34.28%	34.28%
	$R^2$	0.945	0.945	0.945	0.890	0.890	0.890

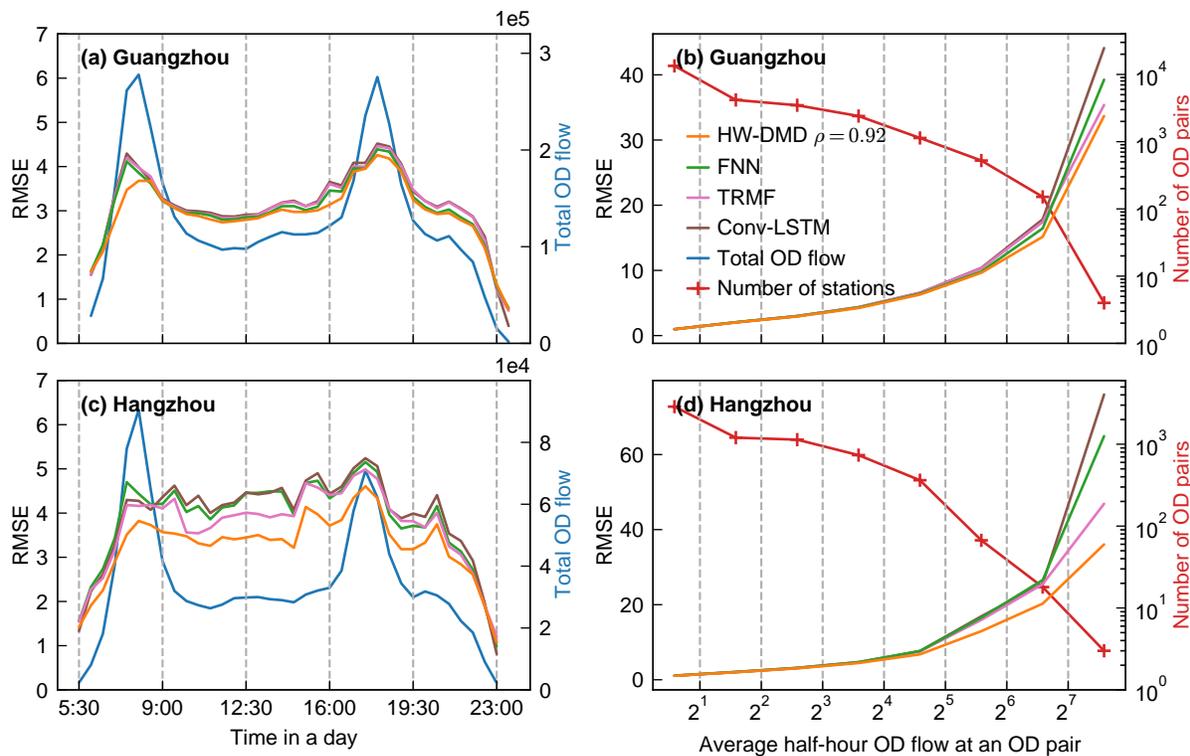
Examining the aggregated boarding flow is important because it reflects if the forecast errors in OD matrices’ are properly distributed, which is crucial when using OD matrices in traffic assignments. Moreover, the boarding flow itself is of interest to many applications. Table 5.2 shows the boarding flow forecasting; all models except SARIMA calculate boarding flow by OD matrices. The two HW-DMD models are the best models in most cases. The only exception is that FNN slightly outperforms HW-DMD for the three-step forecast of the Guangzhou data set. Importantly, HW-DMD is the only model that outperforms SARIMA, a well-established boarding flow forecasting model, in both data sets, showing that the forecast of HW-DMD accurately reflects the marginal distribution of OD matrices.

The magnitude of OD flow in a metro system varies significantly in time and space

**Table 5.2:** Models' performance for boarding flow forecasting.

Method	Criterion	Guangzhou			Hangzhou		
		One-step	Two-step	Three-step	One-step	Two-step	Three-step
HW-DMD $\rho = 0.92$	RMSE	<b>93.99</b>	102.61	107.58	<b>50.08</b>	<b>54.14</b>	<b>56.32</b>
	WMAPE	<b>6.09%</b>	<b>6.68%</b>	6.98%	<b>7.38%</b>	<b>8.05%</b>	<b>8.12%</b>
	$R^2$	<b>0.991</b>	<b>0.989</b>	0.988	<b>0.989</b>	<b>0.988</b>	<b>0.987</b>
HW-DMD $\rho = 1$	RMSE	94.51	<b>102.46</b>	106.55	51.28	55.66	58.45
	WMAPE	6.18%	6.74%	6.98%	7.54%	8.29%	8.43%
	$R^2$	0.991	0.989	0.988	0.989	0.987	0.986
TRMF	RMSE	126.03	127.87	128.65	77.70	81.19	83.12
	WMAPE	7.92%	8.07%	8.13%	10.00%	10.55%	10.81%
	$R^2$	0.983	0.983	0.983	0.975	0.972	0.971
FNN	RMSE	101.93	104.00	<b>106.06</b>	67.16	68.83	70.77
	WMAPE	6.44%	6.58%	<b>6.69%</b>	9.00%	9.22%	9.50%
	$R^2$	0.989	0.989	<b>0.988</b>	0.981	0.980	0.979
Conv-LSTM	RMSE	117.16	121.22	123.40	71.46	75.75	78.07
	WMAPE	6.87%	7.19%	7.35%	8.83%	9.63%	9.98%
	$R^2$	0.985	0.984	0.984	0.978	0.976	0.974
HA	RMSE	136.56	136.56	136.56	88.25	88.25	88.25
	WMAPE	8.38%	8.38%	8.38%	11.09%	11.09%	11.09%
	$R^2$	0.980	0.980	0.980	0.967	0.967	0.967
SARIMA	RMSE	110.23	120.60	126.52	55.59	60.97	64.66
	WMAPE	7.15%	7.65%	7.93%	7.86%	8.28%	8.50%
	$R^2$	0.987	0.985	0.983	0.987	0.984	0.982

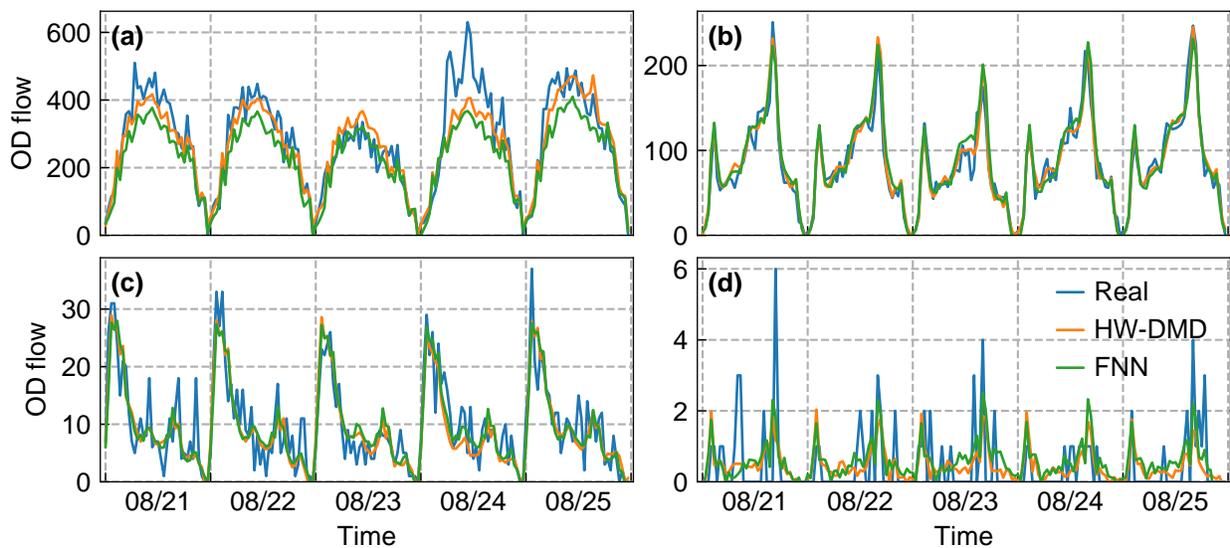
dimensions. Therefore, we further compare HW-DMD with other models under different scenarios. Figure 5.5 (a) and (c) show the forecast RMSE at different times of a day. We can see the RMSE of HW-DMD is the smallest in most time slots, particularly for the Hangzhou data set. Other models, such as Conv-LSTM, perform slightly better in the early morning and late night, but the difference is close, and the total network OD flow of these periods is pretty small. Figure 5.5 (b) and (d) show the forecast RMSE for OD pairs with different flow magnitudes. The forecast RMSE of HW-DMD is considerably lower than other models for high-flow OD pairs (average half-hour OD flow larger than  $2^4$ ). Note the number of OD pairs drops exponentially with the increase of OD flow, showing the superior forecast capability of HW-DMD for highly skewed data.



