



Deep generative model for human mobility behavior

Ye Hong^{a,b,*}, Yatao Zhang^{c,d}, Konrad Schindler^e, Martin Raubal^{a,c}

^a Institute of Cartography and Geoinformation, ETH Zurich, Switzerland

^b Department of Human Geography, Lund University, Sweden

^c Future Resilient Systems, Singapore-ETH Centre, ETH Zurich, Singapore

^d Department of Geography, University College London, England

^e Photogrammetry and Remote Sensing, ETH Zurich, Switzerland

ARTICLE INFO

Keywords:

Human mobility

Activity schedules

Deep generative model

Transport planning

Travel behavior modeling

ABSTRACT

Understanding and modeling human mobility is central to challenges in transport planning, sustainable urban design, and public health. Despite decades of effort, simulating individual mobility remains challenging because of its complex, context-dependent, and exploratory nature. Here, building on the activity-based view of daily mobility, we propose MobilityGen, a diffusion-based generative framework for simulating multi-attribute activity-travel sequences over days to weeks at large spatial scales. By linking behavioral attributes with environmental context, MobilityGen reproduces key patterns such as scaling laws for location visits, activity time allocation, and the coupled evolution of travel mode and destination choices. It reflects spatio-temporal variability and generates diverse and plausible mobility patterns consistent with the built environment. Beyond standard validation, MobilityGen enables analyses that have been difficult with earlier models, including how access to urban space varies across travel modes and how co-presence dynamics shape social exposure and segregation. Together, these results support an integrated, data-driven basis for fine-grained studies of human mobility behavior and its societal implications.

1. Introduction

Human mobility is inherently complex and often appears unstructured. From day to day, individuals visit locations across various spatial scales, using multiple travel modes at different periods of the day to engage in a range of activities. Indeed, movement decisions involve intricate psychological planning (Axhausen and Gärling, 1992) and are influenced by individual factors such as physical ability (Burbidge and Goulias, 2009) and socioeconomic status (Hanson and Hanson, 1981). Despite this complexity, human mobility is fundamentally constrained by physical infrastructures (Jiang et al., 2009), such as road and rail networks, and heavily influenced by the surrounding environment (Handy et al., 2002; Hong et al., 2014). Ample empirical evidence demonstrates that mobility patterns follow simple scaling laws (Brockmann et al., 2006; González et al., 2008; Alessandretti et al., 2018; Schlöpfer et al., 2021), and exhibit striking regularities, for instance regarding spatial location choices (Song et al., 2010b). These collective regularities form the foundation for the realistic simulation of human mobility behaviors. For transport planning, the key challenge is to simulate not only *where* people travel, but also *when* activities occur, and *how* trips are made, while preserving their interdependencies across days.

Despite significant efforts and the importance of the task, simulating mobility behavior remains challenging. Current approaches to synthesizing disaggregated travel demand, such as the eqasim pipeline (Hörl and Balac, 2021), the OASIS framework (Pougala et al.,

* Corresponding author.

E-mail addresses: ye.hong@keg.lu.se (Y. Hong), yatao.zhang@ucl.ac.uk (Y. Zhang), schindler@ethz.ch (K. Schindler), mraubal@ethz.ch (M. Raubal).

<https://doi.org/10.1016/j.trc.2026.105916>

Received 12 January 2026; Received in revised form 19 July 2026; Accepted 3 August 2026

0968-090X/© 2026 The Author(s).

Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

2023), and various other methods based on the principle of utility maximization (Bhat and Singh, 2000; Auld and Mohammadian, 2012), focus on activity participation and temporal constraints, while often simplifying the spatial structure of activity locations and travel. Conversely, state-of-the-art mechanistic models, including the Container model (Alessandretti et al., 2020), the exploration and preferential return (EPR) model (Song et al., 2010a), and more recent variants derived from them (Pappalardo et al., 2015; Barbosa et al., 2015; Zhao et al., 2021), faithfully reproduce scaling laws of location visits, yet their limited modeling capacity restricts them from capturing the interactions between individuals' spatial preferences and their other movement decisions (Pappalardo et al., 2023). Neglecting those interactions can lead to misconceptions about individual space usage. This has far-reaching consequences for downstream applications that depend on accurate spatial modeling, such as sustainable transportation (Xu et al., 2018), human-centered urban design (Zhang et al., 2023), and epidemic prevention (Barreras and Watts, 2024).

Data-driven generative models offer a possible route to simulation that preserves multi-attribute travel behavior at scale. In recent years, learning-based models have revolutionized image and text processing. Notably, natural language processing has experienced remarkable progress through new sequence processing techniques that excel at capturing complex patterns in large text collections, in particular, the transformer architecture (Vaswani et al., 2017) and the denoising diffusion probabilistic model (DDPM) (Sohl-Dickstein et al., 2015; Ho et al., 2020). Their effectiveness in modeling language has facilitated applications across diverse domains concerned with spatial or temporal sequences, such as protein structures (Liu et al., 2024), turbulent flows (Li et al., 2024), or human life events (Savcicens et al., 2024). Inspired by these advances, recent work has begun to adapt deep generative frameworks to human mobility and trajectory synthesis (Chu et al., 2024; Song et al., 2024; Long et al., 2025). These studies highlight the promise of data-driven generation, while leaving open the challenge of jointly modeling the interdependent spatio-temporal and modal structure of daily mobility behavior at scale. Following activity-based modeling and recent data-driven mobility prediction studies (Axhausen and Gärling, 1992; Mo et al., 2022), individual mobility can be represented as a sequence of discrete, time-ordered activity-travel events, each linking concurrent spatial, temporal, and modal attributes. Still, language models are not directly applicable to human mobility behavior. Generalizing deep generative modeling techniques, designed for one-dimensional word sequences, to multifaceted, context-dependent mobility behavior requires careful consideration.

Here, we introduce MobilityGen, which operationalizes this event-based view within a diffusion-based generative framework for daily mobility sequences unfolding over days and weeks at large spatial scales. In this framework, where people go, when activities occur, and how they travel are modeled as coupled behavioral decisions rather than separable components. We use a DDPM to learn the joint structure of multi-attribute event sequences from data. Each event comprises multiple behavioral attributes, modeled as distinct but interdependent learning objectives, with their interactions captured through a shared latent representation that encodes meaningful structure across mode, time, and location. In addition, MobilityGen incorporates contextual features of the built environment, enabling the generation of diverse behavioral variations consistent with observed constraints and context. As our experiments show, MobilityGen preserves national-scale visitation and transition regularities. It also captures behavioral dependencies that span extended temporal horizons, producing synthetic sequences aligned with real-world patterns. Together, these results demonstrate the utility of diffusion-based generative modeling for simulating coupled daily behavior at scale. This enables the generation of comprehensive, fine-grained mobility data and supports applications ranging from testing sustainable transport strategies and assessing accessibility in urban design to analyzing co-presence dynamics that influence social segregation and epidemic response.

2. Methodology

2.1. Event-level mobility representation and generative modeling

To characterize individual mobility and support an event-level model, we analyzed global navigation satellite system (GNSS) trajectories from a smartphone-based travel survey that recorded the mobility behavior of thousands of participants across Switzerland. GNSS traces are segmented into stationary periods, i.e., fixed locations where individuals perform activities, and movement sections corresponding to transits between these activity locations. We then calculate activity attributes (such as time use and travel mode) and project the activity locations onto a hierarchical multi-resolution grid, with the highest resolution being $\approx 500 \times 500 \text{ m}^2$ per cell. Following activity-based modeling and prior data-driven studies (Axhausen and Gärling, 1992; Mo et al., 2022), we represent human mobility as sequences of activity events, each linked to specific attributes such as location, time use, and travel mode, along with contextual information (Fig. 1a). We particularly focus on contexts that characterize the built environment, including the coordinate geometry of locations and points of interest (POIs) that represent urban functional facilities.

The modeling challenge is therefore to learn how these heterogeneous event attributes vary together over time and across contexts. Traditional methods often focus on a single aspect, such as location choices (Alessandretti et al., 2020) or travel mode preferences (Reck et al., 2022), or are limited to describing simple behavioral dependencies (Jiang et al., 2016; Hörl and Balac, 2021). MobilityGen addresses this gap by casting the joint spatial, temporal, and modal structure of activity-travel events as a generative modeling problem, learned with a diffusion-based framework conditioned on contextual information.

MobilityGen employs learnable embedding modules to map raw attributes and contexts to a shared embedding space (Fig. 1b). Learning comprehensive mobility embeddings enables the model to encode the rich multi-attribute dependencies of activity events. We utilize a DDPM (Sohl-Dickstein et al., 2015; Ho et al., 2020), a powerful scheme for learning complex, high-dimensional distributions, to capture the structure of mobility event sequences. In a DDPM, the original sequence is gradually corrupted with tractable random noise, and an encoder-decoder network (in our case, a transformer (Vaswani et al., 2017)) learns the reverse denoising process (Fig. 1c). Once trained, the (reverse) process can generate new, plausible sequences from Gaussian noise (Fig. 1d). Additionally, a transformer encoder derives features from the observed travel sequence, acting as a guiding signal during denoising for generating

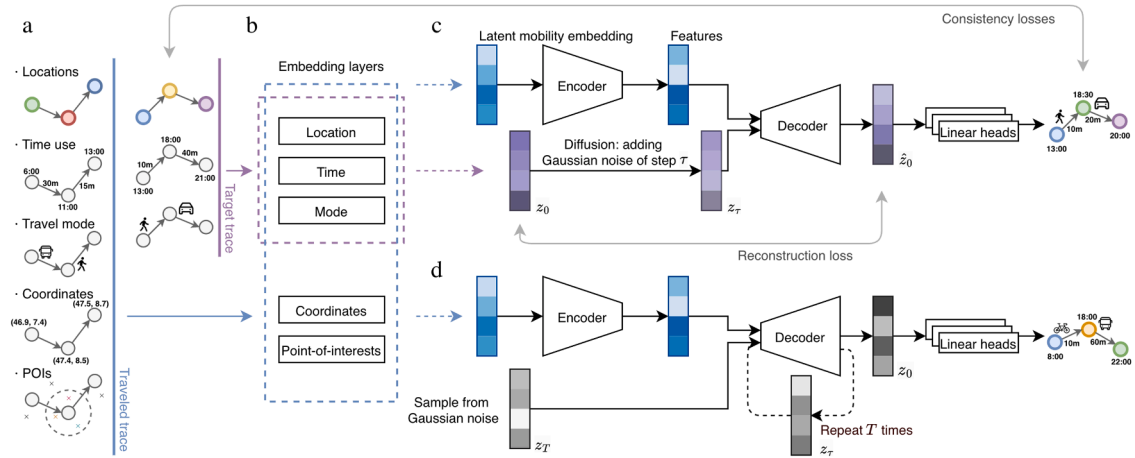


Fig. 1. Modeling individual mobility with an event-level generative framework. **a.** We represent individual mobility as a sequence of chronologically ordered activity events, where each event has associated activity attributes (location, time use, and travel mode) and contextual information (coordinates and nearby POIs). **b.** Raw activity events are encoded into latent mobility embeddings through dedicated embedding modules. **c.** During training, the diffusion process adds Gaussian noise to the target sequence embedding z_0 . A transformer-based decoder learns to reverse this process and reconstruct a denoised embedding $\hat{z}_0$. Features from the traveled sequence are extracted by an encoder and provided as guidance to the decoder. Final activity attributes are predicted via linear output heads. The model is trained by minimizing reconstruction and consistency losses between the generated and original sequences (see Methods). **d.** During inference, we sample a Gaussian noise vector $z_T \sim \mathcal{N}(0, \mathbf{I})$ and iteratively refine it using the decoder, guided by encoder-extracted features. The denoised embedding z_0 is then decoded into the sequence of activity events. Travel mode icons adapted from Flaticon.com.

event embeddings. Finally, linear decoder heads extract the respective activity attributes from the predicted embeddings. Together, these attributes form the final output, namely a new, simulated sequence of activity events for an individual, sampled in a way that respects observed transition and combination patterns across attributes.

2.2. Problem definition

Formally, we focus on conditional mobility generation, aiming to simulate a sequence of future activity events $\mathbf{y} = [y_1, y_2, \dots, y_{k_1}]$ based on a preceding sequence of observed events $\mathbf{x} = [x_1, x_2, \dots, x_{k_2}]$ by the same individual. Each activity event x_i or y_i is represented as a tuple (l, t, d, m) , where l denotes the location identifier, t the start time, d the duration, and m the travel mode. To incorporate spatial context, each event in the observed sequence $\mathbf{x}$ is further associated with geographic coordinates g and land-use context c describing the surrounding environment.

2.3. Diffusion-based modeling of discrete activity sequences

Diffusion models, grounded in non-equilibrium thermodynamics (Sohl-Dickstein et al., 2015), are a class of likelihood-based deep generative models that excel at learning true data distributions and generating new data samples. These models have overcome several limitations of earlier approaches, with recent advancements in reparameterization (Ho et al., 2020) and inference efficiency (Lu et al., 2022) leading to breakthroughs in handling continuous data distributions, such as those in image and audio generation (Yang et al., 2023). This progress has motivated researchers to extend the framework to discrete data, such as text, resulting in the development of discrete diffusion models for categorical distributions (Austin et al., 2021) and embedding diffusion models that incorporate additional discrete-to-continuous mapping steps (Li et al., 2022; Gong et al., 2023). Here, we build on the conditional embedding DDPM by Li et al. (2022) and generalize the framework to accommodate multifaceted mobility behavior. This formulation suits mobility event sequences because it denoises the full target sequence jointly. Joint denoising lets the model capture global sequence coherence, where location, start time, duration, and mode are mutually constrained across a day or multiple days. It also mitigates the train-inference mismatch of autoregressive rollouts, where reliance on ground-truth histories during training and on the model's own predictions at inference can compound errors over long event horizons. Stochastic denoising further provides a natural mechanism for sampling multiple plausible continuations from the same observed history.

The employed diffusion model introduces two Markov chains: a forward chain that progressively destructs data by injecting noise and a reverse chain that reconstructs the original data from noised samples. Formally, given a continuous data point z_0 sampled from a real-world data distribution $q(\mathbf{z})$, i.e., $z_0 \sim q(\mathbf{z})$, the forward Markov chain gradually corrupts z_0 into a standard Gaussian noise

$\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$ in $\tau \in [1, 2, \dots, T]$ steps:

$$q(\mathbf{z}_1, \dots, \mathbf{z}_T | \mathbf{z}_0) := \prod_{\tau=1}^T q(\mathbf{z}_\tau | \mathbf{z}_{\tau-1}) \quad (1)$$

where $\mathbf{z}_1, \dots, \mathbf{z}_T$ are the chain of noisy latent variables. For each forward step, the perturbation is controlled by adding Gaussian noise with a predefined step-dependent mean $\boldsymbol{\mu} = \sqrt{1 - \beta_\tau} \mathbf{z}_{\tau-1}$, and covariance $\boldsymbol{\Sigma} = \beta_\tau \mathbf{I}$, with $\beta_\tau \in (0, 1)$:

$$q(\mathbf{z}_\tau | \mathbf{z}_{\tau-1}) := \mathcal{N}(\mathbf{z}_\tau; \sqrt{1 - \beta_\tau} \mathbf{z}_{\tau-1}, \beta_\tau \mathbf{I}) \quad (2)$$

Given the forward chain, the reverse denoising process aims to obtain $p_\theta(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau)$, with which we can gradually reconstruct the original data $\mathbf{z}_0$ by sampling from $\mathbf{z}_T$ and applying the reverse Markov chain:

$$p_\theta(\mathbf{z}_0, \dots, \mathbf{z}_T) := p(\mathbf{z}_T) \prod_{\tau=1}^T p_\theta(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau) \quad (3)$$

where $p(\mathbf{z}_T) = \mathcal{N}(\mathbf{z}_T; 0, \mathbf{I})$. As the variance β_τ is, by definition, small, the reverse process follows the same functional form as the forward process, i.e., a Gaussian distribution (Sohl-Dickstein et al., 2015):

$$p_\theta(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau) := \mathcal{N}(\mathbf{z}_{\tau-1}; \boldsymbol{\mu}_\theta(\mathbf{z}_\tau, \tau), \boldsymbol{\Sigma}_\theta(\mathbf{z}_\tau, \tau)) \quad (4)$$

where the mean $\boldsymbol{\mu}_\theta(\mathbf{z}_\tau, \tau)$ and variance $\boldsymbol{\Sigma}_\theta(\mathbf{z}_\tau, \tau)$ are parameterized by deep neural networks with learnable parameters θ . Training is performed by maximizing the marginal likelihood of the data $\mathbb{E}[\log p_\theta(\mathbf{z}_0)]$, and the canonical objective is the variational lower bound $\mathcal{L}_{vlb}$ of $\log p_\theta(\mathbf{z}_0)$ (for a detailed derivation of the objective, see Sohl-Dickstein et al., 2015):

$$\mathcal{L}_{vlb} = \mathbb{E}_{q(\mathbf{z}_1, \dots, \mathbf{z}_T | \mathbf{z}_0)} \left[D_{\text{KL}}(q(\mathbf{z}_T | \mathbf{z}_0) \| p(\mathbf{z}_T)) + \sum_{\tau=1}^T D_{\text{KL}}(q(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau, \mathbf{z}_0) \| p_\theta(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau)) - \log p_\theta(\mathbf{z}_0 | \mathbf{z}_1) \right] \quad (5)$$

where $D_{\text{KL}}(\cdot \| \cdot)$ denotes the Kullback-Leibler (KL) divergence, and $q(\mathbf{z}_{\tau-1} | \mathbf{z}_\tau, \mathbf{z}_0)$ is the forward process posterior, which follows a Gaussian distribution with mean $\hat{\boldsymbol{\mu}}(\mathbf{z}_\tau, \mathbf{z}_0)$ and variance $\hat{\boldsymbol{\beta}} \mathbf{I}$ that can be calculated from $\mathbf{z}_0$, $\mathbf{z}_\tau$ and β_τ . The objective can be further simplified by reweighting each KL divergence term to obtain a mean-squared error reconstruction loss:

$$\mathcal{L}_{reconstruct} = \mathbb{E}_{\mathbf{z}_0, \mathbf{z}_\tau, \tau} \|\hat{\mathbf{z}}_\theta(\mathbf{z}_\tau, \tau) - \mathbf{z}_0\|^2 \quad (6)$$

Here, $\hat{\mathbf{z}}_\theta(\mathbf{z}_\tau, \tau)$ represents the prediction of the original data given $\mathbf{z}_\tau$ and τ . Intuitively, this parameterization can be interpreted as training a neural network to predict the original ground truth data from an arbitrarily corrupted version.

