Towards realistic individual activity location demand synthesis using deep generative networks

Ye Hong

Yatao Zhang

Martin Raubal

STRC conference paper 2024 **April 30, 2024**

STRC | **24th Swiss Transport Research Conference**
Monte Verità / Ascona, May 15-17, 2024

Towards realistic individual activity location demand synthesis using deep generative networks

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

Yatao Zhang
Future Resilient Systems
Singapore-ETH Centre
yatzhang@ethz.ch

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

April 30, 2024

Abstract

Traffic modeling and transport planning applications increasingly require a pipeline for synthesizing individual travel demands, including daily activity patterns and their location realizations. Unfortunately, the capacity to simulate realistic location sequences is still constrained by the vast option space and the flexibility in choosing an activity location. Traditional demand synthesis pipelines typically use the space-time prism approach, sampling locations from a constrained candidate location set, resulting in oversimplified location patterns. Deep generative models, such as generative adversarial networks and diffusion networks, have shown their potential, mostly on small-scale datasets and without explicitly modeling the spatio-temporal characteristics of human mobility. Here, we employ state-of-the-art deep generative networks on individual GNSS tracking datasets to simulate location sequences at a large scale, which adheres to location visit patterns at both individual and population levels. The presented approach introduces new possibilities for activity location generation. The framework has the potential to simulate comprehensive activity patterns by incorporating various dimensions of travel behavior and integration with other tasks in the demand generation pipeline, such as population synthesis. Additionally, realistic spatio-temporal trajectories can be applied in related fields requiring knowledge of individual movements.

Keywords

Travel demand; Activity-location; Human mobility; Deep learning; Generative model; Sequence generation

1 Introduction

Realistic individual travel demands are essential for fine-grained traffic modeling and transport planning applications, such as the deployment of agent-based models, which have gained widespread attention among researchers and practitioners (Horni *et al.*, 2016), and the increasingly popular digital twin concept (Zhou *et al.*, 2022). Individual travel demands encompass diverse daily activity patterns and location realizations that accurately resemble individual behavior while collectively following the population distribution. Location realizations can be viewed as spatio-temporal trajectories that describe the movements of individuals and have found applications in various related fields such as energy demand estimation, environmental exposure modeling, and epidemic control (Dodge *et al.*, 2020). Despite its importance, synthesizing realistic location sequences remains challenging, constrained by the vast option space and the inherent flexibility in choosing an activity location.

Previous attempts in location synthesis can be identified across diverse yet related fields, each concentrating on a distinct aspect of the problem. Traditional travel demand synthesis pipelines typically perform location realization conditioned on simulated activity patterns, employing the space-time prism approach to sample locations from a constrained candidate location set (Hörl and Balac, 2021). While creating locations that fulfill activity-travel patterns, the approach involves sampling locations one at a time for a short period (typically one day), overlooking spatial patterns and interdependencies in location visits. Researchers in statistical physics have developed mechanistic mobility models based on human digital traces, simulating location sequences with realistic mobility patterns (Schläpfer *et al.*, 2021). However, modeling the relationship between location choice and other activity-travel behavior aspects is not straightforward, hindering the development of these models for travel demand synthesis. Deep learning approaches have been introduced to location synthesis (Feng *et al.*, 2020), particularly through the use of deep generative models such as generative adversarial networks (GANs) and diffusion networks, which were initially developed for image and text generation. With abundant flexibility and capability, these generative networks are expected to provide realistic location realizations by appropriately modeling their spatio-temporal characteristics; however, existing studies mainly focus on small-scale datasets, and their full potential has yet to be fully realized.

This study showcases the effectiveness of utilizing state-of-the-art deep generative networks for synthesizing individual activity locations based on a large-scale longitudinal GNSS tracking dataset. We introduce two paradigms for simulating new location sequences that adhere to existing location visit behavior: 1) the next location prediction network with sampling strategies, and 2) the conditional diffusion network. Mobility metrics reflecting individual- and population-level location visit patterns and distribution distance measures are utilized to assess the quality of the resulting sequences. Additionally, we discuss promising future steps to extend the proposed methods toward realizing more realistic activity locations and synthesizing complete activity travel demands. The presented approach introduces new possibilities for activity demand generation, and we believe that accurate location synthesis can greatly benefit applications in related fields that require knowledge of individual movements.