**Figure 5.5:** The RMSE of OD flow forecasting at different times and different OD pairs. (a) and (c) shows the RMSE of OD matrix forecasting at every 30-minute interval, along with the total OD flow in the network. Using  $2^i$  as boundaries, we divide OD pairs into groups according to their average half-hour OD flow; the forecast RMSE at each group and the number of OD pairs of each group are shown in (b) and (d).

Finally, we show the real and one-step forecast of OD flow at four representative OD pairs of Guangzhou metro in Figure 5.6. The OD flow exhibits a clear daily periodicity, explaining why HA already works reasonably well. Compared with FNN, HW-DMD is

better at forecasting the fluctuation of high-flow OD pairs, as shown in Figure 5.6 (a) and (b). In Figure 5.6 (a), the forecast of HW-DMD is often lower than the real value; this is hard to avoid since there is a two-lag delay when collecting the real OD flow. More OD pairs in the system are like Figure 5.6 (c) and (d) with a low flow but high noise. Under such high volatility, the forecast by HW-DMD reflects a smooth average value. In fact, the performances of other models are often undermined by noise. The SVD truncation to the data greatly enhances HW-DMD's ability in handling the noise data (Figure 5.1). Overall, HW-DMD achieves a great balance between forecasting and noise reduction, which is particularly hard for such a high-dimensional system with diverse flow magnitudes.

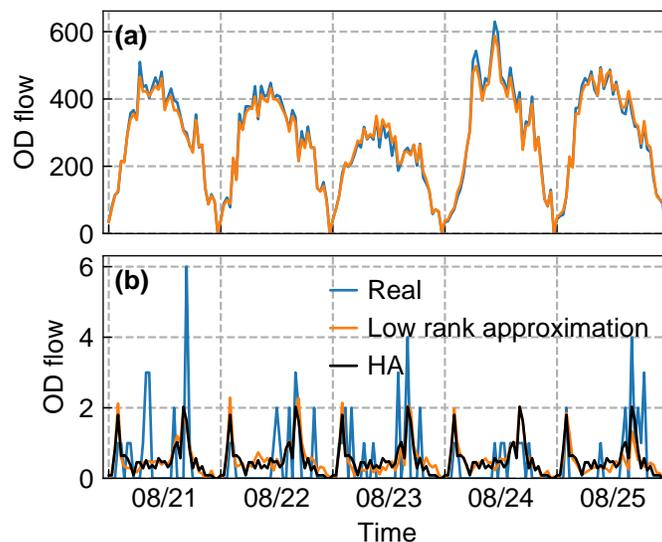


**Figure 5.6:** The real and forecasted time series of four selected OD pairs of Guangzhou metro. (a) is the busiest OD pair in the Guangzhou metro data set. (a) to (d) are in a flow decreasing order.

### 5.7.5 Effect of the low-rank assumption

The demands of majority OD pairs are small and sparse by nature, making it difficult for a forecasting model to distinguish random fluctuation (noise) and intrinsic dynamic patterns. Taking the OD pair shown in Figure 5.6 (d) as an example, the randomness in this OD pair is quite large compared with its average flow (low signal-to-noise ratio). A good forecasting should be robust to the noise while maintaining accurate cumulative effects of OD pairs in total (e.g., the boarding flow). This section evaluates the impact of using the low-rank assumption on forecasting and noise filtering.

According to Section 5.6.2, the forecast of HW-DMD is always on the column space of  $U_Y$ . Therefore, the best possible value of an OD snapshot  $\hat{y}_i$  calculated by HW-DMD is the rank-reduced full-size data, i.e.,  $U_Y U_Y^\top y_i$ , which is the upper bound of an HW-DMD's forecast ability. Figure 5.7 shows how well this low-rank approximation fits the original data. We can see the low-rank approximation keeps most information for the high-demand OD pair of Figure 5.7 (a). In contrast, most fluctuations in the sparse-demand OD pair of Figure 5.7 (b) are truncated. By comparing with HA, we can see the low-rank approximation reflects the average daily pattern of the sparse-demand OD pair, which is a reasonable approximation when considering the cumulative effects of OD pairs. Therefore, the rank truncation is crucial for filtering the noise in a large number of sparse-demand OD pairs.



**Figure 5.7:** Comparing OD flow with its low-rank approximation in two Guangzhou metro OD pairs. (a) and (b) corresponds to the (a) and (d) in Figure 5.6, respectively.

**Table 5.3:** The difference between original data and the low-rank approximation.

Variable	Criterion	Guangzhou	Hangzhou
OD flow	RMSE	2.82	3.00
	WMAPE	28.80%	30.65%
	$R^2$	0.963	0.947
Boarding flow	RMSE	64.15	36.89
	WMAPE	4.69%	5.95%
	$R^2$	0.996	0.994

Table 5.3 further quantitatively evaluates the differences between the original OD data

and its low-rank approximation. The results in Table 5.3 are the forecast upper bound of HW-DMD under the current rank-reduced space. By comparing Table 5.3 with the forecast of HW-DMD in Table 5.1 and Table 5.2, we can see that a significant portion of the forecast error of HW-DMD essentially attributes to the rank truncation, but there is still space to improve the current HW-DMD model (e.g., by higher order, larger  $r_X$ , more regression covariates).

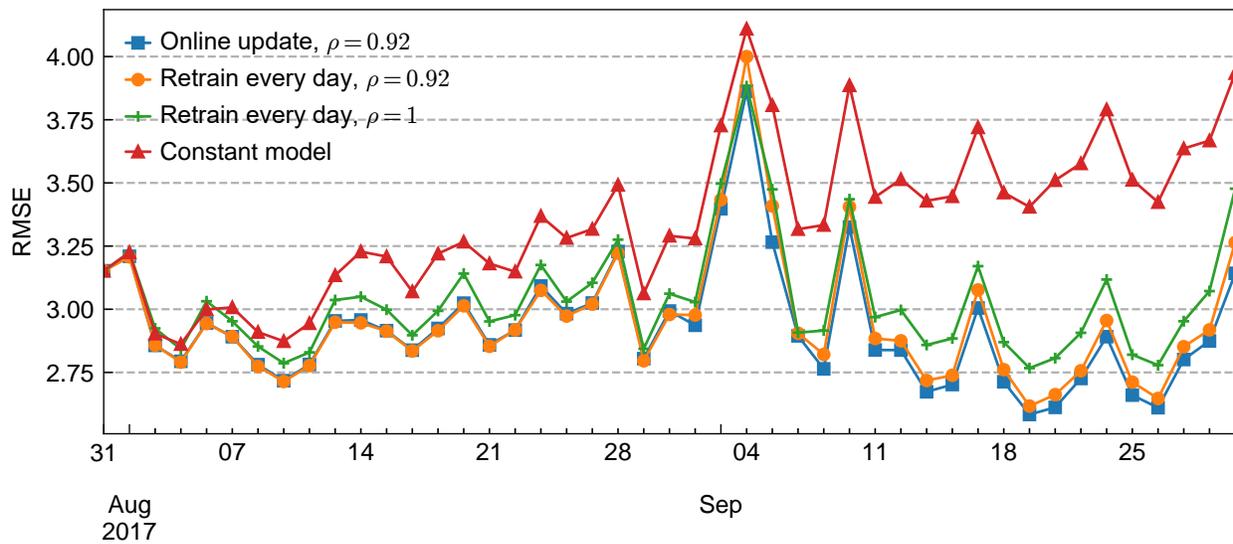
In choosing the rank-reduced space, the two rank parameters in HW-DMD balance the trade-off between forecast accuracy and model complexity. Based on the results of hyperparameter tuning, a further increase in rank  $r_Y$  may result in overfitting (bringing the noise into the rank-reduced target data). We can further slightly improve the forecasting accuracy of HW-DMD by increasing the rank  $r_X$  (related to the rank-reduced input data), but we here prefer a compact model with a smaller  $r_X$  at the cost of slight accuracy loss. Lastly, the current HW-DMD chooses the rank-reduced space purely based on the leading singular values, which may be sensitive and not optimal when encountering significant data anomalies and failures (Duke et al., 2012). Using optimized DMD (Chen et al., 2012) or combined with an anomaly detection algorithm (Scherl et al., 2020) could further improve the current HW-DMD.