To apply continuous diffusion to discrete data, embedding diffusion models introduce an additional embedding process in the forward process and a rounding step in the reverse process. The embedding process transforms discrete tokens into a continuous latent space via a learnable embedding function, enabling the conventional diffusion process to be applied. Given the target sequence $\mathbf{y}$, the input $\mathbf{z}_0$ is obtained in the embedding space as $\mathbf{z}_0 \sim \mathcal{N}(e_\phi(\mathbf{y}), \beta_0 \mathbf{I})$, where $e_\phi(\cdot)$ represents the embedding function parameterized by ϕ . The rounding step translates the diffusion prediction in the embedding space back to discrete tokens: $p_\phi(\mathbf{y} | \mathbf{z}_0)$, whose parameter ϕ is shared with the embedding function. As a result, the neural network is trained end-to-end, optimizing both the reconstruction loss from the diffusion process and the consistency loss from the rounding step:

$$\mathcal{L} = \mathcal{L}_{reconstruct} + \underbrace{\mathbb{E}_{\mathbf{y}, \mathbf{z}_0} [-\log p_\phi(\mathbf{y} | \mathbf{z}_0)]}_{\mathcal{L}_{consistent}} \quad (7)$$

2.4. Neural network architecture

MobilityGen's neural network comprises several components: an embedding module that maps activity attributes into an embedding space, an encoder that extracts information from traveled traces, a decoder that reconstructs target sequences from their corrupted versions, and task-specific linear heads that ensure consistency mapping.

The embedding process converts raw activity attributes of various formats into latent space representations. For categorical attributes with discrete values, such as the location identifier l and travel mode m , we introduce lookup matrices $\mathcal{E}_l$ and $\mathcal{E}_m$ that map attributes to the embedding space, $\mathcal{E}_l: \mathcal{V}_l \rightarrow \mathbb{R}^{d_{emb}}$ and $\mathcal{E}_m: \mathcal{V}_m \rightarrow \mathbb{R}^{d_{emb}}$, where $\mathcal{V}$ represents the set of possible values and d_{emb} denotes the number of hidden dimensions. Intuitively, each row in $\mathcal{E}$ corresponds to the representation of a specific value. For continuous attributes, such as the start time t and duration d , we use small feed-forward networks f_t and f_d to map the value into the hidden space, $f_t: \mathbb{R}^1 \rightarrow \mathbb{R}^{d_{emb}}$ and $f_d: \mathbb{R}^1 \rightarrow \mathbb{R}^{d_{emb}}$. The representation for an event record y_i in the target sequence is obtained by linearly combining the above-introduced transformations:

$$e_\phi(y_i) = \mathcal{E}_l(l_i) + \mathcal{E}_m(m_i) + f_t(t_i) + f_d(d_i) \quad (8)$$

For each record x_k in the traveled sequence, we also incorporate the geometry g_k and surrounding land-use functions c_k . The geometry is modeled using the Space2vec method (Mai et al., 2020), which transforms raw coordinates in a projected coordinate system into vector representations via sine and cosine base functions. Land-use functions are captured using the location-POI embedding method proposed by Hong et al. (2023b), which considers POIs from various categories within a location and applies Latent Dirichlet

Allocation (LDA) to derive their functional descriptions. The relationships between these movement-related contexts and the activity embeddings are learned through feed-forward networks f and linearly combined via a residual connection:

$$\mathbf{e}_\phi(\mathbf{x}_k) = \mathbf{h}(\mathbf{x}_k) + f_{coord}(s2v(g_k) \oplus \mathbf{h}(\mathbf{x}_k)) + f_{poi}(lda(c_k) \oplus \mathbf{h}(\mathbf{x}_k)) \quad (9)$$

$$\text{with } \mathbf{h}(\mathbf{x}_k) = \mathcal{E}_l(l_k) + \mathcal{E}_m(m_k) + f_t(t_k) + f_d(d_k) \quad (10)$$

where $\oplus$ denotes the concatenation operation. The final traveled sequence representation $\mathbf{e}_\phi(\mathbf{x})$ and target sequence representation $\mathbf{e}_\phi(\mathbf{y})$ consist of concatenated activity records in their sequential order. The parameters involved in the embedding process are optimized during network training.

We utilize the transformer encoder-decoder architecture proposed by Vaswani et al. (2017) as the network backbone. The encoder processes the traveled sequence representation through L encoder blocks, each with an identical architecture consisting of a multi-head self-attention layer and a feed-forward layer with two linear operations separated by a non-linear activation function. Residual connections (He et al., 2016), layer normalizations (Ba et al., 2016), and dropout (Hinton et al., 2012) are applied to each layer. The decoder, which processes the corrupted target sequence representation, contains L decoder blocks structured similarly to the encoder. In addition to the two layers in each encoder block, the decoder incorporates an intermediate layer that performs multi-head attention over the encoder stack's output.

The multi-head attention layer applies H scaled dot-product attention operations in parallel. Input representations have model dimension d_{model} . Within each head, queries and keys have dimension d_k and values have dimension d_v , with $d_k = d_v = d_{model}/H$. Following Vaswani et al. (2017), scaled dot-product attention within each head is computed after the head-specific projections as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \cdot \mathbf{V} \quad (11)$$

Multi-head attention linearly projects the input matrices $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ from d_{model} into H head-specific subspaces, applies attention within each head, and concatenates the results:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = (\text{head}_1 \oplus \dots \oplus \text{head}_H) \cdot \mathbf{W}^O \quad (12)$$

$$\text{with } \text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (13)$$

where $\mathbf{W}_i^Q, \mathbf{W}_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_{model} \times d_v}$, and $\mathbf{W}^O \in \mathbb{R}^{H d_v \times d_{model}}$ are learned during training. In the decoder's multi-head attention layers, the queries are derived from the previous decoder block, while the keys and values come from the encoder stack's output. In the self-attention layers of both the encoder and decoder, the keys, values, and queries are all set to the output of the previous block. Additionally, causal masks are applied in the decoder's self-attention operations to prevent access to information from "future" sequence positions, i.e., the event at position k can only attend to information from position k and earlier. Ultimately, the encoder extracts the feature representation x from the traveled mobility sequence, which is then used in the decoder to reconstruct the target trace representation $\hat{\mathbf{z}}_\theta(\mathbf{z}_\tau, \tau, x)$.

Since we aim to generate new multifaceted activity events, the neural network must include multiple task-specific heads in the rounding step to "translate" from the embedding space into respective activity attributes. We introduce linear feed-forward layers $g_l: \mathbb{R}^{d_{emb}} \rightarrow \mathbb{R}^{|\mathcal{V}_l|}$, $g_m: \mathbb{R}^{d_{emb}} \rightarrow \mathbb{R}^{|\mathcal{V}_m|}$, $g_t: \mathbb{R}^{d_{emb}} \rightarrow \mathbb{R}^1$, and $g_d: \mathbb{R}^{d_{emb}} \rightarrow \mathbb{R}^1$ for directly mapping to location identifier, travel mode, start time, and duration, respectively.

The parameters of g_l and g_m are tied to the corresponding embedding modules, and their outputs are processed with a softmax operation to ensure a valid probability distribution over the set of possible values:

$$P(\hat{\mathbf{l}}) = \text{softmax}(g_l(\mathbf{z}_0)) \quad P(\hat{\mathbf{m}}) = \text{softmax}(g_m(\mathbf{z}_0)) \quad (14)$$

where $P(\hat{\mathbf{l}}) \in \mathbb{R}^{k_1 \times |\mathcal{V}_l|}$ and $P(\hat{\mathbf{m}}) \in \mathbb{R}^{k_1 \times |\mathcal{V}_m|}$ are matrices containing the predicted probabilities for all locations and travel modes, respectively. The rounding errors are measured as the average of multi-class cross-entropy loss across the sequence:

$$\mathcal{L}_l = -\frac{1}{k_1} \sum_{i=1}^{k_1} \sum_{j=1}^{|\mathcal{V}_l|} P(\mathbf{l})_{[i,j]} \log(P(\hat{\mathbf{l}})_{[i,j]}) \quad \mathcal{L}_m = -\frac{1}{k_1} \sum_{i=1}^{k_1} \sum_{j=1}^{|\mathcal{V}_m|} P(\mathbf{m})_{[i,j]} \log(P(\hat{\mathbf{m}})_{[i,j]}) \quad (15)$$

where $P(\cdot)_{[i,j]}$ extracts the value at row i and column j , and $P(\mathbf{l})$ and $P(\mathbf{m})$ are the one-hot represented ground truth, e.g., $P(\mathbf{l})_{[i,j]} = 1$ if the chosen location at step i is the j^{th} location, and $P(\mathbf{l})_{[i,j]} = 0$ otherwise.

The rounding losses for start time and duration are measured directly as the mean squared error after the linear heads:

$$\mathcal{L}_t = \|g_t(\mathbf{z}_0) - \mathbf{t}\|^2 \quad \mathcal{L}_d = \|g_d(\mathbf{z}_0) - \mathbf{d}\|^2 \quad (16)$$

Finally, we obtain the consistency loss $\mathcal{L}_{consistent}$ by combining the task-specific rounding losses:

$$\mathcal{L}_{consistent} = \alpha_1 \mathcal{L}_l + \alpha_2 \mathcal{L}_m + \alpha_3 \mathcal{L}_t + \alpha_4 \mathcal{L}_d \quad (17)$$

where the α values balance the contributions of the four attributes. The reconstruction and consistency losses play complementary roles. The reconstruction loss in Eq. (6) is applied to the joint event embedding $\mathbf{z}_0$, which combines location, start time, duration, and mode information, and therefore supervises the full latent event representation directly. The consistency losses in Eq. (17) then act as attribute-level constraints, mapping the reconstructed embedding back to observed activity attributes.

3. Experiment

3.1. Data sources and preprocessing

MobilityGen was trained on a large-scale GNSS-based mobility dataset comprising high-resolution individual trajectories collected across Switzerland. The dataset contains billions of GNSS tracking records from approximately 3700 individuals between 2019 and 2022, and was produced as part of the *Mobility Behaviour in Switzerland* (MOBIS) project, which investigates the behavioral impacts of mobility pricing (Molloy et al., 2023). Starting in autumn 2019, participants provided socio-demographic and travel-related information via a survey and recorded their movements over an eight-week period using a smartphone tracking app. The app collected GNSS samples triggered by motion events detected through onboard sensors (e.g., accelerometer). Following the main study, participants were invited to join the MOBIS-COVID extension, which assessed the impact of the COVID-19 pandemic on travel behavior. Roughly one-third of the original participants continued recording, with 730 individuals contributing data through the end of 2022. For further details on the study design and findings, see Molloy et al. (2023).

We selected 2113 individuals who had been tracked for at least 50 days and whose locations were known at least 60% of the time. This resulted in a dataset with an average tracking duration of 138 days and an average temporal coverage of 88% across individuals. The dataset provides substantial longitudinal depth and multi-attribute annotation across diverse Swiss contexts, supporting learning of the coupled spatio-temporal and modal structure of daily mobility.

Human mobility diaries and activity attributes were derived from raw GNSS tracking records using the standard movement data model (Axhausen, 2007). The tracking app analyzed motion metrics such as speed and acceleration from built-in smartphone sensors to segment GNSS traces into movement *stages* and stationary *staypoints*. Travel mode labels were inferred for each stage from the same motion data. During the tracking study, participants were encouraged to verify or correct the imputed mode labels and to annotate staypoints with their corresponding activity purposes. We identified *activity staypoints* as those where meaningful activities (e.g., home or restaurants) occurred, defined as stays longer than 25 minutes or carrying an activity label other than *waiting*. These staypoints were spatially aggregated into *locations*, representing repeated visits to the same place, using the Trackintel framework (Martin et al., 2023) with parameters $\epsilon = 20$ m and $num_samples = 1$. The resulting locations were projected onto Google's S2Geometry grid system, a hierarchical tessellation of the Earth's surface. We used a progressive gridding strategy from level 10 (coarse) to level 14 (fine, with a spatial resolution of approximately 500×500 m²), selecting a finer level when more than two locations were projected onto the same cell. This process yielded 142,575 location grid cells covering Switzerland, of which 28,741 were actively visited during the study period.

We define a *trip* as the sequence of consecutive movement stages between two activity staypoints and assign its primary travel mode as the one used for the longest stage (one of car, train, bus, tram, bicycle, walk, or other). In activity-based mobility analysis, travel is understood as a means to perform activities at different locations, making the activity itself the central unit of analysis (Axhausen and Gärling, 1992). Accordingly, we define each activity event with the following attributes: *start time* as the departure time of the preceding trip, *duration* as the sum of the trip and stay durations, and *travel mode* as the primary mode used to reach the activity location. The preprocessed dataset consists of 1,079,918 chronologically ordered activity events, each associated with a start time, duration, and travel mode, and spatially projected onto 28,741 unique locations.

For acquiring the built environment information, we collected the POI dataset from OpenStreetMap (OSM) (<http://www.openstreetmap.org>), an open-source project that provides free and easily accessible digital map resources. Historical POI data for Switzerland were acquired from the end of 2022 to align with the time frame of the tracking dataset. The POI dataset was supplemented with OSM building data from the same year. For each building entry, we retained its attribute values but created a new POI by using the center point of its polygon geometry as its spatial representation. After removing redundant and insignificant POIs, the final dataset consisted of 3,104,590 POIs, categorized into 19 first-level and 188 second-level groups. The second-level categories, representing functional type descriptions, were used to obtain urban functions for locations.

We divided the activity event sequences into non-overlapping training, validation, and test sets in a 7:2:1 ratio based on time, with the first 70% of each individual's tracking days designated as training data, the middle 20% as validation data, and the last 10% as test data. We constructed conditional sequence pairs by using mobility over the past three weeks as the traveled sequence and the subsequent activity events as the target sequence, requiring each target to span at least one calendar week. Because individuals differ in activity frequency and observation coverage, the resulting sequences have variable lengths, denoted by k_2 and k_1 for the traveled and target sequences, respectively. This process yielded 590,257, 104,010, and 36,285 traveled-target sequence pairs for training, validation, and testing, respectively. The model is trained on these variable-length sequence pairs using sequence masking, allowing the learning process to reflect differences in individual mobility intensity. Because this temporal split shares individuals between training and test, we additionally evaluate generalization to users entirely absent from training in Appendix D.2.

3.2. Training procedure and computational cost

We optimize the network parameters with Eq. (7) on the training set using the AdamW variant of stochastic gradient descent (Loshchilov and Hutter, 2019), employing a weight decay of 0.01 and an adaptive learning rate that starts at $4e^{-4}$ and is progressively reduced to $2e^{-5}$ following a cosine schedule. Training is conducted over 250,000 steps, with a batch size of 64 traveled-target sequence pairs per step (approximately 32,000 tokens). A linear learning rate warm-up (Vaswani et al., 2017) is applied for the first 5000 steps. To stabilize training, we first optimize the network for 100,000 steps using only the $\mathcal{L}_l$ loss, setting $\alpha_1 = 1$, and $\alpha_2 = \alpha_3 = \alpha_4 = 0$. For the remaining steps, we apply the efficient multiple-gradient descent algorithm for encoder-decoder architec-

tures (Sener and Koltun, 2018) to dynamically adjust the α values, ensuring the parameter optimization process leads to a Pareto optimal solution, where no task can be improved without compromising at least one of the others. We use the *sqr*t noise schedule (Li et al., 2022) to determine the injected noise β_τ at each forward diffusion step and apply loss-weighted sampling to select the noise steps τ during training (Gong et al., 2023). The transformer encoder and decoder each contain $L = 6$ blocks and $H = 8$ heads per attention layer. We set the forward diffusion steps to $T = 2000$, latent embedding size to $d_{emb} = 128$, and the transformer model representation size to $d_{model} = 512$. These hyperparameters were determined through a random search, selecting the best-performing set that minimized the network loss on the validation set.

The complete training process takes ≈ 82 hours using distributed data parallelism across four NVIDIA RTX 4090 GPUs with FP16 half-precision. During inference, the generation length k_1 is specified explicitly; for the main evaluation we set $k_1 = 50$ events, covering 12.7 ± 4.0 calendar days on average. The empirical target sequence is capped to the same length, ensuring that generated and observed sequences are compared over the same number of future activity events. This 50-event horizon is standardized across individuals and applied to all baseline sequence comparisons. Inference uses 200 down-sampled reverse steps, and generating a new target sequence takes 50.4 ± 4.1 milliseconds on a single NVIDIA RTX 3090 GPU. Appendix D.1 reports sensitivity analyses for varying the input traveled sequence length k_2 and the generated target length k_1 .

3.3. Evaluation metrics

We use metrics that summarize high-level mobility patterns at collective and individual levels to quantify the discrepancy between ground truth and generated activity sequences. We consider the following indicators:

- Location visitation frequency (González et al., 2008; Song et al., 2010a). Individuals' preferences in location choices are highly unbalanced and can be quantified by the relationship between the visitation frequency of locations f_k and their rank k , which is reported to follow Zipf's law, i.e., $f_k \sim k^{-\xi}$, with $\xi \sim 1.2$.
- Evolution of radius of gyration (González et al., 2008; Alessandretti et al., 2020). The radius of gyration r_u measures the characteristic distance traveled by individual u and is used to reflect mobility intensity in the spatial dimension. Its evolution after k displacements $r_u(k)$ is defined as:

$$r_u(k) = \sqrt{\frac{1}{k} \sum_{i=1}^k (l^{(i)} - l_u^{cm})^2} \quad (18)$$

where $l^{(i)}$ represents the position of the i^{th} visited location, and $l_u^{cm} = \frac{1}{k} \sum_{j=1}^k l^{(j)}$ is the center of mass of the location sequence after k displacements.

- Temporal mobility entropy (Song et al., 2010b). Closely related to the theoretical predictability of location sequences, temporal mobility entropy S_{temp} captures the full spatio-temporal order of location visits by measuring the probability $P(T'_u)$ of observing ordered subsequences T'_u within the location sequence T_u :

$$S_{temp} = - \sum_{T'_u \in T_u} P(T'_u) \log_2(P(T'_u)) \quad (19)$$

We use the method described in Song et al. (2010b) to estimate the temporal mobility entropy S_{temp} for individuals.

- Mobility motifs (Schneider et al., 2013). Mobility motifs refer to recurring network patterns that emerge when daily location visits are abstracted as networks, where nodes represent activity locations and links denote transitions between them. We segmented the activity events by day and identified motif structures as those occurring in more than 0.5% of all days. The construction of these motifs incorporates both spatial and temporal dimensions of activity participation. Mobility regularity is quantified by the presence and frequency of specific motifs in individuals' schedules (Jiang et al., 2016).
- Evolution of trip package choices (Hong et al., 2023a). The trip package concept was introduced to encapsulate the interactions within multifaceted activity-travel patterns. For each individual, a trip package is constructed when observing the usage of a travel mode to reach an activity location, i.e., a unique mode-location choice. Individuals continuously explore new activity-behavior options, and the number of unique trip packages observed over time quantifies the speed of this exploratory behavior.

4. Results

4.1. Microscopic modeling reproduces location choice behavior

To test whether MobilityGen preserves the core spatial regularities of daily mobility, we begin by evaluating its ability to capture location patterns, which constitute a basic validity check of the generation capability. How individuals select spatial locations for different activities shapes their interactions with physical space and is a fundamental driver of many urbanization-induced challenges (Moro et al., 2021; Nilforoshan et al., 2023; Liu et al., 2023). Here, we focus on the simulated visits to locations and compare them with the real visits from observed sequences.

