Travel Demand Modeling and Estimation for High-Dimensional Mobility

Jeongyun Kim, *Member, IEEE*, Andrea Conti, *Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—The massive amount of data related to spatiotemporal mobility offers new opportunities to understand human mobility with applications in various sectors, including transportation, logistics, and safety. However, the increase in the volume and in the dimension of mobility data makes it challenging to retrieve important information and critical features of spatiotemporal mobility. This paper develops a method to estimate probabilistic occurrences of travel demands considering interactions between origin, destination, and departure time. First, we reveal the important features in the complex structure of mobility data and identify mobility patterns. Then, we derive a data-driven model, accounting for mobility patterns, to estimate and predict travel demands. We quantify the accuracy of the proposed method for a case study using both New York city yellow taxi trip data and for-hire vehicles trip data over the entire city. Results show the accuracy of the proposed method compared to existing approaches.**

***Index Terms*—Intelligent transportation systems, mobility, spatiotemporal pattern, tensor decomposition, travel demand.**

I. Introduction

MOBILITY information is essential for several applications, including network management [1], [2], smart cities [3], [4], intelligent transportation systems [5], [6], and autonomous vehicles [7], [8], [9]. A transportation system leveraging rich mobility information can be a catalyst for improving the quality of life while ensuring efficiency and sustainability [10]. Over the past decade, large quantities of human movement data, such as smart phone data [11], [12], taxi trip data [13], and smart card data [14], [15] have become available. Many researchers have utilized the data to understand human mobility, evaluate transportation systems, and provide ways to improve human mobility. The large quantities of movement data have helped researchers understand spatial-temporal structure of cities in terms of human mobility patterns. At the same time, as data resolution and the number of data attributes, including origin, destination, passenger type, and departure time, have increased, it has become more difficult to uncover major mobility patterns and comprehend the interactions between different data attributes [16]. Further, predicting the high-dimensional urban travel demands is challenging [17] because of complicated dynamics and high computational complexity.

Various travel demand models have been proposed, including the four-step models, activity-based models, and statistical models [18], [19]. These models provide essential information for decision making in transport service design, route planning, and fleet management [20], [21]. As the knowledge of travel demand in terms of origin and destination has become important to plan efficient routes for mobility services, origin-destination (OD) estimation has been actively discussed in the literature. In many cases, OD demand has been estimated in the form of an OD matrix whose row and column indices stand for origin and destination. autoregressive integrated moving average (ARIMA) [22], [23], Poisson models [23], least-square models [24], and Kalman-filter [25] are well-known approaches for estimating time-series travel demand using historical data.

Recently, deep learning methods have been proven effective in discovering human mobility patterns and modeling high-dimensional travel demand [26], [27], [28], [29], [30]. This success is due to their ability to uncover correlations between travel demand and contextual data so that these deep learning methods can enhance the modeling accuracy. However, leveraging this advantage depends on the accessibility to diverse contextual data, which can reduce the model generality of the model. Additionally, computational load must be carefully considered when using deep learning-based approaches.

Dimensionality reduction approaches have been utilized to model high-dimensional travel demands [31], [32], [33]. In particular, decomposition approaches have been applied to discover low-order substructures that capture complex dynamics of travel demands and the dependencies among the substructures [34], [35], [36]. As travel demands can be subject to periodic changes and display strong spatiotemporal correlations, those from different locations and times frequently share similar patterns [37]. In a two-dimensional setting, trip data have been transformed into a spatial unit $\times$ time matrix, and principal component analysis and non-negative matrix factorization have been used to identify spatial and temporal patterns in the matrix [38], [39].

Received 18 July 2023; revised 27 June 2024; accepted 7 July 2024. Date of publication 12 August 2024; date of current version 5 February 2025. The fundamental research described in this paper was supported, in part, by the European Union under the NextGenerationEU Italian NRRP PE00000001 program RESTART, by the National Science Foundation under Grant CNS-2148251, and by federal agency and industry partners in the RINGS program. The material in this paper was presented, in part, at the IEEE International Conference on Communications, Rome, Italy, May 2023 [DOI: 10.1109/ICC45041.2023.10279019]. Recommended for acceptance by M. Segata. *(Corresponding author: Moe Z. Win.)*

Jeongyun Kim is with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: kjy@mit.edu).

Andrea Conti is with the Department of Engineering and CNIT, University of Ferrara, 44122 Ferrara, Italy (e-mail: a.conti@ieee.org).

Moe Z. Win is with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: moewin@mit.edu).

Digital Object Identifier 10.1109/TMC.2024.3435436

In a higher-dimensional setting, more attributes such as origin, destination, day, and passenger type have been considered. The principal substructures of mobility data and the complex dependency between the attributes have been revealed using a non-negative Tucker decomposition model and a non-negative CANDECOMP/PARAFAC (CP) decomposition model [40], [41], [42].

Tensor decomposition approaches have been used to predict future travel demand [43]. Several prediction studies have applied time-series prediction models to the low-dimensional principal patterns [43], [44], [45], [46], [47]. In those works, temporal mobility patterns are first extracted as matrices, and the future temporal mobility patterns are estimated by applying ARIMA [46] and long short-term memory (LSTM) [47]. However, the majority of the OD demand prediction studies over the past decades have focused more on estimating the quantity of travel demands and less on the probabilistic characterization of those demands.

Reliability and resilience become issues in many mobility services, so a more comprehensive understanding of demand stochasticity is required for designing robust transportation systems [48]. Travel demand is inherently variable and uncertain due to human behavior. Therefore, a stochastic approach is expected to provide more reliable transportation strategies. In [49], [50], [51], stochastic travel demand has been estimated using specific statistical models. In [52], [53], domain knowledge regarding travelers' routing behaviors and the physical network topology have been incorporated in the estimation problems. While demand estimation has been extensively studied, existing literature lacks exploration of mobility pattern extraction. Meanwhile, in mobility demand modeling, mobility patterns and their interactions have been identified using a multi-dimensional probabilistic factorization [16]. In [54], to improve the interpretability of transit trip data, a trip activity attribute was considered in mobility pattern modeling.

In travel demand prediction, there is a lack of research in the field of probabilistic approaches, and especially on providing a more comprehensive understanding of demand stochasticity, indicating the need for further investigation. Moreover, interactions among mobility patterns inherently provide key information that can reduce the computational complexity and increase the prediction accuracy. The fundamental questions related to inferring a mobility demand model are listed in the following.

- How can mobility patterns and their temporal interactions be described to gain a comprehensive understanding of demand stochasticity?
- How to leverage low-dimensional patterns to improve the accuracy of the future travel demand prediction?

The answers to these questions enable a clear understanding of requirements for modeling and estimating travel demands in the face of growing complexities and volumes of mobility data. The goal of this paper is to develop a method to estimate probabilistic occurrences of travel demands considering interactions between origin, destination, and departure time. We aim to design an estimation algorithm to infer probabilistic mobility patterns, allowing to reveal the complex structure inherent in the mobility data, thus estimating future travel demands using the probabilistic mobility patterns. We advocate the importance of exploiting temporal interactions among OD patterns to estimate the future demand accurately with computation efficiency.

This paper proposes a probabilistic method for travel demand estimation.[1] We determine probabilistic OD basis patterns and temporal interactions from high-dimensional mobility data, and use them to model uncertainties inherent in travel behaviors. We also exploit the OD basis patterns and temporal interactions in predicting occurrence probability of travel demand. The key contributions of this paper are summarized in the following.

- We propose a method to estimate the travel demand distribution for characterizing the spatial basis patterns and their temporal interactions.
- We develop an efficient prediction method for future travel demands by using the extracted spatiotemporal patterns to achieve high prediction accuracy and high computation efficiency.
- We apply the proposed methods to both the yellow taxi and the for-hire vehicles trip data in New York city for extracting mobility patterns, showing that the proposed prediction method outperforms the existing ones.

The remainder of the paper is organized in the following: Section II presents an approach for modeling and estimating travel demand. Section III presents a method for predicting future travel demand distribution. Section IV provides modeling and estimation results using New York yellow taxi and for-hire vehicle trip data. Finally, Section V gives our conclusions.

Notations: Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{\mathsf{x}}$ and $\boldsymbol{x}$; a random matrix and its realization are denoted by $\mathbf{\mathsf{X}}$ and $\boldsymbol{X}$, respectively. Sets and random sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by X and $\mathcal{X}$, respectively. The indicator function of a set $\mathcal{A}$ is denoted by $\mathbb{1}_{\mathcal{A}}$, i.e., $\mathbb{1}_{\mathcal{A}}(x) = 1$ if $x \in \mathcal{A}$ and $\mathbb{1}_{\mathcal{A}}(x) = 0$ if $x \notin \mathcal{A}$. The transpose of $\boldsymbol{X}$ is denoted by $\boldsymbol{X}^{\mathrm{T}}$.

II. Estimation of Travel Demand Probability Distribution

This section presents a novel method for travel demand estimation. Specifically, we employ a probabilistic factorization approach to uncover OD basis patterns and temporal interactions among the origin basis patterns and destination basis patterns, thus characterizing the complex mobility in high-dimensional and large-scale trip dataset. The extracted OD basis patterns and temporal interactions can be utilized for accurate and efficient prediction of future trips in the following section. The framework of the proposed approach for the travel demand modeling is described in Fig. 1.

[1]A part of this work was presented in a conference version [33].

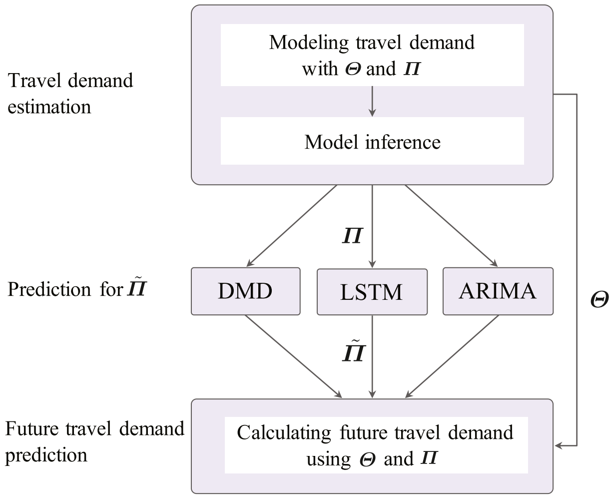


Fig. 1. Framework for estimation and prediction of travel demand.

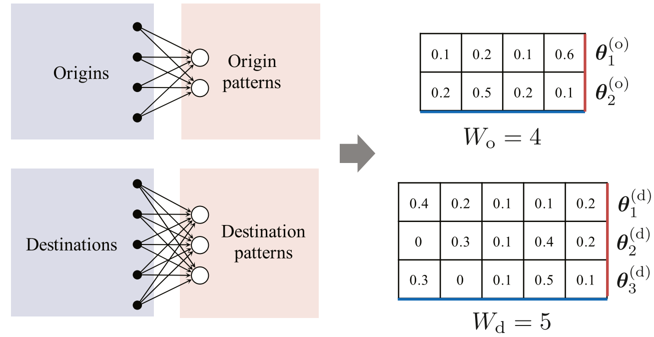


Fig. 2. Example of $\Theta^{(\mathrm{o})}$ and $\Theta^{(\mathrm{d})}$ when $W_\mathrm{o} = 4$, $W_\mathrm{d} = 5$, $K_\mathrm{o} = 2$, and $K_\mathrm{d} = 3$.

