Causal Inference for Interpretable and Robust Deep Learning in Mobility Analysis

Ye Hong

Martin Raubal

STRC Conference Paper 2025 **April 26, 2025**

STRC | **25th Swiss Transport Research Conference**
Monte Verità / Ascona, May 14-16, 2025

Causal Inference for Interpretable and Robust Deep Learning in Mobility Analysis

Ye Hong
Institute of Cartography and Geoinformation
ETH Zurich
hongy@ethz.ch

Martin Raubal
Institute of Cartography and Geoinformation
ETH Zurich
mraubal@ethz.ch

April 26, 2025

Abstract

Deep learning (DL) networks are increasingly utilized in mobility analysis and predictive modeling, yet their intricate internal workings hinder interpretability and complicate robustness assessments, limiting real-world deployment. Recent studies identified causal inference as a promising method for evaluating DL robustness, as it enables the extraction of interpretable and actionable insights. This study introduces a causal intervention framework to assess how mobility-related factors influence DL networks for next-location prediction. We employ mechanistic mobility models to simulate location visit sequences and control behavioral dynamics through targeted interventions in data generation. The modified sequences are analyzed using standardized mobility metrics and processed through pre-trained DL networks to quantify performance variations. Performance deviations highlight three key behavioral factors: (1) sequential patterns in location transitions, (2) individual tendencies for spatial exploration, and (3) heterogeneity in location preferences at both the population and individual levels. We publicly released a modular, open-source Python framework that includes formal data specifications, mobility models for synthetic dataset generation, benchmark DL architectures, and evaluation protocols. These insights contribute to the practical implementation of mobility prediction systems, while the framework establishes a foundation for integrating causal inference to improve DL interpretability and robustness in mobility applications.

Keywords

Mobility behavior; Domain shift; Individual mobility simulation; Next location prediction; Causal intervention.

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). In recent years, the availability of human digital traces and the advancements in data-driven models, particularly deep neural networks, 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 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 leads to a lack of comparability 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 mobility neural networks.

Establishing the behavior and performance connection assists in evaluating network robustness when confronted with unforeseen inputs. The optimization of networks requires a training dataset, making their performance heavily dependent on the quality and representativeness of this data (Yin *et al.*, 2022). However, mobility behavior evolves dynamically over space and time. The 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 periods. Yet, this relationship remains predominantly unexplored.

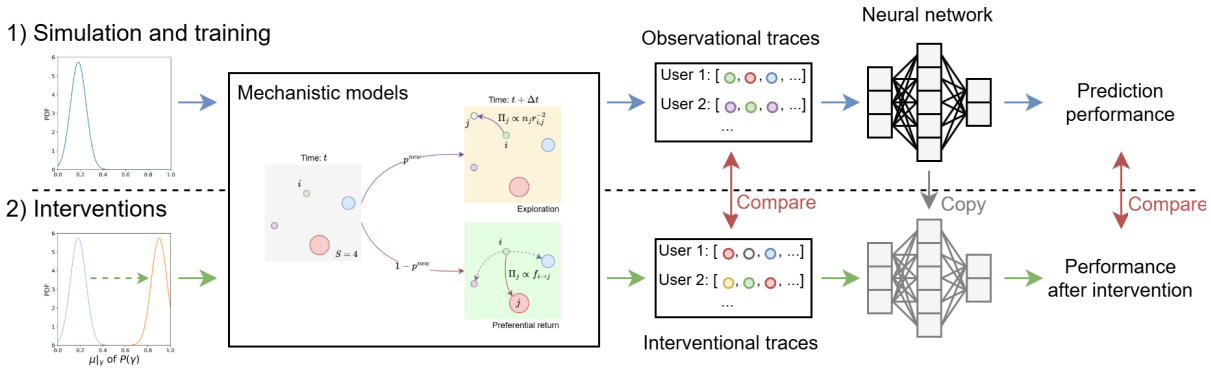
Causal intervention offers a promising tool for generating data from diverse environments,

enabling robustness assessment 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[1]. Specifically, we utilize mechanistic models to generate mobility traces, and employ causal intervention strategies in data generation, allowing for flexible modifications of the defined mobility behavior. We assess the performance of trained neural networks on these synthetic traces for mobility prediction. Results show the impact of behavioral factors and provide benchmarks for mobility prediction networks, with practical applications for evaluating network performance and transferring these networks across environments.

2 Methods

The overall pipeline for assessing the robustness of prediction networks is illustrated in Figure 1.

Figure 1: 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).



