Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Dynamic activity chain pattern estimation under mobility demand changes during COVID-19

Yan Liu\textsuperscript{a, b}, Lu Carol Tong\textsuperscript{c, *}, Xi Zhu\textsuperscript{c}, Wenbo Du\textsuperscript{a}

\textsuperscript{a} School of Electronic and Information Engineering, Beihang University, Beijing 100191, PR China
\textsuperscript{b} Shenyuan Honors College, Beihang University, Beijing 100191, PR China
\textsuperscript{c} Research Institute of Frontier Science, Beihang University, Beijing 100191, PR China

1. Introduction

The disruption induced by major public emergencies may significantly change residents’ activity engagement habits with lasting impacts on individual travel behavior (Mahmassani and Mokhtarian, 2020). During the outbreak of coronavirus disease 2019 (COVID-19), many countries adopted transmission control measures and public health interventions, including city-wide lockdowns, travel bans, social distancing, closure of schools/restaurants/recreational areas, and suspension of public transport (Cobey, 2020; McKenzie and Adams, 2020). Restrictions have been adapted under various stages of the pandemic and led to widely observed behavioral adjustments in terms of activity participation and travel patterns (Beck and Hensher, 2020). The main activity participation habit...
changes are reflected across the spectrum of trip purposes, mode choices, and frequencies. Such activity participation behavior changes are observed across different regions in an extremely complex spatial-temporal pattern. Estimating the structural changes in the mobility demand patterns of regions may help identify the system-wide characteristics of travel behavior, which could lead to reliable forecasts and informed traffic demand management decisions.

The unexpected health emergency has raised worldwide attention to rapid data sharing (Moorthy et al., 2020). Various datasets have been publicly accessible during the pandemic, such as the datasets on the COVID-19 global cases, deaths, and recoveries provided by Johns Hopkins University (Johns Hopkins University, 2020). Valuable datasets related to transportation system aspects include the Google Global Mobility Change Dataset (Google, 2020), Apple Mobility Trend Data (Apple, 2020), and Government Response Event Dataset (Cheng et al., 2020), among others.

To fully utilize these public data and provide useful guidance for transportation decision makers, the construction of a methodological and computational framework for pandemic impacts on activity participation pattern is of urgent necessity (Mahmassani and Mokhtarian, 2020). Such a system can also be applied during the pandemic and post-pandemic phases beyond the current COVID-19 outbreak. However, estimating/predicting multiple-day activity participation patterns is a challenging task in its own right due to the lack of multiple-day survey samples, especially household-level records across regions. In this study, we attempt to design an estimation framework based on both continuously reported aggregated mobility observations and previously collected survey data (as the baseline) in an effort to capture the spatial-temporal correlation between activity participation behavior and mobility demand patterns.

In this paper, to adequately capture the mobility demand evolution, the proposed framework consists of three modules: systematic baseline parameter estimation of activity chains, dynamic change estimation of activity chains, and preference prediction of activity chains. The first module learns activity chains’ system utilities through the computational graph (CG)-based Nested Logit (NL) model and obtains the probabilities of accessing various activity locations. Note that the activity locations refer to location categories, such as residential neighborhoods, workplaces, grocery stores, and parks. The second module aims at capturing the activity chains’ systematic utility changes and probability changes across an NL-based optimization method by considering the daily mobility demand changes of various activity locations. The third part predicts the dynamic demand changes of activity locations by applying the long short-term memory (LSTM) method, and predicts the activity engagement demand through the optimization method in the second module. The mobility demand change pattern in the United States is analyzed, as household survey data are readily available from the Federal Highway Administration, which can easily serve as baseline conditions for further sensitivity analysis. Then, the performance of the proposed approach is examined in various areas during the ongoing COVID-19 pandemic.

During the COVID-19 period, many researchers strive to qualitatively or quantitatively analyze infectious disease spread (Pedersen and Meneghini, 2020; Pei et al., 2020), the effects of public health intervention (Chen and Pan, 2020; Hadijedemtriou et al., 2020; Haug et al., 2020), activity and travel behavior changes (Abdullah et al., 2020; Beck and Hensher, 2020; de Haas et al., 2020; Fatmi et al., 2021; Parady et al., 2020), and mobility demand changes (Arimura et al., 2020; Klein et al., 2020) from different sources of data. However, dynamic activity estimation/prediction of dynamic activity using multi-data sources is still challenging in its own right due to the inconsistency and heterogeneity across different datasets. A wide range of theoretical approaches and analytical methods are used in these studies (Kim, 2021), including epidemic model (Pedersen and Meneghini, 2020; Pei et al., 2020), survey development and analysis (Combs and Pardo, 2021; de Haas et al., 2020), discrete choice model (Fatmi et al., 2021). To offer additional insights into this line, our framework aims to provide a utility-based machine learning perspective to better interpret visiting demand of activity locations/chains on multiple days. The potential contributions of this paper are as follows:

(1) This paper addresses a new class of traveler behavior estimation problems with the support of multiple data sources. To capture the correlation between activity engagement information on regular days and activity location access demand during the emergency, we propose a utility-based conceptual method for estimating deterministic parameters and structural deviations of activity chain utilities. In particular, we hope to effectively utilize potentially inconsistent multiple data sources, i.e., the micro household survey data (individual-based activity chains data) and the macro google mobility data (dynamic aggregated activity locations data).

(2) With the combination of a discrete choice model, an optimization algorithm, and a time series prediction method, this paper constructs an interpretable machine learning application within a dynamic activity chain pattern estimation (DACPE) framework. This unique integration of machine learning tools and transportation domain knowledge could shed some light on capturing spatial and temporal patterns of activity demand, especially on how to recognize the structural deviations and visiting demand of activity locations/chains on multiple days.

(3) By applying the DACPE framework to open datasets, we estimate and compare the changes of activity chain patterns during the pandemic in various areas across the United States. Case studies show that the proposed method can maintain the modeling consistency of individuals’ activity and aggregated location mobility. A cross-region comparison indicates the weekly periodicity and residents’ overall activity engagement trends during COVID-19. On the other hand, the results also identify the various degrees of magnitudes, activity chain preference, and resilience in pandemic recovery phases across different regions. This framework could be useful in mapping complex interactions between the underlying mobility and activity patterns.

The remainder of this paper is organized as follows: Section 2 reviews the related literature on behavioral changes under major public emergencies, activity-based travel demand models, and time series prediction methods on mobility demand. Section 3 introduces the details of datasets prepared for case studies, including “static” household survey data and “dynamic” mobility demand reports. Section 4 describes the conceptual and mathematical formulation of the DACPE framework. The results of the method are
analyzed in Section 5, with the findings of the research in Section 6.

2. Literature review

This review section covers three aspects: individuals’ behavioral changes under major public emergencies, time series prediction methods in the domain of mobility demand as well as activity-based travel demand models. We further explain the motivations of our proposed activity engagement estimation framework during the pandemic.

2.1. Behavioral changes under major public emergencies

Major public emergencies include public health events (e.g., AIV flu, SARS, COVID-19), natural disasters (e.g., earthquakes), and man-made disasters (e.g., wars). These emergencies may impact residents’ lives in all aspects, including their activity engagement habits and travel behavior.

Earlier research efforts focused on traffic planning and optimization problems during public emergencies. To name a few, a qualitative study by Litman (2006) described how to improve emergency transportation services by analyzing the lessons from two hurricanes, and thus help planners make appropriate policies and improve the resilience of the transportation system. In a study conducted by Wang et al. (2007), an optimized resource allocation algorithm based on iterative adjustment was proposed to solve the problem of minimizing the public emergency response time.

With the emerging progress of big data applications, studies have begun to use data-driven methods to investigate people’s behavior changes during public events. Song et al. (2014) constructed datasets on human emergency behavior and mobility and further analyzed the affecting factors of the behavior and mobility pattern changes during a large-scale public emergency. They also proposed a human mobility prediction approach that integrated the hidden Markov model and Markov decision process. Jiang et al. (2018) proposed a multistep-to-multistep prediction model for online urban mobility prediction through the application of recurrent neural networks.

According to an excellent position paper by Axhausen (2020), COVID-19 may promote the formation of a new equilibrium of transportation system in the future, in which the new normal of working from home would lead to reallocation for the workplace and mobility demand pattern changes. Although people may eventually return to normal traveling patterns, it is certain that people’s activity participation behaviors will change substantially for at least a few years (Bhat, 2020). Numerous studies have examined the impact of COVID-19 on people’s behavior from different perspectives (Kabiri et al., 2020; Sun et al., 2020), such as social distance, online shopping, work, and transportation. Unnikrishnan and Figliozzi (2020) surveyed people living in Portland on shopping modes and home deliveries and found that people with higher income and higher-level technology were more likely to make online purchases and spend money on home deliveries. Hu et al. (2020) applied the Bureau of Public Road (BPR) model to different metropolises and highlighted that high-transit metropolises were at higher risk of congestion when people tended to use single-occupancy vehicles instead of public transport as the community reopened in the post-epidemic period. Klein et al. (2020) assessed the individual mobility demand changes in the United States and preliminarily interpreted the impacts of policy interventions such as work-from-home policies and mobility restrictions on interurban and urban mobility. Parady et al. (2020) conducted a survey targeting Kanto residents, utilized regression models to estimate the shopping frequency and used a discrete choice model to estimate the leisure frequency. Thus, factors affecting travel behavior have been analyzed quantitatively.