A. Probabilistic Travel Demand Model

A set of observed trip data $\mathcal{X} \triangleq \{\mathsf{x}_i; i = 1, 2, \ldots, n\}$ is composed with realizations of the independent and identically distributed (i.i.d.) random trips $\mathsf{x}_i \triangleq [\mathsf{o}_i\ \mathsf{d}_i\ \mathsf{t}_i]^\mathrm{T}$ having probability mass function (PMF) $f_\mathsf{x}(\boldsymbol{x})$. The random variables $\mathsf{o}_i, \mathsf{d}_i$, and t_i respectively represent the indices of the origin, destination, and departure time of the i-th trip such that $\mathsf{o}_i \in \{1, 2, \ldots, W_\mathrm{o}\}, \mathsf{d}_i \in \{1, 2, \ldots, W_\mathrm{d}\}$, and $\mathsf{t}_i \in \{1, 2, \ldots, W_\mathrm{t}\}$. The origin and destination are locations where each trip begins and ends. For example, they are specific locations such as bus stops, zones, and cities. The i-th trip data realization is expressed as

$$\boldsymbol{x}_i = [o_i\ d_i\ t_i]^\mathrm{T}, \qquad \forall i = 1, 2, \ldots, n. \tag{1}$$

The spatial and temporal probability distribution of travel demand can be expressed as a $W_\mathrm{o} \times W_\mathrm{d} \times W_\mathrm{t}$ tensor $\boldsymbol{V}$. Each element of $\boldsymbol{V}$ is defined as

$$[\boldsymbol{V}]_{c_\mathrm{o}, c_\mathrm{d}, c_\mathrm{t}} = f_\mathsf{x}\left([c_\mathrm{o}\ c_\mathrm{d}\ c_\mathrm{t}]^\mathrm{T}\right) \tag{2}$$

where $c_\mathrm{o}, c_\mathrm{d}$, and c_t are the indices of the origin, destination, and time of $\boldsymbol{V}$, respectively. The tensor $\boldsymbol{V}$ is composed with W_t OD matrices, and each OD matrix satisfies $\sum_{c_\mathrm{o}=1}^{W_\mathrm{o}} \sum_{c_\mathrm{d}=1}^{W_\mathrm{d}} [\boldsymbol{V}]_{c_\mathrm{o}, c_\mathrm{d}, c_\mathrm{t}} = 1, \forall c_\mathrm{t}$.

We aim to discover the probabilistic OD basis patterns inherent in the trip data $\mathcal{X}$ together with their probabilistic temporal interactions based on a probabilistic factorization approach. To this end, we design a modeling method to satisfy the following conditions for the spatial basis patterns and temporal interactions: i) the sum of the elements of each spatial basis pattern and temporal interaction is 1 and ii) each element is larger than and equal to 0. To satisfy these conditions for achieving the desired result, we employ a probabilistic factorization approach. This approach allows to capture the complicated dependence and high order interactions among origins and destinations. In addition, the revealed OD basis patterns and the temporal interactions are utilized in future travel demand estimation to reduce the computation load significantly, which will be discussed in the following section.

Latent class models are applied in this paper to infer the distribution $f_\mathsf{x}(\boldsymbol{x})$ for establishing a connection between observed multivariate categorical data and a set of latent classes. This connection is essential to capture the interactions between OD basis pattern. To extract meaningful patterns from high-dimensional data, a common approach is to aggregate data into lower-dimensional structures using categorical distribution [55], [56]. Following this approach, the distribution f_x is modeled as a categorical distribution with a parameter Θ, i.e., $g_\mathsf{x}(\boldsymbol{x}; \Theta)$ with

$$\Theta \triangleq \left[\Theta^{(\mathrm{o})}\ \ \Theta^{(\mathrm{d})}\right] \tag{3}$$

where $\Theta^{(\mathrm{o})}$ and $\Theta^{(\mathrm{d})}$ denote the OD basis patterns, respectively. Let K_o and K_d denote the numbers of origin basis patterns and destination basis patterns, respectively, extracted from the observed trip data $\mathcal{X}$. Consequently, the matrices $\Theta^{(\mathrm{o})}$ and $\Theta^{(\mathrm{d})}$ have dimensions $K_\mathrm{o} \times W_\mathrm{o}$ and $K_\mathrm{d} \times W_\mathrm{d}$, respectively. Each row of $\Theta^{(\mathrm{o})}$ and $\Theta^{(\mathrm{d})}$ indicates an origin pattern and a destination pattern, respectively, representing a probability distribution of a trip occurrence that accounts for the importance of the locations. Specifically, the k_o-th row of $\Theta^{(\mathrm{o})}$ is the k_o-th origin pattern and denoted as $\boldsymbol{\theta}_{k_\mathrm{o}}^{(\mathrm{o})}, k_\mathrm{o} = 1, 2, \ldots, K_\mathrm{o}$. Each element of $\boldsymbol{\theta}_{k_\mathrm{o}}^{(\mathrm{o})}$, which is denoted as $\theta_{c_\mathrm{o}, k_\mathrm{o}}^{(\mathrm{o})}$, represents how important the c_o-th cell is in the k_o-th origin pattern. Similarly, $\theta_{c_\mathrm{d} k_\mathrm{d}}^{(\mathrm{d})}$ is the element of $\boldsymbol{\theta}_{k_\mathrm{d}}^{(\mathrm{d})}$. Since each pattern is a probability distribution, we have $\sum_{c_\mathrm{o}=1}^{W_\mathrm{o}} \theta_{c_\mathrm{o} k_\mathrm{o}}^{(\mathrm{o})} = 1$ and $\sum_{c_\mathrm{d}=1}^{W_\mathrm{d}} \theta_{c_\mathrm{d} k_\mathrm{d}}^{(\mathrm{d})} = 1$. For example, consider a set of trip data composed with four origins and five destinations ($W_\mathrm{o} = 4$ and $W_\mathrm{d} = 5$). Since we want to find macroscopic movement patterns based on the trip data, the number of OD basis patterns is set to be smaller than that of the origins and destinations, e.g. $K_\mathrm{o} = 2$ and $K_\mathrm{d} = 3$ in Fig 2. Specifically, the number of basis patterns determines the capacity of the model. More basis patterns can describe the trip patterns in more detail, which may help to better fit to the data but makes difficult to interpret the modeling result and requires higher computation loads.

A core tensor Π is defined as temporal interaction information among $\Theta^{(\mathrm{o})}$ and $\Theta^{(\mathrm{d})}$. Given that there are K_o origin patterns and K_d destination patterns, this results in $K_\mathrm{o} K_\mathrm{d}$

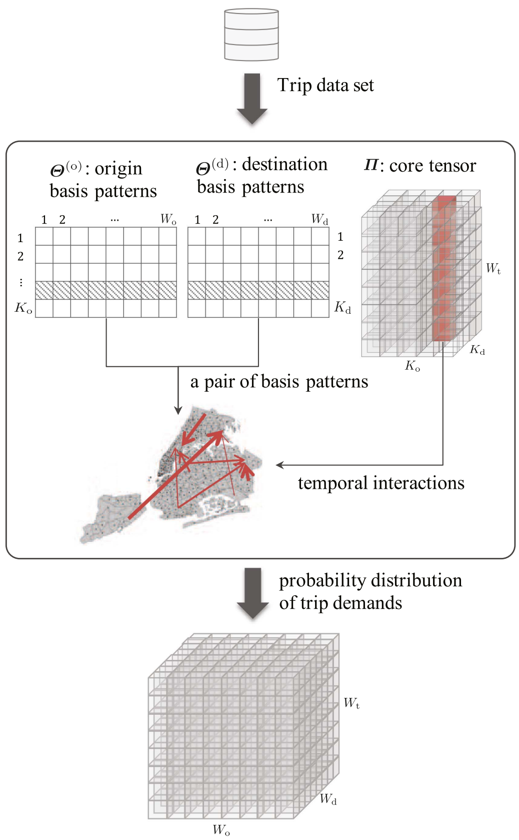


Fig. 3. Probabilistic travel demand modeling.

possible interactions for each time. Over W_t time instants, the number of possible interactions is $K_\mathrm{o}K_\mathrm{d}W_\mathrm{t}$, and the dimension of $\boldsymbol{\Pi}$ is $K_\mathrm{o} \times K_\mathrm{d} \times W_\mathrm{t}$. For each time index c_t, the probability that a trip belongs to the k_o-th origin pattern and the k_d-th destination pattern is defined as $\pi_{c_\mathrm{t}k_\mathrm{o}k_\mathrm{d}}$, which satisfies $\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\pi_{c_\mathrm{t}k_\mathrm{o}k_\mathrm{d}} = 1$.

The occurrence probability of $\mathsf{x}_i = \boldsymbol{x}_i$ is modeled by a parameterized function as

$$g_\mathsf{x}(\boldsymbol{x}_i;\boldsymbol{\Theta}) = \mathbb{P}\{\mathsf{x}_i = \boldsymbol{x}_i;\boldsymbol{\Theta}\} = \sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\pi_{t_ik_\mathrm{o}k_\mathrm{d}}\theta^{(\mathrm{o})}_{o_ik_\mathrm{o}}\theta^{(\mathrm{d})}_{d_ik_\mathrm{d}} \tag{4}$$

which depends on $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$. In (4), $\theta^{(\mathrm{o})}_{o_ik_\mathrm{o}}$ is the probability of $\mathsf{o}_i = o_i$ given that o_i belongs to the k_o-th origin pattern, and $\theta^{(\mathrm{d})}_{d_ik_\mathrm{d}}$ is the probability of $\mathsf{d}_i = d_i$ given that d_i belongs to the k_d-th destination pattern. The proposed probabilistic modeling of travel demand is described in Fig. 3. The temporal occurrence probability of trips belonging to the k_o-th origin pattern and the k_d-th destination pattern is obtained by using the k_o-th origin pattern from $\boldsymbol{\Theta}^{(\mathrm{o})}$, k_d-th destination pattern from $\boldsymbol{\Theta}^{(\mathrm{d})}$, and the corresponding temporal interactions from $\boldsymbol{\Pi}$. Once we have the patterns in $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ as well as their temporal interactions in the core tensor $\boldsymbol{\Pi}$, the tensor $\boldsymbol{V}$ in (2) can be obtained as

$$\boldsymbol{V} \approx \left(\boldsymbol{\Theta}^{(\mathrm{o})}\right)^\mathrm{T}\boldsymbol{\Pi}\,\boldsymbol{\Theta}^{(\mathrm{d})}. \tag{5}$$

B. Model Inference

To find $\boldsymbol{\Theta}$ that maximizes the occurrence probability of $\mathsf{x}_i = \boldsymbol{x}_i$ in (4), we estimate $\boldsymbol{\Theta}$ with respect to the maximum likelihood criterion. For tractable maximum likelihood estimation of $\boldsymbol{\Theta}$ using the expectation-maximization (EM) algorithm, we introduce latent variables $\mathsf{z}_i \in \{[k_\mathrm{o}\ \ k_\mathrm{d}]^\mathrm{T} | k_\mathrm{o} = 1, 2, \ldots, K_\mathrm{o}, k_\mathrm{d} = 1, 2, \ldots, K_\mathrm{d}\}$ on the joint membership across all combinations of the OD basis patterns. From (4), the joint probability of $\mathsf{x}_i = \boldsymbol{x}_i$ and $\mathsf{z}_i = \boldsymbol{z}_i$ for all $i = 1, 2, \ldots, n$ is given by

$$f_{\mathsf{X},\mathsf{Z}}(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta}) = \prod_{i=1}^{n}\prod_{k_\mathrm{o}=1}^{K_\mathrm{o}}\prod_{k_\mathrm{d}=1}^{K_\mathrm{d}}\left[\pi_{t_ik_\mathrm{o}k_\mathrm{d}}\theta^{(\mathrm{o})}_{o_ik_\mathrm{o}}\theta^{(\mathrm{d})}_{d_ik_\mathrm{d}}\right]^{\mathbb{1}_{\mathcal{Z}_{\mathrm{o,d}}}(\boldsymbol{z}_i)} \tag{6}$$

where

$$\mathcal{Z} \triangleq \{\boldsymbol{z}_i : i = 1, 2, \ldots, n\}$$
$$\mathcal{Z}_{\mathrm{o,d}} \triangleq \{\boldsymbol{z}_i \in \mathcal{Z} : \boldsymbol{z}_i = [k_\mathrm{o}\ \ k_\mathrm{d}]^\mathrm{T}\}.$$

The EM algorithm consists of two main steps: the expectation step (E-step) and maximization step (M-step). The EM algorithm first initializes $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$ randomly. Then, E-step and M-step successively approximate $\boldsymbol{\Theta}$ until the convergence criterion $|\boldsymbol{\Theta}^{[m]} - \boldsymbol{\Theta}^{[m-1]}| < \epsilon$ is met, where $\boldsymbol{\Theta}^{[m]}$ is an updated $\boldsymbol{\Theta}$ at the m-th iteration, and ϵ is the convergence tolerance.