2 Methods

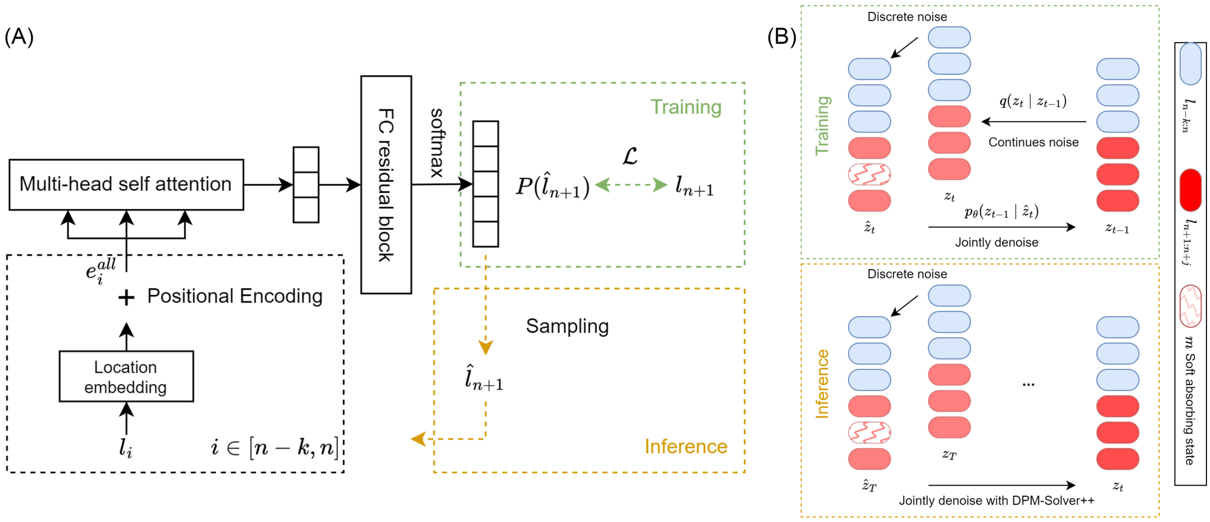
We first define the primary analysis unit of activity location and then introduce the problem of activity location sequence synthesis.

Definition 1 (Activity location) *An activity location l is defined as an area in space where an individual stays for an extended time to perform an activity. l uniquely identifies a location and can be represented as a categorical identifier or the geometry of the location.*

Problem 1 (Location sequence synthesis) *Consider a time-ordered location sequence $(l_i)_{i=n-k}^{n}$ visited by user u^i in a time window with k step preceding the current time step n, the goal is to simulate realistic location sequences $(l_i)_{i=n+1}^{n+j}$ with j steps into the future conditioned on the previous sequence.*

In this study, we define the context time window k to include all historical location visits within the past two weeks, and set j to 50 steps for the length of the simulated location sequence. The discussion on context and simulated sequence length is reserved for future studies. We introduce two deep generative approaches to tackle this problem, with their general framework shown in Figure 1. In the following, we provide a detailed description of each approach.

Figure 1: Overview of the proposed deep generative models for activity location sequence simulation. (A) The MHSA network, trained for next activity location prediction, performs autoregressive generation with sampling strategies. Figure adopted from Hong *et al.* (2023b). (B) The DiffuSeq-v2 model, trained to recover the sequence from corrupted noises, generates locations parallelly in a non-autoregressive way. Figure adopted from Gong *et al.* (2023b).