We first assess the ability of MobilityGen to capture fundamental properties of location choices and compare its performance against two state-of-the-art microscopic mobility models, namely, EPR (Song et al., 2010a) and Container (Alessandretti et al., 2020) (Fig. 2). We begin with the average rank distribution of location visits (Fig. 2a), a widely used measure of preference for activity

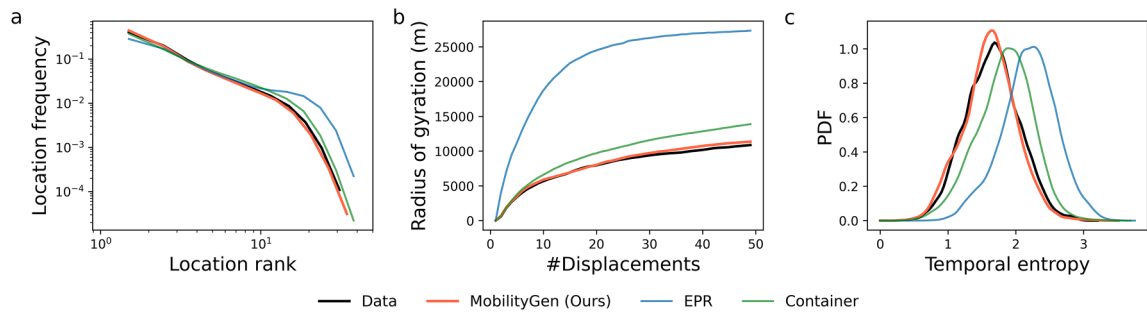


Fig. 2. Evaluating microscopic mobility models with location metrics. **a.** Rank-frequency distribution of visited locations. **b.** Median radius of gyration as a function of the number of displacements. **c.** Distribution of temporal mobility entropy across individuals. All metrics are compared between real data (black), MobilityGen (red), EPR (blue), and Container (green). MobilityGen aligns most closely with empirical distributions across all metrics.

locations (González et al., 2008; Barbosa et al., 2018). Our results indicate that MobilityGen provides a significantly better fit to real data than both the EPR and Container models, as confirmed by the likelihood ratio test (Clauset et al., 2009) (with $P < 10^{-3}$) and the Wasserstein distances (Panaretos and Zemel, 2019) between the rank distributions (see Appendix A.1 and Table A.1). The radius of gyration, which measures the characteristic distance an individual covers (González et al., 2008), serves to quantify the movement intensity in space. We observe that the median evolution of the radius of gyration across individuals, $r_u(k)$, over displacement steps k , is well described by logarithmic growth for both real and generated location sequences (Fig. 2b), consistent with previous findings (Alessandretti et al., 2020). However, when fitting $r_u(k) \propto \alpha \log(k)$, only the value α derived from MobilityGen agrees well with the real data (see Appendix A.1). Finally, we evaluate the temporal mobility entropy that quantifies the predictability of location sequences, considering both the frequency and the order of visits (Song et al., 2010b). Using the likelihood ratio test and the Wasserstein metric, we further confirm that, compared against the EPR and Container models, MobilityGen most accurately approximates the true entropy distribution over the population (with $P < 10^{-3}$, Fig. 2c and see Appendix A.1 and Table A.1). Additionally, we evaluated a representative set of state-of-the-art data-driven mobility models, including deep learning-based sequence prediction models and end-to-end generative models, and further location metrics (see Appendix A.1). Our findings indicate that MobilityGen consistently provides the best fit to real-world data. Evaluations with metrics from language modeling confirm that MobilityGen can simulate location visits that closely resemble the real sequence while avoiding regurgitation of input sequences and maintaining sufficient diversity (see Appendix A.2). Finally, a held-out-individual analysis examines whether these location patterns generalize to users absent from training: visitation-frequency and uncorrelated-entropy patterns remain close to the seen-user reference under high source-location coverage, whereas measures of spatial extent show larger seen–unseen differences and deteriorate further when source-location coverage is lower (Appendix D.2).

To further evaluate the generated mobility patterns, we examine the spatial distribution of location visits. As an illustrative example, we map the observed and synthetic trajectories of a sample individual (Fig. 3a and b). The comparison shows that MobilityGen effectively captures both visit and transition patterns at frequently visited locations (thick lines), while also generating occasional visits to less familiar locations (thin lines). Accurately modeling such exploratory behavior remains a central challenge in human mobility research (Cuttone et al., 2018). To assess how well MobilityGen addresses this challenge at the population level, we compare the aggregated spatial distribution of location visits across Switzerland in the real and simulated data (Fig. 3c and d). Visit frequencies are color-coded from low (yellow) to high (red). MobilityGen closely matches the concentration of frequently visited locations and also generates visits to less frequented areas, reflecting a balance between individual-level exploration and population-level spatial preference. Using the entropy $E = -\sum_k p(f_k) \log p(f_k)$ to quantify the dispersion of location visits f_k , we obtain 0.33 for the real data and 0.53 for the simulated data, indicating greater spatial dispersion in the latter. To analyze spatial dynamics at a finer resolution, we examine three representative cities: Bern, Zurich, and Lucerne (Fig. 3e–j), each characterized by distinct urban functions and mobility patterns. MobilityGen successfully reproduces key urban visitation structures in all three cases. However, spatial dispersion remains higher in the simulated data, with entropy values of 2.49 vs. 1.89 for Bern, 3.99 vs. 3.37 for Zurich, and 2.71 vs. 2.11 for Lucerne. Despite this discrepancy, MobilityGen compares favorably to other microscopic mobility models in replicating the location visit distributions (see Appendix A.1, Table A.2). Complementary visual comparisons are provided at both national and urban levels in Appendix A.1, Figs. A.9 and A.10.

4.2. Structured behavioral patterns across time, mode, and location

To examine the model beyond aggregate location patterns, we evaluate whether MobilityGen captures structured relationships among activity timing, travel mode, and location within unified activity events. We assess these patterns for each sequence in the test set and compare cross-attribute regularities and interactions (Fig. 4).

Incorporating temporal information allows MobilityGen to reason about the scheduling of each activity. This temporal dimension is evident in the activity durations (Fig. 4a) and the number of daily visited locations (Fig. 4b), both of which closely match those observed in the real data. To further examine patterns of activity participation, we construct mobility motifs (Schneider et al., 2013),

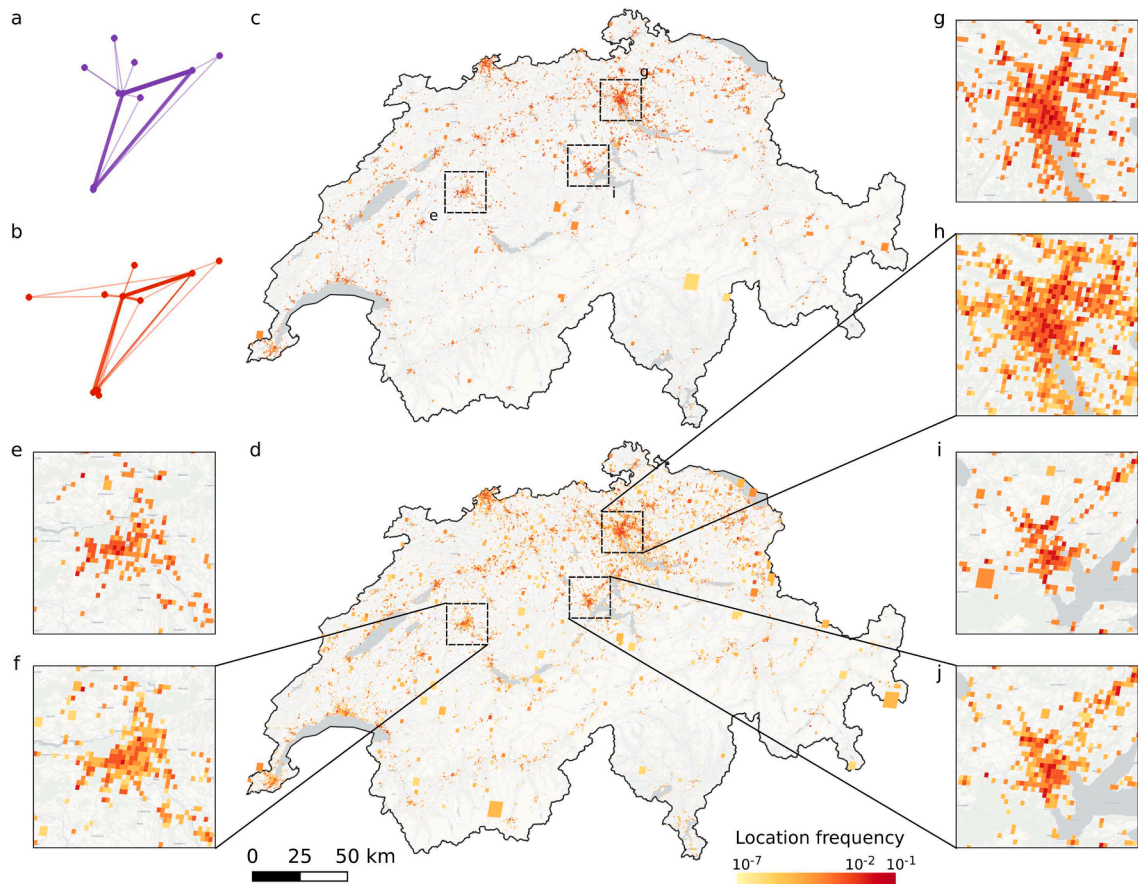


Fig. 3. MobilityGen generates realistic spatial patterns of location visits across individual, urban, and national levels. We compare the spatial distribution of activity locations in real data (a, c, e, g, and i) with those generated by MobilityGen (b, d, f, h, and j). a and b show visits and transitions for a sample individual, where edge width is proportional to transition frequency. The generated sequence reflects routinely visited locations while also capturing exploratory behavior. c and d display aggregated visit frequencies across all individuals in Switzerland, with warmer colors indicating higher visitation intensity on a log scale. Panels e–j provide zoomed-in views for Bern (e, f), Zurich (g, h), and Lucerne (i, j). The generated sequences closely align with observed spatial patterns while exhibiting slightly greater diversity in the locations visited.

which serve as high-level descriptors of regularities in activity location and scheduling. MobilityGen accurately recovers the relative frequencies of both common and rare motifs, including simple two-location transitions and more complex patterns involving up to five locations (Fig. 4c). We further benchmark MobilityGen against two state-of-the-art spatio-temporal mobility simulators, the Diary-based Trajectory Simulator (DITRAS) (Pappalardo and Simini, 2018) and TimeGeo (Jiang et al., 2016), using additional mobility metrics (see Appendix B.1). Across all measures, MobilityGen more faithfully reproduces the interplay between activity timing and location choice, yielding the closest alignment with empirical behavior. The held-out-individual analysis (Appendix D.2) further shows that this advantage persists for individuals absent from training.

Next, we incorporate user-verified travel mode labels into MobilityGen to jointly model mode choice in relation to activity events. Fig. 4d compares the travel mode distributions from the observed and simulated activity sequences, with discrepancies quantified by the Wasserstein distance (see Appendix B.1, Table B.4). Across all scenarios, MobilityGen reconstructs the mode preferences exhibited in real-world behavior. Car and walking, the two most common modes, show excellent agreement between empirical and simulated shares. Notably, MobilityGen captures subtler distinctions among public modes, including bus, train, and tram, accurately reflecting their relative frequencies. We further assess mode-specific behavior by comparing the displacement distributions for car and walking trips (Fig. 4e and f). For car travel, the simulated displacements closely align with the observed distribution (Kolmogorov-Smirnov statistic $KS = 0.02$, with $P < 10^{-3}$). For walking, MobilityGen produces slightly longer displacements ($KS = 0.06$, $P < 10^{-3}$), potentially due to short trips being disproportionately affected by spatial discretization. This effect could be mitigated in future work by adopting a finer location grid, although such refinement may increase computational complexity.

Finally, we assess the evolution of trip packages, defined as combinations of travel modes and activity locations (Hong et al., 2023a), to evaluate MobilityGen's ability to model complex attribute interactions (Fig. 4g). The average number of unique trip packages over time, denoted as $Tp(t)$, follows a sub-linear growth over time t , consistent with previous findings (Hong et al., 2023a). The scaling exponent α in the relation $Tp(t) \propto t^\alpha$ is nearly identical between simulated ($\alpha = 0.85 \pm 0.01$) and empirical sequences

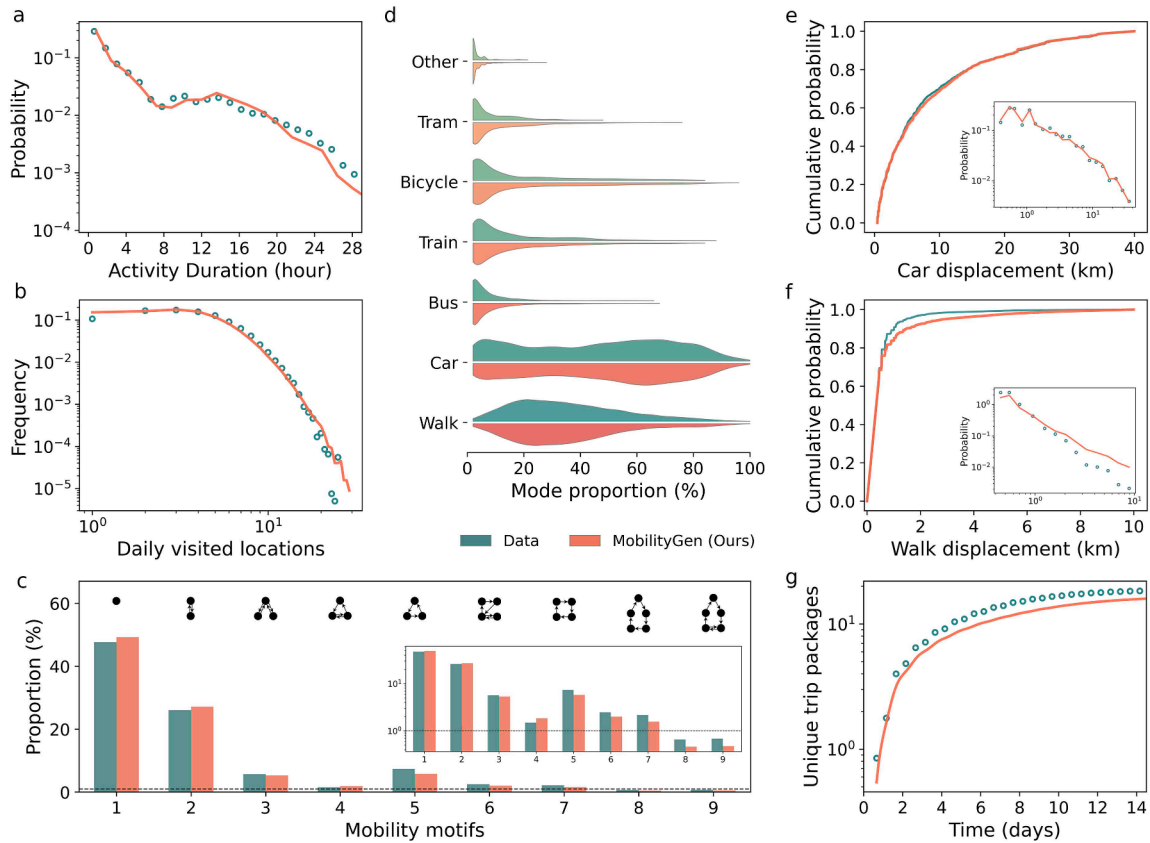


Fig. 4. Generated activity events replicate individual activity-travel patterns across multiple dimensions. **a.** Distribution of activity durations. **b.** Distribution of daily visited locations. **c.** Distribution of mobility motifs, with motif diagrams shown above. The dashed line marks the 1% threshold as a visual guide, and the inset presents the same distribution on a log scale. **d.** Violin plots of modal shares across individuals. **e.** Cumulative distribution of car displacements, with the log-log distribution shown in the inset. **f.** Cumulative distribution of walk displacements, with the log-log distribution shown in the inset. **g.** Evolution of unique trip packages over time, averaged across individuals. Across all dimensions, the activity events generated by MobilityGen (red) closely align with empirical patterns from real-world data (green), capturing key behavioral regularities and attribute interactions.

($\alpha = 0.85 \pm 0.01$), indicating that MobilityGen accurately captures the evolving diversity of travel mode and activity location choices over time. Further ablation studies confirm that the full multi-attribute model achieves a more realistic representation of attribute interactions than simplified variants with fewer modeled attributes (see [Appendix C.2](#)).

To understand how MobilityGen captures attribute interactions, we analyze the structure of its learned embedding space. We apply densMAP ([Narayan et al., 2021](#)) to project the 128-dimensional embeddings into two dimensions for visualization (see [Appendix B.2](#)). The projections reveal that MobilityGen learns semantically meaningful representations: similar travel modes appear near each other, temporal attributes vary smoothly along continuous trajectories, and frequently visited locations are well-separated from infrequent ones. When multiple attributes are jointly embedded, the space further reveals distinct event clusters tied to specific locations, alongside a more diffuse region that likely reflects less location-dependent activities. Together, these patterns indicate that the embedding space offers a compact representation of multi-attribute mobility events and a natural basis for quantitative analysis of structure across mode, time, and location. Notably, this structure emerges from the generative training objective alone, suggesting that MobilityGen encodes behavioral regularities and not merely the marginal output distributions. By providing this shared representation, MobilityGen enables complex mobility dynamics to be inferred directly from behavioral sequences and facilitates diverse mobility analyses.

4.3. Context enhances generalization to novel locations

A key element of the proposed framework is that contextual information can be integrated when modeling individual travel decisions – an aspect well-supported by the literature but not adequately covered by existing microscopic mobility models ([Pappalardo et al., 2023](#)). In MobilityGen, context features are combined with activity attributes through learnable representations, yielding a context-aware mobility model (see [Fig. 1b](#)). We focus on spatial contexts that characterize the physical environment, namely coordinate geometry, which reveals movement directionality ([Zhao et al., 2024](#)); and POIs, which display urban functions and imply

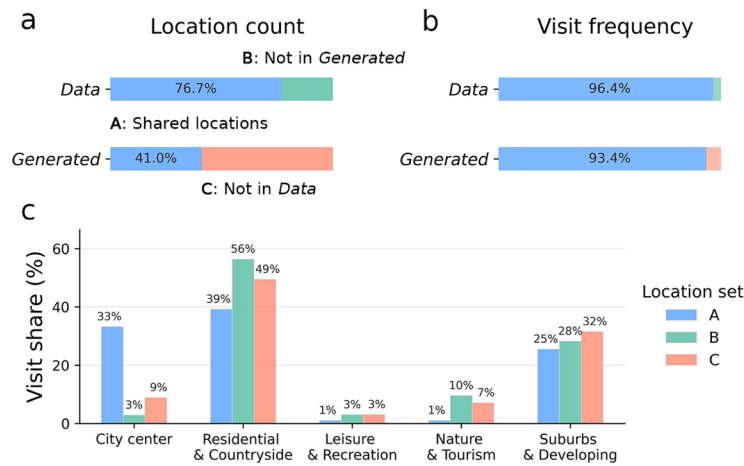


Fig. 5. Point of interest (POI) data supports the identification of locations associated with infrequent activities. **a.** We compare generated sequences with observed data, categorizing locations into those present in both (Set A), those found only in the data (Set B), and those found only in the generated sequences (Set C). **b.** As expected, locations in Set A are visited frequently, whereas those in Sets B and C are rarely visited. **c.** Relative visit frequencies by functional location type for each set. Functional types are inferred from POI distribution (see Methods). Sets B and C exhibit similar visit patterns, in contrast to the distribution observed for Set A.

movement purposes (Hong et al., 2023b). Further contexts could be integrated through appropriate representations; we defer this to future research.

