

# A causal intervention framework for synthesizing mobility data and evaluating predictive neural networks

Ye Hong<sup>a,\*,\*</sup>, Yanan Xin<sup>a,d</sup>, Simon Dirmeier<sup>b</sup>, Fernando Perez-Cruz<sup>c,e</sup>, Martin Raubal<sup>a</sup>

<sup>a</sup> Institute of Cartography and Geoinformation, ETH Zurich, Switzerland

<sup>b</sup> Swiss Data Science Center, ETH Zurich, Switzerland

<sup>c</sup> Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland

<sup>d</sup> Department of Transport and Planning, TU Delft, Netherlands

<sup>e</sup> Bank for International Settlements, Switzerland

## ARTICLE INFO

Dataset link: <https://github.com/irmlma>

### Keywords:

Mobility behavior

Domain shift

Individual mobility simulation

Next location prediction

Causal intervention

## ABSTRACT

Deep neural networks are increasingly utilized in mobility prediction tasks, yet their intricate internal workings pose challenges for interpretability, especially in comprehending how various aspects of mobility behavior affect predictions. This study introduces a causal intervention framework to assess the impact of mobility-related factors on neural networks designed for next location prediction — a task focusing on predicting the immediate next location of an individual. To achieve this, we employ individual mobility models to synthesize location visit sequences and control behavior dynamics by intervening in their data generation process. We evaluate the interventional location sequences using mobility metrics and input them into well-trained networks to analyze performance variations. The results demonstrate the effectiveness in producing location sequences with distinct mobility behaviors, thereby facilitating the simulation of diverse yet realistic spatial and temporal changes. These changes result in performance fluctuations in next location prediction networks, revealing impacts of critical mobility behavior factors, including sequential patterns in location transitions, proclivity for exploring new locations, and preferences in location choices at population and individual levels. The gained insights hold value for the real-world application of mobility prediction networks, and the framework is expected to promote the use of causal inference to enhance the interpretability and robustness of neural networks in mobility applications.

## 1. Introduction

Accurate individual mobility prediction plays a pivotal role in popularizing emerging mobility services (Ma and Zhang, 2022) and serves as a crucial backbone for various intelligent transport system functionalities (Tang et al., 2019). Despite its application potential, predicting individual mobility remains challenging due to the complexity of mobility patterns influenced by diverse behavioral factors and contexts (Hong et al., 2023b). The intricacies in modeling these spatiotemporal dependencies generally hinder the prediction performance of individual mobility (Song et al., 2010b; Barbosa et al., 2018; Wiedemann et al., 2023a). In recent years, the availability of human digital traces and the advancements in data-driven models, particularly deep neural networks capable of capturing spatiotemporal dynamics, have significantly enhanced mobility prediction ability (Wang et al., 2022).

Despite their solid predictive performance, modern neural networks often face criticism for their low interpretability (Manibardo et al.,

2022; Pappalardo et al., 2023), referring to the degree to which humans can comprehend the decision-making process of a model. These networks are commonly regarded as “black boxes” because reconstructing the reasoning behind a particular prediction is challenging. In mobility prediction, the lack of interpretability leads to an unclear understanding of the spatiotemporal patterns captured by the network and, more fundamentally, the influence of behavioral factors (such as location preferences and activity radius (Yuan et al., 2012)) in prediction. This deficiency negatively affects decision-making, policy design, and the perceived reliability and trustworthiness among practitioners (Huang et al., 2020), thereby impeding the seamless integration of mobility prediction networks into real-world applications (Koushik et al., 2020). Furthermore, the scarcity of publicly available individual mobility datasets, primarily due to the privacy-sensitive nature of personal mobility (Wiedemann et al., 2023b), leads to a lack of comparability

\* Corresponding author.

E-mail addresses: [hongy@ethz.ch](mailto:hongy@ethz.ch) (Y. Hong), [yanan.xin@tudelft.nl](mailto:yanan.xin@tudelft.nl) (Y. Xin), [simon.dirmeier@sdsc.ethz.ch](mailto:simon.dirmeier@sdsc.ethz.ch) (S. Dirmeier), [fernando.perezacruz@inf.ethz.ch](mailto:fernando.perezacruz@inf.ethz.ch) (F. Perez-Cruz), [mraubal@ethz.ch](mailto:mraubal@ethz.ch) (M. Raubal).

<https://doi.org/10.1016/j.trip.2025.101398>

Received 10 March 2024; Received in revised form 22 July 2024; Accepted 18 March 2025

Available online 2 May 2025

2590-1982/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

between existing and newly developed prediction models (Graser et al., 2023). Prediction networks are evaluated using datasets that include varying numbers and types of participants, along with differing tracking durations, representing diverse snapshots of the possible mobility behavior (Kulkarni and Garbinato, 2019). Hence, a comprehensive analysis connecting behavior dynamics with prediction performance is imperative to establish benchmark data specifications for evaluating modern neural networks employed in mobility studies.

In addition, establishing the behavior and performance connection assists in evaluating the robustness of these networks when confronted with unforeseen inputs. The optimization of neural networks requires a training dataset, making their performances heavily dependent on the quality and representativeness of this data (Yin et al., 2022). However, mobility behavior evolves dynamically over space and time due to internal needs (e.g., behavioral exploration (Hong et al., 2023a)) and environmental factors (e.g., land use (Acheampong, 2018)). Consequently, the mobility data encountered during application often reflects different behavior than the training data, leading to a discrepancy known as domain shift (He et al., 2020). Enhancing our understanding of performance under various shift scenarios is essential to assessing reliability when applying these networks across diverse geographic regions or time periods. Yet, this relationship remains predominantly unexplored.

Causal intervention offers a promising tool for generating data from diverse environments, enabling the assessment of neural network robustness and providing human-friendly causal explanations for these interventions (Xin et al., 2022). Building upon its advantages, we present a framework for systematically evaluating the impact of mobility behaviors on prediction networks. Specifically, this framework utilizes individual mobility models to generate mobility traces, and employs causal intervention strategies in the data generation process, allowing for flexible modifications of the defined mobility behavior. We subsequently assess the performance of trained neural networks on these synthetic traces for mobility prediction. Our study focuses on next location prediction, which aims to forecast an individual's immediate next location based on their mobility history. The interventions simulated realistic spatial and temporal changes in mobility patterns, leading to performance fluctuations in prediction networks that reflect their robustness when confronted with domain shifts. This framework facilitates assessing the impact of behavioral factors and benchmarking mobility prediction networks, with practical applications for evaluating network performances and transferring these networks across environments. In short, our contributions are summarized as follows:

- We introduce a framework to assess the robustness of mobility prediction networks through causal intervention. This framework enables direct control over behavioral dynamics, quantifying ensuing mobility patterns, and evaluating their influence on network performance.
- We use this framework to simulate spatiotemporal shift scenarios, demonstrating its effectiveness in benchmarking mobility datasets, identifying performance degradation during real-world applications, and improving the interpretability of prediction networks.
- We open-source the framework, enabling straightforward utilization and flexible customization of its components.<sup>1</sup>

## 2. Related work

### 2.1. Mobility behavior and its impact on mobility prediction

Research on individual mobility behavior, with an emphasis on spatiotemporal patterns of activities and trips, has consistently been at the

forefront of mobility studies (Chen et al., 2016), propelling theoretical and methodological advancements within the activity-based analysis framework (Schönfelder and Axhausen, 2016). In this framework, trips are perceived as an induced demand resulting from the necessity to engage in activities at distinct spatial locations (Axhausen and Gärling, 1992), making location selection a pivotal aspect in comprehending individual mobility behavior (Sener et al., 2011). Studies have consistently shown that location choices not only vary across populations (inter-person variability) (Martin et al., 2023b; Ji et al., 2023) but also undergo constant changes over time (intra-person variability) (Susilo and Axhausen, 2014; Hintermann et al., 2023). Furthermore, these decisions are frequently influenced by external factors, some of which, when encountered, can lead to sudden structural changes, such as residential relocation (Ramezani et al., 2021) and large-scale crises (Kellermann et al., 2022). These empirical insights emphasize the need to study activity location choices from a dynamic perspective (Shou and Di, 2018), rather than approaching them solely with static mobility snapshots.

Considering the central role of activity, predicting activity location is a crucial component in mobility ahead planning and optimization (Ma and Zhang, 2022), and has garnered widespread attention (Luca et al., 2021). Accurately inferring the next location is influenced not only by the capability of the employed model (e.g., neural networks) but also by individuals' behavioral patterns. In a pioneering effort to link mobility behavior with location prediction, Song et al. (2010b) proposed entropy as a measure for the theoretical mobility predictability based on visited location sequences of individuals. Their study revealed remarkably consistent predictability across individuals, peaking at approximately 93% in the tested dataset, reflecting inherent travel patterns across various demographic attributes (Song et al., 2010b). Although studies have highlighted their strong correlation with the actual prediction performance (Lu et al., 2013), entropy-related predictability measures remain challenging to interpret (Teixeira et al., 2019). Recent research has started to represent mobility behavior with more straightforward metrics, such as routine and novelty components (Teixeira et al., 2021), usages of transport services (Xu et al., 2022) and recurring patterns in daily mobility (Hong et al., 2023b), and analyze its impact on prediction performance. Nevertheless, our comprehension of these connections is still in its early stages, relying on observed mobility behavior from tracking datasets, which often exhibit limited behavior variability and biased behavior distribution.

### 2.2. Causal intervention and its application in robustness assessment

The field of causal modeling, which investigates causal relationships, provides a systematic framework for generating mobility traces that depict specific behaviors. Establishing a structural causal model (SCM) is essential to describe the causal mechanisms within a system comprising various interconnected factors. In this context, the data generation process is viewed as a causal process, where the input variables of the generation model are considered the causes of the output, with their relationships modeled through SCM (Rahimi et al., 2023). With access to a particular SCM, we can perform interventions on individual factors or combinations thereof, effectively disentangling intricate interactions among variables by unilaterally adjusting the value of a single variable and observing its impact on the generated data (Pearl and Mackenzie, 2018). Utilizing causal interventions to generate controlled results has diverse applications. For instance, in prognostic research, causal intervention has been applied to forecast risks associated with various medical treatments (van Amsterdam et al., 2019), examine the impact of therapy on the cognitive development of prematurely born infants (Silva, 2016), and investigate the effects of demographic variations on brain structure using counterfactual samples (Pawlowski et al., 2020). Similar strategies have also been applied in climate and earth science to unravel the complexities of dynamic spatiotemporal processes (Li et al., 2023; Runge et al., 2023).

<sup>1</sup> The source code is available at <https://github.com/irmlma>.

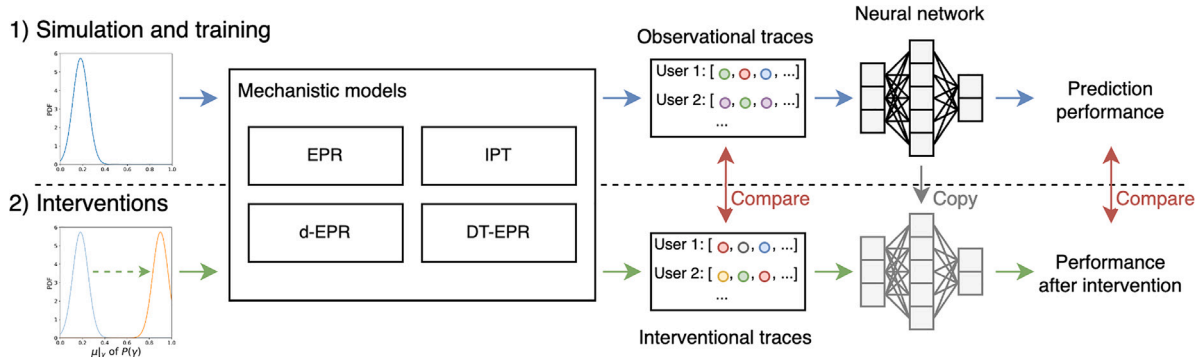


Fig. 1. Framework for evaluating the robustness of prediction networks through causal interventions. We generate location sequences from mechanistic models and feed them into prediction networks to evaluate the prediction performance (blue arrows). This process is repeated for interventional location sequences, obtained by modifying the distribution of behavioral parameters (green arrows). The differences in mobility patterns and prediction performances are compared to assess intervention strengths and network robustness (red arrows). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Causal intervention methods have demonstrated potential in evaluating and enhancing the robustness of neural networks (Schölkopf et al., 2021; Moshkov et al., 2024), where robustness refers to the degree of performance variations under domain shift (Zhou et al., 2023). In this study, we used SCM to abstract the data generation process of mechanistic mobility simulators, which incorporate parameters representing individual behavior to synthesize mobility traces. Noteworthy simulators include the exploration and preferential return (EPR) model (Song et al., 2010a), the location attractiveness model (Yan et al., 2017), and the container model (Alessandretti et al., 2020). These models have significantly advanced in replicating high-level spatiotemporal patterns of individual movements (Pappalardo et al., 2023) and are increasingly applied in large-scale simulations (Xu et al., 2018b; Barbosa et al., 2018). Representing these models as SCMs enables the generation of both in-distribution and out-of-distribution (OoD) data by intervening in input variables and adjusting their strengths. For example, increasing the tendency of individuals to explore new locations in the EPR model allows us to generate mobility data from a more explorative population. As a result, various interventions in mobility models produce a spectrum of data, each reflecting a realistic domain shift, enabling us to assess the robustness of prediction networks in different scenarios.