### 5.7.6 Effect of the online update

The online update algorithm proposed in Section 5.6.3 can update HW-DMD’s parameters daily without storing historical data, which is computationally more efficient. On the Guangzhou metro training set, it takes  $18.7 \pm 0.43$  seconds to train an HW-DMD model, while the online update only takes  $1.0 \pm 0.03$  seconds for each day<sup>3</sup> Besides the training time, we particularly care about if errors will accumulate if we keep using the online update algorithm for a long time. Therefore, we apply the online update algorithm to all the two-month data after the training set of the Guangzhou data set to evaluate its long-term effect. In comparison, we retrain two HW-DMD models (with  $\rho=0.92$  and 1, respectively) every day using all historical data up until the latest. The results are shown in Figure 5.8. We summarize the key findings for Figure 5.8 as follows:

- The RMSE of a constant model gradually increases over time. This indicates the metro system’s dynamics are time-evolving; thus, forecasting models should be updated/retrained regularly for better performance.

<sup>3</sup>We report the mean  $\pm$  standard deviation of 20 runs. Tests were run on a computer with Intel® Core™ i7-8700 Processor and 24 gigabytes of RAM. Other benchmark models have much longer training time than HW-DMD (more than 1 minute for FNN and more than 20 minutes for TRMF and Conv-LSTM).



**Figure 5.8:** The evolution of forecast RMSE under different HW-DMD update methods (Guangzhou data set). Each marker represents the RMSE of forecasted OD flow during a day. Numbered dates in the horizontal axis are Mondays; weekends are skipped.

- The RMSE curve of the online update algorithm clings to the model ( $\rho = 0.92$ ) retrained every day by entire historical data, showing the online HW-DMD update algorithm works consistently well in long-term applications. For a large training set (e.g., after September in Figure 5.8), the online update approach even performs slightly better than retraining.
- Properly reducing the weight for old data improves the forecast. Compare  $\rho = 0.92$  with  $\rho = 1$  for the two retrained models; the benefits of forgetting the old data become more significant as the training data increases.
- The OD flow of certain weekdays can be harder to forecast. Especially for the forecast of September. The RMSEs on Fridays are significantly higher on than other weekdays.

Many forecasting models do not consider the time-evolving dynamics of a metro system. Regular retraining can be prohibitive, especially for complicated models (e.g., deep learning models). This experiment shows the online update algorithm for HW-DMD is a memory-saving and accurate approach to keep an HW-DMD model up to date.

## 5.8 Conclusions and Discussion

This paper proposes a high-order weighted dynamic mode decomposition (HW-DMD) model to solve the real-time short-term OD matrix forecasting problem in metro systems. Experiments show that HW-DMD significantly outperforms common forecasting models under the high-dimensional, sparse, noisy, and skewed OD data. Particularly, we address the delayed data availability problem and the time-evolving dynamics of metro systems, which are often ignored in the literature. The idea of the forgetting rate and online update in dealing with a time-evolving system is also beneficial for other forecasting models. Moreover, the implementation of HW-DMD is simple, and the computation is very efficient, providing a promising solution to general high-dimensional time series forecasting problems.

We discuss several future research directions. (1) Current HW-DMD reshapes OD matrices into vectors for dimensionality reduction. However, performing dimensionality reduction directly on OD matrices may better utilize the column/row-wise correlations and produce more concise models (Chen et al., 2021; Gong et al., 2020). A difficulty in this direction is that the low-rank feature in metro OD matrices is relatively weak because the diagonal elements of metro OD matrices are all zeros. (2) Another future direction is to use a non-linear model instead of the current linear model in the reduced space, such as the deep factor model (Wang et al., 2019b). But a limitation for a non-linear model is that an online update method may be difficult to derive or even impossible. (3) Lastly, current HW-DMD uses external features, such as the boarding flow, simply as covariates. Incorporating more general features (e.g., weather, events) and network structure to improve the HW-DMD is worth investigating.

# Chapter 6

## Final Conclusion & Future Work

### 6.1 Summary of Results

With the increased availability of data in metro systems, there has been much research on understanding travel behavior patterns and developing data-driven planning/operation in metro systems. This thesis incorporates travel behavioral characteristics into the destination inference and the passenger demand forecasting problems in metro systems. Results show that the travel behavior is beneficial to a wide range of planning/operational tasks, from an individual-level trip destination inference, to a station-level passenger demand forecasting, to a network-level OD matrices forecasting.

In Chapter 3, a probabilistic model based on a three-dimensional LDA model is proposed. The travel behavior of each passenger is characterized by a three-dimensional latent tensor, with each dimension representing patterns of time, origin, and destination, respectively. Using the power law property, a station-to-rank preprocessing is proposed to transfer the diverse spatial patterns to similar behavioral regularities, which enhances the accuracy of destination inference. This model has two usages, (1) inferring the destinations of unlinked trips in tap-in only smart card systems. (2) discovering latent travel behavior patterns for passenger clustering. The proposed destination estimation framework is tested on Guangzhou Metro smart card data, in which the ground truth is available. Compared with an individual-history-based model, the topic model not only shows increased accuracy by around 2% but also captures essential latent patterns in passengers' travel behavior. The destination inference can also be regarded as a kind of individual-level forecasting: forecast the most likely destination given the current origin and time. Besides, the proposed topic model is also a generative model that explains the probability of a trip by individual's latent topics, and can be used for travel pattern analysis, passenger

clustering, and trip generation.

Chapter 4 focuses on forecasting the short-term passenger demand at metro stations. Unlike traditional passenger flow forecasting approaches based on time series or machine learning, the returning flow (a covariate) proposed in Chapter 4 is reasoned from passengers' travel behavior. The future returning flow of a metro station can be estimated by the past alighting flow and the return probability parallelogram (RPP), which does not rely on external information. The returning flow explains where and why the passenger demand is generated and thus captures the long-range dependencies between the alighting flow and the boarding flow of a metro station. The effect of returning flow on passenger flow forecasting is tested in the metro smart card data of Guangzhou; the results confirm the returning flow—a single covariate—can consistently improve the passenger flow forecasting under various conditions, including one-step-ahead forecasting, multi-step-ahead forecasting and forecasting under special events. Moreover, paired  $t$ -tests show the improvement is significant for most business-type stations, and the results under different models (SARIMA, SVR, and MLP) further demonstrate the effectiveness of returning flow to general passenger flow forecasting methods.

