A Feature-Based Framework for Evaluating Synthetic Human Mobility

by

Jin Han
B.Sc., Shandong University, 2017

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Xʷsepsəm/Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

A Feature-Based Framework for Evaluating Synthetic Human Mobility

by

Jin Han
B.Sc., Shandong University, 2017

Supervisory Committee

Dr. Kevin Stanley, Supervisor
(Department of Computer Science)

Dr. Brandon Haworth, Departmental Member
(Department of Computer Science)

ABSTRACT

Generating realistic human mobility trajectories is essential for applications in urban analytics, transportation planning, and privacy-preserving data sharing. Evaluating the quality of synthetic data remains challenging. This study introduces a feature-based evaluation framework that characterizes trajectories through a unified set of statistical, geometric, and temporal descriptors. The framework is applied to benchmark GAN- and diffusion-based generative models using three real-world urban datasets with distinct spatial structures. Region-specific fine-tuning enhances realism, while persistent discrepancies in multi-scale entropy coefficients reveal challenges in modeling transitions between dwell and trip states. Incorporating road network information after generation provides limited benefit, suggesting that spatial constraints should be embedded during training. These findings highlight the influence of trajectory length, data quality, and explicit state modeling on generative performance. The study establishes a transparent feature-based approach connecting generative modeling and mobility analysis, supporting the creation of synthetic agents for data-driven urban design and policy evaluation.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank:

my partner, friends, and family, for supporting me in the low moments.

Dr. Kevin Stanley, for mentoring, support, encouragement, and patience.

The city of Victoria, for providing a delightful and inspiring environment.

DEDICATION

To my loved ones, for giving me strength and courage.

Chapter 1

Introduction

Human trajectories provide critical insights for urban planning, traffic management, and epidemic modeling, among many other domains [5]. Traditional data sources such as census records or travel surveys offer coarse-grained visit insights [38, 56], but the widespread adoption of smartphones has enabled the collection of high-resolution, large-scale GPS trajectories [79, 81]. The direct use of real trajectory data is constrained by privacy concerns [18, 16], creating barriers for open data sharing, large-scale benchmarking and deep learning based AI training. In a large mobile phone dataset, just four randomly chosen space-time points (with hourly time granularity and spatial resolution of the phone carrier's antennas) are sufficient to uniquely identify 95% of individuals [44]. Even after coarsening or aggregating the data spatially or temporally, the anonymity gained is limited: uniqueness decays only slowly as approximately according to the 1/10th power of resolution [44]. Mobility traces, even when anonymized, can leak sensitive information such as home or work location, frequently visited places, or participant presence at sensitive sites [41]. Synthetic trajectory generation, where the statistical and behavioral properties of real-world mobility traces are preserved while individual traces are obscured, has been proposed as a solution to this impact [40, 45, 52]. Synthetic data also offers advantages such as controllability, reproducibility, and easier distribution, making it an attractive alternative for both methodological research and downstream applications [62]. To create synthetic mobility traces or trajectories, it is necessary to have both the ability to assess the quality of the trajectories against baseline data (where closer is better) and the ability to ensure the privacy independence of the synthetic trajectories (where further is better). Although numerous generative models have been proposed—ranging from mechanistic approaches [61, 47, 33] to deep learning architectures [24, 52, 77]

and, more recently, large language model–based methods [60, 68]—a fundamental challenge remains: there is no general standard for evaluating whether synthetic trajectories preserve the spatiotemporal structure of the data they are based on. Existing practices often rely on similarity metrics such as Dynamic Time Warping, DTW [6], sequence-matching scores like GeoBLEU [58], or distributional distances such as the Wasserstein distance [54]. These measures capture limited aspects of trajectory behavior and fail to account for broader mobility patterns. Most of these metrics were originally designed for *trajectory prediction* tasks—where the goal is to predict the next N steps based on historical behavior—rather than for *synthetic trajectory generation*, which aims to reproduce the full distribution of plausible trajectories. For instance, in a series of human mobility challenge competitions hosted by ACM SIGSPATIAL[1], DTW and GeoBLEU have been widely used as benchmark metrics for prediction tasks. Crucially these measures evaluate the trajectory quality by how closely the generated trajectories match the training data, potentially embedding training data in the model, compromising privacy. For synthetic trajectory generation tasks, a recent Survey-of-Knowledge (SoK) study [9] highlighted a fundamental gap in evaluating the utility of privacy-preserving generative models. The authors proposed five design goals for trajectory publication mechanisms, among which Goal 4, *Utility*, emphasizes assessing how well released trajectories support downstream analytical tasks without compromising privacy. The study found that no existing evaluation framework adequately measures whether generative models preserve mobility-relevant properties of trajectories. Because the goal of synthetic trajectory generation is to simulate realistic movement behaviors rather than reproduce exact trajectories—traditional distance- or sequence-based metrics are of limit and utilities. A more generalizable, distribution-level evaluation framework capable of capturing the higher-order spatiotemporal structure of human mobility is required to meet this challenge.

To address this gap, I revisit the feature-based framework Representative Features of Geospatial Mobility (ReFGeM) [76], which was originally proposed as a representative set of mobility descriptors rather than a synthetic data evaluation tool. I propose that these features—capturing spatial coverage (convex hull, buffer area), geometric complexity (fractal dimension), and temporal regularity (entropy rate)—are well-suited for assessing population level realism of synthetic trajectories. We employ ReFGeM to analyze a set of recent trajectory generation models, including Noise-TrajGAN (NTG) [9], GeoPointGAN [17], and DiffTraj [83], and find that, despite

methodological advances, their generated trajectories still diverge from real-world mobility as measured using ReFGeM.

Because traces cannot be used without compromising privacy, few public datasets are available for training and testing. The most widely used public dataset, Geolife [80], provides individual-level GPS trajectories but suffers from data sparsity and irregular sampling [3]. Another commonly referenced source, the Foursquare dataset [72], contains point-of-interest (POI) check-ins rather than continuous movement traces, which cannot be directly treated as mobility trajectories, as users typically check in only at locations they find interesting — often omitting long stationary periods such as time spent at home or work. YJMob100K dataset [71] offers large-scale human mobility data with an aggregation based privacy-preserving mechanisms. To ensure anonymity, trajectories are aggregated into 500 m $\times$ 500 m spatial grids and recorded at 30-minute intervals. While this relatively coarse spatial–temporal granularity enables the study of broad urban mobility trends, it limits the ability to capture fine-grained visit sequences or individual-level behavioral regularities. Other trajectory datasets represent vehicular movement rather than human mobility, such as taxi trajectories in Beijing [73] and Roma [7], or bus routes in Rio de Janeiro [19]. Although these datasets are valuable for modeling traffic flow and transportation dynamics, they do not accurately reflect individual human movement patterns, because each vehicle may serve multiple passengers across different trips.