[1]The source code is available at `https://github.com/irmlma`

2.1 Individual mobility models

Individual mobility models generate realistic trajectories based on a predefined set of behavioral parameters, allowing for direct control over the mobility behavior. We introduce a density transition (DT)-exploration and preferential return (EPR) model, which is based on two EPR-based (Song *et al.*, 2010a) models, namely density (d)-EPR (Pappalardo *et al.*, 2015) and individual preferential transition (IPT) (Zhao *et al.*, 2021).

The EPR model introduced two competing mechanisms, namely, exploration and preferential return (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 Δt, where Δt is sampled from its distribution $P(\Delta t)$. They chooses to explore a previously unvisited location with probability p_t^{new}:

$$p_t^{new} = \rho S_t^{-\gamma} \tag{1}$$

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, the d-EPR model assigns population attractiveness factors to locations to model the tendency to visit popular locations. The probability Π_j of selecting location j depends on the travel distance and location attractiveness:

$$\Pi_j \propto n_j r_{i,j}^{-2} \tag{2}$$

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 to location j. 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 a complementary probability $1 - p_t^{new}$. In this case, the IPT model defines the probability of moving to location j to be proportional to the previous visit frequency from location i to j:

$$\Pi_j \propto f_{i \to j} \tag{3}$$

where $f_{i \to j}$ is the empirically observed visitation frequency from i to j, which collectively forms the Markov transition matrix F. We combine the exploration mechanism of d-EPR and the preferential return mechanism of IPT to introduce the DT-EPR model. As a result, for each individual u^i, 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.

2.2 Intervention design

We use empirically estimated behavioral parameters to generate *observational* mobility traces, and 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 scenarios such as spatial shifts when certain locations become more attractive or temporal shifts in mobility behavior between seasons. We perform interventions on the following parameters:

- The exploration tendency p^{new}, affecting whether or not to explore in the next time step (Eq. 1). In EPR-like models, p^{new} is determined by parameters ρ and γ, independently sampled for each individual. We introduce interventions on ρ and γ by altering their distributions, producing pseudo-populations with different exploration behaviors. Additionally, we perform hard interventions on p^{new} by fixing its value to a constant.
- 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). We introduce interventions by manipulating the personalized Markov transition matrix, 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.

2.3 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. Consider a sub-sequence $(L_k)_{k=m}^{n} \in T^i$ visited by individual u^i from time step m to n, the goal is to predict the next visited location, i.e., the location identifier l_{n+1}. We employ LSTM and MHSA-based networks for next location prediction and refer readers to Solomon *et al.* (2021) and Hong *et al.* (2023) for their detailed implementation.

3 Results

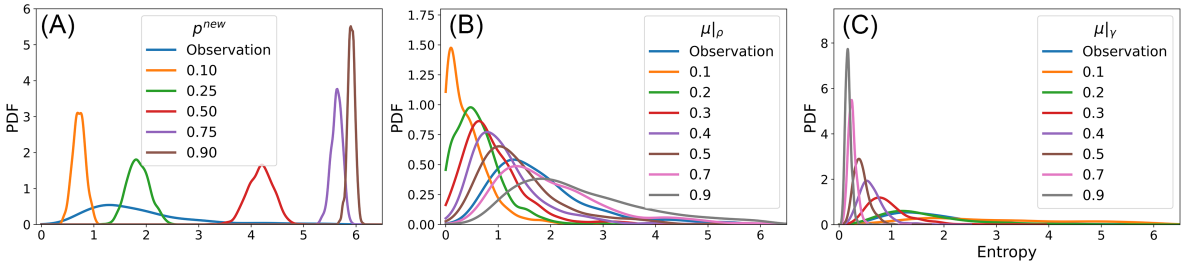
We leverage a smartphone-based travel survey to calibrate the parameters for mechanistic mobility models. The survey was conducted by the Swiss Federal Railways (SBB), known as the SBB Green Class (GC) E-Car pilot study, which aimed to assess the impact of a Mobility-as-a-Service (MaaS) offer on travel 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. Participants were asked to install a commercial application on their smartphones, continuously recording their whereabouts from GNSS signals. The application pre-processed the raw traces to identify *stay points* representing areas where users were stationary, which were later spatially aggregated into *locations*, the basic study units for mechanistic mobility models, using the *Trackintel* library (Martin *et al.*, 2023).