Chapter 5 continues the short-term passenger demand forecasting to an OD matrices level. Major challenges of real-time OD matrices forecasting include (1) high-dimensionality, (2) skewed flow distribution within an OD matrix, (3) delayed availability of recent OD snapshots, and (4) the model maintenance problem caused by time-evolving dynamics. We formulate the OD matrices forecasting problem by a high-order VAR model. To solve the first two challenges, we reconstruct the VAR model as a reduced-order regression model and estimate it using dynamic mode decomposition (DMD). The VAR coefficients estimated by DMD are the best-fit (in terms of Frobenius norm) linear operator for the rank-reduced full-size data. Station boarding demand is used as a covariate in the VAR to replace the available recent OD snapshots. We consider the time-evolving dynamics of metro systems and improve the forecast by exponentially reducing the weights for historical data. A tailored online update algorithm is then developed for the High-order Weighted DMD model (HW-DMD) to update the model coefficients at a daily level, without storing historical data or retraining. Experiments on data from two large-scale metro systems show that the proposed HW-DMD is robust to noisy and sparse data, and significantly outperforms baseline models in forecasting both OD matrices and boarding flow. The online update algorithm also shows consistent accuracy over a long time, allowing us to maintain an HW-DMD model at a much lower cost.

## 6.2 Discussion

### 6.2.1 Practical implications

The probabilistic destination inference model developed in Chapter 3 requires a portion of complete itineraries (either ground truth or an estimated training set, see Section 3.5.5) to learn passengers' travel patterns. Therefore, the target trips of the proposed model are unlinked trips whose destination cannot be estimated by conventional trip-chain-based methods. Besides missing data imputation, destination recommendation is another practical usage of the proposed model. For example, recommending a list of possible destinations to a user when he/she uses a web/mobile navigation service, given his/her travel history, current location, and time. As a generative model, we can also use the model developed in Chapter 3 to generate synthetic trips. The synthetic trips can serve as inputs of agent-based urban simulation (Horni et al., 2016) without confidentiality. As the current model does not capture the order of synthetic trips, a sequential extension (e.g., Mo et al., 2021; Yin et al., 2017) could be developed to model the correlations between trips.

Chapter 4 and Chapter 5 develop models for short-term passenger demand forecasting. The practical implication of these forecasting models is twofold. From passengers' perspective, real-time/projected demand/congestion at stations or trains/buses can help them make travel plans (Cats et al., 2011). The real-time information can be made available through navigation Apps or electronic station boards. In fact, companies like transit app and Google are providing such services. From the perspective of transit agencies, it is possible to dynamically adjust the supply of services according to demand forecasting, making the system more efficient and economical. Traditional public transit uses a static timetable that remains unchanged for months (Ceder, 2016), or only makes some adjustments under special events. However, real-time passenger forecasting will become much more important when the public transit schedule becomes demand responsive with the development of intelligent transportation (Peled et al., 2021).

### 6.2.2 Limitations

Besides chapter-specific limitations as summarized in the conclusion section of each chapter, there are some general limitations in this research. First, the passenger demand forecasting models proposed in Chapter 4 and Chapter 5 require knowing both boarding and alighting demands. However, smart card data in many cities other than Guangzhou do not record alighting information. Data availability is a major problem when applying the

proposed methods to other cities. This limitation could be alleviated with the development of infrastructure and technologies. Second, this research focuses more on verifying ideas than fine-tuning models. Therefore, this research did not incorporate external factors like weather, holidays, and land-use properties. From an academic perspective, we exclude external factors to highlight the effect of the research subject (travel behavior). From a practical perspective, however, the performance of forecasting and destination inference could be further improved by considering these external factors. Third, the inference and forecasting are “upstream” models that should be combined with effective “downstream” models. For example, optimizing transit supply using demand forecasting would rely on dynamic timetable/route design algorithms (e.g., [Iliopoulou and Kepaptsoglou, 2019](#); [Peled et al., 2021](#)), forecasting passenger loads on trains requires combining the OD matrices forecasting with a traffic assignment model. Forecasting is a component in the family of applications in public transportation; there is still a long way to go in integrating these applications for a better public transportation system.

### 6.2.3 Forecasting in transportation

As reviewed in Section 2.2, there has been a surge in research on (short-term) passenger flow forecasting (including this thesis) in recent years. A closely related but even more popular topic is forecasting in a broader transportation context (e.g., traffic flow/speed forecasting, ride-hailing demand forecasting). The surge of research interest in transportation forecasting is because the available data sets, the computational capability, the types of models, and the vision for intelligent transportation have been increasing rapidly. However, a problem occurs that many studies tend to be too homogeneous. It is thus important to discuss what is good research on forecasting in transportation. Good research on transportation forecasting should have at least one of the following qualities:

1. **Demonstrate significant improvements in forecasting accuracy with sufficient evidence.** Accuracy is the most widely used criterion when claiming the contribution of a forecasting model. However, making a fair comparison in forecast accuracy between models is difficult; a critical reason is the lack of widely acknowledged benchmark data sets and forecasting tasks. Forecasting models were proposed for assorted tasks and tested under different data sets. A model designed for a specific task often is not transferable to another study. Moreover, scientific articles usually contain insufficient details to replicate the model and the experiments. Even the same model can have diverse performances under different hyperparameters, train-

ing procedures, or even initialization values, especially for complex models (e.g., deep learning models). Because of the above reasons, there are debates on whether new forecasting methods really improve traffic forecasting (Nair and Dekusar, 2020; Manibardo et al., 2021). Therefore, sufficient evidence must be shown to demonstrate the superiority of a new model/method.

2. **Utilize knowledge/features of the transportation domain and make methodological contributions.** Transportation is a comprehensive domain that includes people, vehicles, and roads. Incorporating domain knowledge into forecasting is meaningful. For example, incorporating behavior characteristics of passengers into metro boarding flow forecasting (Chapter 4) and leveraging network topology in traffic flow forecasting (e.g., Yu et al., 2018; Kwak et al., 2021). The value of this type of research lies not in more accurate forecasting (though they usually are), but in providing a new way of thinking, enhancing the understanding of the system, and unlocking future potentials.
3. **Analyze/resolve problems that have practical significance.** A forecasting model is only a small step in achieving a real-world forecasting project. The “leap forward” in traffic forecasting models has not been widely applied to actual projects; a critical reason is that the improvement in accuracy does not worth the extra computational cost and investment. It is thus important to discuss practical issues, such as the benefit and the lifecycle cost, of a forecasting model. For example, Peled et al. (2021) analyzed the reduction of trip time using forecasting models in demand-responsive transit, Chapter 5 developed an online update algorithm to reduce the model’s maintenance cost. Research in this direction reduces the gap between practice and theory.

Based on the above three aspects, there are a few suggestions that will be helpful for research on transportation forecasting. (1) Make statistical tests instead of presenting a single experiment’s results. (2) Open source code and data for reproducible research (Zheng, 2021). (3) Understand and use the characteristics of the transportation system. (4) Always put the practical requirements in mind while perfecting the model.

### 6.3 Future Research

The future research is summarized by the following two aspects: (1) improvements to the current research and (2) new related directions.