In this study, I employ high-quality private human trajectory datasets, *Interact-VIC* and *Interact-MTL*, collected in Victoria and Montreal, Canada, each covering ten consecutive days at a one-second GPS recording interval [36]. These datasets allow us to train variants of the best-performing generative model, *DiffTraj*, from scratch on Interact-VIC and Interact-MTL, to fine-tune the original Geolife-pretrained model separately for each city, and to train a combined model using trajectories from Geolife (Beijing), Victoria, and Montreal to examine cross-city transferability. To assess model performance, I apply the ReFGeM framework in combination with statistical analyses, demonstrating that it provides consistent and interpretable indicators of generative quality. This design allows us to compare training-from-scratch, fine-tuning, and multi-city training regimes under a unified evaluation framework. I argue that this analysis demonstrates how ReFGeM helps address the Goal 4 (Utility) challenge identified in [9], by establishing both the usability and practical value of synthetic trajectory evaluation. Based on this analysis, I show that:

- Existing generative models show limitations in producing useful trajectories,

particularly in separating and switching between dwells and trips.

- Cross-city transfer is feasible. Models trained on data from multiple cities exhibit stable performance across different urban contexts, indicating that exposure to diverse mobility patterns helps establish a strong foundation for generalization.
- ReFGeM provides interpretable evaluation for trajectory generation scenarios without one-to-one ground-truth pairs, offering feature-level diagnostics that support analysis across model comparisons and varied training regimes.
- The current state of the art generators produce trips that are too short to capture many important behaviors.

Agenda

Chapter 1 introduces the research motivation and problem setting. It discusses the limitations of existing evaluation metrics and the scarcity of public mobility datasets, and provides an overview of the results and the structure of the thesis.

Chapter 2 presents the background required for this work, including the ReFGeM feature set and the Jensen–Shannon divergence score used for trajectory evaluation.

Chapter 3 reviews related work on trajectory generation models and trajectory evaluation methods, with a focus on GAN-based and diffusion-based approaches.

Chapter 4 describes the experimental setup in detail. This chapter covers the datasets, data preprocessing procedures, trajectory generation models, cross-city training and model adaptation strategies, post-processing steps, and the experimental environment.

Chapter 5 presents the experimental results, including comparisons across different trajectory generation models and an evaluation of cross-city transfer performance.

Chapter 6 presents early-stage exploratory experiments on GeoLife and Victoria datasets using limited trajectory generation models, providing preliminary insights into trajectory segmentation and road network integration.

Chapter 7 discusses the results presented in Chapter 5, highlighting key observations, limitations of the proposed approach, and directions for future work, and provides a concluding summary of the thesis.

Appendix A provides additional supporting materials, including details on post-processing, ReFGeM feature distribution plots, model training parameters, and supplementary experiments.

Chapter 2

Background

2.1 ReFGeM Feature Set

The main feature set is based on the work of [76], where the authors proposed a standardized set of Representative Features of Geospatial Mobility (ReFGeM) for characterizing population spatial behavior. In this thesis, I extend the application of a subset of these features to the evaluation of synthetic trajectories. One feature is excluded in our analysis: the convex hull comprising the ten locations with the longest dwell times (CH10). This is because the trajectories in our study are of short duration, whereas CH10 is derived from trajectories spanning at least one week, making it unsuitable for our context. The features employed in this work are described below.

2.1.1 Convex Hull Area

In computational geometry, the convex hull of a set of points is defined as the smallest convex polygon (or polyhedron in higher dimensions) that contains all the points. Given a set of points $P = \{p_1, p_2, \ldots, p_n\}$ in $\mathbb{R}^d$, the convex hull $\mathrm{CH}(P)$ is the set of all convex combinations of points in P:

$$\mathrm{CH}(P) = \left\{ \sum_{i=1}^{n} \lambda_i p_i \;\middle|\; \lambda_i \geq 0,\ \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

Convex hulls have been adopted in trajectory research as geometric descriptors to summarize the spatial extent and structural properties of movement paths. For instance, convex hulls have been applied to identify critical geometric points along individual trajectories, such as turning and curvature points, which can serve as a

basis for trajectory similarity analysis [31]. At the population level, convex hull areas have been used to quantify the spatial extent of mobility modules and to investigate how module size scales with distance from home, revealing universal patterns of spatial expansion in human mobility across urban scales [82]. These studies illustrate that convex hulls provide informative local and global geometric features for analyzing movement patterns and spatial behavior.

2.1.2 Buffer Area

A buffer area refers to a zone of specified distance around a geographic feature, such as a point, line, or polygon. Given a spatial object O and a distance d, the buffer area $B(O, d)$ is defined as the set of all locations whose distance to O is less than or equal to d:

$$B(O, d) = \{x \in \mathbb{R}^2 \mid \text{dist}(x, O) \leq d\}.$$

Buffer areas are widely used to model spatial proximity, accessibility, and interaction regions. In mobility studies, they can represent zones of potential activity around locations of interest, such as the area reachable from a dwell point within a certain radius or the region surrounding transportation nodes for accessibility analysis [46, 74]. They are also commonly used in trajectory evaluation to account for GPS noise or to define flexible spatial constraints when comparing real and synthetic trajectories, especially when extracting activity spaces from GPS tracks [51]. By aggregating points or trajectories within buffer areas, researchers can quantify spatial coverage, local density, or visitation patterns, providing a practical and computationally efficient way to characterize movement behavior.

2.1.3 Fractal Dimension

The fractal dimension is a metric that quantifies the geometric complexity or space-filling capacity of an object or trajectory. A higher fractal dimension indicates greater tortuosity or complexity in the path, whereas a lower fractal dimension corresponds to simpler, more linear trajectories. For example, a straight line has a fractal dimension of 1, a plane has dimension 2, and a classic hexagonal fractal ("hexflake") has a dimension of approximately 1.7712 [42].

One common estimation method is the box-counting approach. Let a spatial object (e.g., a trajectory) be covered with a grid of boxes of side length ϵ, and let

$N(\epsilon)$ be the number of boxes that contain at least part of the object. The box-counting fractal dimension D is then defined as:

$$D = \lim_{\epsilon \to 0} \frac{\log N(\epsilon)}{\log(1/\epsilon)}.$$

In human mobility and trajectory analysis, D provides a quantitative measure of the spatial complexity of movement paths. Trajectories with frequent turns, loops, or extensive spatial coverage yield higher fractal dimensions, whereas straight or minimally tortuous paths have smaller D values. The fractal dimension can be computed for individual trajectories or aggregated at the population level to quantify the space-filling and complexity properties of movement paths and spatial flows [29, 11].