- **E-step:** compute the expected log-likelihood of $\boldsymbol{\Theta}$, denoted by $Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{[m-1]})$, as follows

$$Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{[m-1]}\right) \tag{7}$$
$$= \mathbb{E}_{\mathsf{Z}|\mathsf{X};\boldsymbol{\Theta}^{[m-1]}}\left\{\log f_{\mathsf{X},\mathsf{Z}}\left(\mathcal{X},\mathcal{Z};\boldsymbol{\Theta}\right)|\mathcal{X};\boldsymbol{\Theta}^{[m-1]}\right\}$$
$$= \sum_{i=1}^{n}\sum_{k_\mathrm{o}=1}^{K_\mathrm{o}}\sum_{k_\mathrm{d}=1}^{K_\mathrm{d}}\left(\gamma_i^{k_\mathrm{o}k_\mathrm{d}}\right)^{[m]} \times \left[\log\pi_{t_ik_\mathrm{o}k_\mathrm{d}} + \log\theta^{(\mathrm{o})}_{o_ik_\mathrm{o}} + \log\theta^{(\mathrm{d})}_{d_ik_\mathrm{d}}\right] \tag{8}$$

where

$$\left(\gamma_i^{k_\mathrm{o}k_\mathrm{d}}\right)^{[m]} \triangleq \mathbb{E}_{\mathsf{z}_i|\mathsf{x}_i;\boldsymbol{\Theta}^{[m-1]}}\left\{\mathbb{1}_{\{\boldsymbol{z}_i:\boldsymbol{z}_i=[k_\mathrm{o}\ k_\mathrm{d}]^\mathrm{T}\}}(\boldsymbol{z}_i)|\boldsymbol{x}_i;\boldsymbol{\Theta}^{[m-1]}\right\} = \mathbb{P}\left\{\boldsymbol{z}_i = [k_\mathrm{o}\ \ k_\mathrm{d}]^\mathrm{T}|\boldsymbol{x}_i;\boldsymbol{\Theta}^{[m-1]}\right\}. \tag{9}$$

- **M-step:** update $\boldsymbol{\Theta}$ given by

$$\boldsymbol{\Theta}^{[m]} = \underset{\boldsymbol{\Theta}}{\operatorname{argmax}}\quad Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{[m-1]})$$
$$\text{subject to} \sum_{c_\mathrm{o}=1}^{W_\mathrm{o}}\theta^{(\mathrm{o})}_{c_\mathrm{o}k_\mathrm{o}} = 1, \quad k_\mathrm{o} = 1, 2, \ldots, K_\mathrm{o}$$
$$\sum_{c_\mathrm{d}=1}^{W_\mathrm{d}}\theta^{(\mathrm{d})}_{c_\mathrm{d}k_\mathrm{d}} = 1, \quad k_\mathrm{d} = 1, 2, \ldots, K_\mathrm{d}. \tag{10}$$

Algorithm 1: Inference of $\boldsymbol{\Theta}$ and $\boldsymbol{\Pi}$.

Require: $\mathcal{X}$
Initialize: $\boldsymbol{\Theta}^{[0]}$, $\boldsymbol{\Pi}^{[0]}$
$m \leftarrow 0$
while $|\boldsymbol{\Theta}^{[m]} - \boldsymbol{\Theta}^{[m-1]}| > \epsilon$ **do**
  $m \leftarrow m + 1$
  Update $\gamma_i^{k_\mathrm{o} k_\mathrm{d}}, \forall i, k_\mathrm{o}, k_\mathrm{d}$ using (11)
  Update $\boldsymbol{\Theta}^{[m]}$ using (12a) and (12b)
  Update $\boldsymbol{\Pi}^{[m]}$ using (13)
end while
Return: $\boldsymbol{\Theta}^{[m]}$, $\boldsymbol{\Pi}^{[m]}$

In (8), $\gamma_i^{k_\mathrm{o} k_\mathrm{d}}$ can be updated by applying the Bayes' theorem as follows,

$$(\gamma_i^{k_\mathrm{o} k_\mathrm{d}})^{[m]} = \frac{(\pi_{t_i k_\mathrm{o} k_\mathrm{d}})^{[m-1]} \theta^{(\mathrm{o})}_{o_i k_\mathrm{o}} \theta^{(\mathrm{d})}_{d_i k_\mathrm{d}}}{\sum_{k'_\mathrm{o}=1}^{K_\mathrm{o}} \sum_{k'_\mathrm{d}=1}^{K_\mathrm{d}} (\pi_{t_i k'_\mathrm{o} k'_\mathrm{d}})^{[m-1]} \theta^{(\mathrm{o})}_{o_i k'_\mathrm{o}} \theta^{(\mathrm{d})}_{d_i k'_\mathrm{d}}} \quad (11)$$

where $(\pi_{t_i k_\mathrm{o} k_\mathrm{d}})^{[m-1]}$ denotes the m-th updated value of $\pi_{t_i k_\mathrm{o} k_\mathrm{d}}$. Notice that the problem in (10) is convex, hence the local optimum is the global optimum. For this reason, the optimal solution for the problem can be found by using Lagrange multiplier method, and the new value of $\boldsymbol{\Theta}^{[m]}$ is updated as

$$\left(\theta^{(\mathrm{o})}_{c_\mathrm{o} k_\mathrm{o}}\right)^{[m]} = \frac{\sum_{i=1}^{n} \sum_{k'_\mathrm{o}=1}^{K_\mathrm{o}} \sum_{k'_\mathrm{d}=1}^{K_\mathrm{d}} \mathbb{1}_{\{(o_i, k'_\mathrm{o}): o_i = c_\mathrm{o}, k'_\mathrm{o} = k_\mathrm{o}\}}(o_i, k'_\mathrm{o}) \, (\gamma_i^{k'_\mathrm{o} k'_\mathrm{d}})^{[m]}}{\sum_{i=1}^{n} \sum_{k'_\mathrm{o}=1}^{K_\mathrm{o}} \sum_{k'_\mathrm{d}=1}^{K_\mathrm{d}} \mathbb{1}_{\{k'_\mathrm{o}: k'_\mathrm{o} = k_\mathrm{o}\}}(k'_\mathrm{o}) \, (\gamma_i^{k'_\mathrm{o} k'_\mathrm{d}})^{[m]}} \quad (12a)$$

$$\left(\theta^{(\mathrm{d})}_{c_\mathrm{d} k_\mathrm{d}}\right)^{[m]} = \frac{\sum_{i=1}^{n} \sum_{k'_\mathrm{o}=1}^{K_\mathrm{o}} \sum_{k'_\mathrm{d}=1}^{K_\mathrm{d}} \mathbb{1}_{\{(d_i, k'_\mathrm{d}): d_i = c_\mathrm{d}, k'_\mathrm{d} = k_\mathrm{d}\}}(d_i, k'_\mathrm{d}) \, (\gamma_i^{k'_\mathrm{o} k'_\mathrm{d}})^{[m]}}{\sum_{i=1}^{n} \sum_{k'_\mathrm{o}=1}^{K_\mathrm{o}} \sum_{k'_\mathrm{d}=1}^{K_\mathrm{d}} \mathbb{1}_{\{k'_\mathrm{d}: k'_\mathrm{d} = k_\mathrm{d}\}}(k'_\mathrm{d}) \, (\gamma_i^{k'_\mathrm{o} k'_\mathrm{d}})^{[m]}} \quad (12b)$$

where $(\theta^{(\mathrm{o})}_{c_\mathrm{o} k_\mathrm{o}})^{[m]}$ and $(\theta^{(\mathrm{d})}_{c_\mathrm{d} k_\mathrm{d}})^{[m]}$ denote the m-th updated values of $\theta^{(\mathrm{o})}_{c_\mathrm{o} k_\mathrm{o}}$ and $\theta^{(\mathrm{d})}_{c_\mathrm{d} k_\mathrm{d}}$, respectively. Since each element of $\boldsymbol{\Pi}$ corresponding to the c_t-th time represents the probability that a trip at the time belongs to the k_o-th origin pattern and the k_d-th destination pattern, each element of $\boldsymbol{\Pi}^{[m]}$ is updated based on (9) as

$$(\pi_{c_\mathrm{t} k_\mathrm{o} k_\mathrm{d}})^{[m]} = \frac{1}{n_{c_\mathrm{t}}} \sum_{i_{c_\mathrm{t}}=1}^{n_{c_\mathrm{t}}} (\gamma_i^{k_\mathrm{o} k_\mathrm{d}})^{[m]} \quad (13)$$

where i_{c_t} denotes the index of the observed trip data $\boldsymbol{x}_i$ that has the departure time t_i belonging to c_t, and $n_{c_\mathrm{t}} = \sum_{c_\mathrm{o}=1}^{W_\mathrm{o}} \sum_{c_\mathrm{d}=1}^{W_\mathrm{d}} \sum_{i=1}^{n} \mathbb{1}_{\{(o_i, d_i, t_i): o_i = c_\mathrm{o}, d_i = c_\mathrm{d}, t_i = c_\mathrm{t}\}}(o_i, d_i, t_i)$. The EM algorithm is summarized in Algorithm 1.

III. Prediction of Travel Demand Probability Distribution

In this section, we aim to show that the proposed travel demand estimation method can be effectively utilized in future travel prediction. Since the temporal interactions are extracted into the tensor $\boldsymbol{\Pi}$ with a reduced dimension of $c_\mathrm{t} \times K_\mathrm{o} \times K_\mathrm{d}$, the proposed decomposition-based method enables to reduce computation load of future travel prediction and improve prediction accuracy. We present a computationally efficient prediction method that can be combined with different prediction algorithms including dynamic mode decomposition (DMD)-based, LSTM-based, and ARIMA-based algorithms.

A. Problem Description

Classical prediction methods in time domain rely on the direct use of the historical data $\boldsymbol{V}_{c_\mathrm{t}}$ as inputs for algorithms. In complex urban environment, however, the dimension of each matrix $\boldsymbol{V}_{c_\mathrm{t}}$ is prohibitively large. This requires a high computation load for prediction, which may cause inaccurate prediction under limited computational resources.

To enable an accurate prediction even for large-scale mobility data, we propose an efficient method for predicting the probability distribution of travel demand. In particular, we only use the extracted temporal interactions $\boldsymbol{\Pi}_{c_\mathrm{t}}$ instead of the full matrices $\boldsymbol{V}_{c_\mathrm{t}}$. This approach is based on the intended feature of our spatiotemporal pattern extraction described in Section II, meaning that the spatial basis patterns are static over time. Specifically, the distribution of future travel demand can be obtained by

$$\tilde{\boldsymbol{V}}_{c_\mathrm{t}} = \left(\boldsymbol{\Theta}^{(\mathrm{o})}\right)^\mathrm{T} \tilde{\boldsymbol{\Pi}}_{c_\mathrm{t}} \boldsymbol{\Theta}^{(\mathrm{d})} \quad (14)$$

where $\tilde{\boldsymbol{\Pi}}_{c_\mathrm{t}}$ is the predicted temporal interactions between K_o origin patterns and K_d destination patterns. Based on (14), we can utilize different prediction algorithms (i.e., DMD-based, LSTM-based, and ARIMA-based algorithms) for predicting $\tilde{\boldsymbol{\Pi}}_{c_\mathrm{t}}$ using the matrices $\boldsymbol{\Pi}_{c_\mathrm{t}}, c_\mathrm{t} = 1, 2, \ldots, W_\mathrm{t}$ each with the size of $K_\mathrm{o} \times K_\mathrm{d}$ such that $K_\mathrm{o} K_\mathrm{d} \ll W_\mathrm{o} W_\mathrm{d}$.

B. Dynamic Mode Decomposition-Based Algorithm

DMD is a popular tool of data-driven matrix decomposition technique that is developed using Koopman operator theory [57]. This consists in a modal decomposition algorithm that belongs to the family of singular value decomposition (SVD). This algorithm has a strength in providing insights into the underlying dynamics of the data by extracting spatiotemporal interactions. By identifying dynamic modes and corresponding eigenvalues, it is possible to characterize the complex behavior of sequential travel demands.