Collectively, a wide range of studies has investigated ongoing changes in people’s behavior and mobility during major public emergencies. It should be remarked that most of them focused on qualitative and quantitative causes of activity demand changes based on survey data or day-to-day mobility demand changes on the basis of the mobility data. As one of the major driving forces for transportation and mobility changes, activity participation demand still requires systematic analysis, especially for multiple-day activity chain changes.

2.2. Time series prediction methods for mobility demand

With the objective of predicting the mobility demand change with activity pattern adjustment, it is important to examine the existing time series prediction methods, especially in the traffic state (time, speed, and flow) prediction field (Tang et al., 2020). According to Barros et al. (2015), traffic state prediction methods can be classified as model-driven methods or data-driven methods. Model-driven methods describe the traffic system using prior knowledge; thus, their accuracy can be easily affected by traffic disturbances in complex and changeable real-world traffic environments (Zhang et al., 2021; Zhao et al., 2019).

In contrast, data-driven methods can capture the statistical properties of historical data and provide highly flexible learning results (Antoniou et al., 2013). Early examples of research into data-driven traffic state prediction include the historical average (HA) method, the moving average (MA) method, and the exponential smoothing (ES) method (Liu and Guan, 2004). A widely used model is the autoregressive integrated moving average (ARIMA) (‘I’ stands for ‘integrated’), which is the integrated utilization of the autoregressive (AR) method, the MA method, and their combination, namely, the autoregressive moving average (ARMA) method (Lee and Fambrro, 1999). Compared with other models of the same period, ARIMA showed high accuracy and universality (Moreira-Matias et al., 2013; Williams and Hoel, 2003); thus, some models were developed by extending ARIMA, e.g., KARIMA, which was proposed by Voort et al. (1996) and SARIMA proposed by (Williams and Hoel, 2003). Apart from these models, other linear models adopted for traffic prediction include the linear regression (LR) model and the Kalman filtering (KF) model. The LR model captures the linear or local linear relation between traffic states and historical traffic data (Sun et al., 2004). The KF approach is utilized mainly for the optimal
estimation of the linear dynamic system using noisy observation data, and thus can be applied to short-term traffic state prediction (Anand et al., 2014).

Currently, intelligent transportation systems (ITSs) have been progressively developed to provide efficient traffic management strategies and individual travel advice (Guo et al., 2019). Recognizing that linear models are ill-suited to grasp non-linear trends in complicated transportation systems, many non-linear but complex models have been proposed. The K-nearest neighbors (KNN) method identifies the K traffic states with the most similar features and uses them to predict the next traffic states (Cai et al., 2016; Li et al., 2012). Support vector machine (SVM) and support vector regression (SVR) map the prediction problem into a higher-dimensional space through non-linear transformation, and obtain lots of application in traffic state prediction. Vanaja and Rilett (2004) applied SVR with proposed parameters for travel-time prediction. Yang et al. (2014) proposed an integrated model to optimize SVM parameters through genetic algorithm and then further predict traffic flow through an optimized SVM model. As a probabilistic directed graphic model, Bayesian network (BN) models are able to handle incomplete spatial-temporal data (Castillo et al., 2008; Sun et al., 2006). Li et al. (2020) proposed a partial least squares (PLS)-based traffic state prediction method, which decomposed the correlated traffic data and built a linear combination of the fundamental predictors, proving the efficiency and accuracy through a case study. Tree-based ensemble methods, combining simple regression methods to fit complex non-linear relationships, have been widely acknowledged in the prediction field. Zhang and Haghani (2015) proposed a gradient boosting regression tree method (GBM)-based approach for predicting freeway travel times and realized superior prediction accuracy.

Recently, deep learning (DL) methods have been utilized to solve the traffic state prediction problem and have shown promising estimation performance, of which the most popular methods are convolutional neural networks (CNNs), recurrent neural networks (RNNs), graph neural networks (GNNs), and their combinations (Li et al., 2020). CNN originates from the image processing field and is essentially suitable for modeling the local spatial dependencies of Euclidean spatial data; thus, it has been widely applied in traffic forecasting tasks (Li et al., 2020; Ma et al., 2017; Zhang et al., 2016). RNN is capable of finding short-term dependencies by utilizing the directed self-circulation mechanism effectively (Van Lint et al., 2002). Extended variants, which include LSTM and the gated recurrent unit (GRU), can learn both long- or short-term dependencies and perform better in traffic prediction problems (Fu et al., 2016; Kim et al., 2020). GNNs are more suitable to deal with large graph-structured data and are more suitable for extracting features from non-Euclidean space, such as social media networks and large-scale traffic networks (Yu et al., 2017; Zhao et al., 2019). In addition, the integration of various DNN methods has become an emerging research trend for traffic demand forecasting applications. For example, the FCL-Net, which was proposed by Ke et al. (2017), aimed at integrating convolutional layers, convolutional LSTM units, and LSTM units.

Due to the effectiveness of feature extraction by utilizing multiple-layer architectures, DL methods can fully utilize latent knowledge hidden in big data and improve the accuracy of prediction (Wu et al., 2018b). To achieve a greater level of understanding of the residents’ travel behavior during the pandemic and thus provide practical decision support for traffic management departments, emerging DL methods are adopted to handle the large data sets and forecast the aggregated mobility demand.

### 2.3. Activity-based travel demand models

Travel demand models are intended to estimate and predict future travel patterns (Domencich and McFadden, 1975) and thus provide guidance for transportation planning/management decisions, specifically, urban land-use patterns at the macroscopic level and individual travel mode choices at the microscopic level. Travel demand analytical approaches can be further classified as trip-based models and activity-based models, with the former relying on aggregated traffic information and the latter focusing on individual activity engagement behavior. In practice, activity-based models provide a more holistic understanding of the relationship among multiple trips. With the changing transportation external environment (policy/business), the activity-based model, which can be categorized as constraints-based approaches, utility-maximizing approaches, and computational process approaches (Rasouli and Timmermans, 2013), has been widely used (Miller, 2021). Kim et al. (2017) reviewed modeling approaches and empirical evidence of studies concerning the relationship between social networks and activity-travel behavior. Using household travel survey data, Li and Lee (2017) explored the similarity between activity sequence and language and modeled daily activity patterns through a probabilistic context-free grammar. Based on mobile phone data, Jiang et al. (2016) established a probability-based dynamic choice framework for inferring various activity engagement-related attributes, such as locations, durations, and frequencies. Interestingly, Zhao and Zhang (2017) modeled the activity set as a three-dimensional activity-travel network with space, time, and state factors so that the transfer probability matrix for activity choice and dynamic activity chain choices could be estimated with more accuracy.

Logit models are widely used in traveler behavior models to represent different types of activity choices. In terms of activity location choice models, an early study by Van der Hoorn (1983) applied logit models to capture the temporal dependence of the location and activity choice. Bowman and Ben-Akiva (2001) constructed a nested logit framework for examining interactions among trip choices. By considering a sequence of multidimensional travel decisions (whether to travel, destinations, travel times, and transportation modes), Bhat et al. (2004) established an econometric micro simulator for comprehensive empirical activity analysis. In the agent-based activity scheduling framework proposed by Langerudia et al. (2017), the activity sequence is developed as a series of NL models based on activity type and duration decisions. Västberg et al. (2019) modeled the daily activity-travel pattern through incorporating spatial and temporal information into a dynamic nested logit model, where the trip pattern preference is described as the utility sum of all activity and travel sequence.

Essentially, most activity-based studies estimated or predicted relatively stable activity patterns/schedules for “representative” or typical days. Based on a multi-state super-network, researchers extended the dynamic traffic assignment and dynamic user equilibrium theory into activity-based model (Liu et al., 2016). And a comprehensive set of multi-modal and multi-activity individual activity-
travel scheduling choice problems are addressed (Liao, 2016; Liao et al., 2013). Recently, Liu et al. (2020) proposed a needs-based framework in multi-state super-networks with focuses on linking activity-travel dynamics and traffic equilibria by integrating activity scheduling method and traffic flow evolution theory. In general, the long-term evolution of activity patterns under major interventions has not received sufficient attention due to the limited available sample data along this line as well as privacy concerns for multiday detailed samples (Astroza et al., 2018). Location-based services and emerging big data applications offer a new spectrum of data sources, such as individual-based survey data and location-based mobility data, which bring rich opportunities for advanced human behavioral studies (Huang et al., 2018). In this context, this study aims to propose a multiday activity pattern estimation framework with a focus on the evolution process under a major event. Specifically, we are also interested in how to integrate econometrics-oriented NL model and emerging machine learning methods. We aim at not only providing a new perspective on capturing the correlation of behavioral change patterns and mobility demands, but also shedding more light on practical decisions through multiday activity pattern estimation during public emergencies.

3. Data preparation

Abundant data resources have become a key catalyst for new types of impactful data-intensive research (Foundation, 2020). In the domain of transportation planning, the National Household Travel Survey (NHTS) from the Federal Highway Administration (FHWA, 2017) offers a standard dataset with complete activity information of individuals during regular days. In this study, to model the baseline traveler activity choice behavior, we adopt the activity chain dataset from the NHTS 2017. During the current ongoing pandemic event, various datasets with unprecedented spatial-temporal scales have been made public for research purposes, thereby resulting in a range of interactive web-based data analytical platforms for the public and decision makers (Gao et al., 2020; Zhang et al., 2020). In the broader social science domain, the Google COVID-19 Community Mobility Dataset (Google, 2020) contains six categories of activity locations over the course of almost a year across different countries.