There is space to improve the models presented in this thesis. The models and experiments in this thesis are designed in a relatively simple manner to highlight core contributions. For example, we did not differentiate weekdays and weekends, and we only consider smart card data in a closed metro system without considering other transportation modes. More refined work should be done for practical use, similar models/ideas could be extended to include other data sources (e.g., call detail record data) and transportation modes (e.g., buses). We use a station-to-rank transformation to enhance the word co-occurrences among passengers in Chapter 3; we can try to design a spatial model to replace this transformation, which allows the model to characterize the passenger's spatial patterns. We can also try to use the information of previous trips to build a sequential model to further improve the inference accuracy (e.g., consider the order of activities, [Mo et al., 2021](#)). For Chapter 4 that uses returning flow in the passenger flow forecasting, a more ingenious approach to forecast the future returning flow should be developed. For example, adjusting the current RPP-based forecasting with the recent observations of real returning flow. Moreover, passengers' travel behavior should have a great potential to alert the surge of boarding demand during special events. Besides, it is very interesting and promising to study the effect of travel behavior in enhancing the forecasting in other transport modes with chained travel patterns, such as private vehicles and ride-hailing services. For example, drivers of private vehicles may also have similar chained trips when commuting or in other activities. For the OD matrices forecasting model in Chapter 5, it is worth trying to make forecasts directly on matrices without a vectorization operator, which preserves column/row-wise correlations (e.g., [Chen et al., 2021](#); [Hsu et al., 2021](#)). Because of the hierarchical structure of OD matrices, it is also interesting to explore and eliminate the spatial/temporal correlations the forecasting errors.

In terms of new related directions, There is no doubt that travel behavior has much broader application scenarios. The COVID-19 pandemic has significantly changed public transit and passengers' travel behavior, bringing new prospects and research needs ([Tirachini and Cats, 2020](#)). It is essential to evaluate how users' travel behavior adapts to regulations, policies, and information. Real-time forecasting of passenger loads/congestion in public transit becomes a critical topic with social distancing and capacity limitations. Moreover, deep learning has produced revolutionary breakthroughs in many fields and has also been attempted in public transit, such as forecasting and control (e.g., [Wang and Sun, 2020](#)). An exciting direction closely related to deep learning and travel behavior is mobility synthesis, where deep generative models are used to learn and create artificial but identically distributed mobility data. The synthetic mobility data can compensate for

the limitations (such as privacy concerns and missing values) of real-world data and serve as essential inputs for agent-based urban simulation (e.g., [Feng et al., 2020](#)). Passengers' travel behavior is constantly evolving with the advance of technologies and society, and research must keep up with the development to provide better public transportation.

# Bibliography

- Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle, in: Selected papers of hirotugu akaike. Springer, pp. 199–213.
- Alfatlawi, M., Srivastava, V., 2020. An incremental approach to online dynamic mode decomposition for time-varying systems with applications to eeg data modeling. *Journal of Computational Dynamics* 7, 209–241.
- Alsger, A., Assemi, B., Mesbah, M., Ferreira, L., 2016. Validating and improving public transport origin–destination estimation algorithm using smart card fare data. *Transportation Research Part C: Emerging Technologies* 68, 490–506.
- Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., Hickman, M., 2018. Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies* 87, 123–137.
- Arbabi, H., Mezić, I., 2017. Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator. *SIAM Journal on Applied Dynamical Systems* 16, 2096–2126.
- Assemi, B., Alsger, A., Moghaddam, M., Hickman, M., Mesbah, M., 2020. Improving alighting stop inference accuracy in the trip chaining method using neural networks. *Public Transport* 12, 89–121.
- Avila, A., Mezić, I., 2020. Data-driven analysis and forecasting of highway traffic dynamics. *Nature communications* 11, 1–16.
- Axhausen, K.W., Zimmermann, A., Schönfelder, S., Rindsfuser, G., Haupt, T., 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation* 29, 95–124.
- Barry, J.J., Newhouser, R., Rahbee, A., Sayeda, S., 2002. Origin and destination estimation in new york city with automated fare system data. *Transportation Research Record: 1817*, 183–187.

## BIBLIOGRAPHY

---

- Bhaskar, A., Chung, E., et al., 2014. Passenger segmentation using smart card data. *IEEE Transactions on intelligent transportation systems* 16, 1537–1548.
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.
- Bowman, J.L., Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation research part a: policy and practice* 35, 1–28.
- Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.
- Briand, A.S., Côme, E., Mohamed, K., Oukhellou, L., 2016. A mixture model clustering approach for temporal passenger pattern characterization in public transport. *International Journal of Data Science and Analytics* 1, 37–50.
- Briand, A.S., Côme, E., Trépanier, M., Oukhellou, L., 2017. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. *Transportation Research Part C: Emerging Technologies* 79, 274–289.
- Brunton, S.L., Brunton, B.W., Proctor, J.L., Kaiser, E., Kutz, J.N., 2017. Chaos as an intermittently forced linear system. *Nature communications* 8, 1–9.
- Cantelmo, G., Qurashi, M., Prakash, A.A., Antoniou, C., Viti, F., 2020. Incorporating trip chaining within online demand estimation. *Transportation Research Part B: Methodological* 132, 171–187.
- Cantelmo, G., Viti, F., 2019. Incorporating activity duration and scheduling utility into equilibrium-based dynamic traffic assignment. *Transportation Research Part B: Methodological* 126, 365–390.
- Carrese, S., Cipriani, E., Mannini, L., Nigro, M., 2017. Dynamic demand estimation and prediction for traffic urban networks adopting new data sources. *Transportation Research Part C: Emerging Technologies* 81, 83–98.
- Cats, O., Koutsopoulos, H.N., Burghout, W., Toledo, T., 2011. Effect of real-time transit information on dynamic path choice of passengers. *Transportation Research Record* 2217, 46–54.

## BIBLIOGRAPHY

---

- Ceder, A., 2016. Public transit planning and operation: Modeling, practice and behavior. CRC press.
- Chen, E., Ye, Z., Wang, C., Xu, M., 2020a. Subway passenger flow prediction for special events using smart card data. *IEEE Transactions on Intelligent Transportation Systems* 21, 1109–1120.
- Chen, J., Liu, L., Wu, H., Zhen, J., Li, G., Lin, L., 2020b. Physical-virtual collaboration graph network for station-level metro ridership prediction. *arXiv preprint arXiv:2001.04889* .
- Chen, K.K., Tu, J.H., Rowley, C.W., 2012. Variants of dynamic mode decomposition: boundary condition, koopman, and fourier analyses. *Journal of nonlinear science* 22, 887–915.
- Chen, Q., Li, W., Zhao, J., 2011. The use of LS-SVM for short-term passenger flow prediction. *Transport* 26, 5–10.
- Chen, R., Xiao, H., Yang, D., 2021. Autoregressive models for matrix-valued time series. *Journal of Econometrics* 222, 539–560.
- Cheng, Z., Trépanier, M., Sun, L., 2021a. Incorporating travel behavior regularity into passenger flow forecasting. *Transportation Research Part C: Emerging Technologies* 128, 103200.
- Cheng, Z., Trépanier, M., Sun, L., 2021b. Probabilistic model for destination inference and travel pattern mining from smart card data. *Transportation* 48, 2035–2053.
- Cheng, Z., Trepanier, M., Sun, L., 2022. Real-time forecasting of metro origin-destination matrices with high-order weighted dynamic mode decomposition. *Transportation Science* doi:10.1287/trsc.2022.1128. (in press).
- Chu, K.F., Lam, A.Y., Li, V.O., 2019. Deep multi-scale convolutional lstm network for travel demand and origin-destination predictions. *IEEE Transactions on Intelligent Transportation Systems* 21, 3219–3232.
- Dai, X., Sun, L., Xu, Y., 2018. Short-term origin-destination based metro flow prediction with probabilistic model selection approach. *Journal of Advanced Transportation* 2018.
- Deschaintres, E., Morency, C., Trépanier, M., 2019. Analyzing transit user behavior with 51 weeks of smart card data. *Transportation Research Record* 2673, 33–45.