Among the temporal interaction matrices $\boldsymbol{\Pi}_{c_\mathrm{t}}$, h matrices are used as inputs for the DMD, where $h \le W_\mathrm{t}$. The vectorized form of $\boldsymbol{\Pi}_{c_\mathrm{t}}$ is denoted by $\boldsymbol{y}_{c_\mathrm{t}} = \mathrm{vec}(\boldsymbol{\Pi}_{c_\mathrm{t}}) \in [0, 1]^{K_\mathrm{o} K_\mathrm{d}}$. The h temporal interaction matrices can be expressed by the matrix $\boldsymbol{Y}$ as

$$\boldsymbol{Y} = [\, \boldsymbol{y}_1 \, \boldsymbol{y}_2 \cdots \boldsymbol{y}_h \,]. \quad (15)$$

Algorithm 2: Estimation of the Operator Matrix $\boldsymbol{A}$.

Require: $\boldsymbol{Y}_1, \boldsymbol{Y}_2$
1: Apply SVD of $\boldsymbol{Y}_1$ to get $\boldsymbol{P}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{Q}^{\mathrm{T}}$ using (18)
2: Compute $\tilde{\boldsymbol{A}}$ using (20)
3: Compute the eigendecomposition of $\tilde{\boldsymbol{A}}$ to get $\boldsymbol{\Lambda}$ and $\boldsymbol{W}$ using (21)
4: Compute $\boldsymbol{\Phi}$ using (22)
Return: $\boldsymbol{\Phi}$, $\boldsymbol{\Lambda}$

From (15), we can generate two matrices $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, which can be defined:

$$\boldsymbol{Y}_1 = [\, \boldsymbol{y}_1\, \boldsymbol{y}_2 \cdots \boldsymbol{y}_{h-1} \,] \tag{16a}$$

$$\boldsymbol{Y}_2 = [\, \boldsymbol{y}_2\, \boldsymbol{y}_3 \cdots \boldsymbol{y}_{h} \,]. \tag{16b}$$

The goal of DMD is to estimate the linear operator matrix $\boldsymbol{A}$ with the size of $(K_{\mathrm{o}}K_{\mathrm{d}} \times K_{\mathrm{o}}K_{\mathrm{d}})$ that satisfies the following approximation:

$$\boldsymbol{Y}_2 \approx \boldsymbol{A}\,\boldsymbol{Y}_1. \tag{17}$$

The matrix $\boldsymbol{A}$ can be obtained by solving a least square problem $\min_{\boldsymbol{A}} ||\boldsymbol{Y}_2 - \boldsymbol{A}\boldsymbol{Y}_1||_F^2$. However, in many practical applications, the dimension of OD matrices is large, which makes it difficult to directly solve the least square problem. To overcome this challenge, the DMD algorithm finds the eigenvectors and eigenvalues of the matrix $\boldsymbol{A}$ using Koopman approximation [58].

By applying SVD of $\boldsymbol{Y}_1$ as [59],

$$\boldsymbol{Y}_1 \approx \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^{\mathrm{T}} \tag{18}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{r\times r}$ is a diagonal matrix containing the r dominant singular values of $\boldsymbol{Y}_1$, while $\boldsymbol{P} \in \mathbb{R}^{K_{\mathrm{o}}K_{\mathrm{d}}\times r}$ and $\boldsymbol{Q} \in \mathbb{R}^{(r-1)\times r}$ are orthogonal matrices representing the left and right singular vectors, respectively. The low rank approximation in (18) is used to reduce the computational complexity from using large dimensional trip data. From (17) and using (18), $\boldsymbol{A}$ can be represented as

$$\boldsymbol{A} \approx \boldsymbol{Y}_2\,\boldsymbol{Q}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{P}^{\mathrm{T}}. \tag{19}$$

From (19), the low-rank matrix $\tilde{\boldsymbol{A}}$ can be obtained by projection of $\boldsymbol{A}$ as

$$\tilde{\boldsymbol{A}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\,\boldsymbol{P} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{Y}_2\,\boldsymbol{Q}\,\boldsymbol{\Sigma}^{-1}. \tag{20}$$

Then, we can compute the eigendecomposition of $\tilde{\boldsymbol{A}}$ as

$$\tilde{\boldsymbol{A}}\,\boldsymbol{W} = \boldsymbol{W}\boldsymbol{\Lambda} \tag{21}$$

where $\boldsymbol{W}$ is the eigenvector matrix and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. A dynamic mode matrix $\boldsymbol{\Phi}$, which contains eigenvectors of $\boldsymbol{A}$, can be obtained as [57]

$$\boldsymbol{\Phi} = \boldsymbol{Y}_2\,\boldsymbol{Q}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{W}. \tag{22}$$

The main steps of DMD algorithm that estimates $\boldsymbol{A}$ are summarized in Algorithm 2.

Using the dynamic mode matrix $\boldsymbol{\Phi}$ obtained from Algorithm 2 based on $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_l$, the future temporal interaction at the c_{t}-th time can be estimated as

$$\tilde{\boldsymbol{y}}_{c_{\mathrm{t}}} = \boldsymbol{\Phi}\,\boldsymbol{\Omega}_{c_{\mathrm{t}}-1}\boldsymbol{b} \tag{23}$$

where $\boldsymbol{\Omega}_{c_{\mathrm{t}}} = \xi_{c_{\mathrm{t}}}\mathrm{diag}(e^{\lambda_1 c_{\mathrm{t}}}, e^{\lambda_2 c_{\mathrm{t}}}, \ldots, e^{\lambda_r c_{\mathrm{t}}})$, $\boldsymbol{b} = \boldsymbol{\Phi}^{\dagger}\boldsymbol{y}_1$, and $c_{\mathrm{t}} > h$. The matrix $\boldsymbol{\Phi}^{\dagger}$ denotes the Moore-Penrose inverse of $\boldsymbol{\Phi}$. The eigenvalues in $\boldsymbol{\Lambda}$ are denoted by $\lambda_1, \lambda_1, \ldots, \lambda_r$. The normalization factor $\xi_{c_{\mathrm{t}}}$ ensures that $\mathbf{1}^{\mathrm{T}}\tilde{\boldsymbol{y}}_{c_{\mathrm{t}}} = 1$, where $\mathbf{1}$ is the one-vector. From $\tilde{\boldsymbol{y}}_{c_{\mathrm{t}}}$, the estimated temporal interaction $\tilde{\boldsymbol{\Pi}}_{c_{\mathrm{t}}}$ between $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ at the c_{t}-th time can be obtained.

C. Long Short-Term Memory-Based Algorithm

To predict the future temporal interactions, we use the LSTM network, which is capable of learning the nonlinear dynamics of sequential data in a form of recurrent neural network (RNN). The tensor $\boldsymbol{\Pi}$ not only relates the OD basis patterns in complex space domain, but also involves nonlinear dynamics in time domain. In this regard, the LSTM network can be effectively used to learn both the spatial relations and the temporal interactions in $\boldsymbol{\Pi}$ by exploiting neural networks with both long-term and short-term memories.

For efficient prediction of complex travel demand distribution, the LSTM-based algorithm uses a neural network that consists of multiple hidden LSTM layers and fully connected layers. The number of units per LSTM layer affects how well the model describes the spatial relation inherent in each matrix $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$, while the number of hidden LSTM layers affects the ability to capture hierarchical temporal structure of the tensor $\boldsymbol{\Pi}$. For the given input sequence $\boldsymbol{Y}$ in (15), the LSTM network outputs the predicted temporal interaction at a desired time.

D. Autoregressive Integrated Moving Average-Based Algorithm

Since travel demand is non-stationary in many complex environments, an ARIMA model is adopted to capture the dynamic temporal interactions in $\boldsymbol{\Pi}$. An ARIMA model is expressed as ARIMA(p, l, q) where the parameters p, l, and q denote the order of the autoregressive model, the degree of differencing, and the order of the moving-average model, respectively. For given temporal interactions $\pi_{c_{\mathrm{t}},k_{\mathrm{o}},k_{\mathrm{d}}}, c_{\mathrm{t}} = 1, 2, \ldots, h$, an ARIMA model is given in [60] as

$$\left(\mathbb{I}_0 - \sum_{i=1}^{p}\alpha_i\mathbb{S}_{\mathrm{b}}^i\right)(\mathbb{I}_0 - \mathbb{S}_{\mathrm{b}})^l\,\pi_{c_{\mathrm{t}},k_{\mathrm{o}},k_{\mathrm{d}}} = \zeta + \left(\mathbb{I}_0 + \sum_{i=1}^{q}\beta_i\mathbb{S}_{\mathrm{b}}^i\right)\varepsilon_{c_{\mathrm{t}}} \tag{24}$$

for given k_{o} and k_{d}, where the identity operator and backward shift operator are defined for data sequences in time domain as

$$\mathbb{I}_0\,\pi_{c_{\mathrm{t}},\mathrm{k_o},k_{\mathrm{d}}} = \pi_{c_{\mathrm{t}},k_{\mathrm{o}},k_{\mathrm{d}}} \tag{25a}$$

$$\mathbb{S}_{\mathrm{b}}\,\pi_{c_{\mathrm{t}},\mathrm{k_o},k_{\mathrm{d}}} = \pi_{c_{\mathrm{t}}-1,k_{\mathrm{o}},k_{\mathrm{d}}}. \tag{25b}$$

Applying the backward shift operator l times to $\pi_{c_{\mathrm{t}},\mathrm{k_o},k_{\mathrm{d}}}$ shifts the data backward as $\mathbb{S}_{\mathrm{b}}^l\,\pi_{c_{\mathrm{t}},\mathrm{k_o},k_{\mathrm{d}}} = \pi_{c_{\mathrm{t}}-l,k_{\mathrm{o}},k_{\mathrm{d}}}$. Using (25a) and (25b), the l-th order difference in (24) can be expressed recursively as

$$(\mathbb{I}_0 - \mathbb{S}_\mathrm{b})^l \, \pi_{c_\mathrm{t},k_\mathrm{o},k_\mathrm{d}} = (\mathbb{I}_0 - \mathbb{S}_\mathrm{b})^{l-1} \, (\mathbb{I}_0 - \mathbb{S}_\mathrm{b}) \, \pi_{c_\mathrm{t},k_\mathrm{o},k_\mathrm{d}} \tag{26a}$$

$$= (\mathbb{I}_0 - \mathbb{S}_\mathrm{b})^{l-1} \left(\pi_{c_\mathrm{t},k_\mathrm{o},k_\mathrm{d}} - \pi_{c_\mathrm{t}-1,k_\mathrm{o},k_\mathrm{d}}\right). \tag{26b}$$

In (24), α_i and β_i are the parameters of the autoregressive part and of the moving average part, respectively, while ζ is a constant. The noise $\varepsilon_{c_\mathrm{t}}$ is assumed to follow the independent Gaussian distribution with zero mean and unit variance. To determine the parameters p, l, and q, the step-wise algorithm in [60] is used for a given $\boldsymbol{\Pi}$. Then the optimal parameters α_i, β_i, and ζ are computed to maximize the likelihood for given p, l, and q.

IV. Case Study

This section provides results determined by applying the proposed demand estimation method to two large-scale taxi trip datasets.

A. Data Description

We use two datasets of taxi trip records in New York city: (i) the yellow taxi trip data; and (ii) the high-volume for-hire vehicles (HVFHVs) trip data. The Taxi and Limousine Commission provides access to the taxi trip data starting from 2009 [61]. These data have been frequently used in various works for urban mobility analysis and future demand prediction [62], [63]. Yellow taxis represent the conventional taxi service, whereas HVFHVs provide pre-arranged trip services that include Uber and Lyft. These two trip datasets provide a variety of information including: origin; destination; departure and arrival date/time; trip distance; fares; and payment types. The origins and destinations are identified by 265 taxi zones, of which 69 in Manhattan. We use origin zone ID, destination zone ID, and departure time among the data attributes in both datasets.