2.1.4 Multi-scale Entropy Rate Constants (C1–C5)

The entropy rate measures the average uncertainty or unpredictability of a stochastic process and can be used to characterize the predictability of human trajectories [61]. In mobility analysis, it quantifies the regularity of an individual's movement patterns: a high entropy rate indicates less predictable, more diverse mobility, while a low entropy rate corresponds to regular, predictable trajectories.

To compute entropy rate for trajectories, the space can be discretized into grid cells, converting each trajectory into a string of labels corresponding to visited cells. The entropy rate of this string can then be approximated using data compression techniques, such as the Lempel-Ziv 78 (LZ78) algorithm. The LZ-derived entropy rate H of a string S of length L is defined as:

$$H = \lim_{L \to \infty} \frac{1}{L} \sum_{i=0}^{L-1} \Lambda_i \, \ln L \,, \tag{2.1}$$

where i indexes characters in the string and Λ_i denotes the length of the shortest substring starting at position i that has not appeared previously in the string. This method provides a practical and widely used approach to estimate the entropy rate of real and synthetic trajectories. However, the estimated entropy rate is sensitive to the spatial and temporal resolutions used during trajectory discretization. To address this, Paul et al. [48] proposed a multi-scale extension that defines a sampling-

dependent entropy rate: :

$$H(d,T) = \frac{\log L}{d^2\frac{C_1}{4T^2L} + \frac{C_2}{4T^2L} + 2d\frac{C_3}{4T^2L} + d\frac{C_4}{TL} + \frac{C_5}{TL}}, \tag{2.2}$$

where

$$C_1 = \sum_{i=1}^{n} \frac{1}{v_i^{*2}}, \quad C_2 = \sum_{i=1}^{n} t_{d_i}^2, \quad C_3 = \sum_{i=1}^{n} \frac{t_{d_i}}{v_i^*}, \quad C_4 = \sum_{i=1}^{n} \frac{1}{v_i^*}, \quad C_5 = \sum_{i=1}^{n} t_{d_i}.$$

Here, v_i^* is the apparent speed across the i-th spatial cell, t_{d_i} is the total dwell time within that cell (with side length d), and T denotes the temporal sampling rate. Following the interpretation in [76]: C_1 corresponds to the inverse squared velocity; C_2 to the squared dwell time; C_3 to the ratio of dwell time to velocity (inverse distance); C_4 to velocity; and C_5 to dwell time.

2.2 Jensen Shannon Divergence Score

The Jensen–Shannon (JS) divergence is a symmetric and bounded measure of similarity between two probability distributions [39]. Given two discrete probability distributions P and Q, the JS divergence is defined as

$$\mathrm{JSD}(P \,\|\, Q) = \frac{1}{2}\mathrm{KL}\big(P \,\|\, M\big) + \frac{1}{2}\mathrm{KL}\big(Q \,\|\, M\big),$$

where $M = \frac{1}{2}(P+Q)$ and $\mathrm{KL}(\cdot,|,\cdot)$ denotes the Kullback–Leibler divergence [37]. Unlike the KL divergence, the JS divergence is symmetric and always finite, making it suitable for practical comparisons between empirical distributions. In generative modeling, JS divergence has been widely used as an objective for comparing synthetic with real distributions, for example in the theoretical interpretation of generative adversarial networks [23] and multimodal generative learning [64]. Variants and estimators of JS divergence have also been proposed that extend its applicability in machine learning tasks such as two–sample testing and distribution shift detection [32].

In this thesis, the Jensen–Shannon divergence is employed to evaluate the similarity between real test trajectories and synthetic trajectories generated by multiple trajectory generation models. Specifically, features are first extracted from both real and synthetic trajectories using the ReFGeM representation, which encodes relevant spatial-temporal and semantic information. For each feature, the empirical proba-

bility distributions are computed over the test set and the corresponding synthetic trajectories. The JS divergence is then calculated independently for each feature, providing a quantitative measure of how closely the synthetic data matches the statistical properties of the real data. This feature-level evaluation enables a detailed comparison across different generative models, highlighting strengths and limitations in capturing specific trajectory patterns. Lower JS divergence values correspond to smaller differences between the distributions of synthetic and real trajectories, which is critical for downstream applications such as trajectory prediction, mobility pattern analysis, and urban planning simulations.

Chapter 3

Related Work

3.1 Data Foundations for Mobility Modeling

The evolution of human mobility modeling has been linked to the availability, granularity, and volume of temporal-spatial data. As data acquisition methods have shifted from manual surveys to automated sensing, the resolution of observable mobility patterns has improved, necessitating complex modeling techniques. However, this increase in resolution has been accompanied by escalating privacy risks, motivating the need for synthetic trajectory generation.

Traditional Approaches and Indirect Proxies Early mobility research primarily relied on travel diaries, census data, tax revenue data and household surveys [38, 56]. While these sources provide high-quality semantic information regarding trip purposes and demographics, they suffer from high collection costs, limited sample sizes, and low temporal updates [65]. Such data are typically static and coarse-grained, making them suitable for estimating Origin-Destination (OD) matrices but insufficient for modeling continuous, time-varying trajectory dynamics. Alternative proxies like currency bill circulation [8] have been explored, though their utility is constrained by the inability to trace consistent individual trajectories, as bills frequently change hands. Similarly, CCTV camera footprints [70] provide insights into flow but face challenges in continuity and spatial coverage.

Call Detail Records (CDRs) The widespread adoption of mobile phones has transformed mobility tracking by enabling large-scale mobility analysis. Call Detail Records (CDRs) capture the location of a user's nearest cell tower during calls or

text messages. CDRs offered the first opportunity to study city-wide or country-wide population flows at a massive scale [27, 61]. However, CDRs are inherently sparse and spatially imprecise—limited by the density of cell towers [5]—which restricts their utility for micro-level trajectory generation or fine-grained navigation tasks.

Location-Based Social Networks (LBSNs) With the rise of platforms like Foursquare [72], Twitter, and WeChat, Location-Based Social Network (LBSN) data emerged as a rich source of mobility information. LBSN data consists of "check-ins" that link spatial coordinates with semantic labels (e.g., "Restaurant," "Gym") and social connections [13, 55]. While valuable for POI recommendation and understanding social influence on mobility, LBSN data is often biased towards specific demographics and discontinuous, failing to capture the path taken between check-ins.

GPS and High-Resolution Trajectories The proliferation of GPS-enabled smartphones and connected vehicles has established Global Positioning System (GPS) traces as the standard for modern mobility modeling [79, 81]. Unlike CDRs or check-ins, GPS data provides a continuous or high-frequency sequence of spatial coordinates (x, y, t), capturing precise geometric details such as turns, velocity changes, and pauses. This high granularity presents both opportunities and challenges. While it allows for the reconstruction of realistic movement behaviors, it also introduces high dimensionality and complex spatiotemporal dependencies. Simple probabilistic models often struggle to capture these intricate distributions, paving the way for the deep generative models—such as GANs and Diffusion models—discussed in the following section. However, privacy concerns [18, 16] have limited access to real GPS records, hindering large-scale analysis and benchmark dataset creation. As a result, synthetic trajectory generation has gained traction as a key research area.