We analyze location patterns to assess the influence of contextual factors on the simulation performance. Generated locations are compared to observed data and categorized into three sets: Set A includes locations present in both sequences, while Sets B and C contain locations visited exclusively in the observed and generated sequences (Fig. 5a). Since MobilityGen replicates key location preferences (see Fig. 3), Set A is expected to capture the locations where individuals conduct most of their activities. Indeed, an analysis of the visit frequencies reveals that 96.4% of the true visits and 93.4% of the generated visits occur at locations in Set A (Fig. 5b). In contrast, locations in Sets B and C receive relatively few visits, reflecting individuals' tendencies toward novelty-seeking and the exploration of unfamiliar locations (Arentze and Timmermans, 2005; Song et al., 2010a; Alessandretti et al., 2018).

To understand how MobilityGen models novelty-seeking behavior, we identify the functional types of locations using POI data. Applying k -means clustering to POI representations of locations yields five distinct location classes, whose functional types are inferred from their spatial distributions and the enrichment of specific POI categories (Yao et al., 2017; Niu and Silva, 2021) (see Appendix C.1). This functional classification enables a detailed analysis of location preferences, revealing that visits in Set A predominantly occur in the *City center*, whereas visits in Sets B and C are largely concentrated in *Residential & Countryside*. Moreover, preferences for *Leisure & Recreation* and *Nature & Tourism* sites are markedly higher in Sets B and C than in Set A. Using pairwise Wasserstein distances to quantify the similarity of visitation patterns across sets, we find that Sets B and C are closely aligned (Wasserstein distance: 0.11), while their distributions differ substantially from that of Set A (Set A vs. Set B: 0.57; Set A vs. Set C: 0.56). In other words, MobilityGen captures key patterns of exploratory behavior, despite the inherent difficulty of characterizing novel locations and predicting which specific places individuals are likely to explore.

A comparison with a model that excludes contextual information confirms that incorporating context improves the ability to characterize locations (see Appendix C.2). This improvement is particularly relevant in settings where historical mobility information is limited, such as when individuals seek novel locations for their activities. To benchmark how effectively MobilityGen leverages contextual cues, we conducted comparative experiments against established mobility flow generation models: Gravity (Zipf, 1946), Deep Gravity (Simini et al., 2021), and random forest (Cabanas-Tirapu et al., 2025). These baselines estimate flows between location pairs using features such as distance, population, and selected facility densities (e.g., shops and transport hubs). Flow-based models differ fundamentally from MobilityGen in both problem formulation and modeling capacity. For instance, they cannot capture temporal dynamics or disaggregated mobility. To ensure a fair comparison, we focused exclusively on locations that were not observed during training. The results show that MobilityGen outperforms the flow generation baselines by a substantial margin across multiple flow metrics (see Appendix C.3 and Table C.9). We attribute this improved performance in novel locations to the integration of spatial features, including coordinate geometry and detailed POI information, as well as the joint modeling of contextual and behavioral factors. While further research is needed to fully understand novelty-seeking mobility, our findings support the hypothesis that comprehensive context modeling is necessary to capture the variability in human behavior.

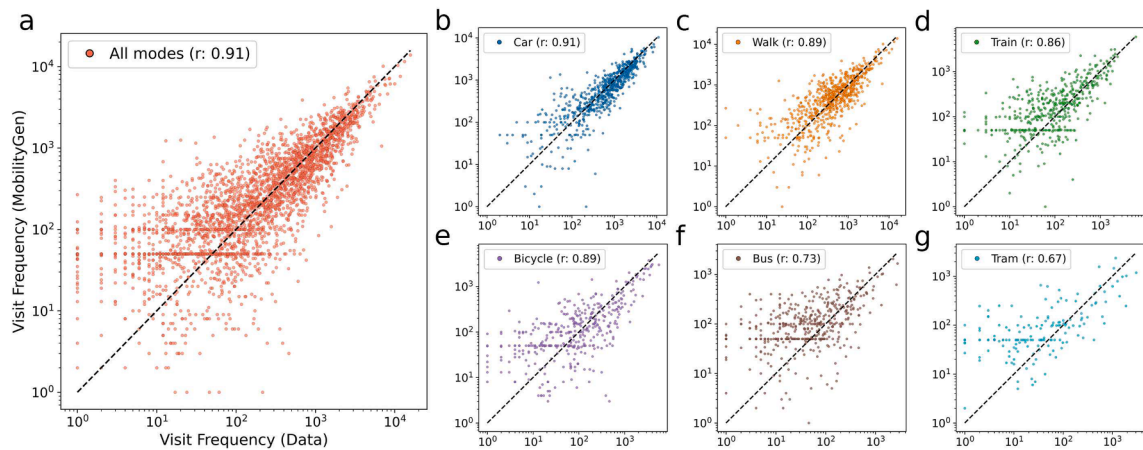


Fig. 6. Reproducing location importance across travel modes. **a.** Visit frequencies for grouped locations in real and simulated activity events across all modes, showing strong overall agreement ($r = 0.91$, $P < 10^{-3}$). **b – g.** Mode-specific comparisons for car (**b**), walk (**c**), train (**d**), bicycle (**e**), bus (**f**), and tram (**g**). Each dot represents a group of locations with the same visitation frequency observed in the data. Diagonal lines denote perfect agreement. Higher agreement is observed for more frequently used modes (e.g., car and walk), while deviations increase for less common modes (e.g., tram).

4.4. Reproducing mode-specific spatial behavior and co-presence dynamics

Beyond standard validation, we ask whether the generated multi-attribute event sequences support analyses that have so far been difficult or infeasible with existing simulators. Our simulations generate realistic daily schedules that jointly capture where people go, how they travel, and when activities occur. Earlier tools often rely on predefined rules or cover only part of this behavioral structure, which has limited their use for studying differences across travel modes or patterns of shared presence in everyday settings. We focus on two representative scenarios with broad relevance to transport planning and urban research: (i) mode-specific spatial visitation, showing how individuals use different travel modes to move through urban space, and (ii) spatio-temporal co-presence, reflecting how individuals co-occur in space and time within shared environments. Both scenarios are inherently multi-attribute: mode-specific visitation requires coherent joint generation of location and travel mode at the event level, and co-presence requires joint realism in location and time within individual schedules.

Physical spaces serve distinct urban functions and inherently vary in importance. Some locations act as core anchors and attract frequent visits, while others are used more occasionally (Schläpfer et al., 2021; Abbiasov et al., 2024). MobilityGen faithfully captures this spatial heterogeneity. To demonstrate this, we compare simulated and real activity events by grouping locations according to their visitation frequency. The results show strong alignment (Pearson $r = 0.91$, $P < 10^{-3}$; Fig. 6a), with important locations consistently retaining their prominence and highlighting the stability of core urban anchors. Crucially, MobilityGen enables analysis of how location importance differs across travel modes, a perspective not available in previous approaches. As shown in Fig. 6b–g, agreement is highest for common modes such as car ($r = 0.91$) and walk ($r = 0.89$), while less frequent modes such as bus ($r = 0.73$) and tram ($r = 0.67$) show greater variation. These differences reveal how urban space is accessed unevenly across modes and underscore the challenge of simulating less frequent ones. Addressing this gap likely requires targeted modeling strategies and richer contextual data, particularly to better represent sustainable modes that are policy-relevant yet remain underutilized in practice. By capturing both stable urban anchors and variation in mode-specific access, MobilityGen enables new analyses of mode-aware demand, supporting applications from infrastructure usage modeling to accessibility assessments and scenario-based policy evaluation.

While spatial visitation captures where individuals go, co-presence patterns reveal whom they encounter, offering complementary insight into the intensity and distribution of interpersonal contact in daily urban life. A key application of co-presence lies in socio-spatial segregation research, where attention has shifted from static residential indicators to individual-level experienced segregation enabled by fine-grained mobility data (Liao et al., 2025). We assess MobilityGen's ability to reproduce these patterns by focusing on income-based segregation. We group individuals into four income quartiles and measure, at each location, the time spent by people from different income groups (Moro et al., 2021) (see Appendix E). These measures are then combined into an individual-level segregation score that reflects each person's exposure to others across income levels. Simulated and observed segregation scores show strong agreement (Pearson $r = 0.70$, $P < 10^{-3}$; Fig. 7), capturing both overall trends and the variability in co-presence-based income segregation. A breakdown by income quartiles confirms that MobilityGen reproduces systematic differences in exposure and substantially outperforms baseline simulators such as TimeGeo and DITRAS, whose separation of space and time limits their ability to capture realistic contact patterns (see Appendix E, Fig. E.15). This case study illustrates MobilityGen's broader value as a generative model of social exposure. By capturing individual-level interactions in daily schedules, it provides a simulation environment to explore how population-level social mixing evolves under different scenarios. This extends beyond segregation analysis to domains such as information diffusion and infectious disease spread, where spatial co-presence fundamentally shapes outcomes.

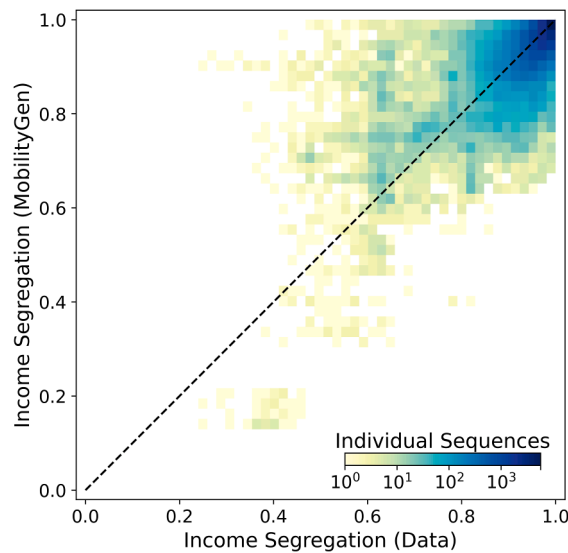


Fig. 7. Reproducing individual-level income segregation from co-presence. Heatmap comparing segregation scores computed from observed and simulated activity events. Cell color indicates the number of individuals sharing each observed-simulated score pair. The diagonal line denotes perfect agreement. MobilityGen shows high agreement with empirical segregation patterns ($r = 0.70$, $P < 10^{-3}$).

5. Discussion and conclusion

Despite growing interest in understanding human movement patterns and their implications, existing simulation models fail to capture the full spatio-temporal variability of mobility behavior. Here, we address this gap by adapting diffusion-based generative modeling to activity-travel event sequences, generating multi-attribute mobility events over days and weeks. We instantiate this approach in MobilityGen, which provides strong alignment with empirical patterns across key mobility indicators and captures coupled structure across time, space, and mode that earlier approaches struggle to represent. Its generative capability is reflected in coherent held-out activity-travel sequences, plausible visits to previously unobserved locations, and emergent individual- and population-level patterns. Because the model produces complete multi-attribute trajectories, it supports downstream analyses such as how urban space is accessed differently across travel modes and how co-presence dynamics shape social exposure, including income-based segregation. Taken together, these capabilities show how deep generative modeling can extend mobility simulation beyond aggregate-statistic alignment to fine-grained analyses of mobility, accessibility, and social exposure.

Our data-driven framework extends event-level representations into a generative modeling setting. By integrating behavioral attributes into a shared latent space, MobilityGen captures empirical dependencies across multi-day sequences that are difficult to specify exhaustively through predefined rules. This positions MobilityGen as a complement to behaviorally grounded activity and transport demand models: where those models encode relationships a priori, MobilityGen learns them from observed sequences. These learned representations are also inspectable, with the embedding space analysis revealing behaviorally meaningful structure as latent geometry, a possible substrate for further behavioral investigation. Looking ahead, MobilityGen could be paired with structural behavioral models in hybrid approaches that combine learned representations with explicit behavioral assumptions, supporting applications such as household activity participation and targeted mobility interventions. Its flexibility also supports richer contextual information, from socio-demographic characteristics to social networks and environmental factors such as weather. The held-out-individual analysis delineates a boundary of the current framework: activity-travel regularities transfer more robustly to individuals absent from training, whereas spatial generation remains sensitive to source-location coverage. Extending the framework toward settings with limited or no observed individual history is therefore a practically important direction, since detailed mobility data are typically available only for small subsets of a population, whereas many applications require mobility estimates at population scale. Together, these extensions position MobilityGen as a bridge between mobility research and disciplines such as social, health, and environmental sciences, a step toward an integrative framework for interdisciplinary mobility modeling.

From a practical standpoint, MobilityGen generates long-term, realistic, and multifaceted activity schedules for individuals using GNSS tracking data, reducing the need for labor-intensive traditional travel surveys. Generative mobility models offer several distinct advantages: they enable the creation of synthetic datasets at scales far beyond those of real-world collections and mitigate privacy concerns when analyzing or sharing disaggregated data. These large, openly accessible datasets, in turn, could promote the standardization and intercomparison of movement analysis methods. Finally, access to fine-grained synthetic mobility data could unlock previously unattainable simulation scenarios and studies in support of sustainable transport, urban planning, public health, and beyond.

CRedit authorship contribution statement

Ye Hong: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization; **Yatao Zhang:** Writing – review & editing, Methodology; **Konrad Schindler:** Writing – review & editing, Supervision; **Martin Raubal:** Writing – review & editing, Supervision, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used GPT-5 in order to improve the language, grammar, and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

No conflict of interest exists.

Acknowledgement

This work is supported by the Hasler Foundation (grant number 21041). The research is partially conducted at the Future Resilient Systems program at the Singapore-ETH Centre, supported by the National Research Foundation (NRF) Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

Code availability

All code to preprocess movement data and train MobilityGen is available at https://github.com/mie-lab/mobility_generation (Apache-2.0 license). Analyses and visualizations were performed in Python and QGIS 3.22.

Data availability

Raw data for the MOBIS dataset are not publicly available due to privacy considerations, but aggregated trip-level data are available from the corresponding author upon request. OpenStreetMap POIs used for context were downloaded from Geofabrik (region: Europe/Switzerland; snapshot: late 2022) and can be downloaded from <https://download.geofabrik.de/>.

Appendix A. Simulating individual preferences in location visits

A.1. Comparison with microscopic mobility models

To evaluate the ability of MobilityGen to replicate fundamental properties of individual location behavior, we compare it with representative microscopic mobility models spanning mechanistic, predictive, and generative approaches. These include: (i) classical mechanistic models such as the exploration and preferential return (EPR) model (Song et al., 2010a) and the Container model (Alessandretti et al., 2020); (ii) sequence prediction models, including a first-order Markov chain (Gambs et al., 2012) and a multi-head self-attention (MHSA)-based neural network (Hong et al., 2023b); and (iii) the deep generative MoveSim model (Feng et al., 2020). The design and implementation details of each model are described in the following sections.

- The EPR model is a well-established microscopic mobility model designed to reproduce scaling laws in the evolution of distinct locations and their visitation frequencies over time. We estimate model parameters from the training data. For each test sequence, we determine the number of unique locations S visited and the corresponding visit counts f_i for each location i , and then simulate 50 displacement steps. At each step, the individual either explores a new location with probability $p^{\text{new}} = \rho S^{-\gamma}$, where ρ and γ are parameters, or returns to a previously visited location with the complementary probability $1 - p^{\text{new}}$. In exploration steps, a new location is determined by sampling a displacement Δr from the empirically estimated jump length distribution $P(\Delta r)$. For return steps, the probability of selecting location j , denoted as Π_j , is proportional to its past visitation frequency, i.e., $\Pi_j \propto f_j$.
- The Container model describes how individuals allocate time and move across nested spatial scales (Alessandretti et al., 2020). The model is built on a hierarchical structure H with L levels, where each level groups locations into increasingly broader spatial containers. Each container is associated with an attractiveness score a , and mobility decisions are governed by an $L \times L$ displacement matrix D . The probability of moving from the current location j to a destination k is given by:

$$p_{D_{j,h}, D_{j,k}} \prod_{l \leq D_{j,k}} a(k_l) \quad (\text{A.1})$$

Here, the first term $p_{D_{j,h}, D_{j,k}}$ denotes the probability of traveling a distance $D_{j,k}$, given that the individual is currently at distance $D_{j,h}$ from their home location h . The second term captures the probability of choosing location k at that level distance, where $a(k_l)$ represents the attractiveness of a container at level l that contains location k . Parameters H , a , and D are estimated for each individual via maximum likelihood estimation on the training data (Alessandretti et al., 2020). We generate 50 displacements per individual for evaluation on the test set.

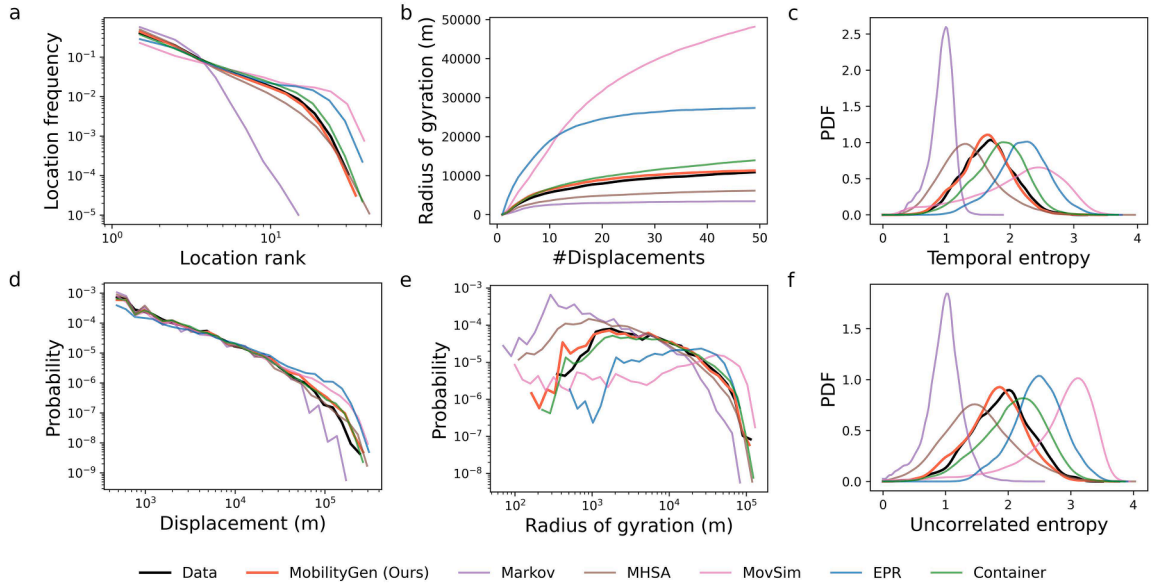


Fig. A.8. Evaluating microscopic mobility models with location metrics. **a.** Rank-frequency distribution of visited locations. **b.** Median radius of gyration as a function of the number of displacements. **c.** Distribution of temporal mobility entropy across individuals. **d.** Distribution of displacements for the entire population. **e.** Distribution of radius of gyration across individuals. **f.** Distribution of uncorrelated mobility entropy across individuals. These metrics are compared between real data (black) and traces generated by MobilityGen (red), Markov (purple), MHSA (brown), MoveSim (pink), EPR (blue), and Container (green).