In this paper, we extract activity chains from the NHTS according to activity locations contained in Google COVID-19 Community Mobility Reports. In other words, Google mobility dataset forms the basis of analyzing the ongoing daily activity change, while NHTS is utilized as an activity chain baseline before the COVID-19. Therefore, we first introduce the Google mobility dataset in the following section.

3.1. Global mobility dataset with information privacy protection measures

We obtain the global mobility demand dataset from Google COVID-19 Community Mobility Reports. This dataset aims to provide mobility insights on how different countries and regions around the globe respond to the policies aimed at combating COVID-19. The accessible information covers daily mobility demand changes in categories of places, namely, retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential neighborhoods. To avoid privacy concerns, the highly aggregated measure for each type of activity location is provided by Google in terms of the percentage change for people visiting that activity location compared to the baseline day. The (right-before pandemic) baseline day established from the 5-week period is the median value on the corresponding day of the week from 3rd Jan–6th Feb 2020. In addition, the advanced anonymization and differential privacy technology that is used in the dataset adds artificial noise to prevent individual identification. A major concern is that the mobility demand dataset reflects only a type of relative trend; hence, we need to solve a non-trivial estimation task without knowing the underlying population volume for each activity location or the exact sampling rates. Overall, it is extremely critical for researchers to fully utilize such valuable mobility datasets with privacy protection measures.

The mobility demand change data from a region can be mathematically described in the following matrix form:

$$D_{L,J} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1J} \\ y_{21} & y_{22} & \cdots & y_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ y_{L1} & y_{L2} & \cdots & y_{LJ} \end{bmatrix},$$  

(1)

where $L$ is the total number of days from the starting observed day and $J$ is the number of categories of locations. The entry in the $i$th row and $j$th column of a matrix $D_{L,J}$ is denoted as $y_{ij}$, which represents the mobility percentage change compared with the baseline day in the $j$th place on the $i$th day, ranging from 0% to 100%.

3.2. Activity chain dataset before the pandemic

To analyze the activity engagement pattern and estimate the change in activity chains through the Google COVID-19 Community Mobility Reports, we also need a normal-year baseline dataset with more detailed information from individual households. For this purpose, we utilize the NHTS 2017 with 923,572 trips from numerous household members across different cities in the United States, with approximately 20 activity types. Each trip is associated with household member attributes (e.g., gender, income, age, educational attainment) as well as trip-level information such as distance, duration, and transportation modes. To match the normal-year activity chain dataset with the aggregated day-by-day global mobility dataset, we extract a more specific set of activities from the NHTS to include only residential neighborhoods, workplaces, grocery stores, and parks, as presented in Table 1. The statistics of the top 10 most
frequent activity chains are presented in Table 2. An activity chain includes several continuous trips of an individual. For example, the HWH consist of 2 trips: from home to work, and from work to home. Therefore, from Table 2, we can calculate the total utilized trips as 206,813. Compared to the total NHTS dataset with 923,572 trips, the extracted percentage is 22.393%. Note that the origin and destination of an activity chain are often the same type of activity location, namely, home, and activity chains HH and HHH represent different individual activity patterns. The activity chain H indicates the individual stays at home without specific activity information for the whole day. On the other hand, the activity chain HHH indicates the individual stays in residential areas for the whole day while doing some specific activities, including work from home (paid) and community activities. Activity chain and corresponding notation are utilized in the subsequent activity estimation methodology. As traveler behavior varies among days of the week, the activity chain data are also categorized as weekday and weekend data in this study. As shown in Table 2, the NHTS dataset is a large-scale dataset with more than 80,000 records on weekdays and more than 20,000 records on weekends, which brings challenges for the estimation of the activity pattern in this study.

3.3. Understanding the context of the problem: Mapping from individual trip chains to aggregated activity location types

In this section, we plan to use an illustrative example to demonstrate how to infer day-to-day activity pattern changes from partial mobility demand change observations during the pandemic. As discussed above, the activity-related household survey data only offers accurate trip information of individuals under normal conditions. Compared to traditional survey data, mobile device location data, such as Google Mobility Data, generate a larger spatial-temporal coverage range with lower cost (Yang et al., 2020). Using the above datasets, this study is motivated by how to fully utilize a combination of aggregated mobility data and baseline individual activity information to estimate the activity adjustments.

In this study, we suppose that activity participation is the essential reason for mobility demands at various location types. In other words, aggregated mobility demand change reflects the adjustment of individuals’ desire for activity engagement over a period. To better understand the context of the problem, the following synthesized example is used to demonstrate the mechanism of influence of the activity participation probability on the aggregated travel probability.

For simplicity, partially based on the Google Mobility Report, we consolidate the same types of activity chains in Table 2 into 5 representative activity chains, namely, HHH, HWH, HGH, HPH, and HWGH. Let us assume there are 20 individuals choosing their activities. Fig. 1 (a) represents the activity engagement pattern before the pandemic. The numbers of individuals who choose these five representative activity chains are 5, 8, 3, 2, and 2. Due to the pandemic, staying at home now becomes a likely activity/option to select, work-related and leisure-related activity probabilities have substantially declined, while the grocery-related activity probability has only slightly decreased, as presented in Fig. 1 (b). The above five activity chains are chosen by 11, 4, 3, 1, and 1 traveler, respectively. As the origin and destination of an activity chain are often the same type of activity location, namely, home, to explain the mobility change mode more clearly, the calculation of the aggregated mobility probability does not consider the origin or destination in this example.

4. Problem statement and methodology

In this section, we formally present the problem statement, which is followed by the solution framework of spatial-temporal activity pattern estimation and forecasting.

4.1. Problem statement and solution framework

In the problem of activity pattern estimation and forecasting that is investigated in this paper, the objective is to infer activity adjustments during public emergencies. Origin-destination (OD) matrices extracted from different data types are widely used for addressing the activity pattern recognition problem. However, in our case, we use the relative location visiting probability change from the baseline day to the current day as one of the critical input measurements, as expressed in Eq. (18); Fig. 1 illustrates an example of a simple mapping process between activity chains and activity location types. Given multiple heterogeneous datasets, which include household survey data and aggregated mobility change reports, this study aims at constructing an interpretable inference framework for capturing the spatial-temporal pattern interactions of activity adjustments through mobility demand evolution trends. Note that compared to the traditional machine learning method which is a black box, the proposed DACPE framework with utility theory is more interpretable as the parameters and activity choices are mathematically formulated. To offer more insights into our modeling framework, we discuss the underlying concepts and guiding principles as follows:

Table 1
Mapping between activity types and different activity locations.

| Activity type | Activity location | Abbreviation | Notation |
|---------------|-------------------|--------------|----------|
| Regular home activities (chores, sleep) | Home | H | AL₁ |
| Work-related activities (work, meeting, travel on business) | Workplace | W | AL₂ |
| Buy goods (groceries, clothes, appliances, fuel) | Grocery | G | AL₃ |
| Recreational activities (visit parks, movies, bars, museums) | Park | P | AL₄ |
Compared to recent studies that rely solely on machine learning methods to capture statistically sound but “locally oriented” spatial and temporal correlations in traffic flow observations (e.g., Ma et al. (2017), Liang et al. (2020)), we are more interested in developing a hybrid utility-based machine learning framework to better identify the inherent “system-level” correlation on location choices through the use of tour-based logit models. For example, if the probability of selecting home-work-shopping-home is reduced, and the relative proportions of other activity chains remain consistent, then the resulting likelihoods of visiting shopping and work locations would be impacted simultaneously.

Following the individual utility maximization hypothesis, we assume that an individual’s activity choice reflects his or her underlying preferences for available activity choices under both regular and irregular circumstances. The utility function of choosing an activity is defined as the sum of a systematic portion for observed individual preferences and a constant portion that represents the impacts of some unobserved characteristics related to the activity choice.

The random portion of an activity choice depends on the alternative specific constants of activity locations, which aim at reflecting all types of external and unobserved factors, including various government policies from different states, epidemic situation as well as public attention to the epidemic. By defining activity-oriented deviation parameters that correspond to the change in the alternative specific constant, we hope to represent the day-to-day variations in activity engagement patterns and aggregated mobility changes through the (interpretable) utility-based behavioral model framework.