3.2 Generative Models for Trajectory Data

Early efforts in trajectory generation relied on mechanistic models, such as the Exploration and Preferential Return (EPR) framework [61], which explains mobility as a balance between exploring new places and returning to familiar ones. Variants of these models introduced refinements such as distinguishing between returners and explorers [47], or incorporating temporal rhythms and social interactions [33, 15]. While these approaches are highly interpretable, they rely on rigid statistical assumptions

and generally fall short in reproducing the complex, non-linear spatiotemporal correlations observed in high-resolution GPS data.

The field has more recently shifted towards data-driven deep generative models. Recurrent neural networks (RNNs) [24] were among the first to treat trajectories as temporal sequences. However, RNNs primarily focus on next-location prediction rather than holistic trajectory synthesis. Consequently, Generative Adversarial Networks (GANs) [40, 45, 52, 17, 69] and Diffusion Probabilistic Models (DPMs) [14, 84, 83, 78] have become the dominant paradigms.

The recent emergence of Large Language Models (LLMs) has also inspired approaches that treat trajectories as textual token sequences [60, 68]. Despite their success in NLP, LLMs face distinct challenges in mobility modeling: tokenizing continuous geospatial coordinates (x, y) often leads to quantization errors or necessitates explicitly modeled vocabularies, struggling to capture fine-grained metric distances [12]. Furthermore, their reliance on massive pre-training corpora makes them computationally prohibitive for localized, data-impoverished environments compared to specialized lightweight generative models.

In this work, we focus on GAN- and diffusion-based models, building upon the implementations presented in [9, 10]. We adopt and extend three representative models: ***Noise-TrajGAN***, ***GeoTrajGAN***, and ***DiffTraj***. Their detailed architectures are described in Section 4.3.1. Unlike recent LLMs that enable one-shot trajectory generation from text prompts [68], the selected models are trained from a trajectory dataset learning full trajectory distributions, incorporating structured latent variables and temporal dependencies to provide trajectory synthesis. Compared to purely mechanistic approaches, they can learn richer spatiotemporal dynamics, while being less data-hungry and less opaque than large language models, making them appropriate in data-impoverished environment [84, 68]

3.2.1 Generative Adversarial Network (GAN) Models

GANs [28] consist of a generator (G) that synthesizes trajectory sequences from noise, and a discriminator (D) that distinguishes them from real data. Existing trajectory GANs can be broadly categorized based on their data representation: *Grid-based* and *Vector-based* approaches.

Grid-based Approaches. These models convert continuous GPS traces into discrete grid representations, treating mobility generation similarly to image generation. Ouyang et al. [45] proposed a non-parametric GAN where trajectories are converted into sparse matrices, with CNNs used to extract hierarchical spatial–temporal features. Feng et al. [25] introduced *MoveSim*, which maps trajectories to a 2D grid and employs a self-attention–based generator. While effective for capturing spatial density, grid-based methods often suffer from resolution loss due to discretization.

Vector-based Approaches. To preserve precise location information, vector-based models process trajectories as continuous sequences (x_t, y_t). Rao et al. [52] developed *LSTM-TrajGAN*, combining LSTM networks with GANs. It encodes user-level features into a latent space, using a many-to-many LSTM generator to synthesize realistic coordinates. Wang et al. [69] proposed a two-stage map-based GAN (TSG), which first generates a coarse grid sequence and then refines it into continuous coordinates conditioned on road network topology.

Privacy-Preserving GANs. A parallel line of research explores the use of GANs for data anonymization. Liu et al. [40] were among the first to conceptualize this direction in their vision paper, proposing *trajGANs* to generate synthetic trajectories that preserve aggregate spatial properties while mitigating re-identification risks. Building on this conceptual foundation, subsequent studies integrated formal privacy guarantees. For instance, *DP-TrajGAN* [75] and *PATE-GAN* [34] modify the training process by clipping gradients and adding Laplacian noise, ensuring theoretically bounded privacy loss. These works highlight the potential of GANs not just as generators of realistic data, but as robust mechanisms for privacy-enhancing data release.

3.2.2 Diffusion-based Models

Diffusion probabilistic models [59] capture the data distribution by reversing a gradual noise-addition process. For trajectory generation, the forward process perturbs trajectory distributions with Gaussian noise, while the reverse process iteratively denoises them to synthesize realistic paths.

Zhu et al. [83] proposed *DiffTraj*, which adapts the U-Net architecture [53]—originally designed for image segmentation—to model spatial–temporal correlations in trajectory data. This model serves as a baseline in our experiments. Moving beyond

image-like architectures, *TrajGDM* [14] employs a graph-based diffusion model. It uses a Transformer-based generator to capture dynamic spatial dependencies and an LSTM-based encoder for continuous trajectory embeddings.

Recent studies have further incorporated environmental constraints. *ControlTraj* [84] introduced a controllable diffusion model conditioned on road-segment sequences, employing a *RoadMAE* autoencoder to enforce topological consistency. Similarly, MA^2Traj [78] integrates origin–destination information via a multi-attribute aggregation module, fusing trip attributes such as distance and speed to guide the reverse diffusion process.

3.3 Evaluation of Trajectories

Even though generative models for trajectory data have evolved rapidly, the way to evaluate their outputs still lacks a unified standard. Recent surveys [63, 43] have shown that most existing approaches essentially reduce evaluation to similarity measures. First, **geometric point-set measures**, such as Hausdorff and Fréchet distances, quantify the spatial proximity between the shapes of two trajectories, effectively capturing the "worst-case" deviation but often ignoring the temporal alignment of points. Second, **sequence alignment metrics** address temporal shifts. Dynamic Time Warping (DTW) and its variants (e.g., EWDTW [4], PDTW [35]) allow for nonlinear mapping between time steps to measure spatiotemporal similarity, though they remain computationally intensive. Third, **token-based metrics** discretize continuous trajectories into symbolic sequences. Methods like Longest Common Subsequence (LCSS) [67] and Edit Distance on Real Sequence (EDR) are robust to noise and outliers. Similarly, GeoBLEU [58], adapted from Natural Language Processing, assesses segment-level coverage by comparing n-gram overlaps of discretized tokens.