### 3. Methodology

The overall pipeline for assessing the robustness of prediction networks is illustrated in Fig. 1. We start by introducing mechanistic generative models for synthesizing individual location sequences (§3.1). These models incorporate parameters to replicate real-world observational mobility behavior. Subsequently, we perform interventions by modifying the parameters, thereby manipulating mobility behaviors and generating new interventional location sequences (§3.2). Lastly, we train mobility prediction networks using observational location sequences and evaluate them on interventional sequences (§3.3). Prediction performance changes reflect these networks' robustness when behavioral interventions are introduced. In this study, the term *location* refers specifically to a geographical place where individuals engage in activities. This excludes other points, such as waypoints on roads (e.g., GPS recordings) or intermediate stops without meaningful activity (e.g., waiting for a bus). We provide a more detailed description of each module in the following sections.

#### 3.1. Individual mobility models

Individual mobility models generate realistic movement trajectories based on a predefined set of behavioral parameters, allowing for direct control over the mobility behavior of the generated population. We

start with the EPR model (Song et al., 2010a) as our baseline SCM and additionally introduce two EPR-based generative models, namely density (d)-EPR (Pappalardo et al., 2015) and individual preferential transition (IPT) (Zhao et al., 2021). We finally propose the density transition (DT)-EPR model that combines the advantages of d-EPR and IPT to obtain more realistic mobility traces. Fig. 2 delineates the mechanisms of DT-EPR and its connection with d-EPR and IPT models.

EPR generates sequences of location visits for individuals and reproduces scaling laws for distinct location numbers and their visitation frequency over time (Barbosa et al., 2018). The core of EPR is the introduction of two competing mechanisms, exploration and preferential return, into classical random-walk models (Brockmann et al., 2006). These two mechanisms account for the tendency of individuals to return to previously visited locations (Song et al., 2010a). Specifically, observing an individual at location  $i$  at time  $t$ , the model assumes that the individual will change their location after a waiting time  $\Delta t$ , where  $\Delta t$  is sampled from its distribution  $P(\Delta t)$ . The individual chooses to explore a previously unvisited location with probability  $p_{t+\Delta t}^{new}$ :

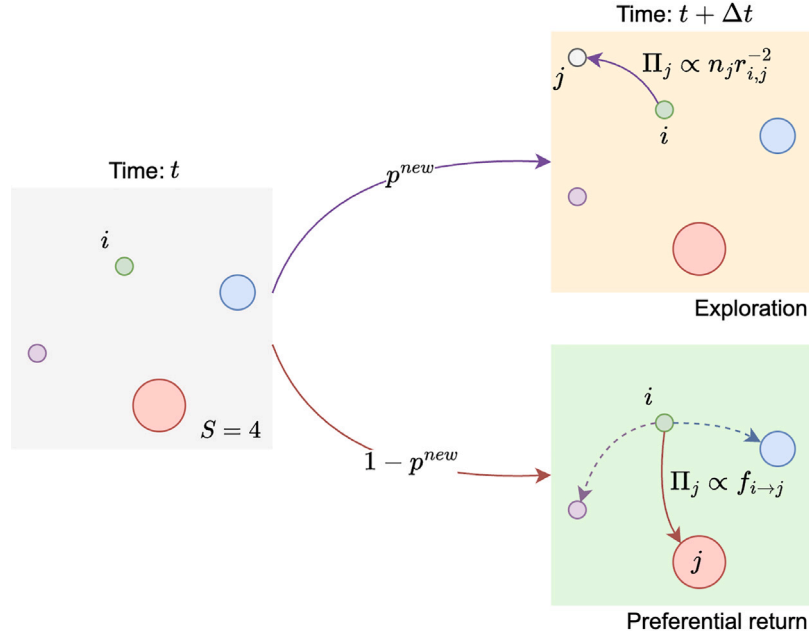
$$p_{t+\Delta t}^{new} = \rho S_t^{-\gamma} \quad (1)$$

where  $0 < \rho \leq 1$  and  $\gamma \geq 0$  are parameters that control the exploration tendency and  $S_t$  denotes the number of distinct location visited until time  $t$ . During this process, a new location is determined by sampling a moving distance  $\Delta r$  from the jump length distribution  $P(\Delta r)$ , with the moving direction chosen uniformly at random. After the move, the number of visited locations increases from  $S_t$  to  $S_t + 1$ . Besides exploring a new location, the individual could return to a visited location with complementary probability  $1 - p_{t+\Delta t}^{new}$ . In this case, the probability of moving to a location  $j$ , denoted as  $\Pi_j$ , is proportional to the number of previous visits to  $j$ , i.e.,  $\Pi_j \propto f_j$ , where  $f_j$  is the visitation frequency of  $j$ .

Later variants of EPR have modified the search for locations to replicate spatial patterns. One such variant is the d-EPR model, which addresses the insufficient reproduction of the evolution of the radius of gyration (Pappalardo et al., 2015). A population attractiveness factor is assigned to each location to model the tendency to visit popular locations (Fig. 2 upper panel). In the model, the probability  $\Pi_j$  of selecting location  $j$  during exploration depends on its travel distance and attractiveness:

$$\Pi_j \propto n_j r_{i,j}^{-2} \quad (2)$$

where  $r_{i,j}$  is the distance between the current location  $i$  and the new location  $j$ , and  $n_j$  denotes the attractiveness, quantified as the empirical visits by all individuals to location  $j$ . Furthermore, a subsequent study has identified that individuals can travel arbitrarily large distances during preferential return, as the location selection of EPR is based solely on empirical visit frequency (Zhao et al., 2021). To



**Fig. 2.** Mechanistic generative model DT-EPR. The individual at location  $i$  visited  $S = 4$  locations with a frequency proportional to the size of the location circle at time  $t$ . At time  $t + \Delta t$ , the individual chooses to either explore a new location with probability  $p^{new}$ , where the next location  $j$  will be chosen based on its population attractiveness  $n_j$  and the travel distance  $r_{i,j}$  (d-EPR mechanism; upper panel), or return to a previously visited location with complementary probability  $1 - p^{new}$ , where the location probability  $\Pi_j$  is proportional to the empirical visit frequency from  $i$  (IPT mechanism; lower panel).

Source: Figure adapted from Song et al. (2010a).

address this limitation, Zhao et al. (2021) proposed the IPT model that constrains the preferential return to be conditioned on the current location, i.e., imposing a first-order Markov process on location choices (Fig. 2 lower panel). The visit probability of location  $j$  is defined as proportional to the number of previous visits from the current location  $i$  to  $j$ :

$$\Pi_j \propto f_{i \rightarrow j} \quad (3)$$

In practice, a personalized Markov transition matrix containing  $f_{i \rightarrow j}$  for each location pair is initialized from empirical location visits and updated during the generation process.

We combine the exploration mechanism of d-EPR and the preferential return mechanism of IPT to introduce a new EPR-based model, which we refer to as DT-EPR (Fig. 2). Consequently, DT-EPR inherits the capacity to capture both population attractiveness and individual preferences in location choices. As a result, for each individual  $u^i$  in the user set  $\mathcal{U} = \{u^1, \dots, u^{|\mathcal{U}|}\}$ , DT-EPR generates a time-ordered trajectory  $T^i = (L_k)_{k=1}^{m_{u^i}}$  composed of  $m_{u^i}$  locations visited by  $u^i$ . A location  $L$  contains spatiotemporal information and is represented as a tuple of  $L = \langle l, p, t \rangle$ , where  $l$  is the location identifier,  $p = \langle x, y \rangle$  represents spatial coordinates in a reference system, e.g., latitude and longitude, and  $t$  is the time of visit. Thus, we construct the location set  $\mathcal{O}^i$  containing the known location identifiers for individual  $u^i$ , and the set  $\mathcal{O} = \{\mathcal{O}^1, \dots, \mathcal{O}^{|\mathcal{U}|}\}$  including all locations in the dataset.

### 3.2. Intervention design

We use empirically estimated behavioral parameters for the DT-EPR model, generating *observational* mobility traces. Since DT-EPR is a parametric simulator, i.e., parameters have real-world interpretations with explainable relationships with other model variables, we can introduce causal interventions to the data-generating process to simulate *interventional* mobility trajectories. Causal interventions can be interpreted as shifts in the observed mobility patterns, representing various scenarios, such as spatial shifts when certain locations become more or less attractive or temporal shifts in mobility behavior between seasons or years.

Considering the modeling mechanisms of the DT-EPR model, we perform interventions on the following parameters to simulate comprehensive behavior shift scenarios:

- The exploration tendency  $p^{new}$ , affecting whether or not to explore in the next time step (Eq. (1)). While the number of distinct locations collectively exhibits sublinear growth (Song et al., 2010a), notable inter-person variability in location choices implies variations in the exploration speed among individuals (Hong et al., 2023a). In EPR-like models,  $p^{new}$  is determined by parameters  $\rho$  and  $\gamma$ , independently sampled for each individual from empirical distributions. We introduce interventions on  $\rho$  and  $\gamma$  by altering their distributions, producing pseudo-populations with different exploration behaviors. Formally, let  $\mathbb{P}_S^{(\rho, \gamma, p^{new})}$  be the joint probability distribution over random variables  $\rho, \gamma, p^{new}$  induced by the SCM  $S$ . We intervene on each of these variables, e.g., for the case of  $\rho$ , we replace the generating mechanism of  $\rho$  with a new one  $\tilde{f}_\rho$ , i.e.,  $\mathbb{P}_S^{(\rho, \gamma, p^{new})} \rightarrow \mathbb{P}_S^{(\tilde{f}_\rho, \gamma, p^{new})}$ . The intervention  $do(\rho = \tilde{f}_\rho(\cdot))$  induces a new probability distribution  $\mathbb{P}_S^{(\rho, \gamma, p^{new})}$  with modified structural equations that we use for sampling interventional data. Additionally, we perform hard interventions on  $p^{new}$  by fixing its value to a constant. This intervention removes the time-dependent modeling of location exploration, falling back to the assumption in the Lévy flight model (Brockmann et al., 2006). Employing the same formalism as above, a hard intervention puts point mass on a specific outcome of a structural equation, i.e.,  $do(p^{new} = a)$ , where  $a$  is a constant value.
- The population attractiveness  $n$ , affecting location choices during exploration (Eq. (2)). We manipulate location attractiveness to simulate changes in the population's spatial preferences. To retain location visitation characteristics, we randomly shuffle empirical visit numbers for a group of locations. The strength of the intervention can be controlled by adjusting the group of locations, e.g., including more locations in the shuffling process introduces a more substantial intervention.
- The empirical individual preference  $f$ , affecting location choices during preferential return (Eq. (3)). As individual mobility behavior is dynamic and varies considerably over time, we introduce



interventions by manipulating the Markov transition matrix for each individual. This is achieved by shuffling the empirical visit numbers for a group of locations, which maintains the overall number of visits while altering the choice probabilities for each location. The strength of the intervention is controlled by selecting the location group to include in the shuffling process.

For each intervention, the DT-EPR model generates interventional mobility trajectories  $\tilde{T}^i = (L_k)_{k=1}^{m_{u^i}}$  for individual  $u^i \in \mathcal{U}$ , which share an identical data format as the observational mobility traces  $T^i$ .

### 3.3. Next location prediction networks

To assess the influence of causal interventions, i.e., the impact of changes in mobility behavior, we evaluate the predictive capability of a neural network trained on observational data but tested on interventional data. We choose next location prediction as the application task. Practically, consider a sub-sequence  $(L_k)_{k=m}^n \in T^i$  visited by individual  $u^i$  in a time window from time step  $m$  to  $n$ , the goal is to predict the location the same individual will visit in the next time step, i.e., the location identifier  $l_{n+1} \in \mathcal{O}$ . While conventional approaches relied on Markov models and matrix factorization methods, recent years have witnessed a growing adoption of neural networks (Luca et al., 2021). In the following, we present a typical pipeline for applying these networks to next location prediction, including generating feature embedding, designing the prediction network, and defining a loss function for parameter optimization.

An effective location prediction method starts by selecting and modeling the information in historical sequences. We include location identifiers, the time for location visits, and information regarding the individual who conducted the visit. To represent these features, we introduce embedding layers that utilize parameter matrices to map the original variables to real-valued embedding vectors. For each location  $L_k$ , vector representations of its location identifier  $l_k$  and time of arrival  $t_k$  are obtained as follows:

$$e_k^l = h^l(l_k; \mathbf{W}^l) \quad e_k^t = h^t(t_k; \mathbf{W}^t) \quad (4)$$

where  $e_k^l$  and  $e_k^t$  are the respective embedding vectors,  $h(\cdot; \cdot)$  denotes the embedding operation, and  $\mathbf{W}$  terms are the learned parameter matrices during training. To capture different periodicity, we separately embed the minute, the hour, and the day of the week from the visit time  $t_k$ . The overall embedding vector  $e_k^{all}$  is obtained by adding the location and temporal embedding vectors:  $e_k^{all} = e_k^l + e_k^t$ . Additionally, we represent the individual  $u^i$  that conducted the travel into a vector  $e^{u^i}$  through a user embedding layer, i.e.,  $e^{u^i} = h^u(u^i; \mathbf{W}^u)$ . Therefore, we obtain the overall embedding vector  $e_k^{all}$  that encodes spatiotemporal features at each time step and the user embedding vector  $e^{u^i}$  corresponding to each location sequence.