We estimated the waiting time distribution $P(\Delta t)$ using a log-normal fit, yielding best-fit parameters $\mu|_{\Delta t} = 0.75$ and $\sigma|_{\Delta t} = 1.49$. To assess exploration tendencies $P(\rho)$ and $P(\gamma)$, we calculated the location exploration speed and fit normal distributions across individuals. This resulted in $\mu|_{\rho} = 0.18$, $\sigma|_{\rho} = 0.07$ and $\mu|_{\gamma} = 0.64$, $\sigma|_{\gamma} = 0.16$, respectively. These parameters were employed to simulate traces for 800 individuals, both in the observational dataset and in each interventional dataset. For each synthetic individual, we independently sample values for ρ and γ, and randomly assign an initial location from their empirically observed top-5 most visited places. The simulation process proceeded until each individual had visited 2000 locations.

3.1 Mobility simulation and intervention

Figure 2 shows the distributions of mobility entropy (Song *et al.*, 2010b), which capture the regularity of mobility patterns. These metrics are presented for both observational and interventional location sequences generated by DT-EPR.

Figure 2: The mobility entropy of observational and interventional location sequences. We show the distributions for (A) hard interventions on p^{new}, (B) interventions on ρ by shifting $\mu|_\rho$ of $P(\rho)$, and (C) interventions on γ by shifting $\mu|_\gamma$ of $P(\gamma)$.



The interventions can effectively and directionally change the underlying mobility pattern, as demonstrated by the shifts in the mobility metric distributions. For example, with an increase in the exploration tendency p^{new}, individuals are encouraged to visit new locations, leading to mobility sequences with higher entropy. Moreover, the impact 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 (Figure 2A), changes induced by exploration parameters ρ and γ are more nuanced and provide more fine-grained control (Figure 2B and C).

3.2 Robustness of location prediction networks

We now evaluate the performance of prediction networks using interventional mobility sequences, which reveal their robustness in out-of-distribution (OoD) scenarios, i.e., when the training and testing data are not generated from the same distribution. Figure 3 displays the variations in Acc@1 scores for interventions on exploration tendency, and Figure 4 depicts variations for interventions on population attractiveness and individual preference. Although similar performance trends are observed for both networks, LSTM consistently outperforms MHSA in OoD settings.

Figure 3: Next location prediction performances for interventions on individuals' exploration tendency. We show the variations in Acc@1 for (A) hard interventions on p^{new}, (B) interventions on ρ by shifting $\mu|_\rho$ of $P(\rho)$, and (C) interventions on γ by shifting $\mu|_\gamma$ of $P(\gamma)$.

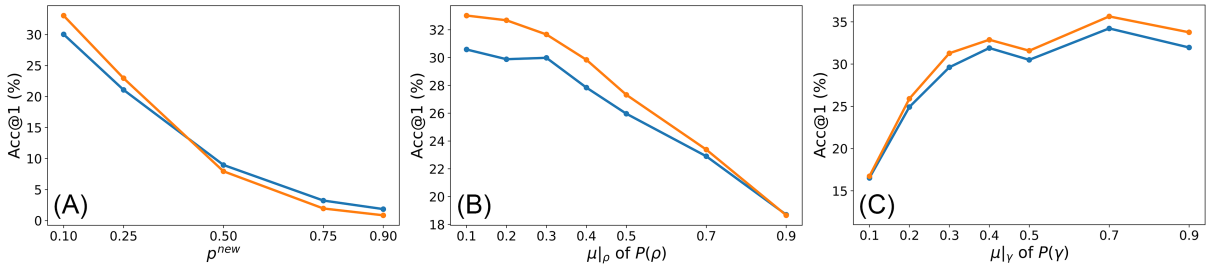
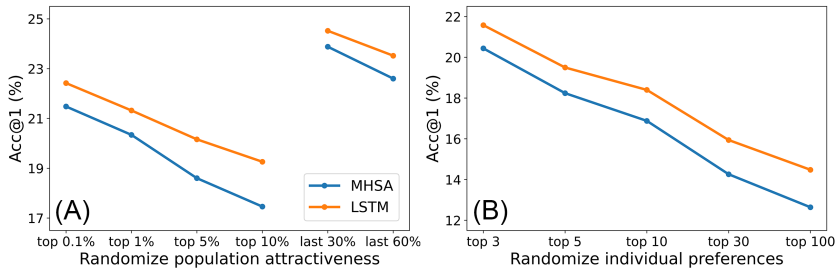


Figure 4: Next location prediction performances for interventions on population attractiveness and individual preference. We show the variations in Acc@1 for (A) randomizing empirical location visits of the dataset, and (B) randomizing empirical location visits for each individual.