Beyond pointwise or sequence-level similarity, several studies evaluate synthetic trajectories by comparing the distributions of aggregated mobility statistics between real and generated data. Statistical distances such as Kullback–Leibler divergence, Wasserstein distance, and Maximum Mean Discrepancy (MMD) are commonly employed to quantify the alignment of these distributions [83, 9]. Standard basic mobility features used in this context include:

- **Trip Distance:** The Euclidean or Manhattan distance between the origin and destination of a single trip, reflecting the scale of travel.

- **Radius of Gyration (r_g):** The root mean square distance of a user's visited locations from their center of mass, characterizing the spatial extent of an individual's activity space [27].
- **Visitation Frequency:** The distribution of visits per location, which typically follows a heavy-tailed power law, indicating the preferential return to familiar places.
- **Origin-Destination (OD) Flows:** The aggregated count of transitions between regions, capturing the macroscopic flow patterns of the population.

While these distributional metrics better reflect population-level realism, they still rely on predefined summary statistics and may overlook higher-order behavioral structure. Both similarity-based and distribution-based approaches treat trajectories primarily as temporal sequences or statistical samples rather than as spatial behaviors. As a result, they fail to capture higher-level mobility characteristics—such as trip purpose, route diversity, or spatial regularity—and provide limited interpretability and diagnostic value, and evaluate the network on its ability to reproduce the training set, rather than its ability to produce novel but realistic trajectories. A more meaningful evaluation should instead incorporate mobility-specific features that reflect the behavioral essence of movement and offer a stronger basis for assessing generative quality.

Zhang et al. [76] introduced the Representative Features of Geospatial Mobility (ReFGeM) framework, which identifies a set of descriptive features specifically designed to capture fundamental aspects of human mobility. The authors argued that features or metrics used to distinguish patterns in spatial behavior should be phenomenologically representative, mathematically rigorous, useful for distinguishing sub-populations, and—most importantly—generalizable. The last criterion is especially demanding, since many spatial metrics suffer from the Modifiable Areal Unit Problem (MAUP) [26], whereby different discretizations of space or variations in sampling frequency can significantly alter their values. To address this, ReFGeM highlights features that are relatively robust, such as the convex hull area and buffer area to characterize an individual's activity range [22, 57], the fractal dimension to quantify the geometric complexity of trajectories, and the entropy rate scaling behavior to capture patterns of regularity [48].

Although ReFGeM was not originally designed for evaluating synthetic data, its feature-based approach provides a more generalized and training data independent

basis for describing human mobility. This formulation offers a foundation for assessing the extent to which synthetic trajectories capture the intrinsic spatiotemporal characteristics of real-world movement, especially if the output is spatially distinct from the training data.

Chapter 4

Experimental Steup

4.1 Datasets

I employ three datasets in the experiments: the Geolife dataset [80], and two datasets from the Interact study [36], Interact-VIC and Interact-MTL. Geolife is publicly available and released by Microsoft. The Interact datasets are collected as part of the INTERACT study [36] and cannot be released due to privacy and ethical constraints. INTERACT is a five-year, four-site, four-wave longitudinal study investigating how changes in urban environments affect health outcomes. For Interact-VIC, I include data from the first two waves of the study, comprising 283 participants aged 18 and above who cycle at least once per month in the Greater Victoria area of British Columbia, Canada. For Interact-MTL, I include data from the first and third waves of the study, excluding Wave 2 to avoid disruptions caused by the COVID-19 pandemic. This subset comprises 218 participants aged 18 and above living on the Island of Montreal, as well as in the City of Laval, Longueuil, Brossard, or Saint-Lambert. All datasets consist of individual-level trajectories with varying spatiotemporal granularity and geographic coverage. Geolife trajectories are recorded every 1–5 seconds or every 5–10 meters using GPS loggers and mobile phones, whereas the Interact trajectories analyzed here are recorded at a fixed one-second interval using a SenseDoc device [2]. In the case of Geolife, the majority of trajectories were collected in Beijing. In this paper I refer to the Geolife dataset as representing Beijing only, disregarding the limited data from other cities, consistent with other studies [9, 10].

4.2 Data Preprocessing

I apply a multi-step preprocessing pipeline to condition each dataset by constraining the geographic extent, removing outliers, aligning sampling rates, segmenting and filtering trajectories, and normalizing them for neural network training. A summary of the basic statistics and descriptions of the datasets is provided in Table 4.1.

Table 4.1: Dataset Description

Dataset	Duration	City	Time Granularity	Spatial Representation	GPS records counts	Participants counts
Geolife	5-30 days	Beijing, China	1–5 seconds	Latitude, Longitude	24,876,978	182
Interact-VIC	10 days	Greater Victoria, Canada	1 second	Latitude, Longitude	84,394,852	283
Interact-MTL	10 days	Greater Montreal, Canada	1 second	Latitude, Longitude	34,310,006	212

Step 1: Filtering by bounding box. For each city, I define a geographic bounding box of the form [min_longitude, max_longitude, min_latitude, max_latitude]. The bounding boxes used in my experiments are: $\text{BBOX}_{\text{Beijing}} = [116.19, 116.56, 39.75, 40.03]$, $\text{BBOX}_{\text{Victoria}} = [-123.540331, -123.271128, 48.391892, 48.57277]$ and $\text{BBOX}_{\text{Montreal}} = [-74.30388, -73.136086, 45.22748, 45.96258]$.

Step 2: Resampling to fixed temporal interval. Each trajectory is resampled to a fixed temporal interval of $\Delta t = 5$ seconds, following the Geolife documentation, which indicates that approximately 91% of trajectories are recorded densely (every 1–5 seconds or every 5–10 meters). The same interval is applied to Interact-VIC and MTL for consistency. Specifically, each trajectory is divided into consecutive time windows of length Δt, the resampling is implemented via `pandas.DataFrame.resample`. In the original Geolife preprocessing, 5-second windows are aligned to midnight of the first day of the trajectory (`origin='start_day'`), which creates many empty windows that are later discarded due to the lack of observations. For the Interact-VIC and MTL dataset, I align the windows to the first timestamp of each trajectory (`origin='start'`), avoiding empty bins and providing a more natural temporal partition. Within each window:

- Numerical attributes (e.g., latitude, longitude, speed) are aggregated by computing the mean;
- Categorical attributes (e.g., user ID, trajectory ID) are taken from the first observation in the window.

Time windows without any observations are discarded. This procedure ensures uniform temporal spacing, suitable for downstream analysis. This resampling uses

window-based aggregation rather than linear interpolation. Resampled points represent average positions within each interval. While the choice of a 5-second interval may not be optimal for all datasets, it is adopted to maintain consistency with prior implementations on Geolife. Determining an optimal resampling interval would require systematic sensitivity analysis and may vary depending on dataset characteristics, which is beyond the scope of this work.