As discussed in Section 3.1, data privacy measures for the protection of personal data bring challenges for estimating specific information. Therefore, we could not fully recover the total volume change in selecting different types of activity chains; instead, the most relevant information that we extract or estimate would be the alternative specific constant change in the logit

| Activity chain                      | Abbreviation | Notation | Total Count (percentage) | Weekday count (percentage) | Weekend count (percentage) |
|-------------------------------------|--------------|----------|--------------------------|---------------------------|---------------------------|
| Home-Home                           | HH           | AC₁      | 29,346 (27.36%)          | 22,722 (26.36%)           | 6,624 (31.43%)            |
| Home-Workplace-Home                 | HWH          | AC₂      | 25,830 (24.08%)          | 23,720 (27.52%)           | 2,110 (10.01%)            |
| Home-Grocery-Home                   | HGH          | AC₃      | 15,326 (14.29%)          | 10,788 (12.52%)           | 4,538 (21.53%)            |
| Home-Home                           | HHH          | AC₄      | 10,717 (9.99%)           | 8,291 (9.62%)             | 2,426 (11.51%)            |
| Home-Grocery-Grocery-Home           | HGGH         | AC₅      | 5,403 (5.04%)            | 3,814 (4.42%)             | 1,589 (7.54%)             |
| Home-Park-Home                      | HPH          | AC₆      | 4,423 (4.12%)            | 2,683 (3.11%)             | 1,740 (8.26%)             |
| Home-Workplace-Home-Home            | HWHH         | AC₇      | 4,205 (3.92%)            | 3,899 (4.52%)             | 306 (1.45%)               |
| Home-Grocery-Home-Home              | HGGH         | AC₈      | 4,065 (3.79%)            | 2,823 (3.28%)             | 1,242 (5.89%)             |
| Home-Workplace-Workplace-Home       | HWWH         | AC₉      | 4,009 (3.74%)            | 3,823 (4.44%)             | 186 (0.88%)               |
| Home-Workplace- Grocery-Home        | HWGH         | AC₁₀     | 3,943 (3.68%)            | 3,631 (4.21%)             | 312 (1.48%)               |

Fig. 1. Illustration of the correlation and mapping between activity chains and activity location types.
model and the resulting selection probability. Thus, in our comparison, we will use the “indirectly estimated activity type selection probability” as the ground truth to evaluate the prediction result of the activity chain selection probability.

(5) We assume that each state across the United States had similar activity participation patterns before COVID-19; thus, baseline parameters of all areas are estimated from the same normal-year baseline dataset of the activity chain.

(6) The mobility demand pattern is affected by the epidemic situation, government policies, and other factors, and generally changes according to a common trend even though the magnitudes of the changes are significantly different. The factors are considered in the random portion in the framework, and the magnitude differences can be captured by the optimization and prediction method. Activity adjustments in various regions are assumed to be estimable and predictable through a machine learning framework in conjunction with the logit-based econometric model.

Based on the above modeling guidelines, we propose a DACPE framework, as illustrated in Fig. 2. It is important to highlight the inherent connections among the three sequential steps of the DACPE framework. First, based on the household survey dataset (activity chain dataset before the pandemic), we construct a utility-based NL model and estimate the baseline parameters of activity chains before public intervention through a CG-based method. The detailed introduction of the CG-based NL model is shown in Section 4.2. At the second step, based on the Google mobility dataset, the mobility demand change is predicted by applying the LSTM method illustrated in Section 4.3. At the third step, based on the systematic parameters \((\beta, \phi, \text{ASC})\) estimated in the first step and changes in real/predicted activity location visiting demands, day-by-day systematic utility changes \((\Delta\text{ASC}_{\text{DYN}}/\Delta\text{ASC}_{\text{PRE}})\) during the pandemic are identified through a non-linear NL-based sequential least-squares quadratic programming (SLSQP) estimation method.

The parameters and variables in the DACPE framework are defined in Table 3, which include the constants, estimation parameters, and other variables.

4.2. Estimation of the baseline parameters

In the proposed DACPE framework, considering possible correlations between alternatives, the NL model is adopted to obtain the deterministic systematic baseline estimation of activity chains before the pandemic. This model consists of a two-layered structure. The upper level of NL includes 4 nests: home-related activity chains, work-related activity chains, grocery-related activity chains, park-related activity chains, that is to say, \(M = 4\). Regarding the lower-level decisions, home-related activity chains include HH and HHH, work-related activity chains include HWH, HWWH, HWHH, and HWGH, grocery-related activity chains include HGH, HGGH, and HGHH, park-related activity chains only include HPH.

The choice probability \(P(AC_m)\) for traveler \(n\) of choosing an alternative \(i\) in the nest \(m\) can be calculated by multiplying the conditional probabilities of the nested alternative by the marginal probability as follows:

\[
P(AC_m) = P(AC_m|J_m) \times P(n|J_m) = \frac{e^{V_{ni}/\theta_m}}{\sum_{i \in J_m} e^{V_{ni}/\theta_m}} \times \frac{e^{V_{nm}/\theta_m} + \Gamma_{nm}}{\sum_{m=1}^{M} e^{V_{nm}/\theta_m} + \theta_m}.
\]

where the first component is the choice conditional probabilities for an activity chain in the \(m^{th}\) nest, and the second part is the
### Table 3
Parameter and variable notation.

| Parameters | Definition |
|------------|------------|
| \( I \)   | Number of activity chains |
| \( J \)   | Number of activity locations |
| \( N \)   | Total number of decision makers |
| \( M \)   | Total number of nests in the NL model |
| \( L \)   | Total number of days from the starting observed day |
| \( Q \)   | Number of characteristics for travelers in the NL model |
| \( K \)   | Number of attributes in the NL model |
| \( F \)   | Total number of features used in the prediction method |
| \( \theta \) | Weight of the L1-norm regularization term |
| \( T \)   | Time steps of the LSTM network |
| \( t \)   | Current time step in an LSTM unit |
| \( N_i \) | Number of the \( j \)th activity location in the \( i \)th activity chain |
| \( AC_i \) | The \( i \)th activity chain |
| \( AL_j \) | The \( j \)th activity location |
| \( c_{qn} \) | The \( q \)th characteristic for the \( n \)th decision maker |
| \( a_{ik} \) | The \( k \)th attribute of the \( i \)th activity chain |
| \( J_m \) | Activity chain set of the \( m \)th nest |

| Indexes    | Definition |
|------------|------------|
| \( i \)   | Index of the activity chain |
| \( j \)   | Index of the activity location |
| \( r \)   | Index of the layer in the CG |
| \( m \)   | Index of the nest |

| Input data | Definition |
|------------|------------|
| \( D_{L,J} \) | Mobility demand change data from a region |
| \( X' \)   | Input sample at time step \( t \) |
| \( Y(AC_{ni}) \) | Discrete variable that denotes the activity chain choice of the \( n \)th decision maker |
| \( P(AC_i) \) | Probability of choosing the \( i \)th activity chain on the baseline day |
| \( P(AL_j) \) | Probability of visiting the \( j \)th activity location on the baseline day |
| \( P'(AC_i) \) | Probability of choosing the \( i \)th activity chain on the considered day \( e \) |
| \( P'(AL_j) \) | Probability of visiting the \( j \)th activity location on the considered day \( e \) |
| \( y^j_t \) | Real mobility demand changes of visiting the \( j \)th activity location at time step \( t \) |
| \( Y_j \)  | Real mobility demand vector at the \( j \)th activity location of the test set |

| Variables | Definition |
|-----------|------------|
| \( \theta_m \) | logsum parameter for the \( m \)th nest |
| \( \beta_{iq} \) | Corresponding coefficient of \( c_{qn} \) |
| \( \phi_{iq} \) | Corresponding coefficient of \( a_{ik} \) |
| \( \gamma_i \) | Alternative specific constant of the \( i \)th activity chain |
| \( N_i \) | Intermediate variable of the \( r \)th layer, the \( i \)th activity chain in the CG |
| \( N_i \) | Intermediate variable which calculates the utility except \( \gamma_i \) of the \( i \)th activity chain in the CG |
| \( P_i \) | Probability of choosing the \( i \)th activity chain in the CG |
| \( ASC_j \) | Alternative specific constant of the \( j \)th activity location |
| \( \Delta ASC_j^e \) | Change in the \( j \)th activity location’s constant on the considered day \( e \) |
| \( \Gamma_{mn} \) | Expected value of the maximum of the \( m \)th nest for the \( n \)th decision maker |
| \( V_{nm} \) | Systematic utility of the \( i \)th activity chain for the \( n \)th decision maker |
| \( P(AC_{ni}|J_m) \) | Choice conditional probabilities for choosing the \( i \)th activity chain in the \( m \)th nest for the \( n \)th decision maker |
| \( P(n|J_m) \) | Marginal choice probability of choosing the \( m \)th nest for the \( n \)th decision maker |
| \( P(AC_{ni}) \) | Probability of choosing the \( i \)th activity chain for the \( n \)th decision maker |
| \( LL \) | Log likelihood function |
| \( i' \) | Input gate variable at time step \( t \) |
| \( C' \) | Memory cell variable at time step \( t \) |
| \( f' \) | Forget gate variable at time step \( t \) |
| \( O' \) | Output gate variable at time step \( t \) |
| \( g' \) | Candidate hidden layer variable at time step \( t \) |
| \( h' \) | Hidden variable at time step \( t \) |
| \( W_i \) | Weights from the input and previous hidden layer to the input gate |
| \( W_f \) | Weights from the input and previous hidden layer to the forget gate |

(continued on next page)
marginal choice probability of choosing the m\textsuperscript{th} nest. $\theta_m$ is a logsum parameter bounded by zero to one, an indicator of the correlation between alternatives in the same nest. As there is only an activity chain in the fourth nest, $\theta_4$ is set to 1. $\Gamma_{nm}$ is the expected value of the maximum of the m\textsuperscript{th} nest, which is computed from the log of the sum of the exponents of the nested utilities:

$$
\Gamma_{nm} = \log \left( \sum_{j \in J_m} e^{V_{nj}/\theta_m} \right). 
$$

### Table 3 (continued)

| Parameters | Definition |
|------------|------------|
| $W_o$      | Weights from the input and previous hidden layer to the output gate |
| $W_g$      | Weights from the input and previous hidden layer to the candidate hidden layer |
| $b_i$      | Biases at the input gate |
| $b_f$      | Biases at the forget gate |
| $b_o$      | Biases at the output gate |
| $b_g$      | Biases at the candidate hidden layer |
| $W_h$      | Weight from the hidden variable to the predicted mobility demand |
| $b_h$      | Bias of the fully connected layer |
| $\hat{y}_j$ | Predicted mobility demand change of visiting the j\textsuperscript{th} activity location at time step t |
| $L_T$      | Loss function of the LSTM network |
| $\hat{Y}_j$ | Predicted mobility demand vector at the j\textsuperscript{th} activity location of the test set |