A location prediction network aims to learn the transition patterns between historical location visits in order to predict the next possible location. Without loss of generality, this process can be viewed as obtaining a (highly non-linear) mapping  $g$  with learnable parameters  $\mathbf{W}^g$  between the visit sequence  $(e_k^{all})_{k=m}^n$ , user information  $e^{u^i}$  and the ground-truth next location  $l_{n+1}$ , i.e.,  $l_{n+1} = g((e_k^{all})_{k=m}^n, e^{u^i}; \mathbf{W}^g)$ . When a new visit sequence is observed, the mapping  $g$  is used to infer the predicted next location  $\hat{l}_{n+1}$ . Various sequential modeling methods can be employed to approximate  $g$ , among which Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997) and Multi-Head Self-Attention (MHSA)-based (Vaswani et al., 2017) networks have gained considerable attention in the field. We implement LSTM and MHSA-based networks for next location prediction and present details of their architectures and realizations in Appendix A.

In practice, for efficient parameter optimization, the prediction network outputs visiting probabilities of directly predicting the next location:

$$P(\hat{l}_{n+1}) = \text{Softmax}(g((e_k^{all})_{k=m}^n, e^{u^i}; \mathbf{W}^g)) \quad (5)$$

where  $P(\hat{l}_{n+1}) \in [0, 1]^{|O|}$  contains visit probabilities of all locations at the next time step, and Softmax denotes the softmax operation, ensuring  $P(\hat{l}_{n+1})$  is a valid probability distribution. With access to the ground truth next location  $l_{n+1}$  during training, we can regard it as a multi-class classification problem, such that the network can be optimized using the multi-class cross-entropy loss  $\mathcal{L}$ :

$$\mathcal{L} = - \sum_{k=1}^{|O|} P(l_{n+1})^{(k)} \log(P(\hat{l}_{n+1})^{(k)}) \quad (6)$$

where  $P(\hat{l}_{n+1})^{(k)}$  represents the predicted probability of visiting the  $k$ th location (the  $k$ th entry in  $P(\hat{l}_{n+1})$ ) and  $P(l_{n+1})^{(k)}$  is the one-hot represented ground truth, i.e.,  $P(l_{n+1})^{(k)} = 1$  if and only if the  $k$ th location corresponds to the ground truth next location.

## 4. Dataset and experiment

### 4.1. Movement data and preprocessing

Individual mobility models rely on behavioral parameters calibrated from real-world data to depict mobility behavior. Smartphone-based travel surveys have emerged as a promising method for acquiring high-quality spatiotemporal travel behavior data (Harding et al., 2021) and are increasingly used for monitoring and studying travel patterns (Joseph et al., 2020; Rout et al., 2021; Mandal et al., 2023). To estimate these behavioral parameters, we leverage a smartphone-based travel survey conducted by the Swiss Federal Railways (SBB), known as the SBB Green Class (GC) E-Car pilot study, which aimed to assess the impact of a Mobility-as-a-Service (MaaS) offer on travel behavior (Martin et al., 2019). The pilot study yielded a large-scale longitudinal GNSS tracking dataset from 139 participants located in Switzerland, spanning from November 2016 to December 2017. As part of the data collection process, participants were asked to install a commercial application on their smartphones, which continuously recorded their whereabouts from GNSS signals. By analyzing motion measurements such as speed and acceleration, the application pre-processed the raw traces to identify two types of mobility events: *stay points* representing areas where users were stationary, and *stages* representing continuous movements using a single travel mode. Post-tracking analysis revealed that the median time between two consecutive GNSS recordings was 13.9 s, indicating a high temporal tracking quality of individual movements throughout the study period.

We perform a series of preprocessing steps on the GC dataset. First, to ensure high-quality tracking data for analysis, we include participants with extended tracking periods (observed for more than 300 days) and high temporal tracking coverage (whereabouts known for more than 60% of the time). Then, locations, the basic movement units in individual mobility models, are derived from stay point sequences. We identify *activities* from stay points with durations exceeding 25 min. Locations are formed by aggregating activity stay points spatially: we use the DBSCAN clustering algorithm with parameters  $\epsilon = 20$  and  $num\_samples = 1$  from the *Trackintel* library to generate dataset-level locations (Martin et al., 2023a). The longitudinal tracking of the GC study provides long-term observations of participants' daily activity location choices, allowing for precise estimation of behavioral parameters and a sufficiently large location choice set during mobility simulations. Since individuals' tendency to explore new locations diminishes significantly over time (Song et al., 2010a; Alessandretti et al., 2018), longitudinal observations reduce the occurrences of location exploration, thereby mitigating its impact on location prediction performance (Cuttone et al., 2018).

#### 4.2. Realization of mobility models

The preprocessed GC dataset is used to estimate the behavioral parameters, including the jump length distribution  $P(\Delta r)$  for vanilla EPR, as well as the waiting time distribution  $P(\Delta t)$  and distributions for exploration tendency ( $P(\rho)$  and  $P(\gamma)$ ) for EPR-like models. Specifically, we consider the log-normal distribution and the power law (including truncated power law) distribution as candidate distributions for determining  $P(\Delta r)$  and  $P(\Delta t)$  (Alessandretti et al., 2017). We use the functions provided by the *powerlaw* library (Alstott et al., 2014) and evaluate the goodness-of-fit using the Akaike information criterion (AIC) and Akaike weights (Zhao et al., 2015). Under AIC, both distributions are best fitted using a log-normal distribution of the probability density function  $P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right)$  with parameters  $\mu|_{\Delta r} = 7.72$ ,  $\sigma|_{\Delta r} = 2.38$ , and  $\mu|_{\Delta t} = 0.75$ ,  $\sigma|_{\Delta t} = 1.49$ , respectively. To estimate exploration tendencies  $P(\rho)$  and  $P(\gamma)$ , we calculate the location exploration speed and fit normal distributions across individuals. We obtain  $\mu|_{\rho} = 0.18$ ,  $\sigma|_{\rho} = 0.07$  and  $\mu|_{\gamma} = 0.64$ ,  $\sigma|_{\gamma} = 0.16$ , respectively. These values are consistent with the numbers presented in Song et al. (2010a).

We simulate traces for 800 individuals in the observational dataset and each interventional dataset. For each synthetic individual, we independently sample  $\rho$  and  $\gamma$  from the distributions  $P(\rho)$  and  $P(\gamma)$ , respectively, and randomly assign them to start at one of their top-5 visited locations empirically observed from GC. The generation process continues until individuals have visited 2000 locations, which approximately matches the tracking length of the GC study.

After obtaining observational location sequences  $T^i$  and interventional location sequences  $\tilde{T}^i$ , we compare their behaviors using mobility metrics:

- The real entropy (Song et al., 2010b) quantifies the regularity of location visit sequences by considering the order and frequency of location visits. It provides insights into the theoretical predictability of mobility (Hong et al., 2023b).
- The mobility motifs (Schneider et al., 2013) refer to recurring sub-structures that arise when representing daily location visits as graphs. These motifs represent common mobility patterns shared among individuals, and the proportion of motifs for an individual reflects the prevalence of these shared patterns (Cao et al., 2019).

We present additional metrics, including radius of gyration (González et al., 2008), number of transited location pairs (Zhao et al., 2021), and location visitation frequency (González et al., 2008), in Appendix B.

#### 4.3. Implementation of prediction networks

Next location prediction networks consider the past seven days as historical length; that is, the goal is to predict the following location an individual will be visiting based on their mobility history in the previous seven days. We adopt a common problem formulation, where prior mobility knowledge is assumed to be available for each individual. Hence, we partition each dataset (i.e., observational and interventional) into non-overlapping train, validation, and test sets based on time. The splitting ratio is set to 3:1:1, where locations occurring in the first 60% of days for each individual are assigned to the train set, and the last 20% are assigned to the test set. We utilize the observational training set to optimize network parameters, and the corresponding validation set to monitor the network loss and performance during training. We conduct a grid search to find the optimal hyper-parameters that minimize the network loss on the same validation set. The considered hyper-parameters, the search ranges, and the final selected values can be found in Appendix A. Finally, we evaluate network performances using the held-out test sets from observational and interventional traces.

We use the following performance metrics to evaluate the network prediction:

- Accuracy@k (Acc@k) measures the correctness by checking if the ground truth location is among the top-k most likely visited locations predicted by the network (in  $P(\hat{l}_{n+1})$  from Eq. (5)). We report Acc@1, Acc@5, and Acc@10.
- F1 score (F1) calculates the harmonic mean of precision and recall, considering the uneven visitation preferences to locations in daily routines. We utilize an F1 score weighted by the visit frequency to emphasize the performance in predicting important locations.
- Mean reciprocal rank (MRR) assesses the relevance of the recommended locations and is defined as the average reciprocal rank at which the ground truth location is identified by the network (in  $P(\hat{l}_{n+1})$  from Eq. (5)).

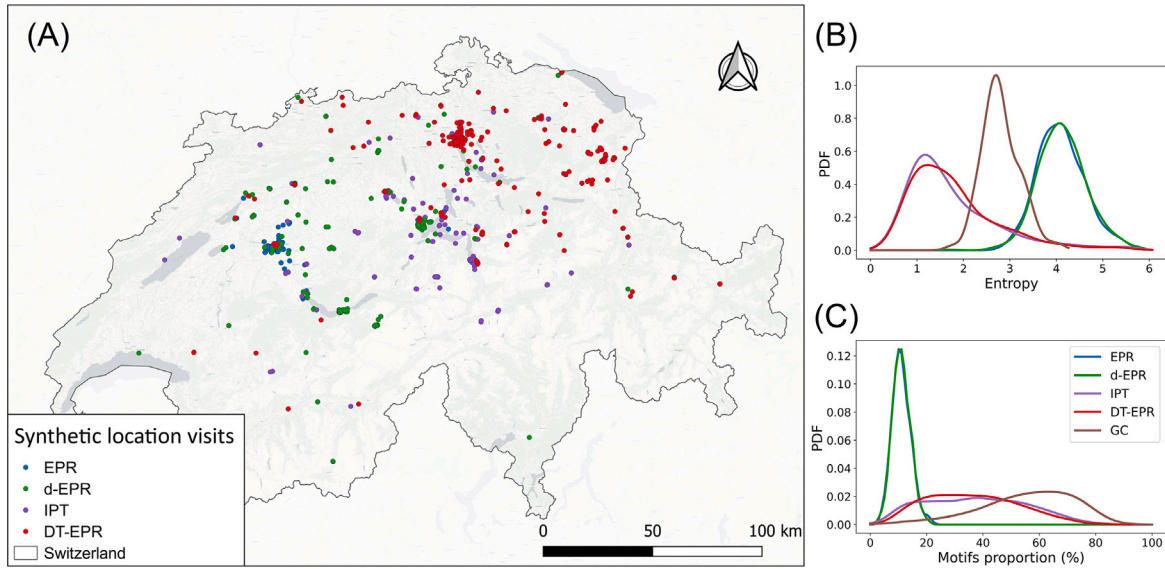
### 5. Results

#### 5.1. Mobility simulation and intervention

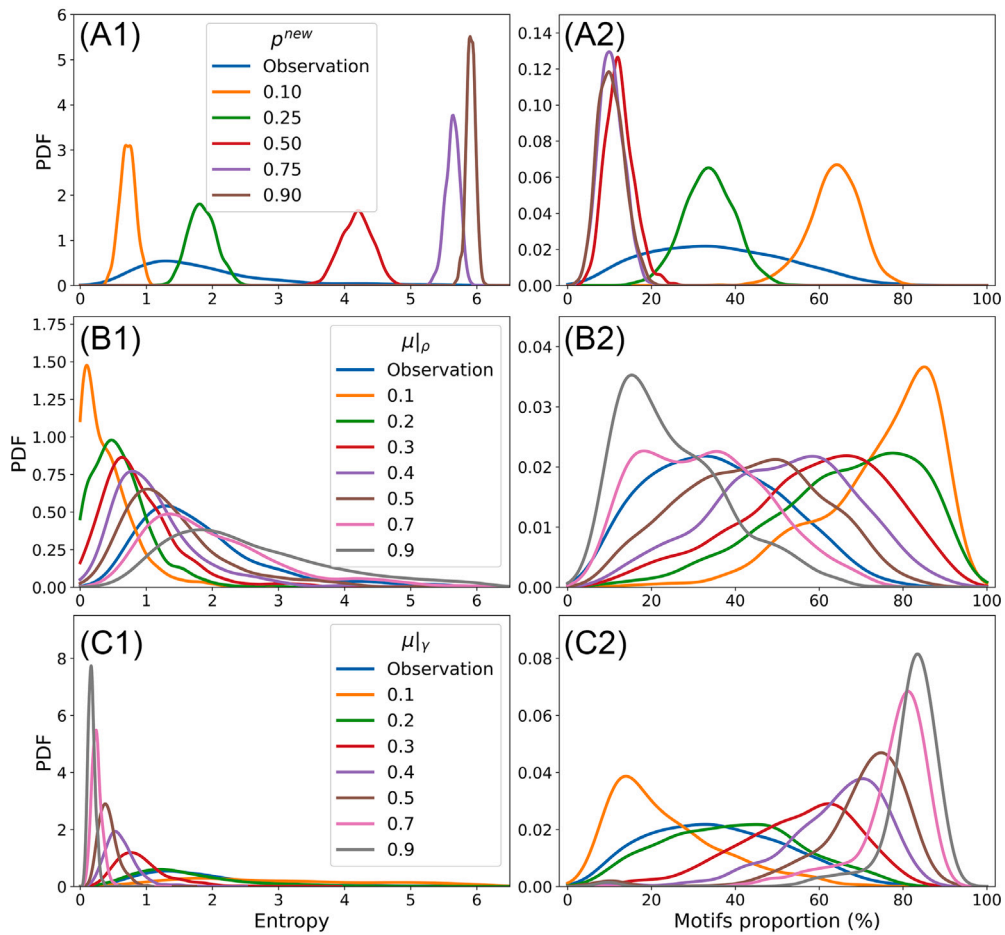
Leveraging individual models and empirically estimating their behavior parameters, we obtain location traces for synthetic individuals. The generation result is illustrated in Fig. 3, where we map a selected individual's locations and plot the entropy and motifs proportion measures for the entire dataset. The spatial distribution of location visits (Fig. 3A) clearly shows the patterns exhibited by EPR-like models. The vanilla EPR model assigns visit probabilities based solely on the geographical distance, thus generating location visits clustered around a central area. While the traces obtained from a d-EPR model are more concentrated in large cities, the locations generated by an IPT model are more dispersed in space, reflecting the empirical knowledge of that specific individual. Finally, the DT-EPR model combines the advantages of both d-EPR and IPT models, producing traces that prioritize locations within urban areas (population attractiveness) and occasionally occur in distant sites (personal preferences). Furthermore, we quantitatively assess the regularities of the synthetic sequences at the dataset level (Fig. 3B and C). EPR and d-EPR exhibit the highest entropy and lowest motifs proportion among all mobility models, indicating that their location visit patterns are irregular and hard to predict. Although the d-EPR model introduces population attractiveness in the exploration phase, it does not alter the process of visiting known locations, which is the primary mechanism underlying mobility's regularity. On the contrary, DT-EPR incorporates the first-order Markov dependence on location visits, producing the most regular location visitation patterns, as evidenced by the lowest entropy and highest motif proportion distributions among all models. Another intriguing observation arises when comparing the distributions of the generated traces with those from the real data. The result from DT-EPR displays lower entropy, signifying higher regularity in location visits, but generally has a lower motifs proportion compared to the real traces (labeled as GC in Fig. 3B and C).