2.1 Next location prediction network with sampling strategies

The remarkable success and widespread adoption of the generative pre-trained transformer (GPT)-2 framework (Radford *et al.*, 2018), along with its subsequent development like ChatGPT, have showcased the transformer networks' capability to autoregressively generate meaningful sentences through pre-training, simply by predicting the next word. As location sequences can be regarded as language sentences, with each location representing a word, location sequence simulation is analogous to language modeling tasks and can be tackled by pre-training a transformer network for next location prediction. Therefore, we implement a next location prediction network following the model described in Hong *et al.* (2023b) and utilize sampling strategies to balance diversity and reduce sampling unlikely locations in the inference stage (Figure 1A). Formally, during training, network parameters are optimized using maximum likelihood estimation with the cross-entropy loss $\mathcal{L}$ given the ground truth next location l_{n+1}:

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

where $\mathcal{O}$ is the set containing all locations, $P(\hat{l}_{n+1})^{(k)}$ represents the predicted probability of visiting the k-th location and $P(l_{n+1})^{(k)}$ is the one-hot represented ground truth, i.e.,

$P(l_{n+1})^{(k)} = 1$ if the true next location is the k-th location, and $P(l_{n+1})^{(k)} = 0$ otherwise. During inference, the next location $\hat{l}_s$ (with $n+1 \leq s \leq n+j$) is sampled from $P(\hat{l}_s)$ and sequentially appended to the end of the original sequence to obtain the next prediction until the desired sequence length is reached. Thus, $P(\hat{l}_s)$ depends on the context location sequence $l_{n-k:n}$ and the preceding location predictions $\hat{l}_{n+1:s-1}$:

$$P(\hat{l}_s) = P(\hat{l}_s \mid \hat{l}_{n+1:s-1}, l_{n-k:n}) = P(\hat{l}_{n+1} \mid l_{n-k:n}) \prod_{i=n+2}^{s} P(\hat{l}_i \mid \hat{l}_{n+1:i-1}, l_{n-k:n}) \tag{2}$$

We sample $\hat{l}_s$ from $P(\hat{l}_s)$ using a combination of top-k and nucleus sampling strategies (Holtzman *et al.*, 2020). The former selects solely from the top-k most likely location predictions, while the latter selects exclusively from the most likely location predictions whose cumulative probability exceeds a predefined threshold p. We chose $k = 200$ and $p = 0.99$ in this study.

2.2 Conditional diffusion network

In recent years, diffusion models have emerged as a new paradigm for deep generative models (Ho *et al.*, 2020). Theoretically underpinned by non-equilibrium thermodynamics and score-matching networks, they have overcome several limitations of previous approaches and have led to significant breakthroughs in image and audio generations. Although the diffusion model's original formulation is defined in continuous space, researchers have adapted it to handle discrete entities such as texts (Li *et al.*, 2022). Subsequent studies extend this framework to conditional diffusion processes, enabling control over the generated sequences, making it suitable for sequence-to-sequence tasks (Gong *et al.*, 2023a,b). As location sequences can be represented in the same format as tests, we adjust the conditional diffusion network DiffSeq-v2 (Gong *et al.*, 2023b) for tackling the location sequence simulation task (Figure 1B).

The standard diffusion framework contains forward and reverse processes. Formally, given a data point sampled from a real-world data distribution $\mathbf{z}_0 \sim q(\mathbf{z})$, the forward process gradually corrupts $\mathbf{z}_0$ into a standard Gaussian noise $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$ in $t \in [1, 2, ..., T]$ steps. For each forward step, the perturbation is controlled by adding Gaussian noise, i.e., $q(\mathbf{z}_t \mid \mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t\mathbf{I})$, with $\beta_t \in (0, 1)$ as predefined variance schedules. Given the forward process, the reverse denoising process aims to recover $p_\theta(\mathbf{z}_{t-1} \mid \mathbf{z}_t)$

through learning a deep neural network f_θ[1], which can gradually reconstruct the original data $\mathbf{z}_0$ via sampling from $\mathbf{z}_T$. Among other modifications, diffusion-LM extends this standard framework by including an embedding function to map discrete text into a continuous space (Li *et al.*, 2022); DiffSeq applies Gaussian noise solely to the output sequence (i.e., $(l_i)_{i=n+1}^{n+j}$), preserving information in the context sequence to serve as a guiding signal (Gong *et al.*, 2023a); and DiffSeq-v2 further introduces discrete noise to the output sequence at a certain probability and leverages DPM-Solver++ to accelerate the inference speed (Figure 1B) (Gong *et al.*, 2023b).