---

Fig. 3. Illustration of the CG-based NL model.
where $V_{ni}$ is the systematic utility of the $i^{th}$ activity chain for the $n^{th}$ traveler, with a form of

$$V_{ni} = γ_i + \sum_{q=1}^{q_{max}} β_{iq} \times c_q + \sum_{k=1}^{k_{max}} φ_k \times a_k,$$

where $a_k$ is the $i^{th}$ activity chain’s $k^{th}$ attribute (travel distance, travel time, etc.), $φ_k$ is the corresponding coefficient of the $k^{th}$ attribute, and $γ_i$ is the $i^{th}$ activity chain’s constant, $c_q$ is the $m^{th}$ characteristic of the traveler $n$ (respondent age, gender, educational attainment, etc.), and $β_{iq}$ is the parameter which defines the direction and magnitude of the incremental bias due to the $q^{th}$ characteristic $c_q$. Obviously, the needs of the entire household are important for determining the activity pattern of each family member. To achieve a more comprehensive validation model, family-related attributes (e.g., household income, count of household members, number of workers in household, age of family members) should be taken into consideration.

The value of $γ_i$ includes the mode-specific constants of various activity locations:

$$γ_i = \sum_{j=1}^{J} N_i^j \times ASC_j,$$

where $N_i^j$ is the number of the $j^{th}$ activity location in the $i^{th}$ activity chain, $J$ is the total number of activity locations (i.e., $J = 4$ as shown in Table 1) and $ASC_j$ is the alternative specific constant of the $j^{th}$ activity location. In our model, without loss of generality, the travel distance is used as the major attribute of each alternative. The alternative specific constants of four activity locations are set as $ASC_1$, $ASC_2$, $ASC_3$, and $ASC_4$, which represent the alternative specific constants of home, work, grocery, and park, respectively. As it is impossible to estimate all the parameters in the NL model, we consider the utility of activity chain HH as the baseline, and thus, set $β_{iq}$ and $ASC_1$ to zero.

To consistently and accurately estimate the parameters in large-scale datasets (i.e., the NHTS dataset and non-linear architectures, we further construct a utility-based NL model through an interpretable CG-based framework. A CG is defined as a directed graph where each node represents a computation and the edge between nodes corresponds to the dependency between computations. Based on the symbolic rules, automated differentiation (AD) (Abadi et al., 2016) is a set of computer techniques that automatically construct a procedure for evaluating numerical gradients accurately. With the underlying mathematical tool performing AD, CG is an effective modeling platform which is able to deal with large-scale data and capture nonlinear relationship. As CG-based NL model is just one part of the DACPE framework, interested readers are referred to papers by Wu et al. (2018a) and Ma et al. (2020) on the underlying modeling platform which is able to deal with large-scale data and capture nonlinear relationship. As CG-based NL model is just one part of the DACPE framework, interested readers are referred to papers by Wu et al. (2018a) and Ma et al. (2020) on the underlying computational graph building blocks. With the connection across main elements of the NL model (namely Eqs. (2)–(6)), the forward propagation process of the CG-based activity chain choice model is illustrated in Fig. 3. In this directed graph, there are 68 input nodes, including 30 constants calculated by the NHTS 2017, 10 coefficients of activity chains, 20 coefficients of individual characters, 4 logsum parameters of nests, and 4 mode-specific constants of activity locations. In addition, there are 92 intermediate CG nodes, which represent elementary decomposition functions and play a role of transitions between inputs and outputs. After the calculation of the activity chain selection probability, we can obtain the objective function as the output node. In the activity chain choice model, variables are estimated by maximizing the log likelihood function, as expressed in Eq. (6):

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} Y(AC_{ni}) \ln P(AC_{ni}),$$

where $Y(AC_{ni})$ is the $n^{th}$ traveler’s activity chain choice, which is a discrete variable that equals 1 if the traveler chooses the $i^{th}$ activity chain and 0 otherwise; $P(AC_{ni})$ represents the probability of choosing the $i^{th}$ activity chain for the $n^{th}$ decision maker, and $N$ is the total number of travelers.

In the process of estimating activity-related variables, TensorFlow (Abadi et al., 2016) is used to construct the CG and provide AD techniques for deriving the gradients. Specifically, the derivate of the variable with respect to the negative log likelihood function can be calculated by the nodes and links depicted in Fig. 3. Among all the nodes in Fig. 3, $ASC_j, β_{iq}, φ_k$ are variables under estimation, $a_k$ and $c_q$ are constants calculated by NHTS 2017, $N_i$ and $N_j$ represent other intermediate variables in the CG. The description of the derivate process is based on the chain rule-based calibration, for the estimating variable $ASC_2$, the partial derivative of $LL$ with respect to the parameter $ASC_2$ is expressed as Eq. (7).

In the numerical optimization field, the Brodyen–Fletcher–Goldfarb–Shanno (BFGS) method is one of the most effective matrix-update or quasi Newton methods for solving unconstrained non-linear optimization problems. By applying the BFGS (Liu and Nocedal, 1989) optimizer as the optimization solver, we calibrate non-convex functions and deliver consistent statistical estimates through the inverse Hessian matrix $H^{-1}$ that was obtained through the algorithm. For the estimating variable $ASC_2$, the calculation formulas of the standard error and t-statistic are presented as Eq. (8) and Eq. (9).

$$\frac{\partial LL}{\partial ASC_2} = \frac{\partial LL}{\partial P_i} \frac{\partial P_i}{\partial ASC_1} + \frac{\partial LL}{\partial ASC_2} + \frac{\partial LL}{\partial ASC_3} + \frac{\partial LL}{\partial ASC_4} \frac{\partial ASC_2}{\partial ASC_1},$$

$$std.err = \sqrt{(H^{-1})_{ii} / N},$$
\[ t\text{.ratio} = \frac{ASC_2 \text{std.err}}{\text{std.err}} \]  

### 4.3. Mobility demand prediction

With the aim of predicting the day-to-day activity chain patterns during the pandemic through the Google mobility dataset and thus providing input predicted mobility data for the dynamic estimation of activity chain pattern, we introduce explainable time series prediction method for predicting the complex mobility demand pattern during public intervention. LSTM is a widely used method that is able to handle complex time series data for both long and short durations (Hochreiter and Schmidhuber, 1997). As temporal dependencies are the core components of our problem, by considering both the accuracy and efficiency, LSTM is chosen as the prediction method for the demand at various activity locations, as illustrated in Fig. 4. Based on the predicted multiple-day mobility dataset, day-to-day systematic utility changes of activity chains are predicted via the underlying estimation method introduced in Section 4.4.

The LSTM network can be essentially regarded as a non-linear optimization model to minimize the loss function \( L_T \) (Kim et al., 2020), which is defined as the mean square error of \( y_{tj} \) and \( \hat{y}_{tj} \):

\[ L_T = \sum_{t=1}^{T} (y_{tj} - \hat{y}_{tj})^2, \]

where \( y_{tj} \) and \( \hat{y}_{tj} \) represent the real mobility demand and the predicted mobility demand, respectively, of visiting the \( j \)th activity location at time step \( t \).

An LSTM unit is typically comprised of a memory cell, an input gate, a forget gate, and an output gate. At the current time step \( t \), given the input sample \( X_t \), the input gate variable \( i_t \), the memory cell variable \( C_t \), the forget gate variable \( f_t \), the output gate variable \( O_t \), the candidate hidden layer variable \( g_t \) and the hidden variable \( h_t \) are expressed in Eqs. (11)–(16), respectively.

\[
i_t = \sigma(W_i [h_{t-1}, X_t] + b_i), \]

\[ C_t = f_t \ast C_{t-1} + i_t \ast g_t \]

\[ f_t = \sigma(W_f [h_{t-1}, X_t] + b_f), \]

\[ O_t = \sigma(W_o [h_{t-1}, X_t] + b_o), \]

\[ g_t = \tanh(W_g [h_{t-1}, X_t] + b_g), \]

\[ h_t = O_t \ast \tanh(C_t), \]

Activation functions \( \tanh(\cdot) \) and \( \sigma(\cdot) \) are the key to capturing non-linear patterns in an LSTM unit. The operator \( \ast \) denotes the Hadamard product of matrices with the same dimensions. \( W_i, W_f, W_o, \text{and} W_g \) are weights from the input and previous hidden layer to the input gate, the forget gate, the output gate, and the candidate hidden layer, respectively. \( b_i, b_f, b_o, \text{and} b_g \) are the biases at the input gate, the forget gate, the output gate, and the candidate hidden layer, respectively.

The prediction result \( \hat{y}_{tj} \) is calculated by a fully connected layer from the hidden variable \( h_t \):

\[ \hat{y}_{tj} = W_h h_t + b_h \]

where \( W_h \) is the weight from hidden variable \( h_t \) to the predicted mobility demand \( \hat{y}_{tj} \), and \( b_h \) denotes the bias of the fully connected layer.