Next, we introduce causal interventions to the data-generation process and assess their direction and impact using mobility metrics. Fig. 4 displays the distributions of mobility entropy and motifs proportion for both observational and interventional location sequences from DT-EPR. The latter are generated by implementing interventions on individuals' exploration tendencies. Additional metrics for describing the impact of these interventions can be found in Appendix B. The interventions on population-level attractiveness and individual-level preference preserve the general visitation characteristics; hence, high-level mobility metrics cannot accurately reflect their impact.

The interventions can effectively and directionally change the underlying mobility pattern, as demonstrated by the shifts in the mobility metric distributions, shown in Fig. 4. For example, with an increase in the exploration tendency  $p^{new}$ , individuals are encouraged to visit new locations, resulting in an increasingly higher number of transited location pairs (Appendix B) and leading to mobility sequences with higher entropy and lower motifs proportion (Fig. 4A). Moreover, the impact



**Fig. 3.** Synthetic location visits generated by EPR-like individual mobility models. (A) Spatial distribution of location visits for an exemplary user. (B) Mobility entropy and (C) motifs proportion distributions of the generated population. Map data © OpenStreetMap contributors, © CARTO.



**Fig. 4.** The entropy and motifs proportion distributions of observational and interventional location sequences. We show the metric distributions for (A) hard interventions on  $p^{new}$ , (B) interventions on  $\rho$  by shifting  $\mu|_{\rho}$  of  $P(\rho)$ , and (C) interventions on  $\gamma$  by shifting  $\mu|_{\gamma}$  of  $P(\gamma)$ .

on the generated location sequences can be compared among the different interventions. While intervening on the exploration tendency

$p^{new}$  significantly alters the mobility patterns (Fig. 4A), changes induced by exploration parameters  $\rho$  and  $\gamma$  are more nuanced and provide more



**Table 1**

Next location prediction performances for observational location sequences from EPR-like models. The mean and standard deviation across five runs with different random parameter initializations are reported.

Networks	Datasets	Acc@1 (%)	Acc@5 (%)	Acc@10 (%)	F1 (%)	MRR (%)
LSTM	EPR	1.1 $\pm$ 0.03	2.8 $\pm$ 0.04	3.9 $\pm$ 0.02	0.3 $\pm$ 0.01	2.2 $\pm$ 0.03
	d-EPR	1.2 $\pm$ 0.04	3.2 $\pm$ 0.1	4.6 $\pm$ 0.1	0.4 $\pm$ 0.04	2.5 $\pm$ 0.05
	IPT	17.3 $\pm$ 0.3	25.8 $\pm$ 0.3	28.8 $\pm$ 0.2	11.2 $\pm$ 0.1	21.4 $\pm$ 0.2
	DT-EPR re-generated	25.1 $\pm$ 0.2	35.1 $\pm$ 0.2	38.5 $\pm$ 0.2	17.8 $\pm$ 0.2	30.0 $\pm$ 0.2
	DT-EPR trained	55.0 $\pm$ 0.1	62.9 $\pm$ 0.1	64.6 $\pm$ 0.05	46.2 $\pm$ 0.2	58.6 $\pm$ 0.1
MHSA	EPR	0.8 $\pm$ 0.01	2.4 $\pm$ 0.02	3.7 $\pm$ 0.02	0.3 $\pm$ 0.01	2.0 $\pm$ 0.01
	d-EPR	0.9 $\pm$ 0.01	2.8 $\pm$ 0.02	4.2 $\pm$ 0.04	0.4 $\pm$ 0.01	2.2 $\pm$ 0.01
	IPT	15.0 $\pm$ 0.2	23.7 $\pm$ 0.4	27.3 $\pm$ 0.4	10.6 $\pm$ 0.1	19.3 $\pm$ 0.2
	DT-EPR re-generated	24.4 $\pm$ 0.1	33.8 $\pm$ 0.2	37.5 $\pm$ 0.2	18.7 $\pm$ 0.04	29.0 $\pm$ 0.1
	DT-EPR trained	55.8 $\pm$ 0.02	62.7 $\pm$ 0.04	64.4 $\pm$ 0.03	47.3 $\pm$ 0.04	59.1 $\pm$ 0.02

fine-grained control (Fig. 4B and C). In summary, mobility metrics facilitate the description and comparison between observational and interventional location sequences, allowing us to assess the direction and strength of the interventions on mobility behavior.

### 5.2. Next location prediction

Thanks to the mobility simulation models and the causal intervention process, we can assess the performance of predictive networks when confronted with distribution shifts. We begin by comparing the performance of LSTM and MHSA on mobility sequences synthesized by different variants of the EPR model (Table 1), thereby uncovering the influence of generative mechanisms. These networks are trained with location sequences from DT-EPR and evaluated on datasets obtained using the same observational parameters but different ablations of its components. We also assess their performances on a new DT-EPR population with different initializations (e.g., for each individual a newly sampled start location,  $p$ , and  $\gamma$ ; see §4.2 for details of the mobility model realization). This variant is denoted as *DT-EPR re-generated* in Table 1, with the network performance differences reflecting the effect of induced randomness in the mobility model. When evaluated on the training population (*DT-EPR trained* in Table 1), MHSA demonstrates a stronger capability to infer the next location compared to LSTM, as indicated by the higher performance indicators Acc@1, F1, and MRR. However, LSTM outperforms MHSA when confronted with all other populations, i.e., when predicting for location sequences on which the network was not trained, demonstrating superior generalization ability. In addition, both networks exhibit the worst performance when provided with sequences obtained from the vanilla EPR model, indicating the absence of clear transition patterns in these sequences – a property overlooked by the EPR mechanism. As a further enhancement to the model, IPT enables the generation of location sequences with transition patterns that a predictive network can utilize to infer the next location. Moreover, incorporating population preferences during exploration imposes essential structural patterns in location visit sequences, as evidenced by the increase in prediction performance between EPR and d-EPR, as well as IPT and DT-EPR. These structural patterns are not apparent in the entropy and motifs proportion distributions (Fig. 3), yet they contribute to increasing the accuracy of next location prediction (Table 1).

### 5.3. Robustness of location prediction networks

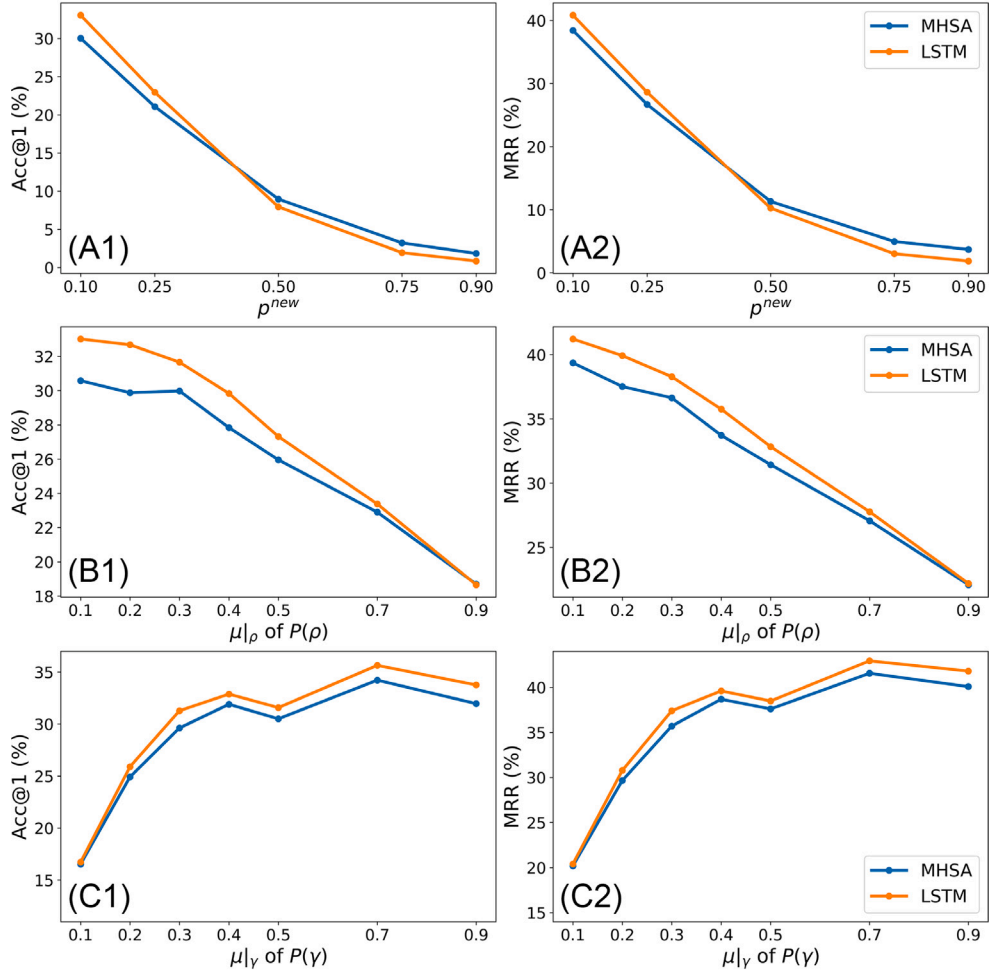
We now evaluate networks' performance using interventional mobility sequences, which reveal their robustness in OoD scenarios, i.e., when the training and testing data are not generated from the same underlying distribution. Fig. 5 displays the variations in Acc@1 and MRR scores of the LSTM and MHSA networks for interventions on exploration tendency, and Fig. 6 depicts the performance variation plot

for interventions on population attractiveness and individual preference. The complete performance results for all conducted interventions are presented in Appendix C. The prediction networks remain consistent with those described in the previous section, i.e., they are trained using observational location sequences from DT-EPR. Consequently, the performance metrics in OoD scenarios can be cross-compared with those presented in Table 1. Although similar performance trends are observed for both networks, LSTM consistently outperforms MHSA in OoD settings, confirming the observation from the previous results.