## BIBLIOGRAPHY

---

- Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.
- Ding, C., Duan, J., Zhang, Y., Wu, X., Yu, G., 2017. Using an ARIMA-GARCH modeling approach to improve subway short-term ridership forecasting accounting for dynamic volatility. *IEEE Transactions on Intelligent Transportation Systems* 19, 1054–1064.
- Du, B., Liu, C., Zhou, W., Hou, Z., Xiong, H., 2018. Detecting pickpocket suspects from large-scale public transit records. *IEEE Transactions on Knowledge and Data Engineering* 31, 465–478.
- Du, B., Zhou, W., Liu, C., Cui, Y., Xiong, H., 2019. Transit pattern detection using tensor factorization. *INFORMS Journal on Computing* 31, 193–206.
- Duke, D., Soria, J., Honnery, D., 2012. An error analysis of the dynamic mode decomposition. *Experiments in fluids* 52, 529–542.
- Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank. *Psychometrika* 1, 211–218.
- Fan, Z., Arai, A., Song, X., Witayangkurn, A., Kanasugi, H., Shibasaki, R., 2016. A collaborative filtering approach to citywide human mobility completion from sparse call records, in: *International Joint Conference on Artificial Intelligence*, pp. 2500–2506.
- Farrahi, K., Gatica-Perez, D., 2009. Learning and predicting multimodal daily life patterns from cell phones, in: *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 277–280.
- Feng, J., Li, Y., Zhang, C., Sun, F., Meng, F., Guo, A., Jin, D., 2018. Deepmove: Predicting human mobility with attentional recurrent networks, in: *Proceedings of the 2018 world wide web conference*, pp. 1459–1468.
- Feng, J., Yang, Z., Xu, F., Yu, H., Wang, M., Li, Y., 2020. Learning to simulate human mobility, in: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3426–3433.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics* 82, 540–554.

## BIBLIOGRAPHY

---

- Gambs, S., Killijian, M.O., del Prado Cortez, M.N., 2012. Next place prediction using mobility markov chains, in: Proceedings of the first workshop on measurement, privacy, and mobility, pp. 1–6.
- Gao, Y., Guo, Z., Long, Y., Cui, Z., Li, X., 2022. Passengers' travel behavior before and after the adjustment of regular bus collinear sections: A case study in the incipient phase of metro operation in xiamen. *Travel Behaviour and Society* 26, 221–230.
- Ghaemi, M.S., Agard, B., Trépanier, M., Partovi Nia, V., 2017. A visual segmentation method for temporal smart card data. *Transportmetrica A: Transport Science* 13, 381–404.
- Gong, Y., Li, Z., Zhang, J., Liu, W., Zheng, Y., 2020. Online spatio-temporal crowd flow distribution prediction for complex metro system. *IEEE Transactions on Knowledge and Data Engineering* 34, 865–880.
- Gong, Y., Li, Z., Zhang, J., Liu, W., Zheng, Y., Kirsch, C., 2018. Network-wide crowd flow prediction of sydney trains via customized online non-negative matrix factorization, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 1243–1252.
- González, M.C., Hidalgo, C.A., Barabási, A.L., 2008. Understanding individual human mobility patterns. *Nature* 453, 779–782.
- Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H., 2018. Estimation of population origin–interchange–destination flows on multimodal transit networks. *Transportation Research Part C: Emerging Technologies* 90, 350–365.
- Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H., Attanucci, J.P., 2013. Automated inference of linked transit journeys in london using fare-transaction and vehicle location data. *Transportation Research Record* 2343, 17–24.
- Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. *Transportation Research Part C: Emerging Technologies* 64, 1–16.
- Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, Z., Zhao, J., 2017. Measuring regularity of individual travel patterns. *IEEE Transactions on Intelligent Transportation Systems* 19, 1583–1592.

## BIBLIOGRAPHY

---

- Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101, 5228–5235.
- Guangzhou Metro, 2019. URL: <http://www.gzmtr.com/ygwm/gsgk/gsjjs/>. Accessed 2019-11-02.
- Guo, J., Xie, Z., Qin, Y., Jia, L., Wang, Y., 2019. Short-term abnormal passenger flow prediction based on the fusion of SVR and LSTM. *IEEE Access* 7, 42946–42955.
- Han, G., Sohn, K., 2016. Activity imputation for trip-chains elicited from smart-card data using a continuous hidden markov model. *Transportation Research Part B: Methodological* 83, 121–135.
- Han, Y., Wang, S., Ren, Y., Wang, C., Gao, P., Chen, G., 2019. Predicting station-level short-term passenger flow in a citywide metro network using spatiotemporal graph convolutional neural networks. *ISPRS International Journal of Geo-Information* 8.
- Hao, S., Lee, D.H., Zhao, D., 2019. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. *Transportation Research Part C: Emerging Technologies* 107, 287–300.
- Hasan, S., Schneider, C.M., Ukkusuri, S.V., González, M.C., 2013. Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics* 151, 304–318.
- Hasan, S., Ukkusuri, S.V., 2014. Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies* 44, 363–381.
- He, L., Agard, B., Trépanier, M., 2020. A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method. *Transportmetrica A: Transport Science* 16, 56–75.
- He, L., Trépanier, M., 2015. Estimating the destination of unlinked trips in transit smart card fare data. *Transportation Research Record* , 97–104.
- Hemati, M.S., Williams, M.O., Rowley, C.W., 2014. Dynamic mode decomposition for large and streaming datasets. *Physics of Fluids* 26, 111701.
- Horni, A., Nagel, K., Axhausen, K. (Eds.), 2016. *Multi-Agent Transport Simulation MAT-Sim*. Ubiquity Press, London.

## BIBLIOGRAPHY

---

- Hsu, N.J., Huang, H.C., Tsay, R.S., 2021. Matrix autoregressive spatio-temporal models. *Journal of Computational and Graphical Statistics* , 1–32.
- Hu, J., Yang, B., Guo, C., Jensen, C.S., Xiong, H., 2020. Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks, in: 2020 IEEE 36th International Conference on Data Engineering (ICDE), IEEE. pp. 1417–1428.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O’Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeeen, F., 2020. forecast: Forecasting functions for time series and linear models. URL: <http://pkg.robjhyndman.com/forecast>. r package version 8.12.
- Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: principles and practice. 2nd ed., OTexts, Melbourne, Australia.
- Iliopoulou, C., Kepaptsoglou, K., 2019. Combining its and optimization in public transportation planning: state of the art and future research paths. *European Transport Research Review* 11, 1–16.
- Jiao, P., Li, R., Sun, T., Hou, Z., Ibrahim, A., 2016. Three revised kalman filtering models for short-term rail transit passenger flow prediction. *Mathematical Problems in Engineering* 2016, 9717582.
- Jung, J., Sohn, K., 2017. Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. *IET Intelligent Transport Systems* 11, 334–339.
- Ke, J., Qin, X., Yang, H., Zheng, Z., Zhu, Z., Ye, J., 2021. Predicting origin-destination ride-sourcing demand with a spatio-temporal encoder-decoder residual multi-graph convolutional network. *Transportation Research Part C: Emerging Technologies* 122, 102858.
- Kieu, L.M., Bhaskar, A., Chung, E., 2015. A modified density-based scanning algorithm with noise for spatial travel pattern analysis from smart card AFC data. *Transportation Research Part C: Emerging Technologies* 58, 193–207.
- Kitamura, R., 1990. Panel analysis in transportation planning: An overview. *Transportation Research Part A: General* 24, 401–415.
- Kwak, S., Geroliminis, N., 2020. Travel time prediction for congested freeways with a dynamic linear model. *IEEE Transactions on Intelligent Transportation Systems* 22, 7667–7677.