- **Markov model.** Classical models for location prediction represent individual mobility as a Markov chain (Ashbrook and Starner, 2002), where locations are treated as states and transitions between them are encoded in a Markov transition matrix. We implement a first-order Mobility Markov Chain (1-MMC) (Gambis et al., 2012), in which the probability of visiting the next location depends only on the current one. During generation, the next location is sampled from the top three most likely locations based on the learned transition probabilities. This process is repeated until the generated sequence reaches the target length of 50 steps.
- **MHSA-based model.** Recent advances in next location prediction leverage deep sequential models that represent locations as latent embeddings and learn transition dynamics via attention-based architectures (Luca et al., 2021). We train an MHSA network to autoregressively generate location sequences. At each step, the next location is sampled using a combination of top- k and nucleus sampling (Holtzman et al., 2020). The top- k strategy restricts sampling to the k most probable candidates, while nucleus sampling limits it to the smallest set of locations whose cumulative probability exceeds a predefined threshold p . We set $k = 200$ and $p = .99$. The selected location is appended to the sequence and used as input for the next prediction step. This procedure is repeated for 50 steps.
- **MoveSim.** As a baseline for generative modeling of individual location choices, we implement MoveSim, a sequence-to-sequence model based on generative adversarial networks (GANs). The model consists of a generator that simulates location sequences using mobility priors, and a discriminator that distinguishes between real and generated sequences. These networks are trained jointly in an adversarial framework using a min-max objective. Unlike the original implementation, we exclude auxiliary features such as historical transitions, physical distances, and functional similarity, as they require pairwise location matrices that are computationally infeasible to construct given the large number of unique locations in our dataset.

We evaluate the performance of all models using established location-based metrics. In addition to those reported in the main text, we include the jump length distribution (Brockmann et al., 2006) and the uncorrelated entropy (Song et al., 2010b). Results are shown in Fig. A.8. To assess model fit, we compute the log-likelihood ratio, comparing the likelihood of the observed data under each model relative to MobilityGen. We additionally calculate the Wasserstein distance to quantify discrepancies between empirical and model-generated distributions for each metric (Table A.1). Overall, MobilityGen provides the closest approximation to the empirical data across all evaluated metrics, except for the jump length distribution, where it tends to overestimate the frequency of longer displacements (Fig. A.8d). Moreover, the median evolution of the radius of gyration across individuals, $r_u(k)$, over displacement steps k , is well described by logarithmic growth (Fig. A.8b). The fit $r_u(k) \propto \alpha \log(k)$, with parameters α derived from MobilityGen ($\alpha_{\text{MobilityGen}} = 2967 \pm 29$), aligns most closely with the real data ($\alpha_{\text{true}} = 2875 \pm 35$), while the baselines exhibit notably different values ($\alpha_{\text{EPR}} = 6431 \pm 185$, $\alpha_{\text{Container}} = 3784 \pm 68$, $\alpha_{\text{Markov}} = 796 \pm 26$, $\alpha_{\text{MHSA}} = 1541 \pm 12$, $\alpha_{\text{MoveSim}} = 14,976 \pm 462$).

Beyond matching marginal distributions, we also examine whether each model reproduces the geographic concentration and dispersion of visits at population scale. Fig. A.9 compares aggregated visitation intensity across Switzerland for the empirical data and each microscopic model, while Fig. A.10 provides the corresponding urban-scale comparisons for Bern, Zurich, and Lucerne. To summarize these maps quantitatively, we report the entropy of the visit-frequency distribution across locations (Table A.2), where

Table A.1

Model performance in replicating location visit distributions. The log-likelihood ratio measures the likelihood that the real data were generated by each model relative to MobilityGen, with positive values favoring MobilityGen (all $P < 10^{-3}$, except for $P(f_{\Delta r})$ in MHSA, where $P = .66$). The Wasserstein distance assesses the dissimilarity between generated and real location metric distributions, with higher values indicating greater divergence. Results are shown for number of visits per location (f_k), jump length ($f_{\Delta r}$), radius of gyration (f_{r_g}), uncorrelated entropy (S_{unc}), and temporal entropy (S_{temp}); best-performing values are shown in bold.

	Log-likelihood					Wasserstein distance				
	$P(f_k)$	$P(f_{\Delta r})$	$P(f_{r_g})$	$P(S_{unc})$	$P(S_{temp})$	f_k	$f_{\Delta r}$	f_{r_g}	S_{unc}	S_{temp}
EPR	114,195	166,326	31,364	74,711	30,298	2.62	15,882	12,825	0.60	0.54
Container	12,504	−6568	946	2741	3549	0.68	1146	3578	0.20	0.18
Markov	796,846	2697	16,530	329,477	2,728,649	2.03	2223	10,102	0.91	0.73
MHSA	66,983	−115	4352	8355	9616	0.85	1491	5137	0.35	0.30
MoveSim	268,850	35,384	79,922	48,838	22,837	5.04	8128	34,096	0.92	0.51
MobilityGen (Ours)	—	—	—	—	—	0.26	2318	575	0.08	0.04

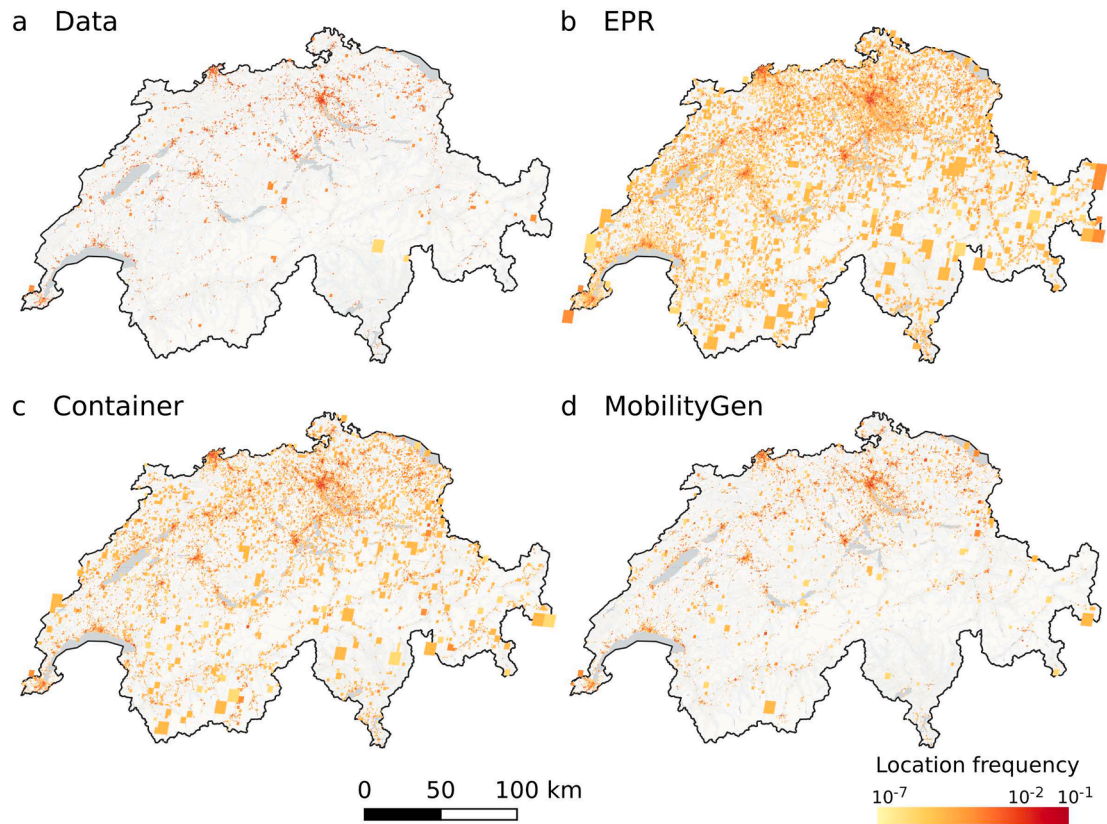


Fig. A.9. Comparison of spatial visitation patterns across models. Aggregated visit frequencies from real data (a) are compared with those generated by three baseline mobility models: EPR (b), Container (c), and MobilityGen (d). While all models recover broad national-scale trends, MobilityGen more accurately reproduces both high-density urban visitation and spatial dispersion into surrounding regions.

higher values indicate broader spatial dispersion. The baselines either concentrate visits too strongly or disperse them too widely relative to the data, whereas MobilityGen yields dispersion values that are closer to the empirical patterns at both national and city scales, while still showing a tendency toward slightly higher dispersion in several cases.

Collectively, the results show that MobilityGen more accurately captures key dimensions of location behavior, including visitation frequency, travel distance, and spatio-temporal spread, than traditional microscopic models.

A.2. Quantifying similarity and diversity in individual location sequences

To complement classical mobility metrics, we evaluate the simulated location sequences using metrics adapted from language modeling, which quantify both their similarity to and diversity from the real sequences. Similarity is measured using the Bilingual

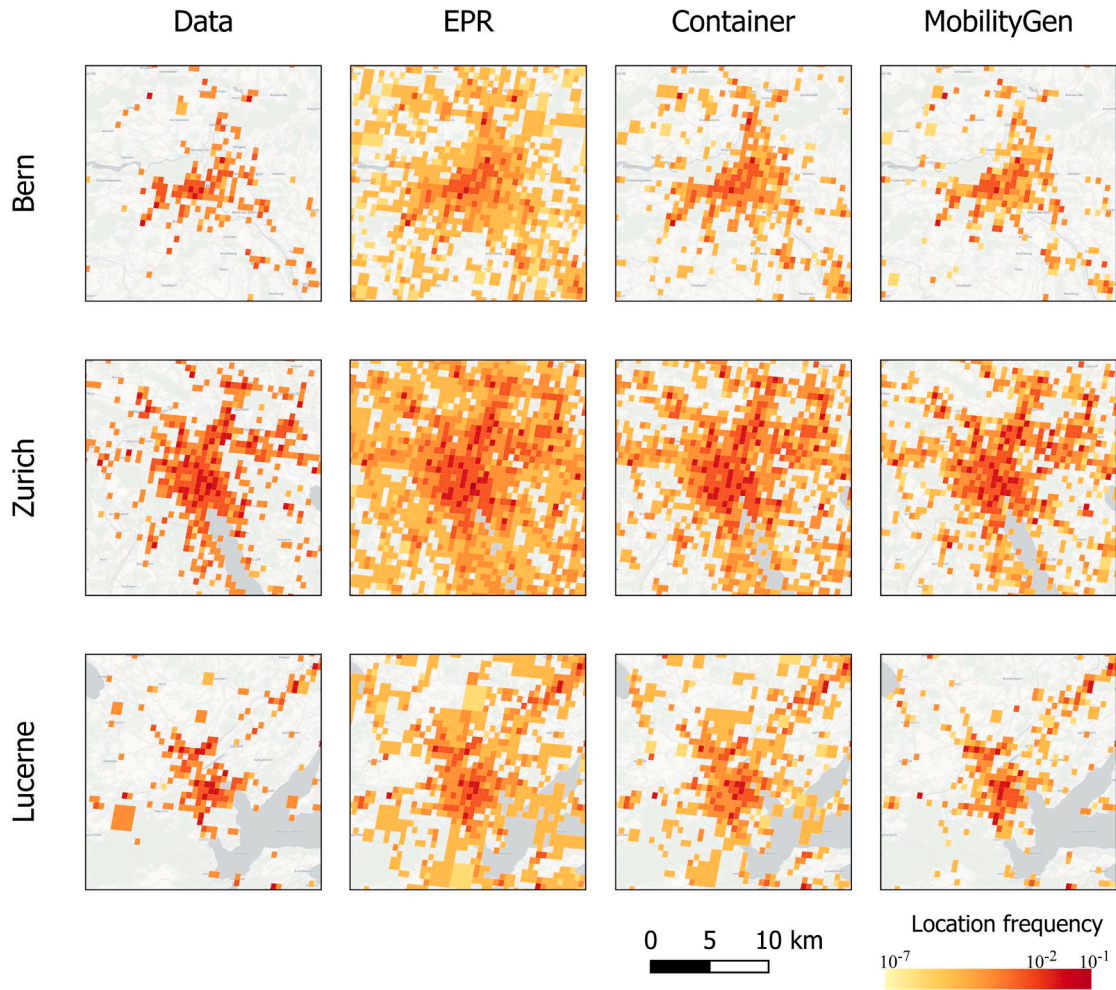


Fig. A.10. Urban-level comparison of spatial visitation patterns across models. Real and simulated visit frequencies are compared across Bern (top row), Zurich (middle row), and Lucerne (bottom row). Consistent with national-level patterns, MobilityGen best captures fine-grained spatial visitation within dense urban cores and across peripheral areas.

Table A.2

Spatial dispersion in visitation patterns. Entropy values are reported for the real data and for location traces generated by each microscopic mobility model. Higher entropy indicates broader spatial dispersion of visits. Values below those of the real data suggest insufficient reproduction of exploratory behavior, while values above indicate misalignment with population-level location preferences.

	National	Bern	Zurich	Lucerne
EPR	1.20	3.89	5.04	3.50
Container	0.91	3.13	4.60	3.18
Markov	0.17	1.03	2.07	1.34
MHSA	0.50	2.79	4.05	2.62
MoveSim	0.67	3.21	4.30	2.58
MobilityGen (Ours)	0.53	2.49	3.99	2.71
Data	0.33	1.89	3.37	2.11

Evaluation Understudy (BLEU) score (Papineni et al., 2002), an n-gram-based metric that compares each generated sequence to its corresponding ground truth, with higher values indicating greater similarity. We implement the smoothed sentence-level version of BLEU at n-gram levels 1 through 4 (BLEU-1 to BLEU-4), commonly used in sequence generation tasks. To evaluate intra-sequence diversity, we use the distinct unigram (dist-1) metric (Gong et al., 2023), which measures the proportion of unique locations within a sequence. Lower dist-1 values indicate higher repetition and less diversity. At the sequence level, we compute the diverse 4-gram (div-4) metric (Deshpande et al., 2019), which captures the proportion of distinct 4-grams in a sequence. Higher div-4 values indicate

Table A.3

Sequence similarity and diversity with language modeling metrics. The BLEU score quantifies similarity between generated and real location sequences, with higher values indicating better alignment. Distinct unigrams (dist-1) and diverse 4-grams (div-4) measure sequence diversity at the intra-sequence and inter-sequence levels, respectively, where higher values indicate greater variability.

	BLEU	dist-1	div-4
Markov	0.28	0.08	0.02
MHSA	0.24	0.21	0.29
MoveSim	0.002	0.51	0.60
EPR	0.09	0.42	0.77
Container	0.19	0.31	0.57
MobilityGen (Ours)	0.26	0.23	0.34

richer generation diversity. Evaluation results for all implemented microscopic mobility models are presented in Table A.3. Overall, MobilityGen produces location visit sequences that closely mirror real-world patterns, while avoiding input replication and preserving a realistic level of behavioral diversity.

While these metrics are informative, they should be interpreted with caution: (i) In contrast to language modeling, where multiple semantically valid reference sentences may exist, mobility datasets typically provide only one observed realization of an individual's behavior, despite the existence of many plausible alternatives. (ii) Realistic mobility behavior involves a balance between exploration and habitual returns. As such, high diversity is not inherently desirable, and diversity metrics alone may not fully capture the plausibility of generated sequences.

Appendix B. Generating realistic activity event sequences

B.1. Comparison with spatio-temporal mobility simulators

To evaluate the effectiveness of MobilityGen in reproducing spatio-temporal activity-travel patterns, we benchmark its performance against two state-of-the-art mobility frameworks: DITRAS (Pappalardo and Simini, 2018) and TimeGeo (Jiang et al., 2016). Detailed descriptions of their methodologies and implementations are provided below.

- DITRAS decouples temporal pattern generation from spatial location assignment by modeling temporal dynamics with a mobility diary generator and assigning these diaries to physical locations using the density-EPR model (Pappalardo et al., 2015). The diary

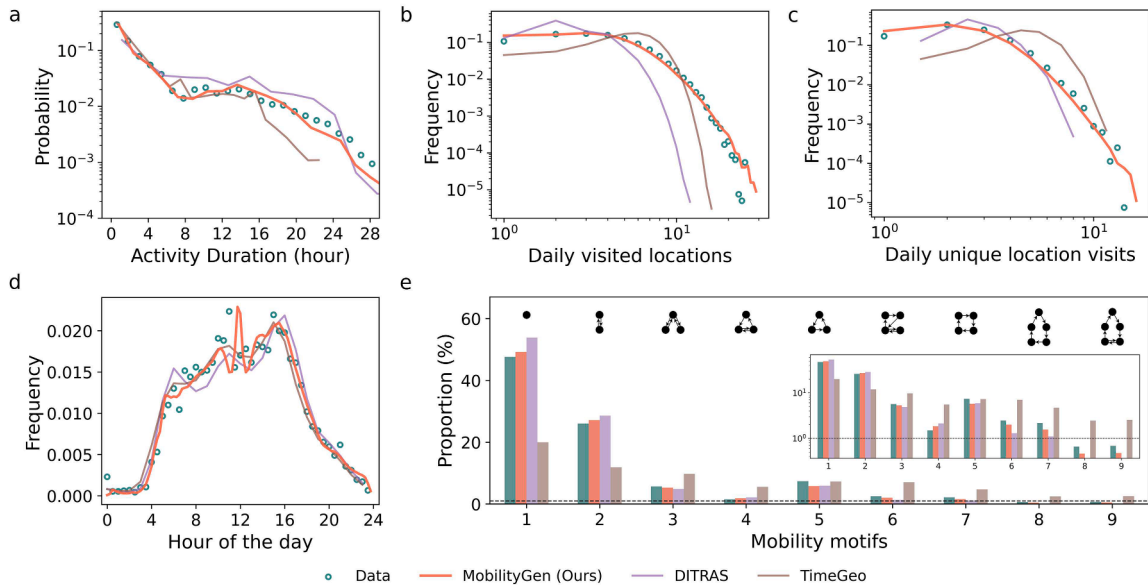


Fig. B.11. Generated events replicate individual activity-travel patterns and interdependencies. **a.** Distribution of activity durations. **b.** Distribution of daily visited locations. **c.** Distribution of daily unique visited locations. **d.** Distribution of activity start times throughout the day. **e.** Distribution of mobility motifs, with motif diagrams shown above. The dashed line marks the 1% threshold as a visual guide, and the inset presents the same distribution on a log scale. Metrics are compared between real data (green) and activity events generated by MobilityGen (red), DITRAS (purple), and TimeGeo (brown).