B. Identification of Trip Patterns

A single trip is represented by a triplet composed with origin zone ID, destination zone ID, and departure time as in (1). Applying the demand modeling method proposed in the previous section to real-world trip data of New York city, we identify basis patterns and temporal interactions and model the trip occurrence probability of the high-dimensional and large-scale urban travel demand. The modeling period spans over 10 days, from March 5-th to March 14-th. In such period, the number of trips by yellow taxis is 2,701,464 and that by HVFHVs is 7,706,061. Within this timeframe, the yellow taxi data encompass 28,102 observed OD pairs that have passengers, with a minimum number of passengers of 1 and a maximum number of 119. Meanwhile, the HVFHVs trip data includes 55,635 observed OD pairs, with a minimum number of passengers of 1 and a maximum number of 49. The time domain is aggregated in 30-minute intervals, which results in a total of $265 \times 265 \times 48 \times 10 = 33,708,000$ possible combinations. The convergence tolerance ϵ of Algorithm 1 is set to 10^{-6}.

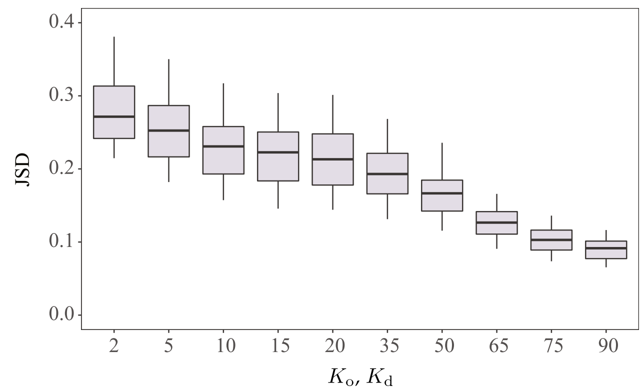


Fig. 4. Impact of the number of basis patterns on modeling accuracy.

1) Performance Evaluation: The accuracy of the proposed model in (14) is evaluated by comparing it with the probability distribution given by the sample mean of the taxi trip data such that

$$[\boldsymbol{V}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}} \approx \frac{1}{\sum_{c'_\mathrm{o}=1}^{W_\mathrm{o}} \sum_{c'_\mathrm{d}=1}^{W_\mathrm{d}} [\boldsymbol{V}']_{c'_\mathrm{o},c'_\mathrm{d},c_\mathrm{t}}} [\boldsymbol{V}']_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}} \tag{27}$$

where

$$[\boldsymbol{V}']_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}} = \sum_{i=1}^{n} \mathbb{1}_{\{(o_i,d_i,t_i):o_i=c_\mathrm{o},d_i=c_\mathrm{d},t_i=c_\mathrm{t}\}}(o_i,d_i,t_i).$$

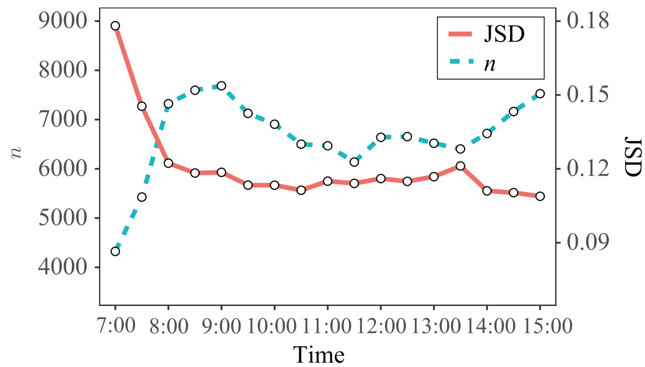
The Jensen-Shannon divergence (JSD) between two probability distributions (14) and (27) is used as a metric to determine the model accuracy, specifically

$$\eta(\boldsymbol{V}_{c_\mathrm{t}}, \tilde{\boldsymbol{V}}_{c_\mathrm{t}}) = \frac{1}{2} \sum_{c_\mathrm{o}=1}^{W_\mathrm{o}} \sum_{c_\mathrm{d}=1}^{W_\mathrm{d}} [\boldsymbol{V}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}} \log\left(\frac{[\boldsymbol{V}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}}}{[\boldsymbol{M}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}}}\right) + \frac{1}{2} \sum_{c_\mathrm{o}=1}^{W_\mathrm{o}} \sum_{c_\mathrm{d}=1}^{W_\mathrm{d}} [\tilde{\boldsymbol{V}}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}} \log\left(\frac{[\tilde{\boldsymbol{V}}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}}}{[\boldsymbol{M}]_{c_\mathrm{o},c_\mathrm{d},c_\mathrm{t}}}\right) \tag{28}$$

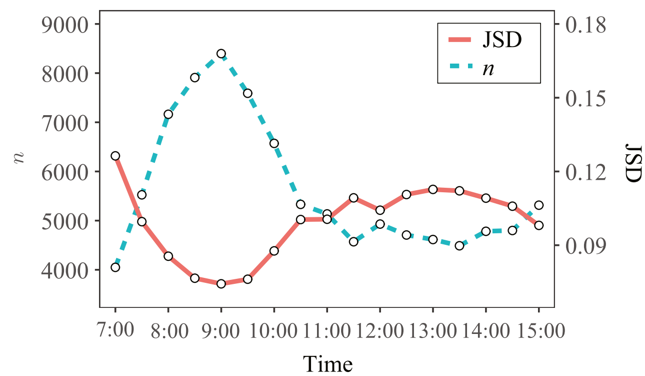
where $\boldsymbol{M} = (\boldsymbol{V} + \tilde{\boldsymbol{V}})/2$. The JSD in (28) measures the distance between the distributions $\boldsymbol{V}_{c_\mathrm{t}}$ and $\tilde{\boldsymbol{V}}_{c_\mathrm{t}}$. It is employed to assess the extent to which $\tilde{\boldsymbol{V}}_{c_\mathrm{t}}$ diverges from $\boldsymbol{V}_{c_\mathrm{t}}$. The JSD is non-negative and symmetric; it becomes zero if and only if two distributions are identical.

Fig. 4 shows how the accuracy of the proposed model changes according to the number of basis patterns. It can be noticed that the JSD decreases with the number of basis patterns. Each box plot has modeling results for all time periods. Since the number of basis patterns represents the variety of sources available for describing travel demands, having a large number of basis patterns enhances the model's ability to describe travel demands in detail.

Fig. 5 shows the impact of the number of trip data n used for modeling to the model accuracy using yellow taxi data and HVFHVs trip data. The modeling target area in Fig. 5(a) is New York city, while that in Fig. 5(b) is Manhattan. In both cases, K_o and K_d are set to 10. In Fig. 5(a), the demand of yellow taxis in New York city rises near 8,000 at 9:00 am and then declines, but



(a) Yellow taxi trip modeling - New York city



(b) HVFHVs trip modeling - Manhattan

Fig. 5. Temporal changes in modeling accuracy and the number of trip data n.

remains above 6,000 during the morning and the early afternoon. In Fig. 5(b), the demand of HVFHVs in Manhattan also rises near 8,000 at 9:00 am and then declines sharply. These two results show that the model accuracy changes in the opposite trend to the changes in n. The JSD decreases with the number of data used. In Fig. 5(b), when more than 8,000 trip data are used, the JSD is below 0.09. Considering New York city has 265 zones and Manhattan has 69, a sufficient number of basis patterns are configured in HVFHVs trip modeling relative to the number of zones. Consequently, this leads to higher modeling accuracy in HVFHVs trip modeling than yellow taxi trip modeling. From Figs. 4 and 5, it can be noticed that the number of basis patterns and the amount of trip data sufficient to describe complex trip behaviors are key factors in improving the accuracy of modeling trip behaviors.

2) Origin and Destination Basis Patterns: To analyze the modeling result in more detail, the OD basis patterns of yellow taxi trips and HVFHVs trips are plotted in Figs. 6 and 7, displaying the basis patterns $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ derived from yellow taxi trip data when $K_{\mathrm{o}} = K_{\mathrm{d}} = 10$. To better visualize the pattern configuration, the values in each column are rescaled to 0-1 so that $\theta'^{(\mathrm{o})}_{c_{\mathrm{o}}k_{\mathrm{o}}} = \theta^{(\mathrm{o})}_{c_{\mathrm{o}}k_{\mathrm{o}}}/\max(\boldsymbol{\theta}^{(\mathrm{o})}_{k_{\mathrm{o}}})$. Note that Manhattan areas exhibit strong patterns in most of $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ in yellow taxi trip data. Divided into various subregions of Upper Manhattan, Lower Manhattan, and Midtown Manhattan, East Manhattan, and West Manhattan, each area contributes to the composition of overall trip patterns. We can find that adjacent neighborhoods in Manhattan are grouped in the same pattern in many cases, which indicates that they share the similar movements. For example, Midtown and Lower Manhattan areas are grouped in $\boldsymbol{\theta}^{(\mathrm{o})}_2$, $\boldsymbol{\theta}^{(\mathrm{d})}_1$, and $\boldsymbol{\theta}^{(\mathrm{d})}_2$. Meanwhile, we can observe that a pattern can include areas that are separately located. For example, trips to and from John F. Kennedy (JFK) International Airport and LaGuardia Airport, which constitute a major movement in New York city, are grouped in $\boldsymbol{\theta}^{(\mathrm{o})}_6$, $\boldsymbol{\theta}^{(\mathrm{o})}_7$, $\boldsymbol{\theta}^{(\mathrm{o})}_9$, $\boldsymbol{\theta}^{(\mathrm{o})}_{10}$, $\boldsymbol{\theta}^{(\mathrm{d})}_6$, and $\boldsymbol{\theta}^{(\mathrm{d})}_{10}$. Additionally, several zones in Brooklyn and Queens, which are near to Manhattan or airports, exhibit distinct patterns in $\boldsymbol{\theta}^{(\mathrm{o})}_9$, $\boldsymbol{\theta}^{(\mathrm{o})}_{10}$, $\boldsymbol{\theta}^{(\mathrm{d})}_9$ and $\boldsymbol{\theta}^{(\mathrm{d})}_{10}$.

Figs. 8 and 9 exhibit clear distinct patterns of HVFHVs trips compared to those observed in yellow taxi trip data. While the basis patterns of yellow taxi trips are predominantly grouped within Manhattan, those of HVFHVs trips extend throughout all five boroughs in New York city, including Manhattan, Brooklyn, Queens, the Bronx, and Staten Island. In March 2019, the percentage of trips having both origin and destination within Manhattan stood at 84.1% for yellow taxi trips and 31.4% for HVFHVs trip trips. This indicates a notable discrepancy, suggesting that Uber and Lyft cater to a larger portion of passengers outside Manhattan, whereas yellow taxis primarily serve within Manhattan. This aligns with what can be inferred from the OD basis patterns shown in Figs. 8 and 9.

All patterns exhibit a group formed by adjacent zones, which indicates that nearby areas share similar trip behaviors, consistent with observations in yellow taxi trip cases shown in Figs. 6 and 7. Additionally, $\boldsymbol{\theta}^{(\mathrm{o})}_4$ and $\boldsymbol{\theta}^{(\mathrm{d})}_4$ encompass Midtown Manhattan areas, LaGuardia airport, and JFK airports, showing a similarity in trip patterns between these regions. For instance, when there is high demand for trips to the airports, there is also increased demand for trips to Midtown Manhattan. In Queens and Brooklyn, each area is divided into regions closed to Manhattan and those that are not. Similarly, within Manhattan itself, lower and upper regions exhibit distinct trip patterns.