Step 3: Splitting long breaks. Some trajectories span extended periods with long inactive intervals, which disrupt the continuity of the data. Trajectories containing gaps longer than 60 seconds are split into segments. Gaps shorter than 60 seconds are ignored.

Step 4: Trajectory length filtering. Trajectories with fewer than 10 points are discarded. Trajectories longer than 200 points are either truncated (for Geolife) to 200 points or split (for Interact-VIC and MTL) into multiple non-overlapping segments of approximately 200 points each. For example, an Interact-VIC trajectory containing 350 points would be divided into one segment of 200 points and another of 150 points. The Geolife truncation strategy follows the original implementation [9]; however, since trajectories in Interact-VIC and MTL are typically much longer (often covering nearly a full day of continuous recording), truncation leads to a loss of over 90% of GPS records, which the splitting strategy avoids. After Step 3, more than half of the trajectories in Interact-VIC and MTL exceed 200 points, while only about 12% of the trajectories in Geolife are longer than 200 points. The cumulative distribution function (CDF) plots of trajectory lengths for each dataset are shown in Figure 4.1.

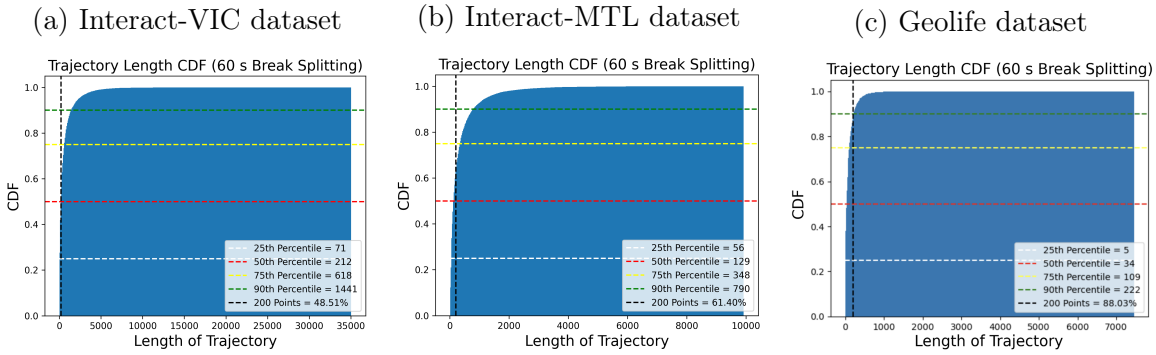


Figure 4.1: Empirical Cumulative Distribution Function (ECDF) plots of trajectory lengths. **(a)** Interact-VIC dataset, **(b)** Interact-MTL dataset, and **(c)** Geolife dataset. The plots compare real data (blue) with synthetic trajectories generated by the Geolife-pretrained model (green) and the Interact-finetuned model (red).

Step 5: Spatial normalization. This step prepares trajectories for model input, following the procedure in [52], which underlies the two GAN-based models. All latitudes and longitudes are standardized relative to the dataset centroid, allowing models to better capture spatial deviation patterns across trajectory points. Each trajectory is normalized by subtracting a reference point and dividing by a scaling factor. The reference point is the midpoint of longitude and latitude:

$$ref_{lon} = \frac{\max(lon) + \min(lon)}{2}, \quad ref_{lat} = \frac{\max(lat) + \min(lat)}{2}.$$

The scaling factor is computed as

$$sf = (\max(lon) - ref_{lon}, \max(lat) - ref_{lat}),$$

and each coordinate (lon, lat) is normalized by

$$(lon, lat) \leftarrow \frac{(lon, lat) - (ref_{lon}, ref_{lat})}{sf}.$$

This spatial normalization is applied only during the model training stage. The subsequent ReFGeM feature computation is still performed on the original latitude and longitude values.

The number of valid GPS records and participants retained after each pre-processing step is summarized in Table 4.2.

Table 4.2: Number of valid GPS records and participants after each pre-processing step for Geolife, Interact-VIC, and Interact-MTL.

	Geolife		**Interact-VIC**		**Interact-MTL**	
	GPS Records	**Participants**	**GPS Records**	**Participants**	**GPS Records**	**Participants**
Step 0: Unprocessed	24,876,978	182	84,394,852	283	34,310,006	218
Step 1: BBox Filtering	16,455,443	179	78,126,029	283	34,310,006	218
Step 2: Fixed 5s Sampling	8,598,197	179	15,835,190	283	8,591,190	218
Step 3: Splitting at 60s	8,598,197	179	15,835,190	283	8,591,190	218
Step 4: Length Filtering	6,486,421	170	15,817,026	280	8,573,093	212

After preprocessing, the trajectory-level statistics are summarized in Table 4.3. Specifically, the Geolife, Interact-VIC, and Interact-MTL datasets contain 69,504, 97,912, and 59,112 trajectories, with average lengths of 93, 162, and 145 points, respectively.

Table 4.3: Summary statistics of extracted trajectories after preprocessing.

Dataset	Number of Trajectories	Avg. Length
Geolife	69,504	93
Interact-VIC	97,912	162
Interact-MTL	59,112	145

4.3 Modeling

4.3.1 Trajectory Generation Models

Noise-TrajGAN [9] is a variant of LSTM-TrajGAN [52]. Unlike the LSTM-TrajGAN, which requires real trajectory segments as input, Noise-TrajGAN uses only Gaussian noise vectors as the generator's input, allowing the model to synthesize trajectories without relying on a real dataset.

GeoTrajGAN [9] builds upon GeoPointGAN [17], a model originally developed for generating spatial point distributions rather than full trajectories. GeoPointGAN employs a classic GAN structure where Gaussian noise is mapped to synthetic geographic points, and a discriminator classifies samples as real or fake. To extend this to trajectory generation, GeoTrajGAN introduces a bidirectional LSTM-based generator to capture temporal dependencies and employs two discriminators: one adapted from GeoPointGAN (with normalization adjustments) and another LSTM-based discriminator to assess the sequential coherence of generated trajectories.

DiffTraj [83] is the first model to leverage diffusion-based generative modeling for trajectory synthesis with built-in privacy preservation. Its architecture follows a UNet [53] backbone, featuring up- and down-sampling modules constructed from ResNet blocks with Conv1D layers. An attention-based transition module [66] connects these components to capture sequential dependencies. The model further supports optional conditional information for guided generation, such as average speed, distance, and start time. Buchholz et.al [10] extended this framework by re-implementing the original models and incorporating several differential privacy (DP) mechanisms [50] to explore the trade-off between privacy and utility. Here, I adopt the original non-DP version of *DiffTraj* as presented in [83], which corresponds to the conditional embedding "8-stat" variant described in [10], in order to focus on evaluating the quality of synthetic trajectories.