![Fig. 4. Architecture of the LSTM network.](image-url)
4.4. Dynamic estimation of pattern changes in daily activity chains

To infer the day-by-day activity engagement demand through partial mobility demand observation of activity locations, we define an activity-oriented deviation parameter $\Delta ASC_e^j$ within the interpretable utility-based NL framework. $\Delta ASC_e^j$ represents the change in the $j$th activity location’s alternative specific constant on the considered day $e$, and the day-to-day activity participation pattern estimation problem is represented as a non-linear constrained optimization problem. The objective function in this estimation problem is expressed as:

$$
\min \sum_{j=1}^{J} \frac{1}{P(AL_j)} |P'(AL_j) - (1 + y_e^j) P(AL_j)| + \theta \sum_{j=1}^{J} |\Delta ASC_e^j|,
$$

where $y_e^j$ is the mobility demand change percent at the $j$th activity location on the considered day $e$, which is provided by the Google Mobility Dataset; $\theta$ is the weight of the L1-norm regularization item, which is added to control the rangeability of $\Delta ASC_e^j$, and $P(AL_j)$ and $P'(AL_j)$ represent the $j$th activity location’s visitation probabilities on the baseline day and the considered day $e$. $P(AL_j)$ and $P'(AL_j)$ can be calculated from the probabilities of activity chains, as expressed in Eq. (19) and Eq. (20). $P(AC_i)$ is the average value of $P(AC_{ni})$ for all decision makers in the activity chain dataset, $P(AC_{ni})$ is determined by Eqs. (2)–(5), and $P'(AC_i)$ is similarly defined by adding $\Delta ASC_j$ to $ASC_j$.

$$
P(AL_j) = \sum_{i=1}^{I} N_i P(AC_i),
$$

$$
P'(AL_j) = \sum_{i=1}^{I} N_i P'(AC_i).
$$

By considering the accuracy and efficiency of alternative algorithms for the dynamic pattern estimation of activity chains, the SLSQP algorithm (Kraft and Schnepper, 1989) is selected and integrated into the proposed DACPE framework.

5. Results

In this section, we conduct experiments to evaluate the performance of the proposed interpretable DACPE framework. Based on the global mobility dataset and the activity chain dataset, we estimate and predict the daily changes in activity chain patterns during the pandemic in various states/areas across the United States using time-series data up to October 2020.

5.1. Mobility demand change analysis

In the Google Global Mobility Dataset, eight representative states/areas from six geographic regions across the United States are chosen as analysis objects, including Massachusetts in New England, New York, and Washington D.C. in the Mid-Atlantic, Florida in the South, Illinois in the Midwest, Arizona, and Texas in the Southwest, California in the West. General travel trends under public emergencies are relatively similar among the regions from February to October 2020. For example, the mobility demand changes of workplaces in selected regions are shown in Fig. 5, which exhibit readily observable periodicity characteristics of one week. As the baselines vary according to the day of the

![Fig. 5. Workplace travel trends from February to October in the selected eight states/areas.](image-url)
week, there is a “dropping-off” trend on weekdays compared to weekends. In general, the pandemic in the United States has gone through four stages until October 2020, namely, an outbreak period, a slow outbreak phase, a slow recovery phase, and stabilization, which correspond to reductions in public gatherings and outside activities, a slow decrease in trips, recovery of daily activities, and a “new normal” in offices. Specifically, before COVID-19 was declared a worldwide pandemic on March 11th by the World Health Organization (WHO), residents’ daily work routines in America were not obviously influenced. During the outbreak and recovery phases of the pandemic from March to June, people were less likely to work outside. As the pandemic entered a plateau from June to early October, the travel rate of workplaces increased to a relatively stable level. Among the eight selected areas, Washington D.C. appears to differ substantially from the others, with a larger decline in work-related activities during the outbreak and a smaller increase during the recovery phases of the pandemic. “Valleys” in the figure correspond to significant declines in mobility demand that were caused by federal holidays such as Washington’s Birthday, Memorial Day, American Independence Day, and Labor Day.

The correlations between mobility demands at various activity locations are investigated through the correlation matrix diagram shown in Fig. 6, and the corresponding Pearson correlations are specified in the picture. We find that changes in workplace, grocery, and park mobility are positively correlated with each other and negatively correlated with changes in home mobility. This can be explained by the fact that during the pandemic, with various policies and interventions adopted by governments, citizens prefer staying at home rather than going out to other places, especially to work places.

### 5.2. Systematic parameter estimation before COVID-19

Based on the activity chain dataset before the pandemic, we estimate baseline parameters as well as corresponding statistical properties (standard error and t-statistic) through the interpretable CG-based NL model, as presented in Table 4. As activity engagement patterns differ significantly between weekdays and weekends, the baseline parameters are estimated separately. In this table, \( \phi_1 \) is the corresponding coefficient of the travel distance attribute; \( r_i \) is the \( i^{th} \) activity chain’s constant (\( i \) ranges from 1 to 10); \( c_{mq} \) is the \( q^{th} \) characteristic of the traveler \( n \) (respondent age less than 30, age from 30 to 50, age more than 50, male, female, thus \( q \) ranges from 1 to 5); \( \beta_{iq} \) is the coefficient of the \( m^{th} \) characteristic \( c_{mq} \) for the \( i^{th} \) activity chain, and \( ASC_i \) is the alternative specific constant of the \( i^{th} \) activity location. To form the baseline case for all the parameters, \( \beta_{1q} \) (which represents the coefficient of the home-home chain’s characteristics) and \( ASC_1 \) (which represents the alternative specific constant of home) are set to zero. There are three nests in our nested logit model (\( m \) ranges from 1 to 4), \( \theta_m \) is a logsum parameter bounded by zero to one, an indicator of the correlation between alternatives in the same nest. Relatively low standard errors and high t-statistic values demonstrate the accuracy and reliability of the parameter estimation. From the parameter estimation result in Table 4, it can be noted that the value of \( \beta_{3q} \) is much higher than those for \( \beta_{1q} \) and \( \beta_{2q} \), which indicates that middle-aged citizens are more likely to participate in various activities. Besides, with the values of \( \beta_{1q} \) in nest 2 higher than \( \beta_{15} \) and \( \beta_{13} \) in nest 3 higher than \( \beta_{14} \), it could be linked to the interpretation that women prefer to adopt grocery-related activities, whereas men are likely to take part in more work-related activities. From the coefficient variance on weekdays and weekends, we observe interesting and interpretable phenomena. For instance, the decrease in \( ASC_2 \) and the increases in \( ASC_3 \) and \( ASC_4 \) on weekends indicate that people prefer shopping and visiting parks on their days off.

To evaluate the model, the negative log likelihood function before and after the calibration (-LL(initial) and -LL(final)), and the likelihood ratio test statistic \( \lambda_{LR} \) are listed in Table 4. LL is expressed as Eq. (6). By defining the null hypothesis \( H_0 \) as the NL model before coefficient calibration and the alternative hypothesis \( H_1 \) model as the model after calibration, the likelihood ratio test value can be defined as:

\[
\lambda_{LR} = 2[\text{LL(final)} - \text{LL(initial)}]
\]  

(21)

Fig. 6. Correlation matrix diagram of various places in the selected eight states/areas.
Table 4
Parameter estimation results for weekdays and weekends

| Variable | Coef. | Std. err. | t-ratio | Coef. | Std. err. | t-ratio |
|----------|-------|-----------|---------|-------|-----------|---------|
| Nest1    |       |           |         |       |           |         |
| $\beta_1$ | 0.2951 | 0.0096  | 30.7819 | 0.3494 | 0.0163  | 21.4730 |
| $\mu_{11}$ | 0 | NA | NA | 0 | NA | NA |
| $\mu_{12}$ | 0.0190 | 0.0101 | -18.9574 | 0.2006 | 0.0126 | -15.9153 |
| $\mu_{41}$ | 0.0073 | 0.0081 | 0.8984 | 0.0100 | 0.0113 | -0.8843 |
| $\mu_{42}$ | -0.0181 | 0.0066 | -15.3848 | 0.1282 | 0.0114 | -11.2831 |
| $\mu_{43}$ | -0.1424 | 0.0079 | -18.0115 | -0.1531 | 0.0118 | -12.9843 |
| $\mu_{44}$ | -0.1431 | 0.0078 | -18.2900 | -0.1857 | 0.0121 | -15.2963 |