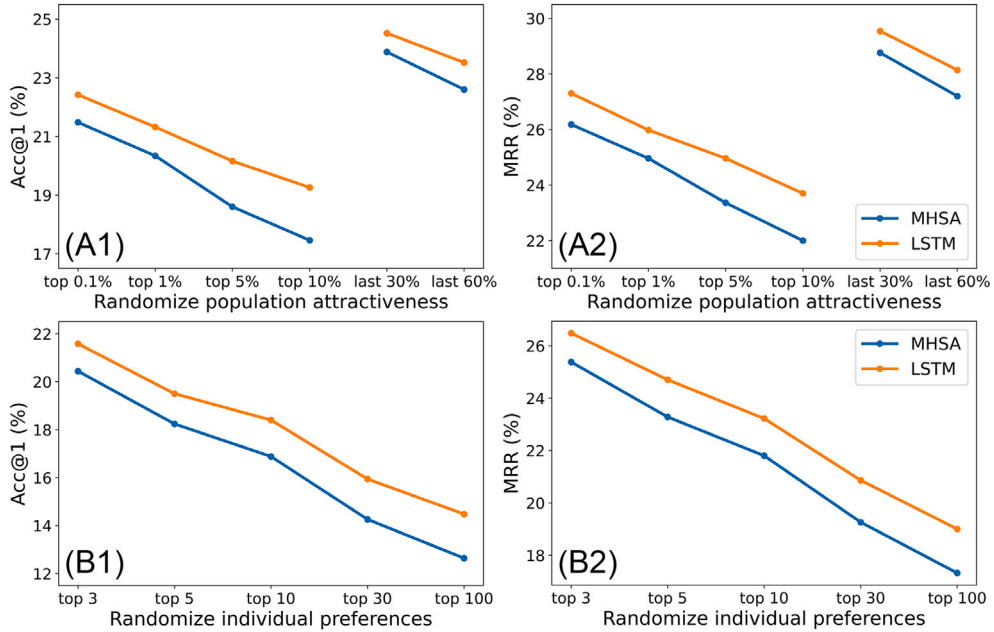
The performance variations for exploration interventions (Fig. 5) generally align with their strengths and directions, as measured using mobility metrics shown in Fig. 4. We also observe non-linear relations between intervention strength and prediction performance. In particular, the hard interventions on  $p^{new}$  significantly influence the prediction capability. Setting  $p^{new} > 0.5$  results in performance indicators of Acc@1 < 10% and MRR < 10%, suggesting that the learned location transition patterns cannot be adequately utilized. Comparatively, interventions on  $\gamma$  and  $\rho$  indirectly affect  $p^{new}$ , which retains the diminishing exploration speed over time. As a result, influences on next location prediction are milder, e.g., the Acc@1 still achieves  $\sim 18\%$  with the strongest implemented interventions ( $\mu|_{\rho} = 0.9$  and  $\mu|_{\gamma} = 0.1$ ). Moreover, we observe the prediction performances are relatively stable for  $\mu|_{\rho} \in [0.1, 0.3]$  and  $\mu|_{\gamma} \in [0.5, 0.9]$ , even though the location sequences continue to exhibit lower mobility entropy and higher motifs proportion (Fig. 4) due to decreasing exploration tendency. This saturation suggests that even if individuals explore new locations at a lower rate, many location visit patterns are inherently stochastic and complex, making them challenging to capture by a trained network.

Interventions on population attractiveness and individual preference reveal how altering visit frequencies for specific locations affects the prediction performances, as shown in Fig. 6. In each intervention, we consider a set of locations and randomly shuffle their empirical visitation numbers within the group; for instance, in the shuffling process, “top 1%” includes the most frequently visited 1% of locations across the population (Fig. 6A), and “top 3” considers the three most visited locations for each individual (Fig. 6B). Both types of interventions substantially impact the ability to correctly predict the next location, with altering the number of visits separately for each individual showing a stronger influence, as evidenced by the higher drop in performance indicators. When compared with the observational datasets (Table 1), even changes in preference for a few most critical locations (e.g., “top 0.1%” for location attractiveness or “top 3” for individual preferences) result in a significant prediction capability decrease. On the contrary, intervening on a large portion of locations that are not frequently visited (i.e., “last 30%” and “last 60%” for location attractiveness) has minimal impact on the prediction performances. These results emphasize the indispensable role of essential locations in shaping daily mobility and reveal their relation with the generalization ability of next location prediction networks.





**Fig. 5.** Next location prediction performances for interventions on individuals' exploration tendency. We show the variations in Acc@1 and MRR for (A) hard interventions on  $p^{new}$ , (B) interventions on  $\rho$  by shifting  $\mu_\rho$  of  $P(\rho)$ , and (C) interventions on  $\gamma$  by shifting  $\mu_\gamma$  of  $P(\gamma)$ .



**Fig. 6.** Next location prediction performances for interventions on population attractiveness and individual preference. We show the variations in Acc@1 and MRR for (A) randomizing empirical location visits of the dataset, and (B) randomizing empirical location visits for each individual.

## 6. Discussion

This study has assessed the robustness of neural networks on the next location prediction task. We implement multiple individual mobility simulation models, on which we conduct causal interventions to synthesize mobility traces when structural changes occur. These interventional mobility traces are then used as input for trained networks and evaluated for location prediction. The examination of mobility simulators revealed that the vanilla EPR fails to generate sequences with realistic visitation order, while IPT tends to produce over-regular sequences by explicitly incorporating the first-order Markov dependence (Fig. 3). These findings highlight the need for further developments in individual mobility models, essential for visit order-sensitive applications, such as mobility prediction (Ma and Zhang, 2022), autonomous vehicles scheduling (Li and Liao, 2020), and smart charging optimization (Cai et al., 2022). Moreover, the comparison between the two sequence modeling networks, LSTM and MHSA-based, reveals their generalization characteristics. While the MHSA-based network performs better when predicting locations for the same population, LSTM demonstrates higher generalization capability when dealing with sequences sampled OoD.

Performing causal interventions on the mobility simulators enables us to control the strength and direction of mobility behavior. These interventions quantitatively reveal the behavioral impact on location predictors and have practical implications for downstream applications. We establish a connection between exploration tendency, mobility behavior, and prediction performance using mobility metrics, providing a benchmark for comparing existing or newly developed mobility prediction networks, potentially on other datasets. This enables future studies to evaluate their population's mobility behavior and determine whether their networks perform better than those reported in this study. Additionally, this connection aids in selecting target populations for predictive networks (Solomon et al., 2021) and preemptively assessing the performance of demographic groups (Baumann et al., 2018). An improved understanding of prediction networks reveals the priorities in enhancing their behavioral robustness, which greatly benefits real-world applications such as enhanced location-based mobility services (e.g., on-demand mobility (Kieu et al., 2020)) and traffic management (Sun and Kim, 2021).

The findings from location interventions indicate that shifts in spatiotemporal preferences significantly affect prediction network performance. However, these preference shifts cannot be differentiated using high-level mobility metrics, as they alter location choices while maintaining the empirical visitation distribution. This observation highlights the need to develop change detection methods for identifying precise mobility behavior change points or periods (Hong et al., 2021; Roy et al., 2023). Moreover, our results reveal that preference changes in the essential locations have a much more significant effect on the prediction network than changes in the less visited ones. Recent studies on intra-person variability of travel behavior suggest that individuals constantly update their important locations over time (Järv et al., 2014; Alessandretti et al., 2018). This implies that real-world deployment of mobility prediction networks requires integrating online learning, where networks are continuously updated as new data arrives (van Lint, 2008; Jiang et al., 2018). In the online learning framework, this study can help identify the point at which the network's performance becomes insufficient and subsequently update network parameters to adapt to evolving mobility behavior.

## 7. Conclusion

Unraveling the role and impact of multifaceted mobility behavior on prediction outcomes is imperative to the real-world application of mobility prediction systems. Here, we present a framework to examine how behavioral factors influence mobility prediction networks through

causal interventions. Using mechanistic mobility models with parameters that capture mobility behaviors, we perform causal interventions on these parameters to generate mobility traces that mirror real-world behavior variations. Quantitative evaluation using mobility metrics demonstrates our capability to effectively and deliberately modify behaviors. Subsequently, we evaluate these interventional traces with well-trained networks for the next location prediction task, and the resulting performance variations indicate the robustness of networks confronting domain shifts. Our results reveal vital behavior factors affecting prediction performance, including sequential location transition patterns, the tendency to explore new locations, and location preferences at both population and individual levels. These findings demonstrate the framework's effectiveness and pave the way for various downstream applications, including cross-comparisons among prediction networks and performance monitoring in dynamically evolving spatiotemporal scenarios.

As one of the pioneering studies to explore the causal impact of mobility behavior on individual mobility prediction, this research sheds light on several directions for future work. The proposed framework can be applied to any mechanistic generation model abstracted as an SCM. Our interventions focus on the EPR model and its extensions, and our results are fundamentally limited by their capability to simulate activity location choices. Future research should investigate other mobility simulation models, such as location attractiveness (Yan et al., 2017) and container (Alessandretti et al., 2020), to examine the impact of alternative behavioral mechanisms on mobility prediction. In addition to factors that describe mobility, quantifying the causal effects of contexts (e.g., urban functions, land use, and weather situations) on mobility behavior is a promising direction. This endeavor would require a mechanistic model that explicitly formulates the relationship between contextual factors and mobility behavior, which could be inferred using causal discovery methods (Runge et al., 2023). Last, counterfactuals offer greater causal reasoning capabilities than interventions, as outlined in the three-layer causal hierarchy of association (layer 1), intervention (layer 2), and counterfactual (layer 3) by Pearl and Mackenzie (2018). Exploring counterfactual explanations for prediction networks provides actionable suggestions that can lead to desirable alternative outcomes (Xin et al., 2022). This aspect is crucial for the real-world deployment of mobility prediction systems, even though its full potential remains to be realized. In conclusion, we expect this study to stimulate innovative approaches that leverage causal inference to improve the interpretability and robustness of neural networks when applied to mobility analysis and prediction.

## CRedit authorship contribution statement

**Ye Hong:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Yanan Xin:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Simon Dirmeier:** Writing – review & editing, Software, Methodology. **Fernando Perez-Cruz:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Martin Raubal:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

## 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.

## Acknowledgments

This work is supported by the Hasler Foundation under the project titled “Interpretable and Robust Machine Learning for Mobility Analysis” (grant number 21041).

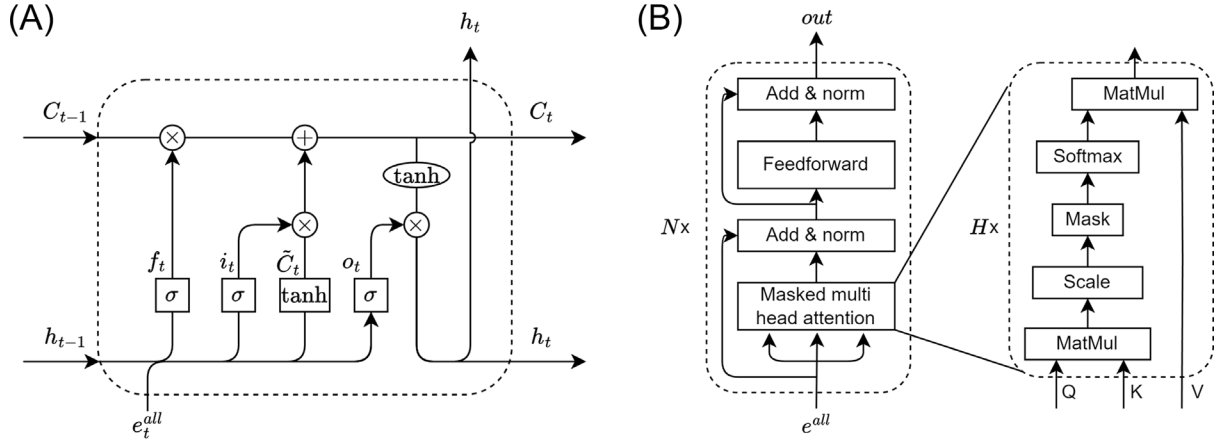


Fig. A.1. Network structure of (A) the LSTM network and (B) the MHSA-based network.

## Appendix A. Prediction network and implementation

In this section, we introduce the fundamental components of LSTM and MHSA-based networks, detailing their implementation and training processes. For specific modifications to enhance performance in the next location prediction task, we refer readers to Solomon et al. (2021) and Hong et al. (2023b), respectively.

**LSTM network.** Initially proposed by Hochreiter and Schmidhuber (1997), LSTM improves over the vanilla recurrent neural network (RNN) for learning temporal dependencies in sequence data by addressing the vanishing gradient issue through its architecture design. The network processes input sequentially and, at each step  $t$ , utilizes three “gates” – forget gate  $f_t$ , input gate  $i_t$ , and output gate  $o_t$  – to learn the dependencies between the current input  $e_t^{all}$  and previous information, stored in a cell state  $C_{t-1}$  and a hidden state  $h_{t-1}$  (Fig. A.1A). This gate-updating process can be formulated as follows:

$$f_t = \sigma((e_t^{all} \oplus h_{t-1}) \cdot W^f) \quad (7)$$

$$i_t = \sigma((e_t^{all} \oplus h_{t-1}) \cdot W^i) \quad (8)$$

$$o_t = \sigma((e_t^{all} \oplus h_{t-1}) \cdot W^o) \quad (9)$$

where  $\oplus$  denote the concatenation operation and  $\sigma$  is the sigmoid function. Then, the current cell state  $C_t$  and hidden state  $h_t$  can be updated as follows:

$$\tilde{C}_t = \tanh((e_t^{all} \oplus h_{t-1}) \cdot W^C) \quad (10)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (11)$$

$$h_t = o_t \odot \tanh(C_t) \quad (12)$$

where  $\odot$  denotes the Hadamard product. Finally,  $C_t$  and  $h_t$  are carried forward to the next step.  $h_t$ , encapsulating all information up to the current step, can also be used to infer predictions.

**MHSA-based network.** First introduced in Vaswani et al. (2017), the MHSA-based network extends the attention mechanism and has become widely adopted in sequential modeling tasks for its efficiency and ability to capture complex dependencies across the entire sequence. In contrast to LSTM, the MHSA-based network processes the input all at once: an embedding matrix  $e^{all}$  is constructed by stacking embedding vectors in sequence order (Fig. A.1B). This input is then passed through a stack of  $N$  identical layers, each containing a masked multi-head attention block and a fully connected feedforward network with two

Table A.1

Hyper-parameter search for next location prediction networks.