The performance variations for exploration interventions (Figure 3) generally align with their strengths and directions, but 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 Acc@1 $< 10\%$, suggesting that the learned location transition patterns cannot be adequately utilized. Comparatively, interventions on γ and ρ indirectly affect p^{new}, which retains the diminishing exploration speed over time. 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 (Figure 2). 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 affects the prediction performances. In the shuffling process, "top 1%"

includes the most frequently visited 1% of locations across the population (Figure 4A), and "top 3" considers the three most visited locations for each individual (Figure 4B). Both types of interventions substantially impact the prediction ability, with altering the number of visits separately for each individual showing a stronger influence, as evidenced by the higher drop in performance indicators. 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.

3.3 Open-source framework

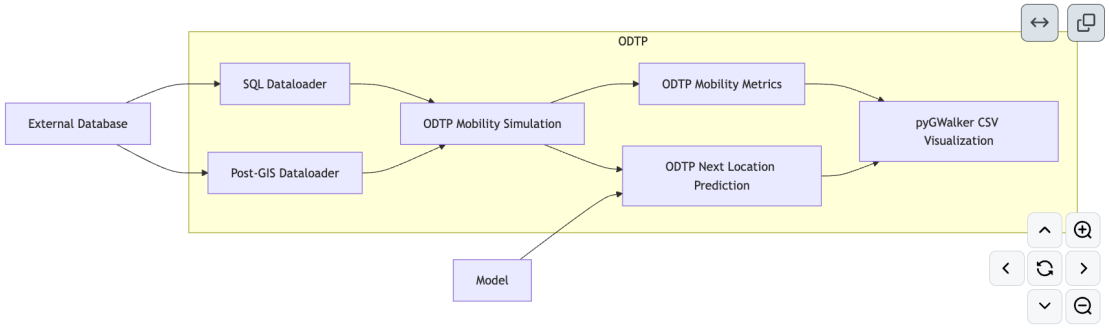
We open-sourced the framework to support the evaluation of deep learning models' robustness in mobility analysis. Our implementation adheres to ODTP guidelines, which aim to enable the study of interventions in mobility systems through a digital representation of the real world (Grübel *et al.*, 2023). This digital twin captures essential characteristics needed to describe and analyze observable processes for specific tasks, ultimately supporting informed decision-making and, when appropriate, the actuation of physical systems based on these insights. ODTP provides a comprehensive framework for designing, managing, executing, and sharing digital twins. It includes command-line and graphical user interfaces, ensuring seamless operation and management. Furthermore, ODTP features a "zoo" repository of modular components, streamlining the efficient creation and recreation of digital twins.

The framework has been open-sourced[2], and the components and structure of the implementation[3] are shown in Figure 5. The pipeline can be hosted locally using Docker Compose, featuring data access controls that support both CSV files for non-sensitive data and a PostGIS data loader for securely importing sensitive movement data via the database. These input data are compatible with existing microscopic mobility simulators (such as DT-EPR), enabling users to define multiple parameter sets for causal interven-

[2]Available at `https://github.com/odtp-org/dt-mobility-causal-intervention`

[3]A complete documentation of the implementation, along with step-by-step instructions for execution, can be found at `https://odtp-org.github.io/odtp-manuals/usecases/mobility-causal-interventions/`

Figure 5: Components and pipeline of the robustness evaluation framework implemented within ODTP.



tion on mobility behaviors. Users can then leverage the mobility metrics component to comprehensively evaluate the mobility behaviors captured in the synthetic sequences. The framework also supports training next location prediction deep neural networks, either from scratch using designated training datasets or by evaluating pre-trained models using specified validation data. To assess robustness, we provide standard performance metrics alongside uncertainty evaluation scores attached to the predictions (Dirmeier *et al.*, 2023). Additionally, we include a visualization component named pyGWalker, which provides a graphic user interface for exploring the result.

4 Conclusion

Unraveling the role and impact of 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, we perform causal interventions on their 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 the tendency to explore new locations and location preferences at both population and individual levels. We open-sourced the framework following ODTP guidelines, ensuring the use of standardized, extensible components that adhere to high software development standards.

5 References

Dirmeier, S., Y. Hong, Y. Xin and F. Perez-Cruz (2023) Uncertainty quantification and out-of-distribution detection using surjective normalizing flows, *arXiv preprint*.

Graser, A., A. Jalali, J. Lampert, A. Weißenfeld and K. Janowicz (2023) Deep Learning From Trajectory Data: a Review of Deep Neural Networks and the Trajectory Data Representations to Train Them, paper presented at the *Proceedings of the Workshop on Big Mobility Data Analytics (BMDA) co-located with EDBT/ICDT 2023 Joint Conference*.