| Nest2    |       |           |         |       |           |         |
| $\beta_1$ | 0.8202 | 0.0099  | 82.7984 | 0.8720 | 0.0132 | 65.9350 |
| $\mu_{11}$ | -0.6938 | 0.0135 | -51.3507 | -0.3984 | 0.0229 | -17.3679 |
| $\mu_{12}$ | 1.3966 | 0.0122 | 114.0269 | 1.2019 | 0.0144 | 83.4318 |
| $\mu_{13}$ | 0.1279 | 0.0109 | 11.7168 | 0.0191 | 0.0107 | 1.7768 |
| $\mu_{14}$ | 0.5151 | 0.0085 | 60.6498 | 0.5199 | 0.0145 | 35.8417 |
| $\mu_{15}$ | 0.3155 | 0.0090 | 34.9877 | 0.3026 | 0.0142 | 21.3196 |
| $\mu_{16}$ | -1.2790 | 0.0228 | -56.1781 | -1.1624 | 0.0181 | -64.2317 |
| $\mu_{17}$ | 0.7721 | 0.0160 | 48.1411 | 0.5638 | 0.0265 | 21.3042 |
| $\mu_{21}$ | -0.4817 | 0.0156 | -80.8259 | -0.3264 | 0.0263 | -12.4244 |
| $\mu_{22}$ | -0.4227 | 0.0123 | -34.2497 | -0.3648 | 0.0126 | -28.9633 |
| $\mu_{23}$ | -0.5659 | 0.0116 | -48.9696 | -0.5602 | 0.0122 | -45.8798 |
| $\mu_{24}$ | -1.2690 | 0.0183 | -69.2833 | -1.0068 | 0.0362 | -27.8106 |
| $\mu_{25}$ | 1.1748 | 0.0150 | 78.3641 | 0.9284 | 0.0251 | 37.0265 |
| $\mu_{26}$ | -0.0983 | 0.0161 | -6.1161 | -0.1565 | 0.0209 | -7.4780 |
| $\mu_{27}$ | 0.0729 | 0.0120 | 6.0786 | 0.0979 | 0.0107 | 9.1688 |
| $\mu_{28}$ | -0.2654 | 0.0124 | -21.3613 | -0.3328 | 0.0125 | -26.5616 |
| $\mu_{29}$ | -1.3422 | 0.0195 | -68.7939 | -1.1262 | 0.0303 | -37.2045 |
| $\mu_{31}$ | 1.0715 | 0.0165 | 64.7978 | 0.8069 | 0.0277 | 29.1280 |
| $\mu_{32}$ | 0.0897 | 0.0165 | 5.4935 | 0.0050 | 0.0176 | -0.2819 |
| $\mu_{33}$ | -0.1543 | 0.0122 | -12.6486 | -0.1892 | 0.0115 | -16.4026 |
| $\mu_{34}$ | -0.0266 | 0.0115 | -2.3069 | -0.1350 | 0.0122 | -11.0963 |

| Nest3    |       |           |         |       |           |         |
| $\beta_1$ | 0.5990 | 0.0063  | 95.4051 | 0.7716 | 0.0118 | 65.4338 |
| $\mu_{11}$ | -1.0010 | 0.0162 | -61.6591 | -0.7744 | 0.0208 | -37.3094 |
| $\mu_{12}$ | 0.4111 | 0.0174 | 23.5834 | 0.4181 | 0.0183 | 22.8089 |
| $\mu_{13}$ | 0.6167 | 0.0116 | 53.2802 | 0.5757 | 0.0125 | 46.0809 |
| $\mu_{14}$ | -0.0632 | 0.0096 | -6.5597 | 0.0662 | 0.0139 | 4.7671 |
| $\mu_{15}$ | 0.0900 | 0.0086 | 10.4589 | 0.1531 | 0.0121 | 12.6184 |
| $\mu_{16}$ | -1.1467 | 0.0218 | -52.6902 | -0.9956 | 0.0182 | -54.6051 |
| $\mu_{17}$ | 0.4972 | 0.0216 | 22.9959 | 0.4833 | 0.0218 | 22.1503 |
| $\mu_{18}$ | 0.7817 | 0.0148 | 52.6469 | 0.6616 | 0.0174 | 38.1214 |
| $\mu_{19}$ | -0.0486 | 0.0089 | -5.4638 | -0.0731 | 0.0081 | -9.0138 |
| $\mu_{20}$ | 0.1808 | 0.0100 | 18.0029 | 0.2225 | 0.0098 | 22.6160 |
| $\mu_{21}$ | -1.4079 | 0.0224 | -62.8436 | -1.2982 | 0.0181 | -71.5293 |
| $\mu_{22}$ | 0.1769 | 0.0208 | 8.5176 | 0.1301 | 0.0220 | 5.9236 |
| $\mu_{23}$ | 0.3049 | 0.0143 | 21.2673 | 0.2234 | 0.0142 | 15.7747 |
| $\mu_{24}$ | -0.5425 | 0.0121 | -44.8481 | -0.4937 | 0.0092 | -53.5646 |
| $\mu_{25}$ | -0.3837 | 0.0103 | -37.3786 | -0.4510 | 0.0100 | -45.2482 |

| Nest4    |       |           |         |       |           |         |
| $\beta_1$ | 1 | NA | NA | 1 | NA | NA |
| $\mu_{11}$ | -0.5891 | 0.0177 | -33.3265 | -0.4221 | 0.0229 | -18.4478 |
| $\mu_{12}$ | -0.2181 | 0.0324 | -6.7230 | -0.0439 | 0.0391 | -1.1228 |

(continued on next page)
The log-likelihood ratio test value $\lambda_{LR}$ of the comparison between the two hypotheses is 87454.75/21035.617 for the estimation of weekdays/weekends. This value is significantly larger than the chi-squared value with 52 degrees of freedom at a 0.001 level of significance (Lavieri and Bhat, 2019), thereby leading to the rejection of $H_0$ and the acceptance of $H_1$, thus implying that the model calibration process is effective.

5.3. Mobility demand prediction during COVID-19

The mobility demand prediction problem in this study aims at predicting the mobility demand of the $j$th activity location at the future time step $t$ (i.e., $y_{jt}$) based on the historical demand at past timesteps. The training dataset consists of 214 days from February 15th to September 15th in 26 states/areas, and the testing dataset remains unchanged with 238 days from February 15th to October 9th in 8 states/areas. Accordingly, we carried out the prediction of different regions and subsequent time periods. To better illustrate the mobility change pattern, we embed additional factors including the workday and holiday information in our LSTM network. The LSTM network with additional factors is denoted as LSTM+. We compare the mobility demand prediction performance of the LSTM+ and LSTM networks (Hochreiter and Schmidhuber, 1997) with those of various baseline methods, namely, the historical average model (HA) (Liu and Guan, 2004), the ARIMA (Ahmed and Cook, 1979), the SVR (Castro-Neto et al., 2009), and the extreme gradient boosting (XGBoost) (Dong et al., 2018).

To evaluate the performance of the model, five metrics are utilized to measure the difference between predicted results and real observations, namely, the root mean square error (RMSE), symmetric mean absolute error (MAE), accuracy (ACC), coefficient of determination ($R^2$), and the adjusted $R^2$, which are defined mathematically as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_{ij} - \hat{y}_{ij})^2}$$

(22)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_{ij} - \hat{y}_{ij}|.$$  

(23)

$$ACC = 1 - \frac{\|Y - \hat{Y}\|_F}{\|Y\|_F}.$$  

(24)

| Table 4 (continued) |
|----------------------|
| Variable | Coef. | Std. err. | t-ratio | Coef. | Std. err. | t-ratio |
| $\beta_{63}$ | $-0.3735$ | $0.0229$ | $-16.3130$ | $-0.3925$ | $0.0249$ | $-15.7710$ |
| $\beta_{64}$ | $-0.5454$ | $0.0086$ | $-63.6089$ | $-0.3521$ | $0.0112$ | $-31.5448$ |
| $\beta_{65}$ | $-0.6352$ | $0.0081$ | $-78.7260$ | $-0.5064$ | $0.0092$ | $-54.9034$ |

Table 5
Prediction results of the LSTM+ network and other baseline methods (average of the selected 8 states/areas)

| Horizon | Method | RMSE | MAE | ACC | $R^2$ | Adjusted $R^2$ |
|---------|--------|------|-----|-----|-------|----------------|
| 1 day   | HA     | 12.0083 | 9.5018 | 0.6630 | $-0.0097$ | $-0.0141$ |
|         | ARIMA  | 12.3405 | 7.0617 | 0.6834 | 0.5270 | 0.5249 |
|         | SVR    | 6.8966 | 3.1194 | 0.7980 | 0.5966 | 0.5948 |
|         | XGBoost| 4.9620 | 2.9845 | 0.8700 | 0.8999 | 0.8995 |
|         | LSTM   | 4.7777 | 2.9856 | 0.8704 | 0.9191 | 0.9187 |
|         | LSTM+  | 3.2849 | 2.6498 | 0.9149 | 0.9448 | 0.9441 |
| 7 day   | HA     | 12.3916 | 9.358193 | 0.650126 | $-0.05805$ | $-0.0628$ |
|         | ARIMA  | 12.34055 | 7.061683 | 0.683435 | 0.527044 | 0.5249 |
|         | SVR    | 7.53546 | 3.899665 | 0.778755 | 0.735749 | 0.7337 |
|         | XGBoost| 7.048088 | 4.076131 | 0.806587 | 0.723452 | 0.7222 |
|         | LSTM   | 6.7045 | 4.1594 | 0.8370 | 0.7944 | 0.7935 |
|         | LSTM+  | 5.1969 | 3.4205 | 0.8624 | 0.8715 | 0.8698 |
where \( n \) is the number of samples in the test set, \( F \) is the total number of features used in the prediction method, \( y'_j \) (without symbol) is the real value at the \( j^{th} \) activity location of the test mobility sample at timestep \( t \), \( \hat{y}'_j \) (\( y'_j \) with circumflex) represents the predicted mobility demand of the \( j^{th} \) activity location at timestep \( t \) and \( \bar{y}_j \) (\( y'_j \) with bar) is the average mobility demand of the \( j^{th} \) activity location. \( \| \cdot \|_F \) is the Frobenius norm of a matrix. Since \( Y_j \) and \( \bar{Y}_j \) are \( n \)-dimensional vectors in our model, the Frobenius norm is the same as the L2 norm.

Table 5 presents the mobility prediction result at workplaces in comparison with those of other traditional methods for one-step ahead prediction and 7-step ahead prediction. Based on the reported results, the LSTM + network offers the best forecasting performance in terms of all evaluation metrics. Thus, our method can precisely capture the complex non-linear mobility change trend, and prepare for the accurate prediction of dynamic activity pattern.