Table B.4

Alignment of behavioral patterns and interdependencies. The Wasserstein distance quantifies dissimilarities between metric distributions derived from simulated and real activity events, with higher values indicating greater divergence. Source sequence results are included as a reference baseline. Results are reported for activity duration (f_d), activity start times (f_t), daily visited locations (f_k^{day}), unique daily visited locations ($f_{|k|}^{\text{day}}$), mobility motifs (f_{motifs}), and travel mode (f_m); best-performing values are shown in bold.

	f_d	f_t	f_k^{day}	$f_{ k }^{\text{day}}$	f_{motifs}	f_m
DITRAS	2.48	1.56	1.51	0.39	0.30	–
TimeGeo	1.92	1.33	1.33	1.33	1.37	–
MobilityGen (Ours)	0.47	0.52	0.29	0.22	0.14	0.048
Source sequence	0.58	1.11	0.66	0.29	0.27	0.043

generator employs a time-aware Markov transition matrix to estimate transition probabilities between “typical” locations (e.g., home) and “non-typical” locations, along with the corresponding durations of stay. Non-typical locations are then assigned using the density-EPR mechanism, an extension of the classic EPR model that incorporates a relevance factor during exploration. We used the DITRAS implementation from scikit-mobility to align with our data format. Based on the training dataset, we estimated the population-level Markov transition matrix and calibrated individual-specific density-EPR parameters. Location relevance was determined by empirical visit frequencies. We adopted a temporal resolution of one hour for simulation.

- TimeGeo. Originally developed to reconstruct high-resolution spatio-temporal mobility trajectories from sparse digital traces, TimeGeo incorporates three individual-specific parameters to characterize temporal movement dynamics: the weekly number of home-based tours, the dwell rate, and the burst rate. These parameters govern time-dependent travel tendencies through a Markov decision process that distinguishes movement patterns based on an individual’s location state. Spatial decisions are modeled using a rank-based EPR mechanism, which selects exploration locations based on their proximity rank relative to the current location, rather than absolute geographic distance. We estimated temporal and spatial parameters for each individual using the training dataset. For each test sequence, the most frequently visited location was designated as the home location, and the model was simulated at a temporal resolution of one hour.

We assess the performance of these models using mobility metrics that characterize activity-travel patterns. To align the hourly outputs of DITRAS and TimeGeo with MobilityGen’s event-sequence formulation, we converted their simulated hourly state sequences into activity event sequences by merging consecutive records assigned to the same location. We then retained the first 50 generated events from each model and compared them with the first 50 real future events, ensuring that all models are evaluated over the same event horizon. In addition to the metrics reported in the main text, we introduce two supplementary measures: the number of unique location visits per day and the distribution of activity start times. Evaluation results are presented in Fig. B.11. To quantify model performance, we compute the Wasserstein distance between the distributions of each metric derived from the generated activity events and those observed in the real data (Table B.4). Metrics calculated from the source sequence are included as a strong reference baseline. However, it is important to account for variability in human activity behavior, as source sequences differ in both the duration and timing of activity events.

MobilityGen consistently outperforms all other models in capturing individual activity-travel patterns and interactions, demonstrating a substantial margin of improvement. Moreover, it simulates event sequences that are as accurate as, and in some cases even more accurate than, those derived from the source sequences.

B.2. Analysis of learned attribute embeddings

We examine the structure of the learned attribute embeddings to uncover how MobilityGen internalizes and organizes behavioral attributes. Each 128-dimensional embedding is projected onto a two-dimensional space using the densMAP method, as implemented in the Python library UMAP (McInnes et al., 2018). We use the default parameters ($\text{min_dist} = 0.1$, $\text{dens_lambda} = 2.0$) and confirm that alternative parameter combinations yield qualitatively consistent embedding structures. Two-dimensional visualizations are generated for both single attributes and composite activity events, computed according to Eq. (8).

Visualizations for travel mode, time, duration, and location are shown in Fig. B.12. Travel modes with similar characteristics are positioned in close proximity within the embedding space; for example, walk and bicycle appear nearby, as do train, bus, and tram. In contrast, the “other” category appears isolated, reflecting its dissimilarity from all other modes. The time and duration embeddings in Fig. B.12b and c form ordered, coherent trajectories, revealing that MobilityGen captures the continuous structure inherent in these attributes. For locations in Fig. B.12d, the model organizes places according to their empirical visit frequencies in the training set, even though this frequency information is not explicitly observable from individual event sequences. This suggests that the model encodes the relative importance of locations based on usage patterns and implicitly captures the characteristic rank-frequency distribution observed in human mobility.

We next examine composite event embeddings (Fig. B.13). Each point represents a mobility event defined by a combination of travel mode, time, duration, and location, with color denoting the visited location. A clear structural distinction emerges: some events

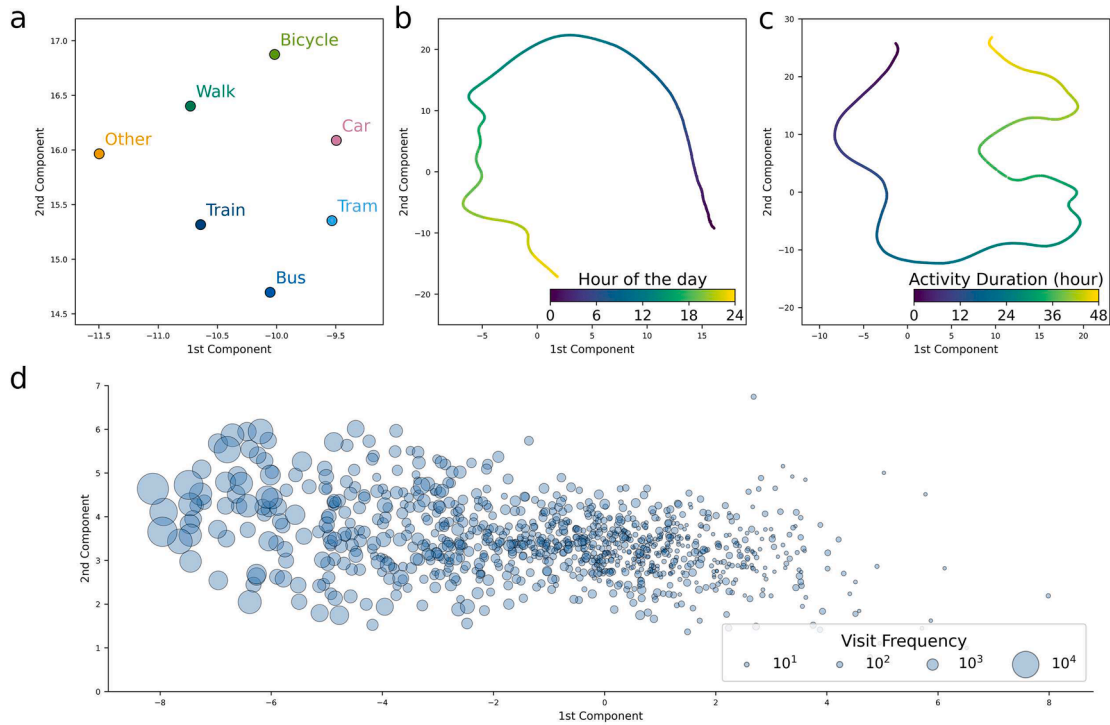


Fig. B.12. Visualization of learned activity embeddings. 2D projections of the 128D attribute embeddings, obtained using the densMAP algorithm ($\text{min_dist}=0.1$, $\text{dens_lambda}=2.0$). **a.** Travel mode embeddings reveal interpretable structure: similar modes (walk and bicycle; bus, tram, and train) are positioned closely, while “other” appears isolated. **b.** Time embeddings form a continuous trajectory that progresses over the 24-hour day, with neighboring hours mapped to adjacent regions. **c.** Duration embeddings follow a similarly ordered curve, reflecting the gradual increase from 0 to 48 hours. **d.** Location embeddings are structured by visit frequency, with marker size denoting log-scaled frequency. Higher-frequency locations cluster toward one end of the space.

form location-based clusters, while others are spatially mixed with no separation by destination. The embedding space thus loosely divides into an outer region of clustered behaviors and a central region of less location-dependent activity. This contrast suggests that the model distinguishes between different types of mobility behavior. We hypothesize that these structural differences reflect underlying activity types; for example, routine visits such as home or work (spatially consistent) versus exploratory or irregular leisure activities (spatially diffuse). Because the embedding space encodes patterns learned directly from event sequences, it provides a valuable basis for analyzing how behavioral variations relate to spatial semantics and for advancing context-aware modeling of mobility behavior.

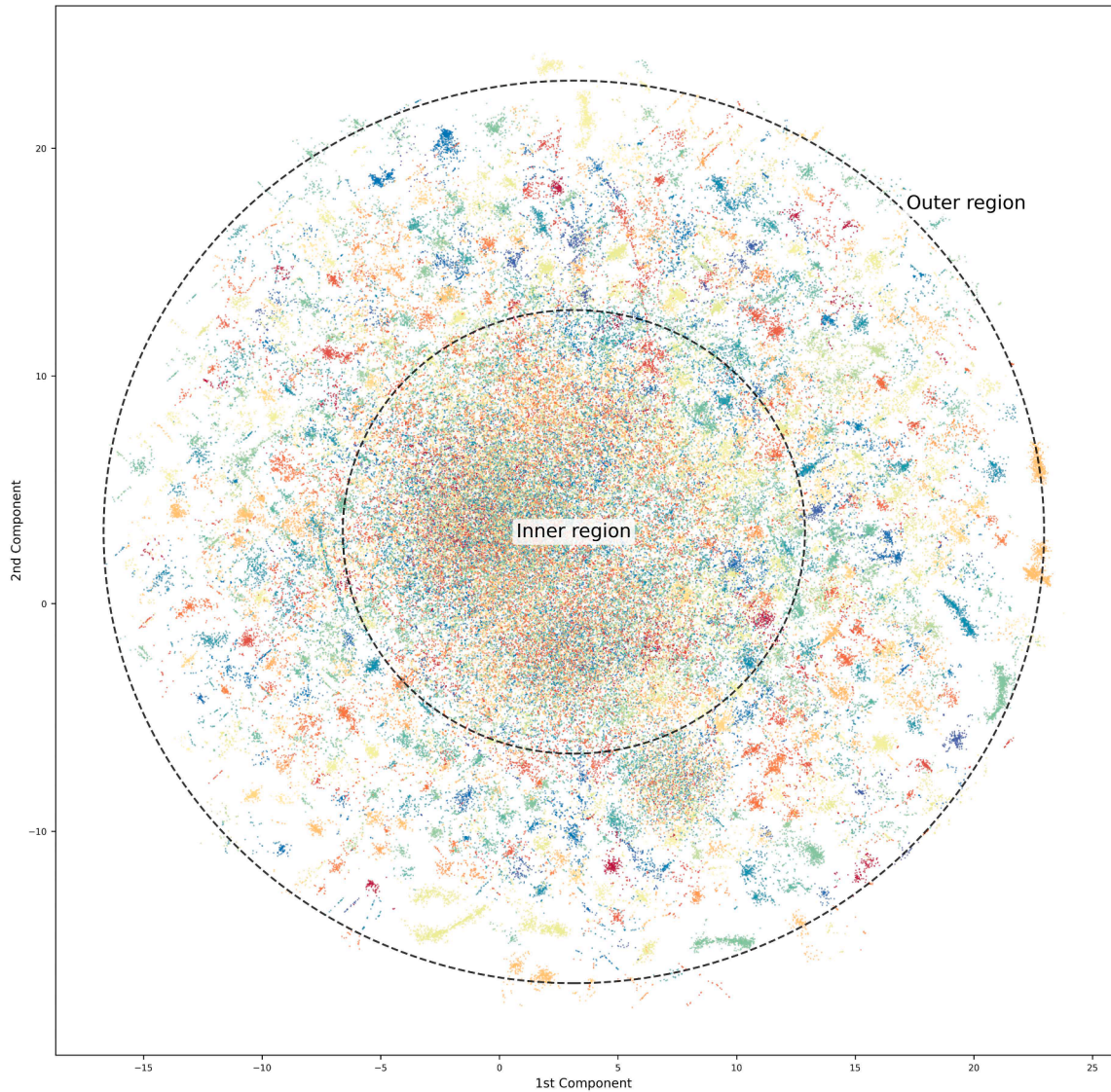


Fig. B.13. Visualization of learned event embeddings. Each point represents a mobility event defined by a unique combination of travel mode, time, duration, and location, with colors indicating the visited location. The embedding space shows a clear structural distinction between two loosely defined regions: an outer region where events form location-based clusters, and an inner region where events are mixed without clear spatial separation. The circular boundaries serve as visual guides to illustrate the transition from disordered to clustered patterns. This contrast suggests that the model distinguishes between different types of mobility behavior.

Appendix C. Role of behavioral and contextual factors

C.1. Identifying urban functions from POI features

POIs reflect the socioeconomic activities in local regions and have been increasingly used to identify urban functions at a fine scale (Yao et al., 2017). A well-established approach to functional identification involves obtaining a POI representation for each study unit, identifying cluster structures from these representations, and labeling the resulting classes based on expert knowledge (Niu and Silva, 2021). Here, we perform k -means clustering on the LDA functional descriptor of each location. Using the silhouette score to assess clustering quality, we determine the optimal number of classes to be five (Fig. C.14a). Visualizing the five-class clustering result using two-dimensional t-SNE reveals that the data points within each class are compact, while those between classes are well-separated (Fig. C.14b). To interpret the urban functions of each location group, we rely on the spatial distribution of classes (Fig. C.14c) and the POI enrichment factor for each class (Table C.5). The latter quantifies the relative occurrence of different categories of POIs. For

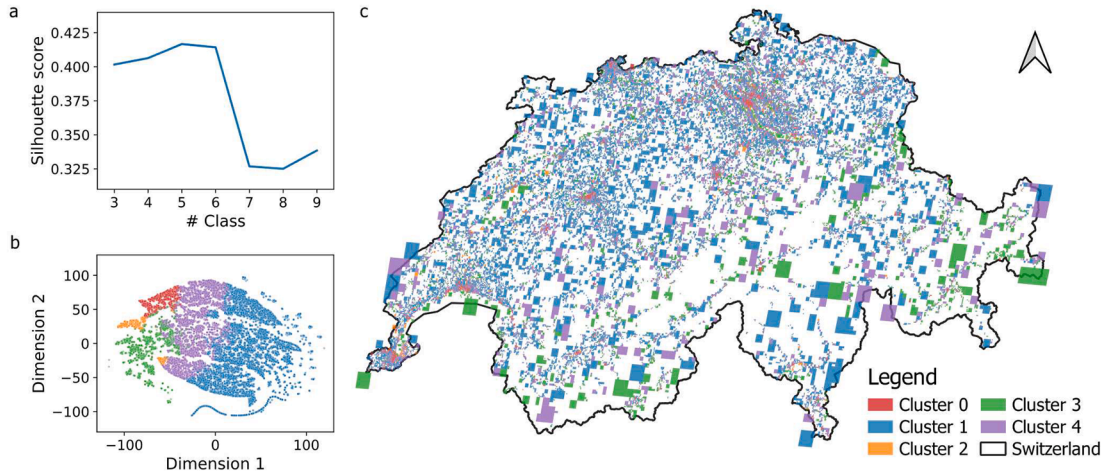


Fig. C.14. Identifying urban functions using location LDA vectors and k -means clustering. **a.** The silhouette score assesses clustering quality, with higher values indicating more compact and better-separated classes. The score suggests that five classes yield the best separation. **b.** Clustering results visualized using 2D t-SNE, where each point represents a location, color-coded by class. **c.** Spatial distribution of the identified location classes, using the same color scheme as in **b.**

Table C.5

POI enrichment across functional clusters. First-level POI categories with more than 1000 instances in the study area were considered. For each cluster, the top three POI types with enrichment factors greater than 1 are highlighted in bold.

POI category	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Accommodation	1.28	0.90	0.41	0.92	0.97
Catering	1.88	0.71	0.44	0.43	0.82
Health	2.31	0.46	0.12	0.04	0.85
Leisure	0.48	1.22	3.26	0.28	1.08
Money	2.21	0.56	0.04	0.06	0.82
Natural	0.04	0.86	0.14	7.15	0.87
Others	1.01	0.83	0.87	1.81	1.03
Place of worship	0.56	1.68	0.22	0.32	0.85
Public	0.90	1.26	0.66	0.13	0.99
Shopping	1.59	0.83	0.37	0.31	0.92
Tourism	0.54	1.05	0.55	3.08	0.97
Traffic	1.00	1.09	0.64	0.50	1.02

each location class i and POI category q , the enrichment factor EF_i^q is defined as:

$$EF_i^q = \frac{N_i^q / N_i}{N^q / N} \quad (\text{C.1})$$

where N_i^q represents the number of category q POIs in location class i , N_i denotes the total number of POIs in location class i , N^q indicates the total number of category q POIs, and N signifies the total number of POIs in the study area.

We label Cluster 0 as the *City center* due to its high prevalence in urban areas and the significant frequency of catering, health, and monetary-related POIs. Locations in Cluster 1 are dispersed outside the city and exhibit a high density of public facilities, leading to their classification as *Residential & Countryside*. Clusters 2 and 3 are designated as *Leisure & Recreation* and *Nature & Tourism*, respectively, due to their notable concentrations of leisure and nature-related POIs. Lastly, locations in Cluster 4 are typically situated near Cluster 0, and given their relatively even distribution of POIs, we classify this cluster as *Suburbs & Developing areas*.

The identification of urban functions is based on the locations' LDA functional descriptors, which are the same data as employed by MobilityGen. Since the process primarily relies on the structure of the data itself, we believe that this result offers a direct illustration of the information used in MobilityGen.

C.2. Effect of representing activity attributes and spatial context

We conduct ablation studies on MobilityGen to assess the contributions of activity attributes and contextual information. Activity attributes serve dual roles in the model: they are embedded jointly with location data to support the learning of mobility behavior (Eq. (8)), and they provide additional supervisory signals through the loss function (Eq. (17)). We begin with a model trained solely on location sequences and progressively incorporate time and travel mode attributes into the pipeline. All other training configurations

Table C.6

Ablation study on activity attributes (location metrics). The full model is compared to reduced variants: one using location only (base), and others including time (base + time use) or travel mode (base + travel mode). Wasserstein distances are reported for number of visits per location (f_k), jump length ($f_{\Delta r}$), radius of gyration (f_{r_g}), uncorrelated entropy (S_{unc}), and temporal entropy (S_{temp}); best-performing values are shown in bold.

	f_k	$f_{\Delta r}$	f_{r_g}	S_{unc}	S_{temp}
Base	0.649 $\pm$ 0.003	3871 $\pm$ 16	901 $\pm$ 21	0.216 $\pm$ 0.001	0.110 $\pm$ 0.0004
Base + time use	0.543 $\pm$ 0.002	2280 $\pm$ 21	625 $\pm$ 19	0.144 $\pm$ 0.001	0.101 $\pm$ 0.0005
Base + travel mode	0.519 $\pm$ 0.002	3493 $\pm$ 15	804 $\pm$ 13	0.136 $\pm$ 0.001	0.088 $\pm$ 0.0003
Full model	0.262 $\pm$ 0.004	2318 $\pm$ 14	574 $\pm$ 17	0.080 $\pm$ 0.001	0.036 $\pm$ 0.001

Table C.7

Ablation study on activity attributes (activity-travel metrics). The full model is compared to reduced variants: one incorporating only time (base + time use) and another incorporating only travel mode (base + travel mode). Wasserstein distances are reported for activity duration (f_d), activity start time (f_t), daily visited locations (f_k^{day}), unique daily visited locations ($f_{|k|}^{\text{day}}$), mobility motifs (f_{motifs}), and travel mode (f_m); best-performing values are shown in bold.

	f_d	f_t	f_k^{day}	$f_{ k }^{\text{day}}$	f_{motifs}	f_m
Base + time use	0.904 $\pm$ 0.01	0.73 $\pm$ 0.01	0.447 $\pm$ 0.002	0.273 $\pm$ 0.001	0.233 $\pm$ 0.01	-
Base + travel mode	-	-	-	-	-	0.113 $\pm$ 0.001
Full model	0.473 $\pm$ 0.001	0.52 $\pm$ 0.01	0.292 $\pm$ 0.01	0.224 $\pm$ 0.002	0.142 $\pm$ 0.004	0.048 $\pm$ 0.002

Table C.8

Ablation study on contextual information. The full model is compared to a variant without contextual inputs (MobilityGen w/o ctx). Locations are grouped based on their presence in real and generated sequences. Wasserstein distances quantify differences in the distribution of visits to functional areas, including Set A (real) vs. Set A (generated) (d_A), Set C vs. Set B (d_B^C), Set C vs. Set A (real) ($d_{A_{\text{observe}}}^C$), and Set C vs. Set A (generated) ($d_{A_{\text{generate}}}^C$). Arrows indicate whether smaller or larger values are preferred.