3) Temporal Interactions Between Patterns: Fig. 10 shows heat maps of $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$ that are extracted in travel demand modeling using yellow taxi trip data from 4:00 am to 9:00 am. We can find the gradual changes in $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$ over time. To figure out changes in mobility during the time period, we need to look into Figs. 6, 7, and 10 together in a comprehensive way. One of the noticeable changes in $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$ is the decrease in the trips to the regions in $\boldsymbol{\theta}^{(\mathrm{d})}_8$, which is composed of West Manhattan. Trips originated from the regions in $\boldsymbol{\theta}^{(\mathrm{o})}_1$ and arriving to the regions in $\boldsymbol{\theta}^{(\mathrm{d})}_4$ increase. The corresponding trips are from Lower Manhattan to East Manhattan. These trips have consistently many number of passengers during morning. Trips from regions in $\boldsymbol{\theta}^{(\mathrm{o})}_6$ to regions in $\boldsymbol{\theta}^{(\mathrm{d})}_9$, from regions in $\boldsymbol{\theta}^{(\mathrm{o})}_7$ to regions in $\boldsymbol{\theta}^{(\mathrm{d})}_1$, from regions $\boldsymbol{\theta}^{(\mathrm{o})}_9$ to regions in $\boldsymbol{\theta}^{(\mathrm{d})}_{10}$, and from regions in $\boldsymbol{\theta}^{(\mathrm{o})}_8$ to regions in $\boldsymbol{\theta}^{(\mathrm{d})}_3$ are the corresponding trips. Analyzing the temporal changes in travel demand is made easier by the significantly reduced dimension of $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$ compared to the original OD matrix. This reduction enables to capture the major temporal changes. In the

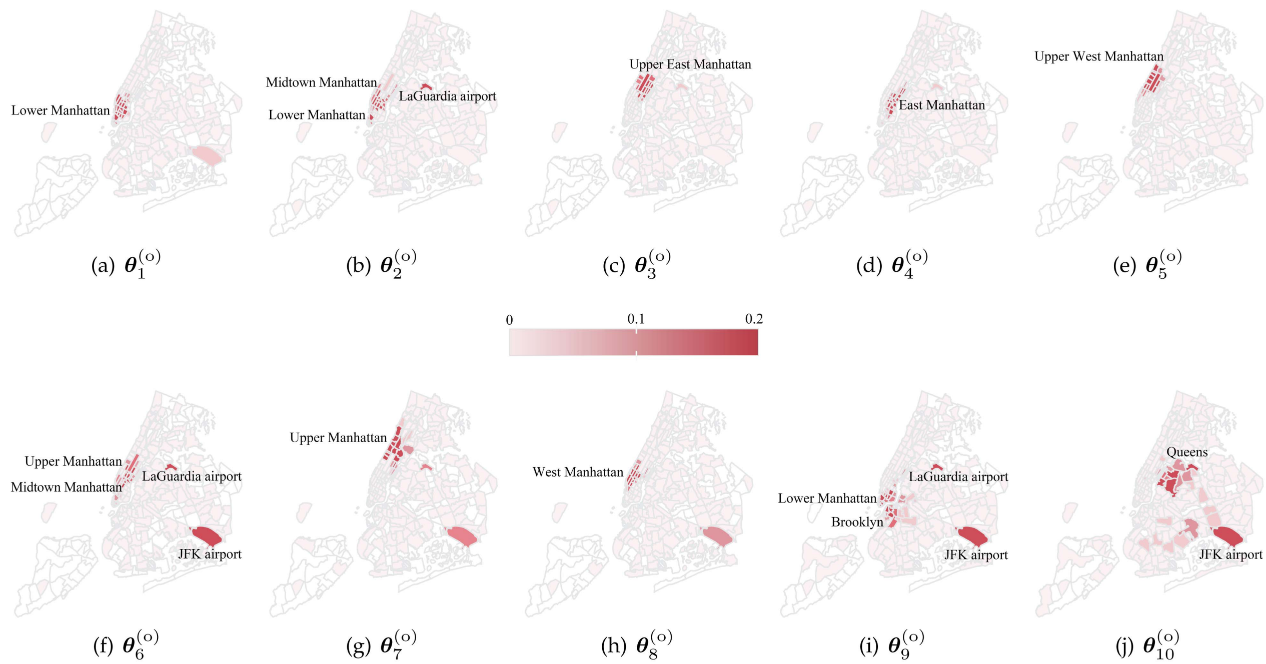


Fig. 6. Geographical representation on $\boldsymbol{\Theta}^{(\mathrm{o})}$ of yellow taxi trip data when $K_{\mathrm{o}} = 10$.

Fig. 7. Geographical representation on $\boldsymbol{\Theta}^{(\mathrm{d})}$ of yellow taxi trip data when $K_{\mathrm{d}} = 10$.

next section, we will explore additional benefits of the reduced dimension of $\boldsymbol{\Pi}_{c_{\mathrm{t}}}$ in future demand prediction.

C. Travel Demand Prediction

From the previous section, it is found that $\boldsymbol{\Pi}$ slowly changes over a period of time. Based on the observed changes, we predict the future $\boldsymbol{\Pi}$ and produce the future OD matrix using $\boldsymbol{\Theta}$ to reduce the computation of predicting the future travel demand probability distribution. The prediction accuracy is measured using the JSD based on the actual future demand distribution in a similar way to how modeling accuracy is measured.

For the DMD-based prediction algorithm and LSTM-based prediction algorithm, h is set to 17 and 4, respectively. The LSTM network is trained and validated using 80% and 20% of 9-day-long data, respectively, and we test the trained the LSTM network using the rest 1-day-long data. The network architecture uses the first two layers of 128 LSTM cells, followed

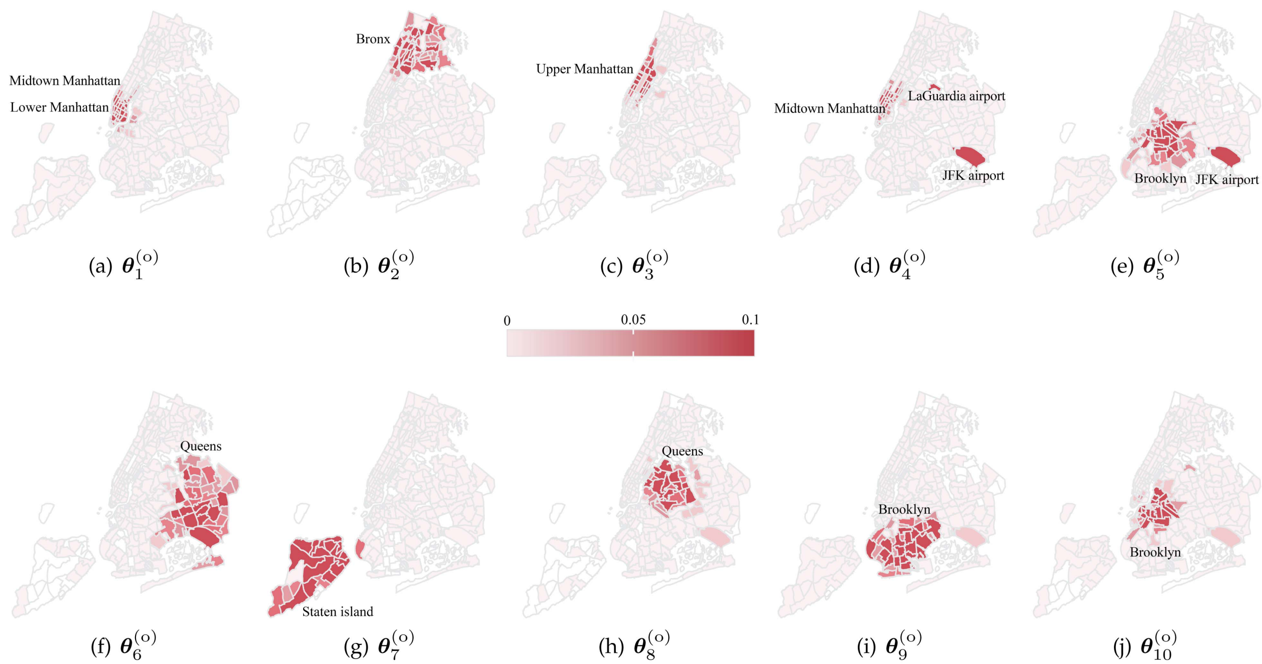
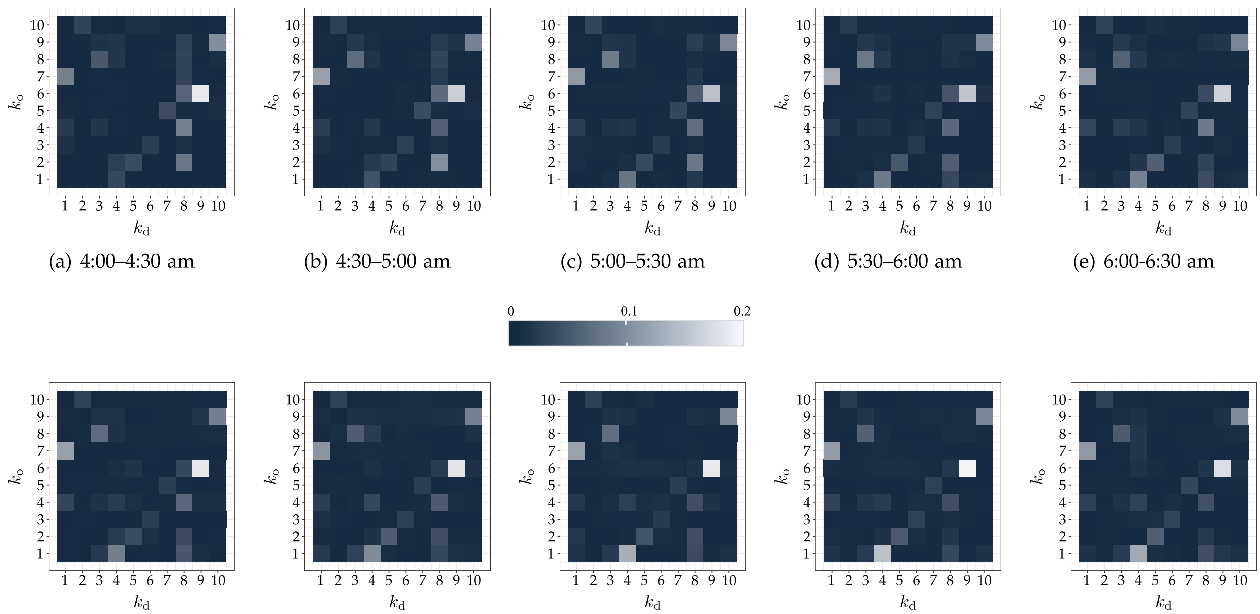


Fig. 8. Geographical representation on $\boldsymbol{\Theta}^{(\mathrm{o})}$ of HVFHVs trip data when $K_{\mathrm{o}} = 10$.

Fig. 9. Geographical representation on $\boldsymbol{\Theta}^{(\mathrm{d})}$ of HVFHVs trip data when $K_{\mathrm{d}} = 10$.

by a fully-connected dense layer of 256 neurons and a final dense output layer containing as many cells as the number of outputs. The LSTM network is trained by the Adam optimizer, employing the mean square error (MSE) as the loss function. The learning rate and number of epochs are set to 0.001 and 1,000, respectively. For the ARIMA-based prediction algorithm, h is set to 29. The parameters p, l, and q for each temporal interaction pair and OD pair are determined using a step-wise algorithm [60].

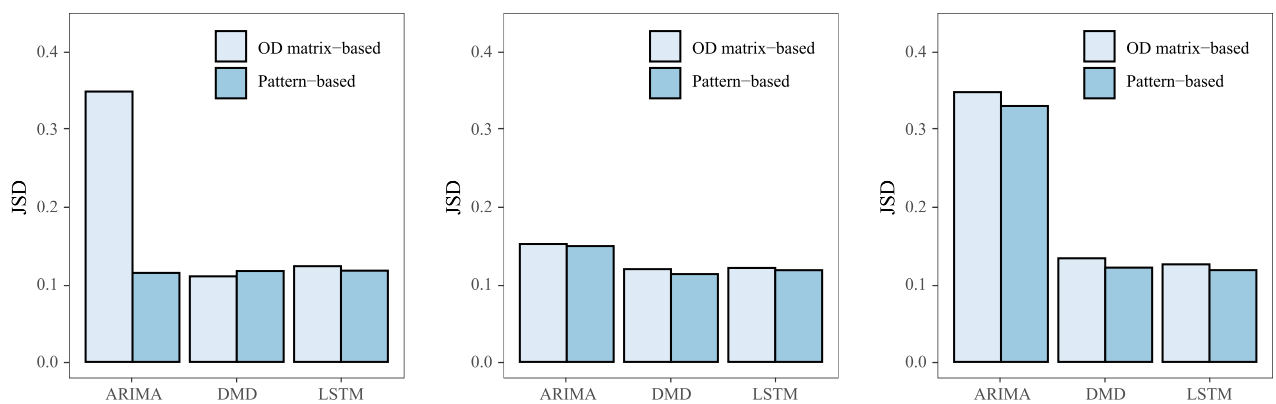
Fig. 11 shows the prediction accuracy when the proposed pattern-based prediction method, combined with ARIMA, DMD, and LSTM, predicts the yellow taxi travel demand in New York city. For the pattern-based prediction method, the travel demand model with $K_{\mathrm{o}} = K_{\mathrm{d}} = 10$ is used. Meanwhile, the baseline method uses the whole OD demand matrices as inputs of the prediction. Since the dimension of the input in the proposed method is only 10×10 and that of the OD matrix-based method is 265×265, the dimension of the input data is only 0.14%



(a) 4:00–4:30 am (b) 4:30–5:00 am (c) 5:00–5:30 am (d) 5:30–6:00 am (e) 6:00-6:30 am

(f) 6:30–7:00 am (g) 7:00–7:30 am (h) 7:30–8:00 am (i) 8:00–8:30 am (j) 8:30–9:00 am

Fig. 10. Grid heat maps of $\boldsymbol{\Pi}$ for $K_{\mathrm{o}} = K_{\mathrm{d}} = 10$.