I employ the Geolife dataset and implement *Noise-TrajGAN*, *GeoTrajGAN*, and the non-DP 8-stat version of *DiffTraj*, using pretrained models provided by [9] and [10].

The 8-stat model incorporates eight additional conditional features: (1) departure time, (2) total distance, (3) total duration, (4) total length, (5) average distance, (6) average speed, (7) start grid ID, and (8) end grid ID. To ensure fairness in comparison, I also implement an unconditional version of *DiffTraj* that takes only noise as input.

During preprocessing, trajectories are resampled at 5-second intervals and truncated to a maximum length of 200 points. Each point is represented as a pair of normalized spatial coordinates within the range [0, 1], consistent with [52]. Sequences are padded with zeros to a fixed length of 200 points, ensuring uniform input dimensions across all samples. Since the GAN-based models and the unconditional *DiffTraj* do not establish a one-to-one mapping with real trajectories, generation simply requires specifying the number of trajectories to synthesize, without additional metadata. For the 8-stat version, conditional information (the "head") is computed for each trajectory and concatenated with the noise vector as auxiliary input. The outputs are synthetic trajectories, each represented as a fixed-length sequence of 200 locations. All hyperparameter configurations are summarized in the Appendix A.3.

I evaluate these models because pretrained implementations are available and they serve as representative baselines highlighted in [9], which identifies key gaps in trajectory generation evaluation. This framework targets these gaps, and using the same models enables a direct assessment of its effectiveness. The analysis could be extended to other advanced models [69, 84, 78] if pretrained versions were available. The primary goal is to demonstrate the applicability and utility of the proposed evaluation schema.

4.3.2 Cross-City Training and Model Adaptation

After implementing the three trajectory generation models, I select the best-performing one—*DiffTraj-8stat*—for further experimentation. I create several additional variants under different training regimes. First, I trained models from scratch on the Interact-VIC and Interact-MTL datasets. Apart from the change in training data, all architectures, hyperparameters, and training settings were kept identical to those of the Geolife-trained model. Second, I constructed fine-tuned variants by initializing from the Geolife-pretrained checkpoint and continuing training on the Interact-VIC and Interact-MTL datasets, using a reduced learning rate (2×10^{-5}) to adapt the models to local mobility patterns while preserving the knowledge learned from Geo-

life. Finally, I trained a multi-city combined model using trajectories from all three datasets: Geolife, Interact-VIC, and Interact-MTL.

I evaluated all six models—the original Geolife-trained model, the VIC- and MTL-trained-from-scratch models, the VIC- and MTL-finetuned models, and the combined multi-city model—on the Geolife, Interact-VIC, and Interact-MTL test sets. I refer to these models as *Geolife_trained*, *VIC_trained*, *MTL_trained*, *VIC_finetuned*, *MTL_finetuned*, and *combine_trained*, respectively.

The objectives of this experiment are threefold: (1) to examine whether higher-quality local data can improve model performance, (2) to assess the cross-city transferability of trajectory generation models, and (3) to demonstrate the utility of ReFGeM for the comparison.

Synthetic test datasets generation

In the original implementation [83, 9], evaluation is performed by feeding the conditional information (the *head*) from the test dataset into the model to generate synthetic trajectories. In my experiments, I treat the test dataset as untouched, avoiding any direct use of test data during generation. I sample from the overall dataset distributions to generate synthetic trajectories. The 8-stat head includes three categorical attributes—departure time (divided into 288 5-minute bins over 24 hours), start grid ID, and end grid ID—and five continuous attributes: total distance (consecutive cumulative Haversine distance), total duration, total length (number of trajectory points), average distance (total distance divided by total length), and average speed (total distance divided by total duration). After inspecting the data, I sample departure time independently from its marginal distribution, sample the start grid ID from its marginal distribution and the end grid ID conditioned on the start grid. For the continuous variables total distance, total duration, and total length, I implement a copula-based sampling method to capture their joint distribution. Average distance and average speed are computed from the sampled values. The sampled head arrays, combined with random noise, are fed into each model to generate synthetic datasets for evaluation. With six models and three datasets, this results in a total of 18 synthetic datasets.

This procedure ensures that the synthetic trajectories do not rely on one-to-one mappings from the test set, better reflecting the real-world generative setting and addressing the *Utility* challenge highlighted in [9].

4.4 ReFGeM Calculation

Given a synthetic or real dataset, I compute the ReFGeM features for each trajectory within that dataset. Both the real-world trajectories and all synthetically generated datasets are represented in geographic coordinates (latitude and longitude). Following the implementation in [76], I first convert them into UTM coordinates (i.e., easting and northing) using the Python library *pyproj*, as distance and spatial binning operations are more straightforward under the UTM system. The grid coordinated is using a grid cell size of 15.625 meters and the cooordinates (x, y) of each cell is calculated using $x = \frac{easting - easting_{min}}{15.625}$, $y = \frac{northing - northing_{min}}{15.625}$,where $easting_{min}$ and $northing_{min}$ are the lower bound of the corresponding bounding box of each city. All subsequent feature calculations are performed on these grid coordinates.

Convex Hull Area. This feature describes an individual's activity space by computing the area of the smallest polygon enclosing all spatial locations, with the outermost points serving as vertices. I employ the QuickHull algorithm via `scipy.spatial`, using the grid coordinates as input.

Buffer Area. To compensate for the sensitivity of convex hull to outliers, the buffer area metric extends each trip by a fixed spatial buffer, here a 15.625m (the minimum grid size for formal and uniform calculation) buffering is applied. All buffers are merged using the `cascaded_union` function in `shapely.ops` to obtain the area. This metric captures the extent of an individual's potential movement area.

Fractal Dimension. This feature quantifies the spatial complexity of a trajectory. The fractal dimension is computed using the box-counting method described in Chapter 3. The trajectory coordinates are first normalized to the unit hypercube, and the number of non-empty boxes $N(\epsilon)$ is counted for multiple scales $\epsilon \in \{0.001, 0.01, 0.1, 1.0\}$. The fractal dimension is then estimated as the negative slope of a linear fit on the log–log plot of $N(\epsilon)$ versus ϵ.

Multi-scale Entropy Rate Constants (C1–C5). This family of features encodes the spatiotemporal complexity of movement. Each trajectory is first converted into a symbolic sequence S, and a Lempel–Ziv–based entropy rate H is computed as in Equation 2.1. Since the marginal dwell times t_{d_i} and apparent speeds v_i^* are not directly observable from the symbolic trajectories, $H(d, T)$ is evaluated across multiple combinations of spatial and temporal scales (d, T). The five coefficients (C_1–C_5) are then obtained by fitting Equation 2.2 using the `scipy.optimize.curvefit` function. These coefficients parameterize a trajectory's entropy across different spa-

tial–temporal resolutions, providing robustness against variations in sampling density and frequency. In my experiments, given the trajectory length of 200 points ($\approx$ 16 minutes) and a 5-second sampling interval, we selected grid sizes of d = [15.625 m, 31.25 m, 62.5 m, 125 m] and sampling intervals T = [5 s, 10 s, 20 s, 40 s, 60 s] to ensure sufficient spatial resolution and to avoid excessively short sequences. For numerical stability during distribution analysis, all fit values smaller than 10^{-9} are treated as zero and grouped into a single bin.