	$d_A \downarrow$	$d_B^C \downarrow$	$d_{A_{\text{observe}}}^C \uparrow$	$d_{A_{\text{generate}}}^C \uparrow$
MobilityGen w/o ctx	0.025 $\pm$ 0.001	0.153 $\pm$ 0.067	0.548 $\pm$ 0.001	0.545 $\pm$ 0.001
MobilityGen	0.009 $\pm$ 0.001	0.105 $\pm$ 0.006	0.560 $\pm$ 0.005	0.558 $\pm$ 0.003

are held constant to ensure a fair comparison. The results are summarized in [Tables C.6](#) and [C.7](#), where model performance is measured using the Wasserstein distance between generated and real sequences across a range of mobility metrics. Each model variant is simulated three times with different random seeds, and we report the mean and standard deviation of the resulting Wasserstein distances. While the base model, trained only on historical location data, already performs strongly relative to classical mobility baselines (see performance numbers in [Table A.1](#)), the inclusion of either time or travel mode data leads to further improvements. The full model, which incorporates all behavioral attributes, achieves the best overall performance, with a slight decline observed only in the jump length metric. These findings support our claim that jointly modeling behavioral attributes results in a more realistic reproduction of individual mobility patterns.

Movement-related context, including location geometry and surrounding land-use functions, is integrated with activity attributes within the embedding space. We hypothesize that this integration enhances MobilityGen's ability to characterize locations and is particularly useful for identifying novel locations for exploration. To evaluate this, we compare the full model with a variant that excludes contextual information. As described in the main text, generated locations are grouped into three sets based on their presence in the observed sequences: Set A includes locations common to both observed and generated data, while Sets B and C consist of locations unique to the observed and generated data, respectively. We then compute the Wasserstein distance to quantify discrepancies in the distribution of visits to functional areas. We expect the model with contextual information to generate more similar visitation patterns for novel locations (i.e., comparisons between Sets B and C), and more distinct patterns between shared and novel locations (i.e., comparisons between Sets A and C), than the model without contextual information. This expectation is supported by the Wasserstein distance results in [Table C.8](#), although the differences between models for Sets B and C are not statistically significant. This is likely due to the relatively low visit frequencies to these locations, which result in greater variability in visitation patterns.

C.3. Comparison with mobility flow prediction models

We benchmark MobilityGen against established flow models to evaluate how effectively it leverages contextual information when predicting mobility flows in unseen locations. We adopt the flow prediction task with the objective to estimate the flow $y(l_i, l_j)$ between each origin-destination (OD) pair l_i and l_j per unit time, given the total outflow O_i from location l_i ([Simini et al., 2021](#)). That is, the task involves learning to distribute the total outflow from a given origin across all potential destinations based on location characteristics. Typical features include inter-location distance, population density, and contextual variables such as land use and POI

Table C.9

Reconstructing mobility flows at novel locations. Simulated flows are compared with real flows at locations not seen during training. For MobilityGen, individual sequences are aggregated into an origin-destination matrix to enable comparison with flow-based models. Evaluation metrics include the common part of commuters (CPC), Pearson correlation coefficient (Pearson r), and Jensen-Shannon divergence (JSD). Arrows indicate the preferred direction for each metric. Entropy quantifies the spatial dispersion of visits to novel locations. Best-performing values are shown in bold.

	CPC $\uparrow$	Pearson r $\uparrow$	JSD $\downarrow$	Entropy
Gravity power-law	0.02	0.19	0.81	9.88
Random Forest	0.02	0.17	0.79	8.75
Deep Gravity	0.08	0.25	0.77	8.33
MobilityGen (Ours)	0.22	0.47	0.69	7.16
Data				6.92

types, which reflect the activity options available at each location. These correspond closely to the contextual inputs incorporated in MobilityGen.

To ensure consistency in temporal coverage and enable a fair comparison, we construct OD flow pairs from the test set of mobility event sequences. Locations are partitioned into two non-overlapping subsets: (i) $\mathcal{S}_{\text{seen}}$, comprising locations observed during the training period, and (ii) $\mathcal{S}_{\text{unseen}}$, containing locations that appear only during testing. Flow models are trained on OD pairs originating from $\mathcal{S}_{\text{seen}}$ and evaluated on flows from $\mathcal{S}_{\text{unseen}}$, providing a direct assessment of generalization to novel locations. For MobilityGen, we aggregate simulated event sequences into OD flow pairs and evaluate performance exclusively on flows originating from $\mathcal{S}_{\text{unseen}}$. We next describe the flow models and feature sets used for comparison:

- Gravity (Zipf, 1946). We implement the singly constrained gravity model, which estimates flow $\hat{y}(l_i, l_j)$ from origin l_i to destination l_j based on population sizes m_i, m_j , and the distance d_{ij} :

$$\hat{y}(l_i, l_j) = O_i P_{ij} = O_i \frac{m_j^{\beta_1} f(d_{ij})}{\sum_k m_k^{\beta_1} f(d_{ik})} \quad (\text{C.2})$$

Here, P_{ij} denotes the normalized probability of travel from l_i to l_j , and $f(\cdot)$ is the deterrence function, specified as either exponential, $f(d) = \exp(\beta_2 d)$, or power-law, $f(d) = d^{\beta_2}$. The model parameters β_1 and β_2 are fitted to the training flows using maximum likelihood estimation. Preliminary experiments suggest slightly better performance using the power-law form.

- Deep Gravity (Simini et al., 2021). This model generalizes the classical gravity framework by introducing hidden layers and nonlinearities, effectively converting it into a feedforward neural network. It incorporates additional geographic context for each location, including land use types, road network lengths, and counts of facilities related to transport, food, health, education, and retail. Each OD pair is described using 39 input features: 18 contextual attributes for both origin and destination, the distance between them, and their population sizes. We follow the original implementation, using the same feature set, architecture, and training procedure.
- Random Forest. We use the same 39 input features as Deep Gravity to train a random forest regressor for predicting OD flow intensities. We employ 1000 estimators (Cabanias-Tirapu et al., 2025) and the default hyperparameters from scikit-learn (Pedregosa et al., 2011).

We evaluate model performance on flows originating from $\mathcal{S}_{\text{unseen}}$ using established mobility flow metrics: common part of commuters, Pearson correlation coefficient, and Jensen-Shannon divergence. Summary results are presented in Table C.9. To further assess spatial distributions, we compute the entropy of disaggregated destination probabilities. These are compared with empirical distributions and those generated by MobilityGen. Across all metrics, MobilityGen consistently achieves the highest performance and most closely reproduces the observed spatial distribution of location visits, outperforming all baseline models. We attribute these gains to the integration of detailed spatial features (e.g., urban function distributions via LDA) alongside the joint modeling of contextual and behavioral factors.

Appendix D. Sensitivity and generalization analyses

D.1. Sensitivity to sequence length

To assess whether MobilityGen's performance depends on the chosen sequence lengths, we conduct sensitivity analyses for both the input traveled sequence length k_2 and the generated target length k_1 . For sensitivity to k_2 , the generated target horizon is fixed at 50 future activity events while the number of observed input events is varied. For sensitivity to k_1 , both generated and empirical target sequences are capped to the same number of future events. In both analyses, we report the corresponding calendar days covered by each event horizon to clarify the effective temporal extent of each setting.

Longer input histories generally improve generation quality (Table D.10). Increasing k_2 from 2 to 50 events substantially reduces the Wasserstein distances across most location and activity-travel metrics, indicating that additional observed history helps the model

recover individual spatial preferences and temporal regularities. The original variable-length traveled setting performs best or close to best for most metrics, suggesting that the model benefits from preserving longer observed histories whenever they are available, and that sufficient input helps infer individual preferences more reliably. This motivates future work on more compact representations of individual preferences, such as learned user embeddings or summary statistics, that could maintain generation quality with shorter observed inputs.

Table D.10

Sensitivity of MobilityGen performance to the input traveled length k_2 . Wasserstein distances are reported for location and activity-travel metrics; lower values indicate better alignment with empirical target sequences. The target horizon is fixed to 50 events, and days covered by the input traveled sequence are reported as mean $\pm$ SD. “Variable” denotes the original variable-length traveled sequence used in the main evaluation.

k_2	Days covered	Location metrics					Activity-travel metrics					
		f_k	$f_{\Delta r}$	f_{r_u}	S_{unc}	S_{temp}	f_d	f_t	f_k^{day}	$f_{ k }^{day}$	f_{motifs}	f_m
2	0.5 $\pm$ 0.4	1.27	27,567	7107	1.00	0.72	0.72	1.57	0.68	1.08	0.89	0.17
5	1.2 $\pm$ 0.8	0.63	15,457	5680	0.42	0.28	0.37	1.55	0.44	0.62	0.52	0.13
20	4.6 $\pm$ 2.0	0.32	5146	1797	0.14	0.06	0.36	0.69	0.41	0.29	0.18	0.08
50	12.2 $\pm$ 3.9	0.32	3171	647	0.12	0.05	0.41	0.65	0.35	0.25	0.14	0.08
Variable	19.9 $\pm$ 2.6	0.26	2318	575	0.08	0.04	0.47	0.52	0.29	0.22	0.14	0.05

The 50-event horizon provides a stable multi-day evaluation setting (Table D.11). At very short horizons, some location metrics yield deceptively small distances because only limited mobility diversity is expressed, while sequence-based metrics such as temporal entropy and motifs become undefined or unstable. At longer horizons, generation remains feasible, but displacement and radius of gyration measures show larger deviations at 100 events, suggesting that maintaining spatial consistency becomes more challenging as the generated horizon increases. Future work could investigate generation strategies that better support longer horizons, for example by training on longer sequences or generating in successive segments.

Table D.11

Sensitivity of MobilityGen performance to the generated target length k_1 . Wasserstein distances are reported for location and activity-travel metrics; lower values indicate better alignment with empirical target sequences capped to the same number of future events. The 50-event horizon is the default setting, and days covered by the target sequence are reported as mean $\pm$ SD. “–” indicates metrics that are undefined or unstable for very short event sequences.

k_1	Days covered	Location metrics					Activity-travel metrics					
		f_k	$f_{\Delta r}$	f_{r_u}	S_{unc}	S_{temp}	f_d	f_t	f_k^{day}	$f_{ k }^{day}$	f_{motifs}	f_m
2	0.8 $\pm$ 0.6	0.05	2736	1440	0.07	–	4.72	4.05	0.28	0.19	–	0.16
5	1.7 $\pm$ 1.0	0.01	2437	768	0.01	0.02	2.70	2.69	0.56	0.32	0.05	0.15
20	4.8 $\pm$ 2.0	0.20	2428	539	0.09	0.06	0.35	0.71	0.49	0.34	0.13	0.11
50	12.7 $\pm$ 4.0	0.26	2318	575	0.08	0.04	0.47	0.52	0.29	0.22	0.14	0.05
100	20.5 $\pm$ 7.3	0.15	4050	1533	0.03	0.02	0.60	0.40	0.27	0.17	0.14	0.10

D.2. Generalization to unseen users

The main evaluation uses a temporal split over users who appear in both the training and test sets. This setting evaluates how well the model continues held-out future trajectories, but performance may partly reflect extrapolation from previously observed individual routines. To assess whether MobilityGen captures mobility structure beyond known individuals, we conduct a held-out-individual experiment in which 150 users are reserved for validation and 150 for testing, with their mobility trajectories removed from the training set. The resulting evaluation contains two regimes. The seen-user setting evaluates held-out future days of training users, as in the main evaluation, while the unseen-user setting uses trajectories from users entirely absent from training. In both regimes, the model is conditioned on the source observation window and generates the next 50 activity-travel events.

A comparison between seen and unseen users must account for coverage of the discrete location space. MobilityGen learns representations for S2 grid cells from locations observed during training; cells absent from the training trajectories receive no positive training examples. Limited coverage can therefore degrade generation quality in addition to the challenge of transferring to a new individual. For each user, we compute source-location coverage as the fraction of unique locations in their source sequence that appeared in the training data; the mean coverage among seen users is 94.8%. We accordingly divide the unseen users at 95% coverage into a high-coverage subgroup ($n = 58$), which most closely matches the seen-user setting in embedding support, and a lower-coverage subgroup ($n = 92$), which tests robustness when a smaller proportion of the user’s source locations was observed during training. Together, the two subgroups comprise all 150 held-out users.

Baseline models are evaluated under the same information constraint as MobilityGen. Each is conditioned only on the source sequence available at test time and generates the next 50 events. For mechanistic baselines, individual parameters are estimated

Table D.12

Comparison between MobilityGen and baseline models for unseen users with high ($\geq 95\%$) and lower ($< 95\%$) source-location coverage. Together, the two subgroups comprise all 150 held-out users; the seen-user continuation setting is included as a reference. Wasserstein distances are reported for location and activity-travel metrics. “—” indicates metrics not defined for baselines that do not generate the corresponding attribute.

Method	Location metrics					Activity-travel metrics					
	f_k	$f_{\Delta r}$	f_{r_u}	S_{unc}	S_{temp}	f_d	f_t	f_k^{day}	$f_{ k }^{\text{day}}$	f_{motifs}	f_m
<i>High source-location coverage ($\geq 95\%$, $n = 58$ users)</i>											
EPR	1.15	9204	5671	0.45	0.35	—	—	—	—	—	—
Container	0.58	3055	1619	0.19	0.20	—	—	—	—	—	—
Markov	1.45	1607	8789	0.83	0.76	—	—	—	—	—	—
MHSA	1.51	12,800	14,102	0.31	0.26	—	—	—	—	—	—
MoveSim	7.21	18,063	22,255	1.01	0.90	—	—	—	—	—	—
DITRAS	0.46	9740	5615	0.14	0.15	1.94	1.72	0.70	0.27	0.50	—
TimeGeo	1.12	13,650	10,487	0.39	0.41	2.89	1.98	1.84	1.57	1.54	—
MobilityGen	0.67	7048	5375	0.15	0.21	0.35	0.94	0.19	0.11	0.17	0.29
<i>Lower source-location coverage ($< 95\%$, $n = 92$ users)</i>											
EPR	1.70	9876	5998	0.47	0.37	—	—	—	—	—	—
Container	0.49	4789	2202	0.11	0.11	—	—	—	—	—	—
Markov	2.14	1828	9608	0.88	0.80	—	—	—	—	—	—
MHSA	2.14	17,112	18,661	0.39	0.28	—	—	—	—	—	—
MoveSim	9.27	38,692	39,841	1.17	0.95	—	—	—	—	—	—
DITRAS	0.55	11,202	6811	0.15	0.17	2.10	1.41	0.81	0.31	0.41	—
TimeGeo	2.11	15,015	11,532	0.52	0.44	2.48	1.57	1.56	1.68	1.46	—
MobilityGen	1.32	20,363	12,687	0.32	0.31	0.92	1.15	0.32	0.18	0.28	0.08
<i>Seen-user continuation reference</i>											
MobilityGen (seen users)	0.65	3236	1146	0.18	0.12	0.78	0.46	0.45	0.29	0.26	0.09

from the source sequence, while population-level priors are fitted exclusively on the seen training users. This protocol prevents any baseline from accessing information unavailable to MobilityGen.

Table D.12 reports the results across the two coverage subgroups. Among the baselines, the learning-based location-only models perform poorly for unseen users, whereas several mechanistic baselines retain comparatively stable spatial performance across coverage levels. In the high-coverage subgroup, MobilityGen remains close to the seen-user reference on visitation frequency and uncorrelated entropy and outperforms DITRAS and TimeGeo across all comparable activity-travel metrics. At the same time, larger differences remain for displacement and radius of gyration, suggesting that measures of spatial extent are sensitive to the shift from seen to unseen users.

In the lower-coverage subgroup, the location metrics deteriorate further, particularly the distance-based measures, and the strongest mechanistic baselines, Container and DITRAS, achieve lower errors across all location metrics. Their smaller changes across coverage subgroups suggest that the deterioration of MobilityGen is partly associated with its dependence on learned location representations. However, because the two subgroups may also differ in their underlying mobility characteristics, the effect cannot be attributed exclusively to embedding coverage. By contrast, MobilityGen continues to outperform DITRAS and TimeGeo across all comparable activity-travel metrics. Activity-travel regularities therefore appear to transfer more robustly than spatial usage patterns.

Overall, the held-out-individual experiment provides a stronger test of MobilityGen’s generative capability and supports two conclusions. First, the model captures population-level activity-travel regularities beyond the continuation of known individual routines. Second, its generalization to unseen users remains limited for spatial behavior and deteriorates substantially when source-location coverage is lower. The gap between the seen-user reference and the high-coverage subgroup suggests that performance in the main setting benefits from user-specific information learned during training, while the additional deterioration in the lower-coverage subgroup is consistent with sensitivity to missing embedding support. These effects point to complementary directions for future work. Pretrained geographic representations could provide informative features for cells absent from the training trajectories, while conditioning on explicit user attributes or representations learned from the source observations could improve adaptation to individual variation.

Appendix E. Deriving experienced income segregation

To measure income segregation, we divide individuals into four income quartiles (equal-sized groups) according to their reported monthly household income: low ($< 4,000$ CHF), lower-middle ($4,000 - 8,000$ CHF), upper-middle ($8,000 - 12,000$ CHF), and high ($> 12,000$ CHF). We then apply the experienced income segregation framework (Moro et al., 2021), described in detail below.

At the location level, segregation is computed from the proportion of time individuals in income quartile q spend at each location l , denoted by τ_{ql} . Assuming that a fully integrated location corresponds to an equal distribution of time spent across quartiles (i.e., $\tau_{ql} = \frac{1}{4}$), the location-level segregation score S_l is defined as: $S_l = \frac{2}{3} \sum_q \left| \tau_{ql} - \frac{1}{4} \right|$. This score, ranging from 0 (perfect integration) to 1 (complete segregation), quantifies the imbalance in visitation by income groups.