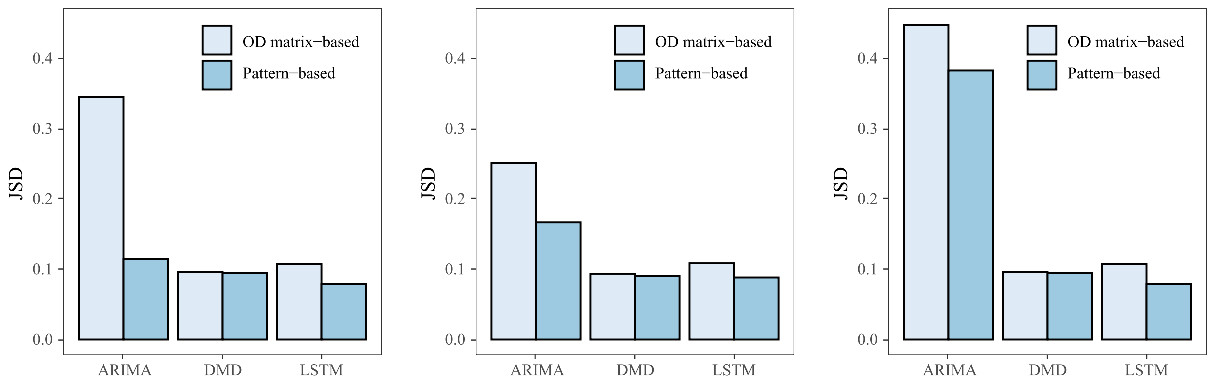
(a) Prediction time horizon = 0.5 hours (b) Prediction time horizon = 3 hours (c) Prediction time horizon = 6 hours

Fig. 11. Prediction accuracy of the proposed pattern-based prediction method using yellow taxi trip data in New York city. The dimension of the input data in the proposed method is only 10×10, while that in OD matrix-based prediction is 265×265.

in the proposed method compared to the conventional prediction method. Nevertheless, the proposed method shows better prediction accuracy in most of the cases. This result indicates that $\boldsymbol{\Theta}^{(\mathrm{o})}$ and $\boldsymbol{\Theta}^{(\mathrm{d})}$ can express complex movements in the reduced dimension. In more detail, the ARIMA shows the worst performance in both prediction methods compared to the DMD and LSTM. While the LSTM shows steady performance across different prediction time horizons, the prediction accuracy of DMD varies on the prediction time horizon. For the short-term prediction in Fig. 11(a), DMD shows the highest prediction accuracy. However, as the prediction time horizon gets longer, the accuracy decreases in Fig. 11(b) and (c).

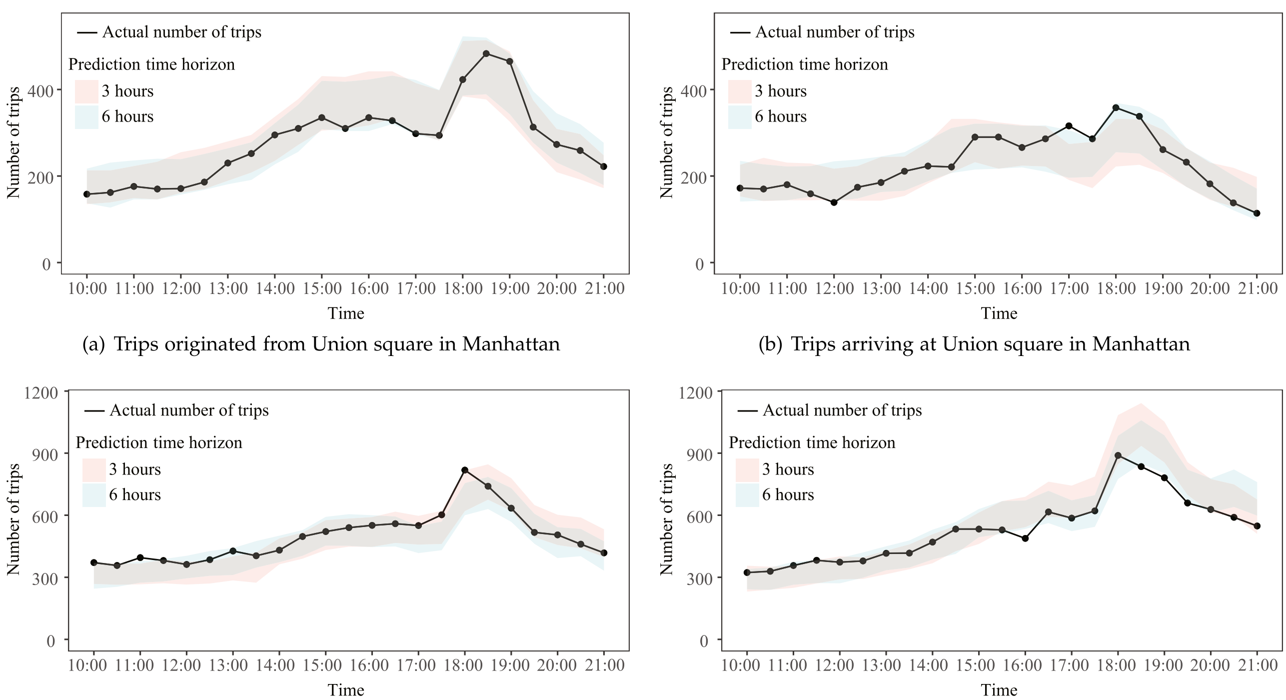
Fig. 12 shows the prediction accuracy result for the HVFHVs travel demand in Manhattan. For the pattern-based prediction method, the travel demand model with $K_{\mathrm{o}} = K_{\mathrm{d}} = 10$ is used. In the baseline prediction method, the dimension of the input is 69×69. In this case, the dimension of the input data is 2.10% in the proposed method compared to the baseline prediction method. The proposed method outperforms the baseline prediction method in all cases, and its performance is more improved compared to Fig. 11. The reason of the improvement in the prediction performance is that the relative number of the patterns used increases compared to the dimension of original OD matrices. In addition, when the proposed prediction method is combined with LSTM, it shows significant improvement and the best performance.

The probability distribution of predicted travel demand can be used as a model for generating future travel demand, considering the variability and uncertainty in travel demand. The model generates numerous travel demand scenarios essential



(a) Prediction time horizon = 0.5 hours (b) Prediction time horizon = 3 hours (c) Prediction time horizon = 6 hours

Fig. 12. Prediction accuracy of the proposed pattern-based prediction method using HVFHVs trip data in Manhattan. The dimension of the input data in the proposed method is only 10×10, while that in OD matrix-based prediction is 69×69.



(a) Trips originated from Union square in Manhattan (b) Trips arriving at Union square in Manhattan

(c) Trips originated from Harlem in Manhattan (d) Trips arriving at Harlem in Manhattan

Fig. 13. Prediction of the number of trips of specific OD pairs in yellow taxi travel demand.

for evaluating transportation operational strategies, such as taxi dispatching and bus scheduling. In Fig. 13, 10,000 trip scenarios per a time period are generated using the probability distribution of yellow taxi travel demand predicted by the proposed LSTM-based prediction model. In each scenario, trips are generated and then distributed across OD pairs according to the predicted probability distribution. Fig. 13 showcases the resulting number of trips for four specific OD pairs, considering prediction time horizons of 3 hours and 6 hours, while the actual number of trips is represented by the black line. Fig. 13(a) and (b) show outcomes for trips originating from and arriving at Union square, located in Midtown Manhattan, while Fig. 13(c) and (d) present results for trips originating from and arriving at Harlem, also located in Upper Manhattan. Observation reveals fluctuations in the numbers of trips from and to both Union square and Harlem over different time periods, with peaks typically occurring at 18:00 or 18:30, followed by a decline. This result demonstrates that the probability distribution of predicted travel demand accurately captures major temporal travel patterns including peak demand periods. Consequently, this validates the accuracy of the predicted future probability distribution, underlining the practicality of the proposed modeling and prediction method.

V. Conclusion

This paper has proposed a novel method for characterizing and predicting the probability distribution of large-scale trip data. The method reveals OD basis patterns and their temporal interactions by applying probabilistic tensor decomposition approach. To infer a probability distribution of travel demand, a latent class model is adopted and an EM algorithm is designed. In future travel demand prediction, the OD basis patterns and their temporal interactions are utilized to enable an accurate prediction for large-scale mobility data. By predicting the future temporal interactions and obtaining the probability distribution of future travel demand, it is shown that the computation load is reduced significantly. The DMD-based, LSTM-based, and ARIMA-based prediction algorithms are combined with the proposed prediction method. We quantify the accuracy of the proposed method for a case study using yellow taxi trip data and HVFHVs trip data of New York city. The impact of factors affecting modeling accuracy was investigated, and it was observed that the modeling accuracy increases as the number of OD basis patterns and the amount of data used increase. In addition, our proposed prediction method outperforms the conventional prediction method in terms of the prediction accuracy while reducing the need of both memory and computation. This high prediction accuracy indicates that the OD basis patterns extracted in the travel demand modeling can capture complicated mobility patterns such as those in urban areas.

References

[1] J. Jeong et al., "Mobility prediction for 5G core networks," *IEEE Commun. Standards Mag.*, vol. 5, no. 1, pp. 56–61, Mar. 2021.

[2] C. A. Gómez-Vega, Z. Liu, C. A. Gutiérrez, M. Z. Win, and A. Conti, "Efficient deployment strategies for network localization with assisting nodes," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 6272–6287, May 2024.

[3] T. Singh, A. Solanki, S. K. Sharma, A. Nayyar, and A. Paul, "A decade review on smart cities: Paradigms, challenges and opportunities," *IEEE Access*, vol. 10, pp. 68 319–68 364, 2022.

[4] M. Batty et al., "Smart cities of the future," *Eur. Phys. J. Special Top.*, vol. 214, pp. 481–518, 2012.

[5] J. Kim, S. Tak, J. Lee, and H. Yeo, "Integrated design framework for on-demand transit system based on spatiotemporal mobility patterns," *Transp. Res. Part C, Emerg. Technol.*, vol. 150, 2023, Art. no. 104087.

[6] Y. Kim, H.-Y. Tak, S. Kim, and H. Yeo, "A hybrid approach of traffic simulation and machine learning techniques for enhancing real-time traffic prediction," *Transp. Res. Part C, Emerg. Technol.*, vol. 160, 2024, Art. no. 104490.

[7] C. Wu, A. R. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, "Flow: A modular learning framework for mixed autonomy traffic," *IEEE Trans. Robot.*, vol. 38, no. 2, pp. 1270–1286, Apr. 2022.

[8] P. Koopman and M. Wagner, "Autonomous vehicle safety: An interdisciplinary challenge," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 90–96, Spring 2017.

[9] C. Wu, A. M. Bayen, and A. Mehta, "Stabilizing traffic with autonomous vehicles," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 6012–6018.