The embedding that maps discrete locations into continuous space can either be learned from scratch alongside the diffusion process or pre-trained to incorporate prior information. We utilize the word2vec skip-gram framework (Mikolov *et al.*, 2013) to obtain the initial location embeddings, ensuring that spatially proximate locations are also close in the embedding space, thereby capturing spatial relation information. For each location, we consider its 8 closest neighbors as positive pairs and randomly sample 20 other locations as negative pairs.

3 Results

We chose a large-scale individual tracing dataset to show the capacity of the deep generative networks for location sequence simulation. The dataset is collected as part of the *Mobility Behaviour in Switzerland* (MOBIS) research project and the subsequent MOBIS-COVID study (Molloy *et al.*, 2022), which aim to assess the behavioral impact of mobility pricing and COVID-19 pandemic, respectively, in Switzerland. As part of the research project, ~ 3700 participants were asked to record their whereabouts for eight weeks starting in Autumn 2019 with a GNSS-based smartphone tracking app. After the main MOBIS study, participants were invited to continue the tracking for the MOBIS COVID study. We randomly selected 500 high-quality participants, each tracked for over 50 days with consistently high temporal tracking quality, for further analyses. Using the trackintel framework (Martin *et al.*, 2023), we generated locations from their raw GNSS traces, subsequently projecting them into Google's S2Geometry grid system at level 13. This process yielded 39,177 locations covering Switzerland, with 14,881 visited by our participants. Our focus lies solely on simulating sequences within these visited locations.

[1]We use the standard BERT model (Devlin *et al.*, 2019) as the backbone network.

We split processed tracking data into non-overlapping train, validation, and test sets in a ratio of 6:2:2 based on time, such that sequences occurring in the first 60% of tracking days for each user are regarded as train and the last 20% of days as test. We utilize the training set to optimize parameters of deep generative networks employing the AdamW optimizer (Loshchilov and Hutter, 2019), with an initial learning rate of $1e^{-3}$, while the validation set is employed to monitor network losses. An early stopping strategy is adopted to pause the learning if the validation loss stops decreasing for 3 epochs. Then, the learning rate is multiplied by 0.1, and training is continued from the model with the lowest validation loss. This early stopping process is repeated 3 times. We finally evaluate the model performances using the held-out test set.

We implement a set of baseline methods for performance comparison, including 1) exploration and preferential return (EPR) model (Song *et al.*, 2010), a mechanistic model designed to replicate location visitation patterns[2]; 2) Markov model (Gambs *et al.*, 2012), a next location prediction baseline that captures basic location transition patterns; and 3) MoveSim model (Feng *et al.*, 2020), a GAN-based location simulation method that has demonstrated success in small-scale datasets. We compare real and simulated location sequences by computing mobility metrics (Alessandretti *et al.*, 2020), focusing on high-level similarities of mobility patterns. The distributions of mobility metrics are shown in Figure 2, with the differences between real and simulated traces quantified using Jensen–Shannon divergence, as presented in Table 1.