Grübel, J., C. Vivar Rios, M. Balać, Y. Xin, R. M. Franken, S. Ossey, M. Raubal, K. W. Axhausen and O. Riba Grognuz (2023) "CH on the move": Introducing the Prototype Digital Twin of The Swiss Mobility System, paper presented at the *Swiss Transport Research Conference (STRC '23)*.

He, T., J. Bao, R. Li, S. Ruan, Y. Li, L. Song, H. He and Y. Zheng (2020) What is the Human Mobility in a New City: Transfer Mobility Knowledge Across Cities, paper presented at the *Proceedings of the 2020 World Wide Web Conference on World Wide Web (WWW '20)*, 1355–1365, New York, NY, USA.

Hong, Y., Y. Zhang, K. Schindler and M. Raubal (2023) Context-aware multi-head self-attentional neural network model for next location prediction, *Transportation Research Part C: Emerging Technologies*, **156**, 104315.

Huang, X., D. Kroening, W. Ruan, J. Sharp, Y. Sun, E. Thamo, M. Wu and X. Yi (2020) A survey of safety and trustworthiness of deep neural networks: Verification, testing, adversarial attack and defence, and interpretability, *Computer Science Review*, **37**, 100270.

Koushik, A. N., M. Manoj and N. Nezamuddin (2020) Machine learning applications in activity-travel behaviour research: a review, *Transport Reviews*, **40** (3) 288–311.

Kulkarni, V. and B. Garbinato (2019) 20 years of mobility modeling & prediction: Trends, shortcomings & perspectives, paper presented at the *Proceedings of the 27th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '19)*, 492–495.

Ma, Z. and P. Zhang (2022) Individual mobility prediction review: Data, problem, method

and application, *Multimodal Transportation*, **1** (1) 100002.

Manibardo, E. L., I. Laña and J. D. Ser (2022) Deep Learning for Road Traffic Forecasting: Does it Make a Difference?, *IEEE Transactions on Intelligent Transportation Systems*, **23** (7) 6164–6188.

Martin, H., H. Becker, D. Bucher, D. Jonietz, M. Raubal and K. W. Axhausen (2019) Begleitstudie SBB Green Class - Abschlussbericht, *Arbeitsberichte Verkehrs- und Raumplanung*, **1439**.

Martin, H., Y. Hong, N. Wiedemann, D. Bucher and M. Raubal (2023) Trackintel: An open-source python library for human mobility analysis, *Computers, Environment and Urban Systems*, **101**, 101938.

Pappalardo, L., E. Manley, V. Sekara and L. Alessandretti (2023) Future directions in human mobility science, *Nature Computational Science*, **3** (7) 588–600.

Pappalardo, L., F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti and A.-L. Barabási (2015) Returners and explorers dichotomy in human mobility, *Nature Communications*, **6** (1) 8166.

Solomon, A., A. Livne, G. Katz, B. Shapira and L. Rokach (2021) Analyzing movement predictability using human attributes and behavioral patterns, *Computers, Environment and Urban Systems*, **87**, 101596.

Song, C., T. Koren, P. Wang and A.-L. Barabási (2010a) Modelling the scaling properties of human mobility, *Nature Physics*, **6** (10) 818–823.

Song, C., Z. Qu, N. Blumm and A.-L. Barabasi (2010b) Limits of Predictability in Human Mobility, *Science*, **327** (5968) 1018–1021.

Tang, Y., N. Cheng, W. Wu, M. Wang, Y. Dai and X. Shen (2019) Delay-Minimization Routing for Heterogeneous VANETs With Machine Learning Based Mobility Prediction, *IEEE Transactions on Vehicular Technology*, **68** (4) 3967–3979.

Wang, S., J. Cao and P. S. Yu (2022) Deep Learning for Spatio-Temporal Data Mining: A Survey, *IEEE Transactions on Knowledge and Data Engineering*, **34** (8) 3681–3700.

Xin, Y., N. Tagasovska, F. Perez-Cruz and M. Raubal (2022) Vision paper: causal inference for interpretable and robust machine learning in mobility analysis, paper presented

at the *Proceedings of the 30th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '22)*, 1–4.

Yin, X., G. Wu, J. Wei, Y. Shen, H. Qi and B. Yin (2022) Deep Learning on Traffic Prediction: Methods, Analysis, and Future Directions, *IEEE Transactions on Intelligent Transportation Systems*, **23** (6) 4927–4943.

Zhao, C., A. Zeng and C. H. Yeung (2021) Characteristics of human mobility patterns revealed by high-frequency cell-phone position data, *EPJ Data Science*, **10** (1).