[10] U. Nations, *UNECE Nexus: Sustainable Mobility and Smart Connectivity*. New York, NY, USA: UN, 2021.

[11] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transp. Res. Part C, Emerg. Technol.*, vol. 58, pp. 162–177, 2015.

[12] S. Çolak, A. Lima, and M. C. González, "Understanding congested travel in urban areas," *Nat. Commun.*, vol. 7, no. 1, pp. 1–8, 2016.

[13] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *J. Transport Geography*, vol. 43, pp. 78–90, 2015.

[14] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt, "Detecting the dynamics of urban structure through spatial network analysis," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 11, pp. 2178–2199, 2014.

[15] A. Alsger, A. Tavassoli, M. Mesbah, L. Ferreira, and M. Hickman, "Public transport trip purpose inference using smart card fare data," *Transp. Res. Part C, Emerg. Technol.*, vol. 87, pp. 123–137, 2018.

[16] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transp. Res. Part B, Methodol.*, vol. 91, pp. 511–524, 2016.

[17] D. Alghamdi, K. Basulaiman, and J. Rajgopal, "Multi-stage deep probabilistic prediction for travel demand," *Appl. Intell.*, vol. 52, pp. 11214–11231, 2022.

[18] K. W. Axhausen and T. Gärling, "Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems," *Transport Rev.*, vol. 12, no. 4, pp. 323–341, 1992.

[19] C. R. Bhat and F. S. Koppelman, *Activity-Based Modeling of Travel Demand*. Berlin, Germany: Springer, 1999.

[20] J. Kim and H. S. Mahmassani, "Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories," *Transp. Res. Procedia*, vol. 9, pp. 164–184, 2015.

[21] M. Saberi, H. S. Mahmassani, T. Hou, and A. Zockaie, "Estimating network fundamental diagram using three-dimensional vehicle trajectories: Extending edie's definitions of traffic flow variables to networks," *Transp. Res. Rec.*, vol. 2422, no. 1, pp. 12–20, 2014.

[22] X. Li et al., "Prediction of urban human mobility using large-scale taxi traces and its applications," *Front. Comput. Sci.*, vol. 6, no. 1, pp. 111–121, 2012.

[23] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi–passenger demand using streaming data," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, Sep. 2013.

[24] M. Bierlaire and F. Crittin, "An efficient algorithm for real-time estimation and prediction of dynamic OD tables," *Operations Res.*, vol. 52, no. 1, pp. 116–127, 2004.

[25] J. Barceló Bugeda, L. Montero Mercadé, M. Bullejos, O. Serch, and C. Carmona Bautista, "A Kalman filter approach for the estimation of time dependent od matrices exploiting bluetooth traffic data collection," in *Proc. TRB 91st Annu. Meeting Compendium Papers DVD*, 2012, pp. 1–16.

[26] Z. Zhou, K. Yang, Y. Liang, B. Wang, H. Chen, and Y. Wang, "Predicting collective human mobility via countering spatiotemporal heterogeneity," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 4723–4738, May 2024.

[27] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, "Real-time prediction of taxi demand using recurrent neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 8, pp. 2572–2581, Aug. 2018.

[28] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series extreme event forecasting with neural networks at Uber," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1–5.

[29] Y. Bai, Z. Sun, B. Zeng, J. Deng, and C. Li, "A multi-pattern deep fusion model for short-term bus passenger flow forecasting," *Appl. Soft Comput.*, vol. 58, pp. 669–680, 2017.

[30] H. Yao et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 2588–2595.

[31] T. Djukic, G. Flötteröd, H. Van Lint, and S. Hoogendoorn, "Efficient real time od matrix estimation based on principal component analysis," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst.*, 2012, pp. 115–121.

[32] X. Li, J. Kurths, C. Gao, J. Zhang, Z. Wang, and Z. Zhang, "A hybrid algorithm for estimating origin-destination flows," *IEEE Access*, vol. 6, pp. 677–687, 2018.

[33] J. Kim, A. Conti, and M. Z. Win, "Inferring spatiotemporal mobility patterns from multidimensional trip data," in *Proc. IEEE Int. Conf. Commun.*, Rome, Italy, May 2023, pp. 3358–3363.

[34] D. Skillicorn, *Understanding Complex Datasets: Data Mining With matrix Decompositions*. London, U.K.: Chapman and Hall/CRC, 2007.

[35] Y. Han and F. Moutarde, "Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization," *Int. J. Intell. Transp. Syst. Res.*, vol. 14, pp. 36–49, 2016.

[36] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transp. Res. Part C, Emerg. Technol.*, vol. 28, pp. 15–27, 2013.

[37] Y. Yu, Y. Zhang, S. Qian, S. Wang, Y. Hu, and B. Yin, "A low rank dynamic mode decomposition model for short-term traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6547–6560, Oct. 2021.

[38] J. Reades, F. Calabrese, and C. Ratti, "Eigenplaces: Analysing cities using the space–time structure of the mobile phone network," *Environ. Plan. B, Plan. Des.*, vol. 36, no. 5, pp. 824–836, 2009.

[39] C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Liò, "Collective human mobility pattern from taxi trips in urban area," *PLoS One*, vol. 7, no. 4, 2012, Art. no. e34487.

[40] F. Zhang, D. Wilkie, Y. Zheng, and X. Xie, "Sensing the pulse of urban refueling behavior," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 13–22.
[41] J. Wang, F. Gao, P. Cui, C. Li, and Z. Xiong, "Discovering urban spatio-temporal structure from time-evolving traffic networks," in *Proc. Asia-Pacific Web Conf.*, Springer, 2014, pp. 93–104.
[42] Z. Fan, X. Song, and R. Shibasaki, "CitySpectrum: A non-negative tensor factorization approach," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 213–223.
[43] Z. Li, H. Yan, C. Zhang, and F. Tsung, "Long-short term spatiotemporal tensor prediction for passenger flow profile," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5010–5017, Oct. 2020.
[44] J. Ren and Q. Xie, "Efficient OD trip matrix prediction based on tensor decomposition," in *Proc. 18th IEEE Int. Conf. Mobile Data Manage.*, 2017, pp. 180–185.
[45] M. Bhanu, S. Priya, S. K. Dandapat, J. Chandra, and J. Mendes-Moreira, "Forecasting traffic flow in big cities using modified tucker decomposition," in *Proc. Int. Conf. Adv. Data Mining Appl.*, Springer, 2018, pp. 119–128.
[46] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017.
[47] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: John Wiley & Sons, 2015.
[48] Y. Yang, Y. Fan, and R. J. Wets, "Stochastic travel demand estimation: Improving network identifiability using multi-day observation sets," *Transp. Res. Part B, Methodol.*, vol. 107, pp. 192–211, 2018.
[49] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J. Amer. Stat. Assoc.*, vol. 91, no. 433, pp. 365–377, 1996.
[50] M. L. Hazelton, "Some comments on origin–destination matrix estimation," *Transp. Res. Part A, Policy Pract.*, vol. 37, no. 10, pp. 811–822, 2003.
[51] H. Shao, W. H. Lam, A. Sumalee, A. Chen, and M. L. Hazelton, "Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts," *Transp. Res. Part B, Methodol.*, vol. 68, pp. 52–75, 2014.
[52] H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, "Estimation of origin-destination matrices from link traffic counts on congested networks," *Transp. Res. Part B, Methodol.*, vol. 26, no. 6, pp. 417–434, 1992.
[53] M. J. Maher, X. Zhang, and D. Van Vliet, "A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows," *Transp. Res. Part B, Methodol.*, vol. 35, no. 1, pp. 23–40, 2001.
[54] Z. Zhao, H. N. Koutsopoulos, and J. Zhao, "Discovering latent activity patterns from transit smart card data: A spatiotemporal topic model," *Transp. Res. Part C, Emerg. Technol.*, vol. 116, 2020, Art. no. 102627.
[55] A. Agresti, *Categorical Data Analysis*. Berlin, Germany: Springer, 2002.
[56] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[57] J. H. Tu, "On dynamic mode decomposition: Theory and applications," *J. Comput. Dyn.,* vol. 1, pp. 391–241, 2014.
[58] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Philadelphia, PA, USA: SIAM, 2016.
[59] N. Mohan, K. Soman, and S. S. Kumar, "A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model," *Appl. Energy*, vol. 232, pp. 229–244, 2018.
[60] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," *J. Stat. Softw.*, vol. 27, no. 3, pp. 1–22, 2008. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v027i03
[61] Taxi and Limousine Commission, "Taxi and Limousine commission trip record data," 2019. [Online]. Available: https://www.nyc.gov/site/tlc/index.page
[62] F. Rodrigues, I. Markou, and F. C. Pereira, "Combining time-series and textual data for taxi demand prediction in event areas: A deep learning approach," *Inf. Fusion*, vol. 49, pp. 120–129, 2019.
[63] B. Du, X. Hu, L. Sun, J. Liu, Y. Qiao, and W. Lv, "Traffic demand prediction based on dynamic transition convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1237–1247, Feb. 2021.

Jeongyun Kim (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in civil and environmental engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2014, 2016, and 2021, respectively.

She is a postdoctoral fellow with Wireless Information and Network Sciences Laboratory, the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. Her research is in the area of modeling urban mobility and developing solutions to the challenges related to the autonomous urban mobility system. Her current research focuses on urban mobility analytics using large-scale movement data and modeling autonomy control system using machine learning.

Dr. Kim was the recipient of the Excellence Paper Award from the Journal of Korean Society of Transportation in 2020, the Excellence Award in Seoul Innovation Challenge Research in 2017, and the Excellence Paper Presentation Award from the Korean Society for Railway in 2016. She was the recipient of the Research Fellowship for Korean-Swiss Science and Technology Program.

Andrea Conti (Fellow, IEEE) is a professor and founding director with the Wireless Communication and Localization Networks Laboratory at the the University of Ferrara, Italy. Prior to joining the University of Ferrara, he was with CNIT and with IEIIT-CNR.

In Summer 2001, he was with the Wireless Systems Research Department, AT&T Research Laboratories. Since 2003, he has been a frequent visitor to the Wireless Information and Network Sciences Laboratory, the Massachusetts Institute of Technology, where he presently holds the Research Affiliate appointment. His research interests involve theory and experimentation of wireless communication and localization systems. His current research topics include network localization and navigation, distributed sensing, adaptive diversity communications, and quantum information science.

Dr. Conti has served as editor for IEEE journals and chaired international conferences. He was elected chair of the IEEE Communications Society's Radio Communications Technical Committee and is co-founder of the IEEE Quantum Communications & Information Technology Emerging Technical Subcommittee. He received the HTE Puskás Tivadar Medal, the IEEE Communications Society's Fred W. Ellersick Prize, and the IEEE Communications Society's Stephen O. Rice Prize in the field of Communications Theory. He is an elected fellow of IET, and a member of Sigma Xi. He has been selected as an IEEE distinguished lecturer.

Moe Z. Win (Fellow, IEEE) is the Robert R. Taylor professor at the Massachusetts Institute of Technology (MIT) and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories and with the NASA Jet Propulsion Laboratory.

His research encompasses fundamental theories, algorithm design, and network experimentation for a broad range of real-world problems. His current research topics include ultra-wideband systems, network localization and navigation, network interference exploitation, and quantum information science. He has served the IEEE Communications Society as an elected member-at-large on the Board of Governors, as elected Chair of the Radio Communications Committee, and as an IEEE distinguished lecturer. Over the last two decades, he held various editorial positions for IEEE journals and organized numerous international conferences. He has served on the SIAM Diversity Advisory Committee.

Dr. Win is an elected fellow of the AAAS, the EURASIP, and the IET. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). His publications, co-authored with students and colleagues, have received several awards. Other recognitions include the MIT Frank E. Perkins Award (2024), the MIT Everett Moore Baker Award (2022), the IEEE Vehicular Technology Society James Evans Avant Garde Award (2022), the IEEE Communications Society Edwin H. Armstrong Achievement Award (2016), the Cristoforo Colombo International Prize for Communications (2013), the Copernicus Fellowship (2011) and the *Laurea Honoris Causa* (2008) from the Università degli Studi di Ferrara, and the US Presidential Early Career Award for Scientists and Engineers (2004).