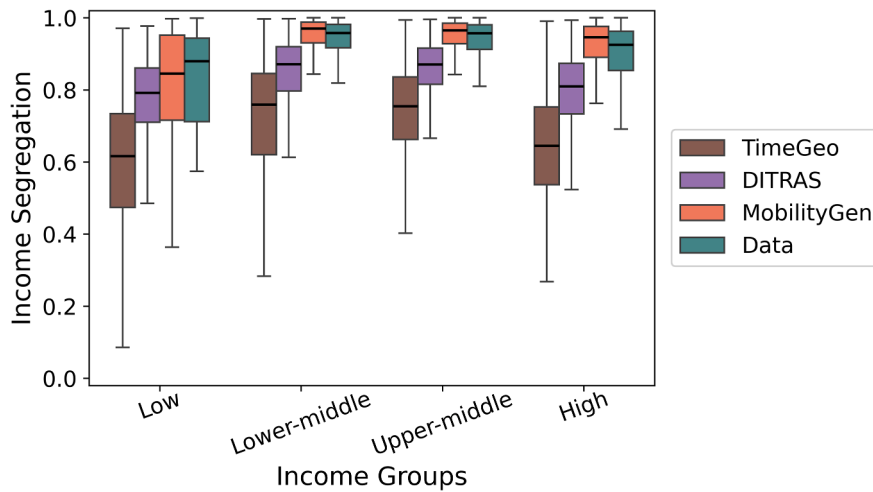


Fig. E.15. Variation in income segregation across groups. Boxplots of individual-level income segregation scores for four income groups (low, lower-middle, upper-middle, and high), comparing observed data (teal) with sequences generated by TimeGeo (brown), DITRAS (purple), and MobilityGen (red). MobilityGen closely reproduces the empirical distributions and outperforms the baseline simulators in capturing variation in segregation levels across groups.

To measure segregation at the individual level, we calculate each person's relative exposure to income quartiles based on the locations they visit. Let τ_{ul} represent the proportion of time individual u spends at location l . Their exposure to quartile q , denoted by τ_{uq} , is then: $\tau_{uq} = \sum_l \tau_{ul} \tau_{ql}$. Using this, the individual-level segregation score S_u is defined analogously to the location-level score:

$$S_u = \frac{2}{3} \sum_q \left| \tau_{uq} - \frac{1}{4} \right| \quad (\text{E.1})$$

As with S_l , a score of $S_u = 0$ indicates perfectly balanced exposure across income groups, while $S_u = 1$ reflects exclusive exposure to a single group.

This formulation approximates co-presence by averaging the time spent at each location across income groups, rather than relying on exact co-occurrence events. As such events are extremely sparse in real mobility data, this time-invariant approach provides a practical way to estimate individual-level interaction patterns. We compute individual-level income segregation scores for each real and simulated event sequence in the testing set. Since multiple sequences may originate from the same individual (and thus the same income group), and often involve repeated visits to similar locations, the resulting scores do not reflect population-wide segregation within the study area. Instead, they serve as a validation metric to assess how well MobilityGen reproduces the co-presence patterns observed in the real data.

References

- Abbasov, T., Heine, C., Sabouri, S., Salazar-Miranda, A., Santi, P., Glaeser, E., Ratti, C., 2024. The 15-minute city quantified using human mobility data. *Nat. Hum. Behav.* 8, 445–455. <https://doi.org/10.1038/s41562-023-01770-y>
- Alessandretti, L., Aslak, U., Lehmann, S., 2020. The scales of human mobility. *Nature* 587 (7834), 402–407. <https://doi.org/10.1038/s41586-020-2909-1>
- 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. <https://doi.org/10.1038/s41562-018-0364-x>
- Arentze, T.A., Timmermans, H.J.P., 2005. Information gain, novelty seeking and travel: a model of dynamic activity-travel behavior under conditions of uncertainty. *Transp. Res. A Policy Pract.* 39 (2), 125–145. <https://doi.org/10.1016/j.tra.2004.08.002>
- Ashbrook, D., Starner, T., 2002. Learning significant locations and predicting user movement with GPS. In: *Proceedings of the 6th International Symposium on Wearable Computers (ISWC '02)*, pp. 101–108. <https://doi.org/10.1109/ISWC.2002.1167224>
- Auld, J., Mohammadian, A.K., 2012. Activity planning processes in the agent-based dynamic activity planning and travel scheduling (ADAPTS) model. *Transp. Res. A Policy Pract.* 46 (8), 1386–1403. <https://doi.org/10.1016/j.tra.2012.05.017>
- Austin, J., Johnson, D.D., Ho, J., Tarlow, D., van den Berg, R., 2021. Structured denoising diffusion models in discrete state-spaces. In: *35th Conference on Neural Information Processing Systems (NeurIPS '21)*, pp. 17981–17993.
- Axhausen, K.W., 2007. Definition of movement and activity for transport modelling. In: *Handbook of Transport Modelling*. Emerald Group Publishing Limited, pp. 329–343. <https://doi.org/10.3929/ethz-a-005278091>
- Axhausen, K.W., Gärling, T., 1992. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transp. Rev.* 12 (4), 323–341. <https://doi.org/10.1080/01441649208716826>
- Ba, J.L., Kiro, J.R., Hinton, G.E., 2016. Layer Normalization. [arXiv:1607.06450](https://arxiv.org/abs/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. <https://doi.org/10.1016/j.physrep.2018.01.001>
- Barbosa, H., de Lima-Neto, F.B., Evsukoff, A., Menezes, R., 2015. The effect of recency to human mobility. *EPJ Data Sci.* 4 (1), 21. <https://doi.org/10.1140/epjds/s13688-015-0059-8>
- Barreras, F., Watts, D.J., 2024. The exciting potential and daunting challenge of using GPS human-mobility data for epidemic modeling. *Nat. Comput. Sci.* 4 (6), 398–411. <https://doi.org/10.1038/s43588-024-00637-0>
- Bhat, C.R., Singh, S.K., 2000. A comprehensive daily activity-travel generation model system for workers. *Transp. Res. A Policy Pract.* 34 (1), 1–22. [https://doi.org/10.1016/S0965-8564\(98\)00037-8](https://doi.org/10.1016/S0965-8564(98)00037-8)

- Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature* 439 (7075), 462–465. <https://doi.org/10.1038/nature04292>
- Burbridge, S., Goulias, K., 2009. Active travel behavior. *Transp. Lett.* 1 (2), 147–167. <https://doi.org/10.3328/TL.2009.01.02.147-167>
- Cabanas-Tirapu, O., Danús, L., Moro, E., Sales-Pardo, M., Guimerà, R., 2025. Human mobility is well described by closed-form gravity-like models learned automatically from data. *Nat. Commun.* 16 (1), 1336. <https://doi.org/10.1038/s41467-025-56495-5>
- Chu, C., Zhang, H., Wang, P., Lu, F., 2024. Simulating human mobility with a trajectory generation framework based on diffusion model. *Int. J. Geogr. Inf. Sci.* 38 (5), 847–878. <https://doi.org/10.1080/13658816.2024.2312199>
- Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. *SIAM Rev.* 51 (4), 661–703. <https://doi.org/10.1137/070710111>
- Cuttone, A., Lehmann, S., González, M.C., 2018. Understanding predictability and exploration in human mobility. *EPJ Data Sci.* 7 (1), 2. <https://doi.org/10.1140/epjds/s13688-017-0129-1>
- Deshpande, A., Aneja, J., Wang, L., Schwing, A.G., Forsyth, D.A., 2019. Fast, diverse and accurate image captioning guided by part-of-speech. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR '19)*, pp. 10695–10704. <https://doi.org/10.1109/CVPR.2019.01095>
- Feng, J., Yang, Z., Xu, F., Yu, H., Wang, M., Li, Y., 2020. Learning to simulate human mobility. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*, pp. 3426–3433. <https://doi.org/10.1145/3394486.3412862>
- Gambis, S., Killijian, M.-O., del Prado Cortez, M.N., 2012. Next place prediction using mobility Markov chains. In: *Proceedings of the First Workshop on Measurement, Privacy, and Mobility (MPM '12)*, pp. 1–6. <https://doi.org/10.1145/2181196.2181199>
- Gong, S., Li, M., Feng, J., Wu, Z., Kong, L., 2023. DiffuSeq: sequence to sequence text generation with diffusion models. In: *11th International Conference on Learning Representations (ICLR '23)*.
- González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature* 453 (7196), 779–782. <https://doi.org/10.1038/nature06958>
- Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. *Am. J. Prev. Med.* 23 (2), 64–73. [https://doi.org/10.1016/S0749-3797\(02\)00475-0](https://doi.org/10.1016/S0749-3797(02)00475-0)
- Hanson, S., Hanson, P., 1981. The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics. *Econ. Geogr.* 57 (4), 332–347. <https://doi.org/10.2307/144213>
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '16)*, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580*
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. In: *34th Conference on Neural Information Processing Systems (NeurIPS '20)*, pp. 6840–6851.
- Holtzman, A., Buys, J., Du, L., Forbes, M., Choi, Y., 2020. The curious case of neural text degeneration. In: *8th International Conference on Learning Representations (ICLR '20)*.
- Hong, J., Shen, Q., Zhang, L., 2014. How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales. *Transportation* 41, 419–440. <https://doi.org/10.1007/s11116-013-9462-9>
- 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. C Emerg. Technol.* 146, 103979. <https://doi.org/10.1016/j.trc.2022.103979>
- Hong, Y., Zhang, Y., Schindler, K., Raubal, M., 2023b. Context-aware multi-head self-attentional neural network model for next location prediction. *Transp. Res. C Emerg. Technol.* 156, 104315. <https://doi.org/10.1016/j.trc.2023.104315>
- Hörl, S., Balac, M., 2021. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transp. Res. C Emerg. Technol.* 130, 103291. <https://doi.org/10.1016/j.trc.2021.103291>
- Jiang, B., Yin, J., Zhao, S., 2009. Characterizing the human mobility pattern in a large street network. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 80 (2), 021136. <https://doi.org/10.1103/PhysRevE.80.021136>
- Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., González, M.C., 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proc. Natl. Acad. Sci.* 113 (37), E5370–E5378. <https://doi.org/10.1073/pnas.1524261113>
- Li, T., Biferale, L., Bonaccorso, F., Scarpolini, M.A., Buzzicotti, M., 2024. Synthetic Lagrangian turbulence by generative diffusion models. *Nat. Mach. Intell.* 6 (4), 393–403. <https://doi.org/10.1038/s42256-024-00810-0>
- Li, X., Thickstun, J., Gulrajani, I., Liang, P.S., Hashimoto, T.B., 2022. Diffusion-LM improves controllable text generation. In: *36th Conference on Neural Information Processing Systems (NeurIPS '22)*, pp. 4328–4343.
- Liao, Y., Gil, J., Yeh, S., Pereira, R.H.M., Alessandretti, L., 2025. Socio-spatial segregation and human mobility: a review of empirical evidence. *Comput. Environ. Urban Syst.* 117, 102250. <https://doi.org/10.1016/j.compenvurbsys.2025.102250>
- Liu, Y., Wang, S., Dong, J., Chen, L., Wang, X., Wang, L., Li, F., Wang, C., Zhang, J., Wang, Y., Wei, S., Chen, Q., Liu, H., 2024. De novo protein design with a denoising diffusion network independent of pretrained structure prediction models. *Nat. Methods* 21, 2107–2116. <https://doi.org/10.1038/s41592-024-02437-w>
- Liu, Z., Liu, C., Mostafavi, A., 2023. Beyond residence: a mobility-based approach for improved evaluation of human exposure to environmental hazards. *Environ. Sci. Technol.* 57 (41), 15511–15522. <https://doi.org/10.1021/acs.est.3c04691>
- Long, Q., Yuan, Y., Li, Y., 2025. A universal model for human mobility prediction. In: *Proceedings of the 31st ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '25)*, pp. 894–905. <https://doi.org/10.1145/3690624.3709236>
- Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. In: *7th International Conference on Learning Representations (ICLR '19)*.
- Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., Zhu, J., 2022. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. In: *36th Conference on Neural Information Processing Systems (NeurIPS '22)*, pp. 5775–5787.
- Luca, M., Barlacchi, G., Lepri, B., Pappalardo, L., 2021. A survey on deep learning for human mobility. *ACM Comput. Surv.* 55 (1), 7:1–7:44. <https://doi.org/10.1145/3485125>
- Mai, G., Janowicz, K., Yan, B., Zhu, R., Cai, L., Lao, N., 2020. Multi-scale representation learning for spatial feature distributions using grid cells. In: *8th International Conference on Learning Representations (ICLR '20)*.
- Martin, H., Hong, Y., Wiedemann, N., Bucher, D., Raubal, M., 2023. Trackintel: An open-source Python library for human mobility analysis. *Comput. Environ. Urban Syst.* 101, 101938. <https://doi.org/10.1016/j.compenvurbsys.2023.101938>
- McInnes, L., Healy, J., Saul, N., Grossberger, L., 2018. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* 3 (29), 861. <https://doi.org/10.21105/joss.00861>
- Mo, B., Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2022. Individual mobility prediction in mass transit systems using smart card data: an interpretable activity-based hidden Markov approach. *IEEE Trans. Syst.* 23 (8), 12014–12026. <https://doi.org/10.1109/TITS.2021.3109428>
- Molloy, J., Castro, A., Götschi, T., Schoeman, B., Tchervenkov, C., Tomic, U., Hintermann, B., Axhausen, K.W., 2023. The MOBIS dataset: a large GPS dataset of mobility behaviour in Switzerland. *Transportation* 50 (5), 1983–2007. <https://doi.org/10.1007/s11116-022-10299-4>
- Moro, E., Calacci, D., Dong, X., Pentland, A., 2021. Mobility patterns are associated with experienced income segregation in large US cities. *Nat. Commun.* 12 (1), 4633. <https://doi.org/10.1038/s41467-021-24899-8>
- Narayan, A., Berger, B., Cho, H., 2021. Assessing single-cell transcriptomic variability through density-preserving data visualization. *Nat. Biotechnol.* 39 (6), 765–774. <https://doi.org/10.1038/s41587-020-00801-7>
- Nilforoshan, H., Looi, W., Pierson, E., Villanueva, B., Fishman, N., Chen, Y., Sholar, J., Redbird, B., Grusky, D., Leskovec, J., 2023. Human mobility networks reveal increased segregation in large cities. *Nature* 624 (7992), 586–592. <https://doi.org/10.1038/s41586-023-06757-3>
- Niu, H., Silva, E.A., 2021. Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London. *Comput. Environ. Urban Syst.* 88, 101651. <https://doi.org/10.1016/j.compenvurbsys.2021.101651>
- Panaretos, V.M., Zemel, Y., 2019. Statistical Aspects of Wasserstein Distances. *Annu. Rev. Stat. Appl.* 6 (1), 405–431. <https://doi.org/10.1146/annurev-statistics-030718-104938>

- Papineni, K., Roukos, S., Ward, T., Zhu, W.-J., 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL '02), pp. 311–318. <https://doi.org/10.3115/1073083.1073135>
- Pappalardo, L., Manley, E., Sekara, V., Alessandretti, L., 2023. Future directions in human mobility science. *Nat. Comput. Sci.* 3 (7), 588–600. <https://doi.org/10.1038/s43588-023-00469-4>
- Pappalardo, L., Simini, F., 2018. Data-driven generation of spatio-temporal routines in human mobility. *Data Min. Knowl. Discov.* 32 (3), 787–829. <https://doi.org/10.1007/s10618-017-0548-4>
- 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. <https://doi.org/10.1038/ncomms9166>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pougala, J., Hillel, T., Bierlaire, M., 2023. OASIS: Optimisation-based activity scheduling with integrated simultaneous choice dimensions. *Transp. Res. C Emerg. Technol.* 155, 104291. <https://doi.org/10.1016/j.trc.2023.104291>
- Reck, D.J., Martin, H., Axhausen, K.W., 2022. Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transp. Res. D Transp. Environ.* 102, 103134. <https://doi.org/10.1016/j.trd.2021.103134>
- Savcisen, G., Eliassi-Rad, T., Hansen, L.K., Mortensen, L.H., Lilleholt, L., Rogers, A., Zettler, I., Lehmann, S., 2024. Using sequences of life-events to predict human lives. *Nat. Comput. Sci.* 4 (1), 43–56. <https://doi.org/10.1038/s43588-023-00573-5>
- Schläpfer, M., Dong, L., O'Keefe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., West, G.B., 2021. The universal visitation law of human mobility. *Nature* 593 (7860), 522–527. <https://doi.org/10.1038/s41586-021-03480-9>
- 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. <https://doi.org/10.1098/rsif.2013.0246>
- Sener, O., Koltun, V., 2018. Multi-task learning as multi-objective optimization. In: 32nd Conference on Neural Information Processing Systems (NeurIPS '18), pp. 525–536.
- Simini, F., Barlacchi, G., Luca, M., Pappalardo, L., 2021. A deep gravity model for mobility flows generation. *Nat. Commun.* 12 (1), 6576. <https://doi.org/10.1038/s41467-021-26752-4>
- Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: Proceedings of the 32nd International Conference on Machine Learning (ICML '15), pp. 2256–2265.
- Song, C., Koren, T., Wang, P., Barabási, A.-L., 2010a. Modelling the scaling properties of human mobility. *Nat. Phys.* 6 (10), 818–823. <https://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. <https://doi.org/10.1126/science.1177170>
- Song, Y., Ding, J., Yuan, J., Liao, Q., Li, Y., 2024. Controllable human trajectory generation using profile-guided latent diffusion. *ACM Trans. Knowl. Discov. Data* 19 (1), 22:1–22:25. <https://doi.org/10.1145/3701736>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS '17), pp. 5998–6008.
- Xu, Y., Çolak, S., Kara, E.C., Moura, S.J., González, M.C., 2018. Planning for electric vehicle needs by coupling charging profiles with urban mobility. *Nat. Energy* 3 (6), 484–493. <https://doi.org/10.1038/s41560-018-0136-x>
- Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.-H., 2023. Diffusion models: a comprehensive survey of methods and applications. *ACM Comput. Surv.* 56 (4), 105:1–105:39. <https://doi.org/10.1145/3626235>
- Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* 31 (4), 825–848. <https://doi.org/10.1080/13658816.2016.1244608>
- Zhang, S., Zhen, F., Kong, Y., Lobsang, T., Zou, S., 2023. Towards a 15-minute city: a network-based evaluation framework. *Environ. Plan. B Urban Anal. City Sci.* 50 (2), 500–514. <https://doi.org/10.1177/23998083221118570>
- 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), 5. <https://doi.org/10.1140/epjds/s13688-021-00261-2>
- Zhao, P., Wang, H., Liu, Q., Yan, X.-Y., Li, J., 2024. Unravelling the spatial directionality of urban mobility. *Nat. Commun.* 15 (1), 4507. <https://doi.org/10.1038/s41467-024-48909-7>
- Zipf, G.K., 1946. The P1 P2/D hypothesis: on the intercity movement of persons. *Am. Sociol. Rev.* 11 (6), 677–686. <https://doi.org/10.2307/2087063>