Network	Hyper-parameter	Search range	Optimal value
LSTM	Embedding dim.	{32, 64, 128}	64
	Hidden dim.	{64, 128, 256, 512}	128
MHSA-based	#Heads $H$	{2, 4, 8}	8
	#Layers $N$	{2, 4, 6}	4
	Embedding dim.	{32, 64, 128}	64
	Feed-forward dim.	{64, 128, 256, 512}	256

linear layers and a ReLU activation function. Residual connections (He et al., 2016) and layer normalizations (Ba et al., 2016) are applied to each layer. The multi-head attention block uses scaled dot-product attention to compute output values based on queries and sets of key-value pairs, efficiently implemented by packing into matrices  $Q$ ,  $K$ , and  $V$  (Vaswani et al., 2017):

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{\dim}}\right) \cdot V \quad (13)$$

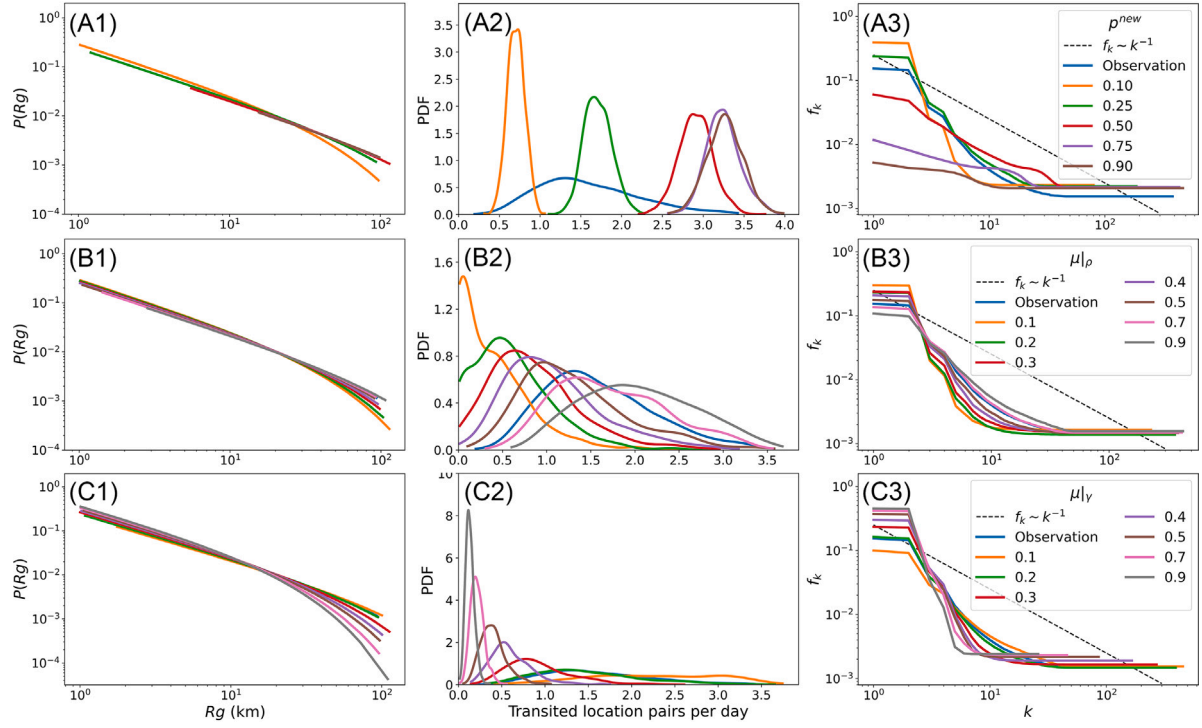
where  $\dim$  is the size of the query vector. Multi-head attention is constructed by scaling the matrices  $Q$ ,  $K$ , and  $V$ , and concatenating the results of  $H$  attention functions:

$$\text{MultiHead}(Q, K, V) = (\text{head}_1 \oplus \dots \oplus \text{head}_H) \cdot W^O \quad (14)$$

$$\text{where } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (15)$$

In each multi-head attention block, the key, value, and query matrices are identical and correspond to the output of the previous block. In addition, forward-masking operations are included to prevent attention operations from accessing information from future time steps.

**Implementation and training details.** During training, we minimize Eq. (6) with the Adam optimizer (Kingma and Ba, 2015) on batches of training data samples. We use a batch size 256, and zeros are appended to the end of each input sequence until its length matches the longest sequence, ensuring consistent length within each batch. The initial learning rate is set to  $1e^{-3}$ , and an L2 penalty of  $1e^{-6}$  is applied to the network loss. We adopt an early stopping strategy, which drops the learning rate by 0.1 if the validation loss does not decrease for 3 consecutive epochs. This early stopping process is repeated 3 times.



**Fig. B.2.** The radius of gyration ( $R_g$ ), number of transited location pairs, and location visitation frequency distributions of observational and interventional location sequences. We show the metric distributions for (A) hard interventions on  $p^{new}$ , (B) interventions on  $\rho$  by shifting  $\mu|_{\rho}$  of  $P(\rho)$ , and (C) interventions on  $\gamma$  by shifting  $\mu|_{\gamma}$  of  $P(\gamma)$ .

The tuned hyper-parameters and their search ranges are shown in Table A.1. We determine the optimal set of values as the one that minimizes the network loss on the validation set of the observational traces from DT-EPR.

#### Appendix B. Additional metrics for describing mobility behavior

Apart from the mobility metrics elucidated in the main text, we introduce additional metrics to characterize mobility behavior within both observational and interventional location visitation sequences:

- The radius of gyration (González et al., 2008) measures the characteristic distance traveled by an individual and is commonly used to reflect their movement intensity in the spatial dimension. Its dataset distribution reveals the heterogeneity in user movement (Xu et al., 2018a).
- The number of transited location pairs (Zhao et al., 2021) measures the number of links in an individual mobility graph, where locations are represented as nodes and travels between locations are represented as edges.
- The location visitation frequency distribution (González et al., 2008) is known to adhere to Zipf's law, where individuals tend to spend the majority of their time at a small number of frequently visited locations. The exponent of this power law, which characterizes the shape of the distribution, reveals the frequency of important location visits and can serve as an indicator of mobility regularity.

Fig. B.2 illustrates alterations in mobility patterns resulting from interventions on the exploration behavior, assessed using the previously mentioned metrics. Notably, these interventions induced changes in the

population's mobility intensity, quantified by the radius of gyration. Although these interventions had a marginal impact on the prevalence of individuals traveling shorter distances, they substantially reduced the likelihood of encountering individuals with extensive travel patterns as exploration declined (lower  $p^{new}$  and  $\mu|_{\rho}$ , and higher  $\mu|_{\gamma}$ ). As expected, a reduction in exploration inclination led to fewer transited location pairs per day (Fig. B.2 middle panel) and a significant decrease in the visitation frequency of the most frequently visited locations (Fig. B.2 right panel).

#### Appendix C. Network performances on interventional location sequences

The complete performance evaluation results for LSTM and MHSA-based networks on all simulated interventional location sequences are presented in Table C.2 and Table C.3, respectively. We report the mean and standard deviation across five network optimization runs with different random parameter initializations. Additionally, we note that generative mechanisms in the mobility model introduce variations in synthesized location sequences (e.g., randomness from mobility model initializations; see §5.2 and Table 1), affecting mobility predictability and prediction performance. Therefore, we suggest focusing on the overall variation trend for interventions rather than exact performance numbers.

#### Data availability

Raw data for the GC dataset are not publicly available due to confidentiality agreements with the participants under GDPR. The source code is available at <https://github.com/irmlma>.



**Table C.2**

Performance of the LSTM model in predicting the next location on interventional location sequences. The mean and standard deviation across five runs with different random parameter initializations are reported.

Intervention	Strength	Acc@1 (%)	Acc@5 (%)	Acc@10 (%)	F1 (%)	MRR (%)
$p^{new}$	0.1	33.0 $\pm$ 0.6	49.8 $\pm$ 0.6	55.3 $\pm$ 0.6	28.9 $\pm$ 0.5	40.8 $\pm$ 0.6
	0.25	23.0 $\pm$ 0.3	34.8 $\pm$ 0.1	38.7 $\pm$ 0.1	16.2 $\pm$ 0.1	28.6 $\pm$ 0.3
	0.5	7.9 $\pm$ 0.1	12.4 $\pm$ 0.03	14.4 $\pm$ 0.05	3.6 $\pm$ 0.1	10.3 $\pm$ 0.04
	0.75	1.9 $\pm$ 0.1	3.7 $\pm$ 0.1	4.9 $\pm$ 0.1	0.6 $\pm$ 0.04	3.0 $\pm$ 0.1
	0.9	0.9 $\pm$ 0.04	2.3 $\pm$ 0.1	3.5 $\pm$ 0.1	0.3 $\pm$ 0.03	1.9 $\pm$ 0.1
$\mu _p$	0.1	33.0 $\pm$ 0.7	50.7 $\pm$ 0.7	56.5 $\pm$ 0.4	29.2 $\pm$ 0.6	41.2 $\pm$ 0.7
	0.2	32.7 $\pm$ 0.4	48.0 $\pm$ 0.3	53.0 $\pm$ 0.6	28.3 $\pm$ 0.3	39.9 $\pm$ 0.2
	0.3	31.7 $\pm$ 0.5	45.8 $\pm$ 0.4	50.4 $\pm$ 0.3	25.9 $\pm$ 0.4	38.3 $\pm$ 0.3
	0.4	29.8 $\pm$ 0.4	42.4 $\pm$ 0.4	47.1 $\pm$ 0.3	23.5 $\pm$ 0.4	35.8 $\pm$ 0.3
	0.5	27.3 $\pm$ 0.3	38.9 $\pm$ 0.3	42.9 $\pm$ 0.2	20.5 $\pm$ 0.3	32.8 $\pm$ 0.2
	0.7	23.4 $\pm$ 0.1	32.4 $\pm$ 0.1	35.7 $\pm$ 0.1	16.2 $\pm$ 0.1	27.8 $\pm$ 0.1
	0.9	18.6 $\pm$ 0.05	25.8 $\pm$ 0.1	28.3 $\pm$ 0.1	11.5 $\pm$ 0.2	22.2 $\pm$ 0.1
$\mu _r$	0.1	16.7 $\pm$ 0.1	24.2 $\pm$ 0.1	26.9 $\pm$ 0.1	10.2 $\pm$ 0.1	20.4 $\pm$ 0.1
	0.2	25.9 $\pm$ 0.3	36.0 $\pm$ 0.2	39.4 $\pm$ 0.3	18.6 $\pm$ 0.2	30.8 $\pm$ 0.2
	0.3	31.3 $\pm$ 0.3	44.2 $\pm$ 0.2	48.2 $\pm$ 0.6	24.8 $\pm$ 0.1	37.4 $\pm$ 0.2
	0.4	32.9 $\pm$ 0.4	47.1 $\pm$ 0.2	51.7 $\pm$ 0.2	27.1 $\pm$ 0.2	39.6 $\pm$ 0.3
	0.5	31.6 $\pm$ 0.4	46.3 $\pm$ 0.4	51.2 $\pm$ 0.3	26.9 $\pm$ 0.5	38.5 $\pm$ 0.3
	0.7	35.6 $\pm$ 0.7	51.4 $\pm$ 0.4	56.7 $\pm$ 0.7	31.6 $\pm$ 0.7	42.9 $\pm$ 0.6
	0.9	33.8 $\pm$ 0.6	51.2 $\pm$ 0.5	57.0 $\pm$ 0.9	30.0 $\pm$ 0.5	41.8 $\pm$ 0.4
Attractiveness	top 0.1%	22.4 $\pm$ 0.2	32.6 $\pm$ 0.2	35.8 $\pm$ 0.2	15.5 $\pm$ 0.3	27.3 $\pm$ 0.1
	top 1%	21.3 $\pm$ 0.2	30.9 $\pm$ 0.1	34.4 $\pm$ 0.2	14.6 $\pm$ 0.1	26.0 $\pm$ 0.1
	top 5%	20.2 $\pm$ 0.2	30.1 $\pm$ 0.3	33.4 $\pm$ 0.1	13.5 $\pm$ 0.2	25.0 $\pm$ 0.2
	top 10%	19.3 $\pm$ 0.3	28.4 $\pm$ 0.2	31.5 $\pm$ 0.2	12.6 $\pm$ 0.2	23.7 $\pm$ 0.2
	last 30%	24.5 $\pm$ 0.3	35.0 $\pm$ 0.1	38.5 $\pm$ 0.1	17.5 $\pm$ 0.2	29.5 $\pm$ 0.2
	last 60%	23.5 $\pm$ 0.2	33.1 $\pm$ 0.3	36.3 $\pm$ 0.1	16.5 $\pm$ 0.1	28.1 $\pm$ 0.2
Preference	top 3	21.6 $\pm$ 0.2	31.7 $\pm$ 0.2	35.4 $\pm$ 0.1	14.9 $\pm$ 0.2	26.5 $\pm$ 0.2
	top 5	19.5 $\pm$ 0.3	30.4 $\pm$ 0.2	34.1 $\pm$ 0.2	13.5 $\pm$ 0.2	24.7 $\pm$ 0.2
	top 10	18.4 $\pm$ 0.2	28.3 $\pm$ 0.3	31.8 $\pm$ 0.4	12.5 $\pm$ 0.1	23.2 $\pm$ 0.2
	top 30	15.9 $\pm$ 0.3	26.2 $\pm$ 0.2	29.8 $\pm$ 0.3	10.7 $\pm$ 0.2	20.9 $\pm$ 0.2
	top 100	14.5 $\pm$ 0.5	23.7 $\pm$ 0.3	27.2 $\pm$ 0.3	9.4 $\pm$ 0.3	19.0 $\pm$ 0.3

**Table C.3**

Performance of the MHSA-based network in predicting the next location on interventional location sequences. The mean and standard deviation across five runs with different random parameter initializations are reported.