## BIBLIOGRAPHY

---

- Kwak, S., Geroliminis, N., Frossard, P., 2021. Traffic signal prediction on transportation networks using spatio-temporal correlations on graphs. *IEEE Transactions on Signal and Information Processing over Networks* 7, 648–659.
- Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. *Biometrika* 98, 901–918.
- Le Clainche, S., Vega, J.M., 2017. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems* 16, 882–925.
- Li, Y., Wang, X., Sun, S., Ma, X., Lu, G., 2017. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transportation Research Part C: Emerging Technologies* 77, 306–328.
- Li, Z., Sergin, N.D., Yan, H., Zhang, C., Tsung, F., 2020. Tensor completion for weakly-dependent data on graph for metro passenger flow prediction, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4804–4810.
- Liu, J., Zheng, F., van Zuylen, H.J., Li, J., 2020. A dynamic od prediction approach for urban networks based on automatic number plate recognition data. *Transportation Research Procedia* 47, 601–608.
- Liu, L., Qiu, Z., Li, G., Wang, Q., Ouyang, W., Lin, L., 2019a. Contextualized spatial-temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 3875–3887.
- Liu, Y., Liu, Z., Jia, R., 2019b. DeepPF: A deep learning based architecture for metro passenger flow prediction. *Transportation Research Part C: Emerging Technologies* 101, 18–34.
- Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.J., 2017. Understanding commuting patterns using transit smart card data. *Journal of Transport Geography* 58, 135–145.
- Ma, X., Wu, Y.J., Wang, Y., Chen, F., Liu, J., 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies* 36, 1–12.
- Ma, X., Zhang, J., Du, B., Ding, C., Sun, L., 2018. Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems* 20, 2278–2288.

## BIBLIOGRAPHY

---

- Ma, Z., Koutsopoulos, H.N., Liu, T., Basu, A.A., 2020. Behavioral response to promotion-based public transport demand management: Longitudinal analysis and implications for optimal promotion design. *Transportation Research Part A: Policy and Practice* 141, 356–372.
- Manibardo, E.L., Laña, I., Del Ser, J., 2021. Deep learning for road traffic forecasting: Does it make a difference? *IEEE Transactions on Intelligent Transportation Systems* .
- Mo, B., Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2021. Individual mobility prediction in mass transit systems using smart card data: An interpretable activity-based hidden markov approach. *IEEE Transactions on Intelligent Transportation Systems* .
- Mohamed, K., Côme, E., Baro, J., Oukhellou, L., 2014. Understanding passenger patterns in public transit through smart card and socioeconomic data, in: *ACM SIGKDD Workshop on Urban Computing*.
- Mohamed, K., Côme, E., Oukhellou, L., Verleysen, M., 2016. Clustering smart card data for urban mobility analysis. *IEEE Transactions on intelligent transportation systems* 18, 712–728.
- Morency, C., Trepanier, M., Agard, B., 2007. Measuring transit use variability with smart-card data. *Transport Policy* 14, 193–203.
- MTL Trajet, 2019. URL: <https://ville.montreal.qc.ca/mtltrajet/en/>. Accessed 2019-11-08.
- Munizaga, M., Devillaine, F., Navarrete, C., Silva, D., 2014. Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies* 44, 70–79.
- Munizaga, M.A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile. *Transportation Research Part C: Emerging Technologies* 24, 9–18.
- Mützel, C.M., Scheiner, J., 2021. Investigating spatio-temporal mobility patterns and changes in metro usage under the impact of covid-19 using taipei metro smart card data. *Public Transport* , 1–24.
- Nair, R., Dekusar, A., 2020. Keep it simple stupid! a non-parametric kernel regression approach to forecast travel speeds. *Transportation Research Part C: Emerging Technologies* 110, 269–274.

## BIBLIOGRAPHY

---

- Newman, M.E., 2005. Power laws, pareto distributions and zipf's law. *Contemporary physics* 46, 323–351.
- Ni, M., He, Q., Gao, J., 2016. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems* 18, 1623–1632.
- Noursalehi, P., Koutsopoulos, H.N., Zhao, J., 2018. Real time transit demand prediction capturing station interactions and impact of special events. *Transportation Research Part C: Emerging Technologies* 97, 277–300.
- Noursalehi, P., Koutsopoulos, H.N., Zhao, J., 2021. Dynamic origin-destination prediction in urban rail systems: A multi-resolution spatio-temporal deep learning approach. *IEEE Transactions on Intelligent Transportation Systems* .
- Nunes, A.A., Dias, T.G., e Cunha, J.F., 2016. Passenger journey destination estimation from automated fare collection system data using spatial validation. *IEEE Transactions on Intelligent Transportation Systems* 17, 133–142.
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.L., 2015. Returners and explorers dichotomy in human mobility. *Nature communications* 6, 8166.
- Peled, I., Lee, K., Jiang, Y., Dauwels, J., Pereira, F.C., 2021. On the quality requirements of demand prediction for dynamic public transport. *Communications in Transportation Research* 1, 100008.
- Pelletier, M.P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19, 557–568.
- Proctor, J.L., Brunton, S.L., Kutz, J.N., 2016. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems* 15, 142–161.
- Ren, J., Xie, Q., 2017. Efficient od trip matrix prediction based on tensor decomposition, in: 2017 18th IEEE International Conference on Mobile Data Management (MDM), IEEE. pp. 180–185.
- Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D., et al., 2009. Spectral analysis of nonlinear flows. *Journal of fluid mechanics* 641, 115–127.

## BIBLIOGRAPHY

---

- Sánchez-Martínez, G.E., 2017. Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. *Transportation Research Record* 2652, 1–7.
- Scheiner, J., 2014. The gendered complexity of daily life: effects of life-course events on changes in activity entropy and tour complexity over time. *Travel Behaviour and Society* 1, 91–105.
- Scherl, I., Strom, B., Shang, J.K., Williams, O., Polagye, B.L., Brunton, S.L., 2020. Robust principal component analysis for modal decomposition of corrupt fluid flows. *Physical Review Fluids* 5, 054401.
- Schmid, P.J., 2010. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics* 656, 5–28.
- Sha, S., Li, J., Zhang, K., Yang, Z., Wei, Z., Li, X., Zhu, X., 2020. RNN-based subway passenger flow rolling prediction. *IEEE Access* 8, 15232–15240.
- Shen, L., Shao, Z., Yu, Y., Chen, X., 2021. Hybrid approach combining modified gravity model and deep learning for short-term forecasting of metro transit passenger flows. *Transportation Research Record* 2675, 25–38.
- Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c., 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems* 28, 802–810.
- Song, C., Qu, Z., Blumm, N., Barabási, A.L., 2010. Limits of predictability in human mobility. *Science* 327, 1018–1021.
- Stathopoulos, A., Karlaftis, M.G., 2003. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies* 11, 121–135.
- Sun, L., Axhausen, K.W., 2016. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological* 91, 511–524.
- Sun, L., Axhausen, K.W., Lee, D.H., Huang, X., 2013. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences* 110, 13774–13779.

## BIBLIOGRAPHY

---

- Sun, L., Chen, X., He, Z., Miranda-Moreno, L.F., 2021. Routine pattern discovery and anomaly detection in individual travel behavior. *Networks and Spatial Economics* .
- Sun, S., Yang, D., Feng, G., Guo, J.e., 2020. Adaensemble learning approach for metro passenger flow forecasting. arXiv preprint arXiv:2002.07575 .
- Sun, Y., Leng, B., Guan, W., 2015. A novel wavelet-SVM short-time passenger flow prediction in beijing subway system. *Neurocomputing* 166, 109–121.
- Tan, M.C., Wong, S.C., Xu, J.M., Guan, Z.R., Zhang, P., 2009. An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 10, 60–69.
- Tang, L., Zhao, Y., Cabrera, J., Ma, J., Tsui, K.L., 2018. Forecasting short-term passenger flow: An empirical study on shenzhen metro. *IEEE Transactions on Intelligent Transportation Systems* 20, 3613–3622.
- Tirachini, A., Cats, O., 2020. Covid-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation* 22, 1–21.
- Toqué, F., Côme, E., El Mahrsi, M.K., Oukhellou, L., 2016. Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks, in: 2016 IEEE 19th international conference on intelligent transportation systems (ITSC), IEEE. pp. 1071–1076.
- Toqué, F., Khouadjia, M., Come, E., Trepanier, M., Oukhellou, L., 2017. Short & long term forecasting of multimodal transport passenger flows with machine learning methods, in: IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 560–566.
- Trépanier, M., Habib, K.M., Morency, C., 2012. Are transit users loyal? Revelations from a hazard model based on smart card data. *Canadian Journal of Civil Engineering* 39, 610–618.
- Trépanier, M., Tranchant, N., Chapleau, R., 2007. Individual trip destination estimation in a transit smart card automated fare collection system. *Journal of Intelligent Transportation Systems* 11, 1–14.
- Tsai, T.H., Lee, C.K., Wei, C.H., 2009. Neural network based temporal feature models for short-term railway passenger demand forecasting. *Expert Systems with Applications* 36, 3728–3736.