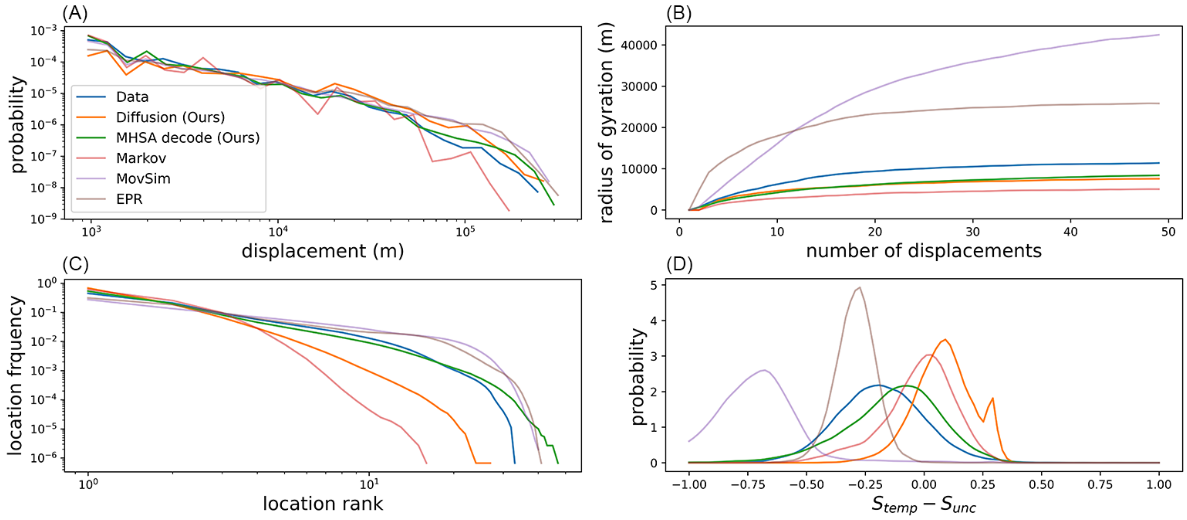
Table 1: Jensen–Shannon divergence between true and simulated traces for displacements (Δr), radius of gyration (R_g), individual location visitation frequency (f_k) and the difference between temporal and uncorrelated entropy (S_{diff}). MHSA decode consistently achieves the best performance (numbers marked in **bold**).

	Δr	R_g	f_k	S_{diff}
EPR (Song *et al.*, 2010)	0.27	0.36	0.17	0.39
MovSim (Feng *et al.*, 2020)	0.31	0.49	0.22	0.72
Markov (Gambs *et al.*, 2012)	0.39	0.36	0.25	0.34
Diffusion (ours)	0.39	0.24	0.23	0.53
MHSA decode (ours)	**0.18**	**0.22**	**0.07**	**0.14**

All generative models, with the exception of the Markov model, produce location traces

[2]More comprehensive mechanistic models, such as the TimeGeo framework (Jiang *et al.*, 2016) and the Container model (Alessandretti *et al.*, 2020), should be included for comparison in future studies.

Figure 2: Quantifying simulated location sequences with mobility metrics. We evaluate the ability of generative models to replicate realistic mobility patterns, showing close agreement between the proposed deep generative approaches and real traces. (A) The distribution of displacements for the population. (B) The median individual radius of gyration evolution against the number of displacements. (C) The visitation frequency of locations versus their rank averaged over individuals. (D) The distribution of the differences between the temporal entropy S_{temp} and the uncorrelated entropy S_{unc} across individuals.



that closely match the real ones in terms of displacement distribution[3] (Figure 2A). The Markov model struggles in replicating displacements over long distances, leading to a slower evolution of the radius of gyration compared to real traces (Figure 2B). Its assumption of capturing only basic sequence patterns (i.e., assuming Markovian property of location visits) resulted in a bias towards selecting frequent locations while overlooking less common ones (Figure 2C). However, the model's approach resulted in a satisfactory approximation of entropy (Figure 2D). In contrast, the EPR model tends to overemphasize location transitions over large distances, as evidenced by the more rapid increase in the radius of gyration and the higher frequencies of visiting less significant locations compared to the real data (Figure 2B and C). These deficiencies become more apparent in the traces generated by MovSim, resulting in poor replications in the entropy distribution (Figure 2D).