Intervention	Strength	Acc@1 (%)	Acc@5 (%)	Acc@10 (%)	F1 (%)	MRR (%)
$p^{new}$	0.1	30.0 $\pm$ 0.5	48.0 $\pm$ 0.9	54.2 $\pm$ 0.8	26.6 $\pm$ 0.3	38.4 $\pm$ 0.7
	0.25	21.1 $\pm$ 0.3	32.7 $\pm$ 0.1	36.9 $\pm$ 0.2	16.1 $\pm$ 0.1	26.7 $\pm$ 0.2
	0.5	9.0 $\pm$ 0.04	13.3 $\pm$ 0.04	15.4 $\pm$ 0.03	5.3 $\pm$ 0.04	11.3 $\pm$ 0.04
	0.75	3.2 $\pm$ 0.03	6.2 $\pm$ 0.1	8.1 $\pm$ 0.1	1.5 $\pm$ 0.04	5.0 $\pm$ 0.04
	0.9	1.8 $\pm$ 0.04	4.9 $\pm$ 0.1	7.0 $\pm$ 0.1	0.9 $\pm$ 0.04	3.7 $\pm$ 0.1
$\mu _p$	0.1	30.6 $\pm$ 0.4	49.1 $\pm$ 0.6	55.8 $\pm$ 0.7	27.8 $\pm$ 0.4	39.4 $\pm$ 0.5
	0.2	29.9 $\pm$ 0.2	45.9 $\pm$ 0.3	51.9 $\pm$ 0.6	26.4 $\pm$ 0.02	37.5 $\pm$ 0.2
	0.3	30.0 $\pm$ 0.7	44.0 $\pm$ 0.5	48.9 $\pm$ 0.6	25.4 $\pm$ 0.6	36.7 $\pm$ 0.7
	0.4	27.8 $\pm$ 0.2	40.0 $\pm$ 0.3	45.2 $\pm$ 0.4	22.8 $\pm$ 0.2	33.7 $\pm$ 0.2
	0.5	25.9 $\pm$ 0.3	37.3 $\pm$ 0.5	41.6 $\pm$ 0.4	20.5 $\pm$ 0.2	31.5 $\pm$ 0.3
	0.7	22.9 $\pm$ 0.2	31.3 $\pm$ 0.3	34.9 $\pm$ 0.2	17.2 $\pm$ 0.2	27.1 $\pm$ 0.2
	0.9	18.7 $\pm$ 0.05	25.4 $\pm$ 0.2	28.1 $\pm$ 0.1	13.0 $\pm$ 0.04	22.1 $\pm$ 0.1
$\mu _r$	0.1	16.5 $\pm$ 0.2	23.8 $\pm$ 0.1	26.8 $\pm$ 0.1	11.4 $\pm$ 0.2	20.2 $\pm$ 0.2
	0.2	24.9 $\pm$ 0.3	34.6 $\pm$ 0.3	38.3 $\pm$ 0.3	19.2 $\pm$ 0.2	29.6 $\pm$ 0.3
	0.3	29.6 $\pm$ 0.5	42.4 $\pm$ 0.1	46.9 $\pm$ 0.1	24.4 $\pm$ 0.5	35.7 $\pm$ 0.3
	0.4	31.9 $\pm$ 0.3	46.0 $\pm$ 0.4	51.1 $\pm$ 0.3	27.3 $\pm$ 0.2	38.7 $\pm$ 0.3
	0.5	30.5 $\pm$ 0.3	45.3 $\pm$ 0.4	50.7 $\pm$ 0.4	26.8 $\pm$ 0.2	37.6 $\pm$ 0.4
	0.7	34.2 $\pm$ 0.8	49.4 $\pm$ 0.7	55.3 $\pm$ 1.2	31.0 $\pm$ 0.7	41.6 $\pm$ 0.8
	0.9	32.0 $\pm$ 0.4	49.0 $\pm$ 0.6	55.5 $\pm$ 0.3	29.0 $\pm$ 0.4	40.1 $\pm$ 0.3
Attractiveness	top 0.1%	21.5 $\pm$ 0.2	31.2 $\pm$ 0.4	34.8 $\pm$ 0.3	16.0 $\pm$ 0.1	26.2 $\pm$ 0.2
	top 1%	20.3 $\pm$ 0.2	29.6 $\pm$ 0.3	33.2 $\pm$ 0.4	15.1 $\pm$ 0.2	24.9 $\pm$ 0.2
	top 5%	18.6 $\pm$ 0.1	28.3 $\pm$ 0.3	32.0 $\pm$ 0.3	13.5 $\pm$ 0.1	23.4 $\pm$ 0.1
	top 10%	17.4 $\pm$ 0.2	26.6 $\pm$ 0.1	30.4 $\pm$ 0.1	12.5 $\pm$ 0.1	22.0 $\pm$ 0.1
	last 30%	23.9 $\pm$ 0.2	34.1 $\pm$ 0.2	37.7 $\pm$ 0.2	18.3 $\pm$ 0.1	28.8 $\pm$ 0.2
	last 60%	22.6 $\pm$ 0.3	32.0 $\pm$ 0.3	35.6 $\pm$ 0.3	17.1 $\pm$ 0.2	27.2 $\pm$ 0.3
Preference	top 3	20.4 $\pm$ 0.1	30.4 $\pm$ 0.3	34.3 $\pm$ 0.3	15.4 $\pm$ 0.1	25.4 $\pm$ 0.2
	top 5	18.3 $\pm$ 0.3	28.5 $\pm$ 0.3	32.4 $\pm$ 0.2	13.7 $\pm$ 0.2	23.3 $\pm$ 0.3
	top 10	16.9 $\pm$ 0.1	26.7 $\pm$ 0.2	30.6 $\pm$ 0.2	12.3 $\pm$ 0.2	21.8 $\pm$ 0.1
	top 30	14.3 $\pm$ 0.2	24.5 $\pm$ 0.3	28.5 $\pm$ 0.1	10.5 $\pm$ 0.2	19.3 $\pm$ 0.2
	top 100	12.6 $\pm$ 0.3	22.1 $\pm$ 0.3	25.9 $\pm$ 0.3	9.2 $\pm$ 0.1	17.3 $\pm$ 0.3

## References

- Acheampong, R.A., 2018. Towards incorporating location choice into integrated land use and transport planning and policy: A multi-scale analysis of residential and job location choice behaviour. *Land Use Policy* 78, 397–409. <http://dx.doi.org/10.1016/j.landusepol.2018.07.007>.
- Alessandretti, L., Aslak, U., Lehmann, S., 2020. The scales of human mobility. *Nature* 587 (7834), 402–407. <http://dx.doi.org/10.1038/s41586-020-2909-1>.
- Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A., 2017. Multi-scale spatio-temporal analysis of human mobility. In: Preis, T. (Ed.), *PLOS ONE* 12 (2), e0171686. <http://dx.doi.org/10.1371/journal.pone.0171686>.
- Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., Baronchelli, A., 2018. Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* 2 (7), 485–491. <http://dx.doi.org/10.1038/s41562-018-0364-x>.
- Alstott, J., Bullmore, E., Plenz, D., 2014. powerlaw: A python package for analysis of heavy-tailed distributions. In: Rapallo, F. (Ed.), *PLoS ONE* 9 (1), e85777. <http://dx.doi.org/10.1371/journal.pone.0085777>.
- van Amsterdam, W.A.C., Verhoeff, J.J.C., de Jong, P.A., Leiner, T., Eijkemans, M., 2019. Eliminating biasing signals in lung cancer images for prognosis predictions with deep learning. *Npj Digit. Med.* 2 (1), 1–6. <http://dx.doi.org/10.1038/s41746-019-0194-x>.
- Axhausen, K.W., Gärling, T., 1992. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transp. Rev.* 12 (4), 323–341. <http://dx.doi.org/10.1080/01441649208716826>.
- Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer Normalization. <http://dx.doi.org/10.48550/arXiv.1607.06450>.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M., 2018. Human mobility: Models and applications. *Phys. Rep.* 734, 1–74. <http://dx.doi.org/10.1016/j.physrep.2018.01.001>.
- Baumann, P., Koehler, C., Dey, A.K., Santini, S., 2018. Selecting individual and population models for predicting human mobility. *IEEE Trans. Mob. Comput.* 17 (10), 2408–2422. <http://dx.doi.org/10.1109/TMC.2018.2797937>.
- Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature* 439 (7075), 462–465. <http://dx.doi.org/10.1038/nature04292>.
- Cai, H., Xin, Y., Martin, H., Raubal, M., 2022. Optimizing electric vehicle charging schedules based on probabilistic forecast of individual mobility. *AGILE: GIScience Ser.* 3, 3. <http://dx.doi.org/10.5194/agile-giss-3-3-2022>.
- Cao, J., Li, Q., Tu, W., Wang, F., 2019. Characterizing preferred motif choices and distance impacts. In: Tang, J. (Ed.), *PLOS ONE* 14 (4), e0215242. <http://dx.doi.org/10.1371/journal.pone.0215242>.
- Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C: Emerg. Technol.* 68, 285–299. <http://dx.doi.org/10.1016/j.trc.2016.04.005>.
- Cuttone, A., Lehmann, S., González, M.C., 2018. Understanding predictability and exploration in human mobility. *EPJ Data Sci.* 7 (1), 2. <http://dx.doi.org/10.1140/epjds/s13688-017-0129-1>.
- González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature* 453 (7196), 779–782. <http://dx.doi.org/10.1038/nature06958>.
- Graser, A., Jalali, A., Lampert, J., Weisenfeld, A., Janowicz, K., 2023. Deep learning from trajectory data: a review of deep neural networks and the trajectory data representations to train them. In: *Proceedings of the Workshop on Big Mobility Data Analytics (BMDA) Co-Located with EDBT/ICDT 2023 Joint Conference*.
- Harding, C., Faghhi Imani, A., Srikuenthiran, S., Miller, E.J., Nurul Habib, K., 2021. Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation* 48 (5), 2433–2460. <http://dx.doi.org/10.1007/s11116-020-10135-7>.
- He, T., Bao, J., Li, R., Ruan, S., Li, Y., Song, L., He, H., Zheng, Y., 2020. What is the human mobility in a new city: Transfer mobility knowledge across cities. In: *Proceedings of the 2020 World Wide Web Conference on World Wide Web. WWW '20*, ACM Press, New York, NY, USA, pp. 1355–1365. <http://dx.doi.org/10.1145/3366423.3380210>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Conference on Computer Vision and Pattern Recognition. CVPR '16*, pp. 770–778. <http://dx.doi.org/10.1109/CVPR.2016.90>.
- Hintermann, B., Schoeman, B., Molloy, J., Schatzmann, T., Tchervakov, C., Axhausen, K.W., 2023. The impact of COVID-19 on mobility choices in Switzerland. *Transp. Res. Part A: Policy Pr.* 169, 103582. <http://dx.doi.org/10.1016/j.tra.2023.103582>.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- Hong, Y., Martin, H., Xin, Y., Bucher, D., Reck, D.J., Axhausen, K.W., Raubal, M., 2023a. Conserved quantities in human mobility: From locations to trips. *Transp. Res. Part C: Emerg. Technol.* 146, 103979. <http://dx.doi.org/10.1016/j.trc.2022.103979>.
- Hong, Y., Xin, Y., Martin, H., Bucher, D., Raubal, M., 2021. A clustering-based framework for individual travel behaviour change detection. In: *11th International Conference on Geographic Information Science - Part II. GIScience '21*, vol. 208, p. 4. <http://dx.doi.org/10.4230/LIPIcs.GIScience.2021.II.4>.
- Hong, Y., Zhang, Y., Schindler, K., Raubal, M., 2023b. Context-aware multi-head self-attentional neural network model for next location prediction. *Transp. Res. Part C: Emerg. Technol.* 156, 104315. <http://dx.doi.org/10.1016/j.trc.2023.104315>.
- Huang, X., Kroening, D., Ruan, W., Sharp, J., Sun, Y., Thamo, E., Wu, M., Yi, X., 2020. A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability. *Comput. Sci. Rev.* 37, 100270. <http://dx.doi.org/10.1016/j.cosrev.2020.100270>.
- Järv, O., Ahas, R., Witlox, F., 2014. Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transp. Res. Part C: Emerg. Technol.* 38, 122–135. <http://dx.doi.org/10.1016/j.trc.2013.11.003>.
- Ji, Y., Gao, S., Huynh, T., Scheele, C., Triveri, J., Kruse, J., Bennett, C., Wen, Y., 2023. Rethinking the regularity in mobility patterns of personal vehicle drivers: A multi-city comparison using a feature engineering approach. *Trans. GIS* 27 (3), 663–685. <http://dx.doi.org/10.1111/tgis.13043>.
- Jiang, R., Song, X., Fan, Z., Xia, T., Chen, Q., Miyazawa, S., Shibasaki, R., 2018. Deep-UrbanMomentum: An online deep-learning system for short-term urban mobility prediction. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 32, (1). <http://dx.doi.org/10.1609/aaai.v32i1.11338>.
- Joseph, L., Neven, A., Martens, K., Kweka, O., Wets, G., Janssens, D., 2020. Measuring individuals' travel behaviour by use of a GPS-based smartphone application in Dar es Salaam, Tanzania. *J. Transp. Geogr.* 88, 102477. <http://dx.doi.org/10.1016/j.jtrangeo.2019.102477>.
- Kellermann, R., Sivizaca Conde, D., Rößler, D., Kliever, N., Dienel, H.-L., 2022. Mobility in pandemic times: Exploring changes and long-term effects of COVID-19 on urban mobility behavior. *Transp. Res. Interdiscip. Perspect.* 15, 100668. <http://dx.doi.org/10.1016/j.trip.2022.100668>.
- Kieu, L.-M., Ou, Y., Truong, L.T., Cai, C., 2020. A class-specific soft voting framework for customer booking prediction in on-demand transport. *Transp. Res. Part C: Emerg. Technol.* 114, 377–390. <http://dx.doi.org/10.1016/j.trc.2020.02.010>.
- Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations. ICLR '15*.
- Koushik, A.N., Manoj, M., Nezamuddin, N., 2020. Machine learning applications in activity-travel behaviour research: a review. *Transp. Rev.* 40 (3), 288–311. <http://dx.doi.org/10.1080/01441647.2019.1704307>.
- Kulkarni, V., Garbinato, B., 2019. 20 years of mobility modeling & prediction: Trends, shortcomings & perspectives. In: *Proceedings of the 27th International Conference on Advances in Geographic Information Systems. SIGSPATIAL '19*, pp. 492–495. <http://dx.doi.org/10.1145/3347146.3359110>.
- Li, X., Feng, M., Ran, Y., Su, Y., Liu, F., Huang, C., Shen, H., Xiao, Q., Su, J., Yuan, S., Guo, H., 2023. Big Data in Earth system science and progress towards a digital twin. *Nat. Rev. Earth & Environ.* 4 (5), 319–332. <http://dx.doi.org/10.1038/s43017-023-00409-w>.
- Li, Q., Liao, F., 2020. Incorporating vehicle self-relocations and traveler activity chains in a bi-level model of optimal deployment of shared autonomous vehicles. *Transp. Res. Part B: Methodol.* 140, 151–175. <http://dx.doi.org/10.1016/j.trb.2020.08.001>.
- van Lint, J.W.C., 2008. Online learning solutions for freeway travel time prediction. *IEEE Trans. Intell. Transp. Syst.* 9 (1), 38–47. <http://dx.doi.org/10.1109/TITS.2008.915649>.
- Lu, X., Wetter, E., Bharti, N., Tatem, A.J., Bengtsson, L., 2013. Approaching the limit of predictability in human mobility. *Sci. Rep.* 3 (1), 2923. <http://dx.doi.org/10.1038/srep02923>.
- Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L., 2021. A survey on deep learning for human mobility. *ACM Comput. Surv.* 55, 7:1–7:44. <http://dx.doi.org/10.1145/3485125>.
- Ma, Z., Zhang, P., 2022. Individual mobility prediction review: Data, problem, method and application. *Multimodal Transp.* 1 (1), 100002. <http://dx.doi.org/10.1016/j.multra.2022.100002>.
- Mandal, A., Johansson, C., Lindelöw, D., 2023. Exploring walking from the perspective of theory of planned behavior. *Transp. Res. Interdiscip. Perspect.* 22, 100931. <http://dx.doi.org/10.1016/j.trip.2023.100931>.
- Manibardo, E.L., Laña, I., Ser, J.D., 2022. Deep learning for road traffic forecasting: Does it make a difference? *IEEE Trans. Intell. Transp. Syst.* 23 (7), 6164–6188. <http://dx.doi.org/10.1109/tits.2021.3083957>.
- Martin, H., Becker, H., Bucher, D., Jonietz, D., Raubal, M., Axhausen, K.W., 2019. Begleitstudie SBB green class - abschlussbericht. *Arbeitsberichte Verkehrs- Und Raumplan.* 1439. <http://dx.doi.org/10.3929/ethz-b-000353337>.
- Martin, H., Hong, Y., Wiedemann, N., Bucher, D., Raubal, M., 2023a. Trackintel: An open-source Python library for human mobility analysis. *Comput. Environ. Urban Syst.* 101, 101938. <http://dx.doi.org/10.1016/j.compenvurbsys.2023.101938>.
- Martin, H., Wiedemann, N., Reck, D.J., Raubal, M., 2023b. Graph-based mobility profiling. *Comput. Environ. Urban Syst.* 100, 101910. <http://dx.doi.org/10.1016/j.compenvurbsys.2022.101910>.
- Moshkov, N., Bornholdt, M., Benoit, S., Smith, M., McQuin, C., Goodman, A., Senft, R.A., Han, Y., Babadi, M., Horvath, P., Cimini, B.A., Carpenter, A.E., Singh, S., Caicedo, J.C., 2024. Learning representations for image-based profiling of perturbations. *Nat. Commun.* 15 (1), 1594. <http://dx.doi.org/10.1038/s41467-024-45999-1>.
- Pappalardo, L., Manley, E., Sekara, V., Alessandretti, L., 2023. Future directions in human mobility science. *Nat. Comput. Sci.* 3 (7), 588–600. <http://dx.doi.org/10.1038/s43588-023-00469-4>.

- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., Barabási, A.-L., 2015. Returners and explorers dichotomy in human mobility. *Nat. Commun.* 6 (1), 8166. <http://dx.doi.org/10.1038/ncomms9166>.
- Pawlowski, N., Coelho de Castro, D., Glocker, B., 2020. Deep structural causal models for tractable counterfactual inference. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems. NIPS '20*, vol. 33, pp. 857–869.
- Pearl, J., Mackenzie, D., 2018. *The Book of Why: The New Science of Cause and Effect*, first ed. Basic Books, Inc., ISBN: 978-0-465-09760-9.
- Rahimi, S., Moore, A.B., Whigham, P.A., Dillingham, P., 2023. Counterfactual reasoning in space and time: Integrating graphical causal models in computational movement analysis. *Trans. GIS* <http://dx.doi.org/10.1111/tgis.13100>.
- Ramezani, S., Hasanazadeh, K., Rinne, T., Kajosaari, A., Kytä, M., 2021. Residential relocation and travel behavior change: Investigating the effects of changes in the built environment, activity space dispersion, car and bike ownership, and travel attitudes. *Transp. Res. Part A: Policy Pr.* 147, 28–48. <http://dx.doi.org/10.1016/j.trra.2021.02.016>.
- Rout, A., Nitowski, S., Ladle, A., Galpern, P., 2021. Using smartphone-GPS data to understand pedestrian-scale behavior in urban settings: A review of themes and approaches. *Comput. Environ. Urban Syst.* 90, 101705. <http://dx.doi.org/10.1016/j.compenvurbsys.2021.101705>.
- Roy, A., Nelson, T., Turaga, P., 2023. Functional data analysis approach for mapping change in time series: A case study using bicycle ridership patterns. *Transp. Res. Interdiscip. Perspect.* 17, 100752. <http://dx.doi.org/10.1016/j.trip.2022.100752>.
- Runge, J., Gerhardus, A., Varando, G., Eyring, V., Camps-Valls, G., 2023. Causal inference for time series. *Nat. Rev. Earth Environ.* 4 (7), 487–505. <http://dx.doi.org/10.1038/s43017-023-00431-y>.
- Schneider, C.M., Belik, V., Couronné, T., Smoreda, Z., González, M.C., 2013. Unravelling daily human mobility motifs. *J. R. Soc. Interface* 10 (84), 20130246. <http://dx.doi.org/10.1098/rsif.2013.0246>.
- Schölkopf, B., Locatello, F., Bauer, S., Ke, N.R., Kalchbrenner, N., Goyal, A., Bengio, Y., 2021. Toward causal representation learning. *Proc. IEEE* 109 (5), 612–634. <http://dx.doi.org/10.1109/JPROC.2021.3058954>.
- Schönfelder, S., Axhausen, K.W., 2016. *Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel*. Routledge.
- Sener, I.N., Pendyala, R.M., Bhat, C.R., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. *J. Transp. Geogr.* 19 (2), 294–303. <http://dx.doi.org/10.1016/j.jtrangeo.2010.03.013>.
- Shou, Z., Di, X., 2018. Similarity analysis of frequent sequential activity pattern mining. *Transp. Res. Part C: Emerg. Technol.* 96, 122–143. <http://dx.doi.org/10.1016/j.trc.2018.09.018>.
- Silva, R., 2016. Observational-interventional priors for dose-response learning. In: *Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS '16*, vol. 29, pp. 1569–1577.
- Solomon, A., Livne, A., Katz, G., Shapira, B., Rokach, L., 2021. Analyzing movement predictability using human attributes and behavioral patterns. *Comput. Environ. Urban Syst.* 87, 101596. <http://dx.doi.org/10.1016/j.compenvurbsys.2021.101596>.
- Song, C., Koren, T., Wang, P., Barabási, A.-L., 2010a. Modelling the scaling properties of human mobility. *Nat. Phys.* 6 (10), 818–823. <http://dx.doi.org/10.1038/nphys1760>.
- Song, C., Qu, Z., Blumm, N., Barabási, A.-L., 2010b. Limits of predictability in human mobility. *Science* 327 (5968), 1018–1021. <http://dx.doi.org/10.1126/science.1177170>.
- Sun, J., Kim, J., 2021. Joint prediction of next location and travel time from urban vehicle trajectories using long short-term memory neural networks. *Transp. Res. Part C: Emerg. Technol.* 128, 103114. <http://dx.doi.org/10.1016/j.trc.2021.103114>.
- Susilo, Y.O., Axhausen, K.W., 2014. Repetitions in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman Index. *Transportation* 41 (5), 995–1011. <http://dx.doi.org/10.1007/s11116-014-9519-4>.
- Tang, Y., Cheng, N., Wu, W., Wang, M., Dai, Y., Shen, X., 2019. Delay-minimization routing for heterogeneous VANETs with machine learning based mobility prediction. *IEEE Trans. Veh. Technol.* 68 (4), 3967–3979. <http://dx.doi.org/10.1109/TVT.2019.2899627>.
- Teixeira, D.d.C., Almeida, J.M., Viana, A.C., 2021. On estimating the predictability of human mobility: the role of routine. *EPJ Data Sci.* 10 (1), 1–30. <http://dx.doi.org/10.1140/epjds/s13688-021-00304-8>.
- Teixeira, D.d.C., Viana, A.C., Alvim, M.S., Almeida, J.M., 2019. Deciphering predictability limits in human mobility. In: *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. SIGSPATIAL '19*, pp. 52–61. <http://dx.doi.org/10.1145/3347146.3359093>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS '17*, vol. 30, pp. 5998–6008.
- Wang, S., Cao, J., Yu, P.S., 2022. Deep learning for spatio-temporal data mining: A survey. *IEEE Trans. Knowl. Data Eng.* 34 (8), 3681–3700. <http://dx.doi.org/10.1109/TKDE.2020.3025580>.
- Wiedemann, N., Hong, Y., Raubal, M., 2023a. Predicting visit frequencies to new places. In: *12th International Conference on Geographic Information Science. GIScience '23*, vol. 277, pp. 84:1–84:6. <http://dx.doi.org/10.4230/LIPIcs.GIScience.2023.84>.
- Wiedemann, N., Martin, H., Suel, E., Hong, Y., Xin, Y., 2023b. Influence of tracking duration on the privacy of individual mobility graphs. *J. Locat. Based Serv.* 1–19. <http://dx.doi.org/10.1080/17489725.2023.2239190>.
- Xin, Y., Tagasovska, N., Perez-Cruz, F., Raubal, M., 2022. Vision paper: causal inference for interpretable and robust machine learning in mobility analysis. In: *Proceedings of the 30th International Conference on Advances in Geographic Information Systems. SIGSPATIAL '22*, pp. 1–4. <http://dx.doi.org/10.1145/3557915.3561473>.
- Xu, Y., Belyi, A., Bojic, I., Ratti, C., 2018a. Human mobility and socioeconomic status: Analysis of Singapore and Boston. *Comput. Environ. Urban Syst.* 72, 51–67. <http://dx.doi.org/10.1016/j.compenvurbsys.2018.04.001>.
- Xu, Y., Çolak, S., Kara, E.C., Moura, S.J., González, M.C., 2018b. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Energy* 3 (6), 484–493. <http://dx.doi.org/10.1038/s41560-018-0136-x>.
- Xu, Y., Zou, D., Park, S., Li, Q., Zhou, S., Li, X., 2022. Understanding the movement predictability of international travelers using a nationwide mobile phone dataset collected in South Korea. *Comput. Environ. Urban Syst.* 92, 101753. <http://dx.doi.org/10.1016/j.compenvurbsys.2021.101753>.
- Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., Lai, Y.-C., 2017. Universal model of individual and population mobility on diverse spatial scales. *Nat. Commun.* 8 (1), 1639. <http://dx.doi.org/10.1038/s41467-017-01892-8>.
- Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., Yin, B., 2022. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Trans. Intell. Transp. Syst.* 23 (6), 4927–4943. <http://dx.doi.org/10.1109/tits.2021.3054840>.
- Yuan, Y., Raubal, M., Liu, Y., 2012. Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Comput. Environ. Urban Syst.* 36 (2), 118–130. <http://dx.doi.org/10.1016/j.compenvurbsys.2011.07.003>.
- Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S., 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Sci. Rep.* 5 (1), 9136. <http://dx.doi.org/10.1038/srep09136>.
- Zhao, C., Zeng, A., Yeung, C.H., 2021. Characteristics of human mobility patterns revealed by high-frequency cell-phone position data. *EPJ Data Sci.* 10 (1), <http://dx.doi.org/10.1140/epjds/s13688-021-00261-2>.
- Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2023. Domain generalization: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 45 (4), 4396–4415. <http://dx.doi.org/10.1109/TPAMI.2022.3195549>.