### 5.4. Dynamic activity pattern estimation and prediction during COVID-19

Given observed/predicted mobility demands at various activity locations, we finally estimate/predict the activity chain probability in the DACPE framework. Based on the hypotheses of the model, the mobility demand change percentages at various activity locations can be calculated from the estimation result. Fig. 7 compares the estimated mobility demand change percentage and the ground truth from the Google Mobility Dataset. Note that the ground truth and the estimated results are well aligned. The high degree of consistency between the estimated result and the ground truth demonstrates the effectiveness of our utility-based estimation method.

In Fig. 8, the estimation and prediction results of representative areas are represented by dashed lines and solid lines, respectively. It is evident that the predicted result matches well with the estimated outcome, which demonstrates the effectiveness of the proposed prediction methods. From the cross-comparative analysis of activity participation probability changes across the United States, we can roughly deduce that residents’ behavioral changes have similar trends in multiple areas. The weekly periodicity of the activity probability in Massachusetts, New York, and Illinois is more readily identifiable than that in other states. There are intersections of work-related and home-home activity chain probability changes in all areas during the outbreak period. However, whereas in the pandemic recovery phases across the United States, most areas have entered a new normal of work, residents in Washington D.C. still prefer staying at home and working online. In addition, we can infer that there are small aftershocks reshaping activity engagement trends in Florida, Arizona, Texas, and California from the rebound of the activity probability curve.

From the probability changes of activity chains, we can further identify individual behavioral pattern adjustments from various areas. During the pandemic, the probabilities of work-related activity chains underwent intense drops, a slowly decreasing period, a recovery phase, and a plateau, while the probabilities of stay-at-home activity chains were in the opposite direction. Residents’ desire for park-related activities has substantially decreased, and grocery-related activity seems to be unaffected. Furthermore, occasional peaks and valleys can be explained by the sharp change of activity chain patterns on the national legal holidays. For instance, on the Day of Mourning, the number of people staying at home increased (+38%), working decreased significantly (-70%), and the other two items did not change much. Although the pandemic outbreak or holidays cause sudden changes of mobility as illustrated in Fig. 5 and Fig. 7, the adoption of effective utility theory, robust parameter estimation process and accurate mobility prediction method make the framework able to capture the sharp changes and provide reliable activity chain estimation/prediction result.

Overall, our proposed framework provides a new perspective for understanding the essential reasons for mobility demand changes, and the consistency in the results across various areas demonstrates the robustness and transferability of the proposed framework. Based on multiple heterogeneous data sources, a reliable inference of unobserved behavioral changes can be realized for policy adaptation and management decisions. Concretely, local government can identify risky activities/locations and make dedicated policies with the activity pattern estimation/prediction result. For example, we may observe that grocery-related activities hardly decline as some shopping are inevitable, in which case strict prevention management and online purchasing should be encouraged (Chen and Pan, 2020; Hu et al., 2021). In addition, as the recovery of work-related activities and transportation, policies like physical distancing, talk reducing and mask wearing should be promoted in order to effectively prevent infections in closed environments such as workplace, conference room, carriages, and underground platforms (Lu et al., 2021; Tirachini and Cats, 2020; Yin et al., 2021). Moreover, non-pharmaceutical interventions like school/park/company closures and bans on public events can be implemented according to the activity preference of local residents (Hadjidemetriou et al., 2020; Haug et al., 2020; Yamamoto et al., 2021).

### 6. Conclusions

Activity participation is a fundamental driver of transportation demand. During the COVID-19 crisis, researchers have been searching for ways to develop a deeper understanding of activity/travel behavior changes and mobility demand changes. To address
the knowledge gap for dynamic activity estimation/prediction using heterogeneous datasets, we first utilize baseline household survey data (individual-based activity chain data) and Google mobility data (dynamic group-based activity location data). And then a new dynamic activity chain pattern estimation framework for obtaining day-to-day location visiting information and identifying behavioral adjustment patterns during the pandemic and recovery phases is developed.

Citizens in different regions may have different activity/travel patterns affected by the epidemic (Arimura et al., 2020; Shakibaei et al., 2021). Based on the activity chain baseline dataset in other regions, we can re-estimate behavioral parameters when estimating their daily activity pattern changes during the pandemic. This paper also examines visiting probability for various activity chains/locations, and thus shed light on macroscopic management policies across regions (Chen and Pan, 2020; Haug et al., 2020).

Furthermore, due to the transferable characteristic of the interpretable machine learning framework, the proposed approach can be easily applied to other scenarios with integrated disaggregated survey data and widely accessible aggregated data.

In our future research, we will consider in-depth validation to enhance the reliability of the model across different conditions by incorporating more new knowledge and observations of other planners and researchers. In addition, the proposed framework will be enhanced by adopting a broader context of macro-micro transportation demand estimation/prediction. First, by integrating various types of data in different domains, such as valuable pandemic datasets (e.g., the COVID-19 cases/deaths/recoveries, government response events), and survey datasets that include individual and family-need-related attributes (e.g., population, income, household income, count of household members, number of workers in household, age of family members), more continuous and discrete variables can be added into our framework, and therefore we can better understand and predict mobility demand changes and activity engagement patterns. Second, by modifying the dynamic estimation and time series prediction mechanism, our model can be extended to short-term local major events (e.g., concerts, sporting events, large exhibitions, education conventions) and capture individuals’ behavioral adjustments in special periods. Last but not the least, besides the considered circumstance of commuting-oriented urban transportation, it will be useful to construct a state-wide/national model based on the current framework to explore long-distance intercity travel patterns considering macro-economic trends, trip purposes, and transportation modes.

Fig. 7. Percentage changes of activity locations in representative states/areas using the Google Mobility dataset as the benchmark, ((a)-(h) ACC: 0.8734, 0.8760, 0.9604, 0.9286, 0.8590, 0.9259, 0.9223, 0.9285).
CRediT authorship contribution statement

Yan Liu: Methodology, Visualization, Writing – review & editing. Lu Carol Tong: Conceptualization, Methodology, Writing – review & editing. Xi Zhu: Investigation, Writing – review & editing. Wenbo Du: Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the National Key Research and Development Program of China (2019YFF0301400), the National Natural Science Foundation of China (61961146005, 71801006).

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In: 12th symposium on operating systems design and implementation (16), 265–283.
Abdullah, M., Dias, C., Muley, D., Shahin, M.d., 2020. Exploring the impacts of COVID-19 on travel behavior and mode preferences. Transport. Res. Interdisciplinary Perspect. 8, 100255. https://doi.org/10.1016/j.trip.2020.100255.
Ahmed, M.S., Cook, A.R., 1979. Analysis of freeway traffic time-series data by using box-jenkins techniques. Transport. Res. Board 722, 1–9.
Anand, A., Ramadurai, G., Vanajakshi, L., 2014. Data fusion-based traffic density estimation and prediction. J. Intell. Transport. Syst. 18 (4), 367–378. https://doi.org/10.1080/15472450.2013.806844.
Antoniou, C., Koutsopoulos, H.N., Yannis, G., 2013. Dynamic data-driven local traffic state estimation and prediction. Transport. Res. Part C: Emerg. Technol. 34, 89–107. https://doi.org/10.1016/j.trc.2013.05.012.
Apple, 2020. Mobility Trend Data. https://www.apple.com/covid19/mobility.
Yang, Z., Mei, D., Yang, Q., Zhou, H., Li, X., 2014. Traffic flow prediction model for large-scale road network based on cloud computing. Math. Problems Eng. 2014, 1–8. https://doi.org/10.1155/2014/926251.

Yin, Y., Li, D., Zhang, S., Wu, L., 2021. How Does Railway Respond to the Spread of COVID-19? Countermeasure Analysis and Evaluation Around the World. Urban Rail Transit 7 (1), 29–57. https://doi.org/10.1007/s40864-021-00140-x.

Yu, B., Yin, H., Zhu, Z., 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.

Zhang, J., Zheng, Y., Qi, D., Li, R., Yi, X., 2016. DNN-based prediction model for spatio-temporal data. In: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 1–4. https://doi.org/10.1145/2996913.2997016.

Zhang, L., Ghader, S., Pack, M.L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., Hu, S., 2020. An interactive COVID-19 mobility impact and social distancing analysis platform. medRxiv. https://doi.org/10.1101/2020.04.29.20095472.

Zhang, M., Chen, S., Sun, L., Du, W., Cao, X., 2021. Characterizing flight delay profiles with a tensor factorization framework. Engineering 7 (4), 465–472. https://doi.org/10.1016/j.eng.2020.08.024.

Zhang, Y., Haghani, A., 2015. A gradient boosting method to improve travel time prediction. Transport. Res. Part C: Emerg. Technol. 58, 308–324. https://doi.org/10.1016/j.trc.2015.02.019.

Zhao, L., Song, Y., Zhang, C., Liu, Y.u., Wang, P.u., Lin, T., Deng, M., Li, H., 2019. T-gcn: A temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst. 21 (9), 3848–3858. https://doi.org/10.1109/TITS.697910.11109/TITS.2019.2935152.

Zhao, S., Zhang, K., 2017. Observing individual dynamic choices of activity chains from location-based crowdsourced data. Transport. Res. Part C: Emerg. Technol. 85, 1–22. https://doi.org/10.1016/j.trc.2017.09.005.