The two proposed deep generative modeling frameworks, the MHSA network with decoding strategies and the conditional denoising diffusion model, exhibit the closest agreement with the real evolution of the radius of gyration (Figure 2B). While the diffusion network tends

[3] We note that the current simulation can only differentiate displacements larger than $\sim 10^3$m due to the projection into the level 13 S2Geometry grid. While enhancing spatial resolution is feasible with a more fine-grained projection, it comes at the cost of increased computational burden.

to overemphasize the more frequently visited locations, MHSA closely mirrors the real location visitation frequency (Figure 2C). Moreover, the MHSA network effectively generates location sequences with realistic entropy distributions (Figure 2D). The evaluation results of the Jensen–Shannon divergence demonstrate that traces generated by MHSA consistently achieve the closest match to the collected data across all mobility metrics (Table 1). While the diffusion network exhibits satisfactory performances in the radius of gyration and individual location visitation frequency, it falls short in replicating realistic displacement and entropy distributions. In summary, straightforward integration of next location prediction networks with sampling strategies can already produce realistic location sequences, while the current popular diffusion network requires further development for mobility applications.

4 Conclusion and outlook

We have introduced two deep generative modeling approaches for synthesizing individual activity location sequences. The quality assessment results using high-level mobility metrics revealed that the proposed methods could generate realistic location visit patterns, outperforming various baselines at individual and population levels. In particular, employing a next-location prediction network with sampling strategies synthesizes sequences that closely match the real traces, while the conditional diffusion network excels at reproducing the evolution of the radius of gyration but still requires further adjustments to capture location visit patterns. As one of the initial attempts towards realistic activity pattern synthesis using deep learning, we envision several future directions worth exploring:

- Location sequence quality assessment. The implemented human mobility metrics offer a thorough evaluation of location visit patterns; however, they may not straightforwardly reflect sequential location choices (Hong *et al.*, 2023a). In light of this, mobility motifs (Schneider *et al.*, 2013) or trajectory similarity metrics (Tao *et al.*, 2021) should be considered for a comprehensive evaluation.
- Activity-travel demand generation. In addition to capturing location patterns, deep generative models have the potential to learn the intricate interplay between various dimensions of travel behavior and simulate complete activity-travel schedules simultaneously. Such joint learning can benefit from multi-modal and multi-task formulations currently under active development in related domains (Chen *et al.*, 2024).

- Context integration. The environmental context, such as urban land use functions and residential surroundings, profoundly impacts individual activity-travel schedules. Leveraging the abundant flexibility and capacity of deep learning networks, these travel-related contexts can be incorporated into the framework through appropriate context modeling and network architecture design (Hong *et al.*, 2023b).
- Controlled synthesis through considering social demographic attributes. A comprehensive activity travel demand requires replicating individual-level socio-demographic attributes for subsequent travel behavior analysis. This population synthesis task is increasingly approached using deep generative modeling (Kim and Bansal, 2023). Exploring methods to integrate population with activity demand generation into a unified deep learning framework is a promising direction.

References

Alessandretti, L., U. Aslak and S. Lehmann (2020) The scales of human mobility, *Nature*, **587** (7834) 402–407.

Chen, C., H. Ding, B. Sisman, Y. Xu, O. Xie, B. Z. Yao, S. D. Tran and B. Zeng (2024) Diffusion Models for Multi-Task Generative Modeling, paper presented at the *12th International Conference on Learning Representations (ICLR '24)*.

Devlin, J., M.-W. Chang, K. Lee and K. Toutanova (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Dodge, S., S. Gao, M. Tomko and R. Weibel (2020) Progress in computational movement analysis – towards movement data science, *International Journal of Geographical Information Science*, **34** (12) 2395–2400.

Feng, J., Z. Yang, F. Xu, H. Yu, M. Wang and Y. Li (2020) Learning to simulate human mobility, paper presented at the *26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Gambs, S., M.-O. Killijian and M. N. del Prado Cortez (2012) Next place prediction using mobility Markov chains, paper presented at the *First Workshop on Measurement, Privacy, and Mobility (MPM '12)*, 1–6, Bern, Switzerland.