## BIBLIOGRAPHY

---

- Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Kutz, J.N., 2014. On dynamic mode decomposition: Theory and applications. *Journal of Computational Dynamics* 1, 391–421.
- Union Internationale des Transports Publics (UITP), 2018. World metro figure 2018–statistic brief. <https://www.uitp.org/publications/world-metro-figures/>. Accessed: 2020-03-20.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: *Advances in neural information processing systems*, pp. 5998–6008.
- Viallard, A., Trépanier, M., Morency, C., 2019. Assessing the evolution of transit user behavior from smart card data. *Transportation Research Record* 2673, 184–194.
- Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies* 13, 211–234.
- Wang, H., Li, L., Pan, P., Wang, Y., Jin, Y., 2019a. Early warning of burst passenger flow in public transportation system. *Transportation Research Part C: Emerging Technologies* 105, 580–598.
- Wang, J., Kong, X., Zhao, W., Tolba, A., Al-Makhadmeh, Z., Xia, F., 2018. Stloyal: A spatio-temporal loyalty-based model for subway passenger flow prediction. *IEEE Access* 6, 47461–47471.
- Wang, J., Sun, L., 2020. Dynamic holding control to avoid bus bunching: A multi-agent deep reinforcement learning framework. *Transportation Research Part C: Emerging Technologies* 116, 102661.
- Wang, J., Zhang, Y., Wei, Y., Hu, Y., Piao, X., Yin, B., 2021. Metro passenger flow prediction via dynamic hypergraph convolution networks. *IEEE Transactions on Intelligent Transportation Systems* .
- Wang, S., Miao, H., Chen, H., Huang, Z., 2020. Multi-task adversarial spatial-temporal networks for crowd flow prediction, in: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 1555–1564.

## BIBLIOGRAPHY

---

- Wang, W., Attanucci, J., Wilson, N., 2011. Bus passenger origin-destination estimation and related analyses using automated data collection systems. *Journal of Public Transportation* 14, 131–150.
- Wang, Y., Smola, A., Maddix, D., Gasthaus, J., Foster, D., Januschowski, T., 2019b. Deep factors for forecasting, in: *International conference on machine learning*, PMLR. pp. 6607–6617.
- Wang, Y., Yin, H., Chen, H., Wo, T., Xu, J., Zheng, K., 2019c. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling, in: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1227–1235.
- Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association* 58, 236–244.
- Wei, Y., Chen, M.C., 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research Part C: Emerging Technologies* 21, 148–162.
- Williams, B.M., 2001. Multivariate vehicular traffic flow prediction: evaluation of arimax modeling. *Transportation Research Record* 1776, 194–200.
- Williams, B.M., Hoel, L.A., 2003. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering* 129, 664–672.
- Wu, Y., Cheng, Z., Sun, L., 2021. Individual mobility prediction via attentive marked temporal point processes. *arXiv preprint arXiv:2109.02715* .
- Xiong, X., Ozbay, K., Jin, L., Feng, C., 2020. Dynamic origin–destination matrix prediction with line graph neural networks and kalman filter. *Transportation Research Record* 2674, 491–503.
- Xue, G., Liu, S., Gong, D., 2020. Identifying abnormal riding behavior in urban rail transit: A survey on “in-out” in the same subway station. *IEEE Transactions on Intelligent Transportation Systems* .
- Xue, G., Liu, S., Ren, L., Ma, Y., Gong, D., 2022. Forecasting the subway passenger flow under event occurrences with multivariate disturbances. *Expert Systems with Applications* 188, 116057.

## BIBLIOGRAPHY

---

- Yang, J., Dong, X., Jin, S., 2020. Metro passenger flow prediction model using attention-based neural network. *IEEE Access* 8, 30953–30959.
- Ye, J., Zheng, F., Zhao, J., Ye, K., Xu, C., 2021. Multi-view trgru: Transformer based spatiotemporal model for short-term metro origin-destination matrix prediction. *arXiv preprint arXiv:2108.03900* .
- Yin, M., Sheehan, M., Feygin, S., Paiement, J.F., Pozdnoukhov, A., 2017. A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems* 19, 1682–1696.
- Yu, B., Yin, H., Zhu, Z., 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting, in: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 3634–3640.
- Yu, H.F., Rao, N., Dhillon, I.S., 2016. Temporal regularized matrix factorization for high-dimensional time series prediction, in: *Advances in neural information processing systems*, pp. 847–855.
- Yu, Y., Zhang, Y., Qian, S., Wang, S., Hu, Y., Yin, B., 2020. A low rank dynamic mode decomposition model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* .
- Zhang, D., Xiao, F., Shen, M., Zhong, S., 2021a. DNEAT: A novel dynamic node-edge attention network for origin-destination demand prediction. *Transportation Research Part C: Emerging Technologies* 122, 102851.
- Zhang, F., Yuan, N.J., Wang, Y., Xie, X., 2015. Reconstructing individual mobility from smart card transactions: a collaborative space alignment approach. *Knowledge and Information Systems* 44, 299–323.
- Zhang, H., Rowley, C.W., Deem, E.A., Cattafesta, L.N., 2019a. Online dynamic mode decomposition for time-varying systems. *SIAM Journal on Applied Dynamical Systems* 18, 1586–1609.
- Zhang, J., Che, H., Chen, F., Ma, W., He, Z., 2021b. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transportation Research Part C: Emerging Technologies* 124, 102928.

## BIBLIOGRAPHY

---

- Zhang, J., Chen, F., Cui, Z., Guo, Y., Zhu, Y., 2020. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems* 22, 7004–7014.
- Zhang, J., Chen, F., Wang, Z., Liu, H., 2019b. Short-term origin-destination forecasting in urban rail transit based on attraction degree. *IEEE Access* 7, 133452–133462.
- Zhang, P., Ma, Z., Weng, X., 2021c. Detecting invalid associations between fare machines and metro stations using smart card data. *Journal of Advanced Transportation* 2021.
- Zhao, J., Qu, Q., Zhang, F., Xu, C., Liu, S., 2017. Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems* 18, 3135–3146.
- Zhao, J., Rahbee, A., Wilson, N.H., 2007. Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering* 22, 376–387.
- Zhao, Y., Ma, Z., Jiang, X., Koutsopoulos, H.N., 2021. Short-term metro ridership prediction during unplanned events. *Transportation Research Record* 2676, 132–147.
- Zhao, Y., Ma, Z., Yang, Y., Jiang, W., Jiang, X., 2020a. Short-term passenger flow prediction with decomposition in urban railway systems. *IEEE Access* 8, 107876–107886.
- Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2018a. Detecting pattern changes in individual travel behavior: A bayesian approach. *Transportation research part B: methodological* 112, 73–88.
- Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2018b. Discovering latent activity patterns from human mobility, in: *The 7th ACM SIGKDD International Workshop on Urban Computing*.
- Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2018c. Individual mobility prediction using transit smart card data. *Transportation Research Part C: Emerging Technologies* 89, 19–34.
- Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2020b. Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model. *Transportation Research Part C: Emerging Technologies* 116, 102627.
- Zheng, Z., 2021. Reasons, challenges, and some tools for doing reproducible transportation research. *Communications in Transportation Research* 1, 100004.

Zúñiga, F., Muñoz, J.C., Giesen, R., 2021. Estimation and prediction of dynamic matrix travel on a public transport corridor using historical data and real-time information. *Public Transport* 13, 59–80.