Gong, S., M. Li, J. Feng, Z. Wu and L. Kong (2023a) DiffuSeq: Sequence to Sequence Text Generation with Diffusion Models, paper presented at the *11th International Conference on Learning Representations (ICLR '23)*.

Gong, S., M. Li, J. Feng, Z. Wu and L. Kong (2023b) DiffuSeq-v2: Bridging Discrete and Continuous Text Spaces for Accelerated Seq2Seq Diffusion Models, paper presented at the *2023 Conference on Empirical Methods in Natural Language Processing (EMNLP '23)*.

Ho, J., A. Jain and P. Abbeel (2020) Denoising Diffusion Probabilistic Models, paper presented at the *34th Conference on Neural Information Processing Systems (NeurIPS '20)*.

Holtzman, A., J. Buys, L. Du, M. Forbes and Y. Choi (2020) The Curious Case of Neural Text Degeneration, paper presented at the *8th International Conference on Learning Representations (ICLR '20)*.

Hong, Y., Y. Xin, S. Dirmeier, F. Perez-Cruz and M. Raubal (2023a) Revealing behavioral impact on mobility prediction networks through causal interventions, https://arxiv.org/abs/2311.11749.

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

Horni, A., K. Nagel and K. W. Axhausen (2016) *The multi-agent transport simulation MATSim*, Ubiquity Press, London.

Hörl, S. and M. Balac (2021) Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data, *Transportation Research Part C: Emerging Technologies*, **130**, 103291.

Jiang, S., Y. Yang, S. Gupta, D. Veneziano, S. Athavale and M. C. González (2016) The TimeGeo modeling framework for urban mobility without travel surveys, *Proceedings of the National Academy of Sciences*, **113** (37) E5370–E5378.

Kim, E.-J. and P. Bansal (2023) A deep generative model for feasible and diverse population synthesis, *Transportation Research Part C: Emerging Technologies*, **148**, 104053.

Li, X., J. Thickstun, I. Gulrajani, P. S. Liang and T. B. Hashimoto (2022) Diffusion-LM

Improves Controllable Text Generation, paper presented at the *36th Conference on Neural Information Processing Systems (NeurIPS '22)*.

Loshchilov, I. and F. Hutter (2019) Decoupled Weight Decay Regularization, paper presented at the *7th International Conference on Learning Representations (ICLR '19)*.

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

Mikolov, T., K. Chen, G. Corrado and J. Dean (2013) Efficient estimation of word representations in vector space, paper presented at the *1st International Conference on Learning Representations (ICLR '13)*.

Molloy, J., A. Castro, T. Götschi, B. Schoeman, C. Tchervenkov, U. Tomic, B. Hintermann and K. W. Axhausen (2022) The MOBIS dataset: a large GPS dataset of mobility behaviour in Switzerland, *Transportation*.

Radford, A., K. Narasimhan, T. Salimans and I. Sutskever (2018) Improving language understanding by generative pre-training.

Schläpfer, M., L. Dong, K. O'Keeffe, P. Santi, M. Szell, H. Salat, S. Anklesaria, M. Vazifeh, C. Ratti and G. B. West (2021) The universal visitation law of human mobility, *Nature*, **593** (7860) 522–527.

Schneider, C. M., V. Belik, T. Couronné, Z. Smoreda and M. C. González (2013) Unravelling daily human mobility motifs, *Journal of The Royal Society Interface*, **10** (84) 20130246.

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

Tao, Y., A. Both, R. I. Silveira, K. Buchin, S. Sijben, R. S. Purves, P. Laube, D. Peng, K. Toohey and M. Duckham (2021) A comparative analysis of trajectory similarity measures, *GIScience & Remote Sensing*, **58** (5) 643–669.

Zhou, M., J. Li, R. Basu and J. Ferreira (2022) Creating spatially-detailed heterogeneous synthetic populations for agent-based microsimulation, *Computers, Environment and Urban Systems*, **91**, 101717.