4.5 Experimental Environment

All experiments were conducted on high-performance computing platforms to ensure reproducibility and efficiency. The *Noise-TrajGAN* and *GeoTrajGAN* experiments were run on Google Colab using a 40 GB NVIDIA A100 GPU. All *DiffTraj*-related model training—including the unconditional version, *Interact-trained* version and *Interact-finetuned* —was performed on the Digital Alliance Canada Fir cluster using a single 80 GB NVIDIA H100 GPU. Trajectory generation and ReFGeM feature computations were also carried out on the Fir cluster. All experiments were implemented in Python, using the following major packages: *pandas 1.5.3*, *numpy 1.26.4*, *scikit-learn 1.3.2*, *scipy 1.10.1*, *pyproj 3.6.1*, *folium 0.7.0*, *geopandas 0.14.4*, *shapely 2.0.7*, *matplotlib 3.7.2*, *osmnx 1.9.4*, and *networkx 3.3*. All parameter configurations used for fine-tuning are provided in Appendix C.

Chapter 5

Results

5.1 Trajectory Generation Model Comparison

Figure 5.1 illustrates the qualitative differences among the generated trajectories. The two GAN-based models exhibit mode collapse: once the generator discovers a trajectory pattern that consistently fools the discriminator, it tends to reproduce this pattern instead of exploring the full range of possible behaviors. This leads to limited diversity and prevents the models from capturing the complex mobility patterns present in the real dataset. Although [9] reports that Noise-TrajGAN was trained under multiple configurations—including different loss functions (WGAN-GP and modified TrajLoss), alternative noise shapes, optimizers, and learning rates—none of these settings mitigated mode collapse in our experiments. For GeoTrajGAN, I also replaced the original BCE loss with a WGAN with Lipschitz Penalty (WGAN-LP) loss [49], but the improvement remained marginal. Details are provided in Appendix A.4. The weak performance of both GAN-based models suggests that, even with a complex training process and large datasets, trajectory synthesis via standard GANs remains unstable. In contrast, both versions of DiffTraj—unconditional and 8-stat—demonstrate improved trajectory realism. They better align with the underlying road network in Beijing, producing paths that are more plausible for human movement. The 8-stat DiffTraj shows the highest fidelity, benefiting from additional conditional information such as departure time, total distance, average speed, and start/end locations, which guide the generator to produce trajectories that reflect more realistic behavioral patterns.

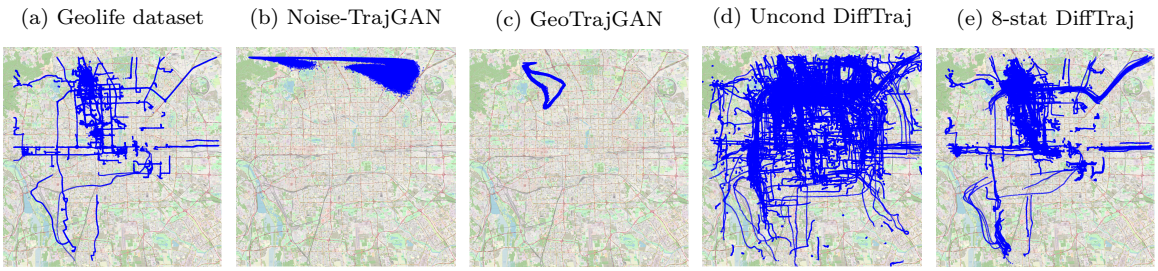


Figure 5.1: Random samples of 5000 trajectories from the Geolife dataset and synthetic trajectories generated by various models. The background shows the road network of Beijing. **(a)** Geolife dataset, **(b)** Noise-TrajGAN, **(c)** GeoTrajGAN, **(d)** Unconditional DiffTraj, and **(e)** 8-stat DiffTraj.

To quantitatively assess these differences, I compute ReFGeM features for both real and synthetic datasets and compare their distributions using the Jensen–Shannon (JS) divergence, which ranges from 0 to 1 and measures distributional similarity, with lower values indicating a closer match to real data. As shown in Table 5.1, the 8-stat DiffTraj model achieves the lowest JS divergence and thus the closest alignment with the Geolife trajectories. Feature C2 is the only case in which the unconditional DiffTraj scores lower. The unconditional variant ranks second overall. Noise-TrajGAN and GeoTrajGAN produce higher JS divergence—typically 3–5 times larger—reflecting the mode collapse behavior. Although the DiffTraj models consistently outperform the GAN-based models across all features, they still exhibit clear limitations. Some samples contain short stationary segments concentrated near dataset boundaries, and others, when conditioned on start and end points, produce trajectories that cross physical obstacles such as lakes. These results show that even the strongest models fall short of generating plausible mobility patterns and point to the need for additional spatial and contextual constraints to support realistic synthetic trajectory generation.

Table 5.1: ReFGeM feature comparison using Jensen-Shannon (JS) score between synthetic models and the Geolife dataset. The best score for each feature is highlighted in **bold**, while the second-best is indicated with underline.

Feature	Uncond DiffTraj	8-stat DiffTraj	Noise-TrajGAN	GeoTrajGAN
C1	0.2585	**0.2568**	0.4515	0.6062
C2	**0.4056**	0.4144	0.6504	0.6413
C3	0.1403	**0.1232**	0.7594	0.7336
C4	0.2586	**0.2355**	0.4755	0.4585
C5	0.2542	**0.2335**	0.7052	0.6002
Buffer Area	0.1349	**0.1154**	0.8326	0.8305
Convex Hull	0.1755	**0.1584**	0.8258	0.8271
Fractal Dimension	0.1753	**0.1619**	0.7408	0.7343

These results also confirm that diffusion-based models are more suitable for human trajectory generation, demonstrating both higher fidelity and better behavioral consistency compared to GAN-based approaches. However, the JS divergence of feature C2 (squared dwell time) is higher—around 0.4, compared to approximately 0.2 for most other features—indicating that all models still struggle to accurately capture dwell behaviors.

5.2 Cross-City Transfer Performance

I computed the Jensen–Shannon (JS) divergence of ReFGeM features by comparing real test trajectories with synthetic trajectories generated by six models: *Geolife_trained*, *VIC_trained*, *VIC_finetuned*, *MTL_trained*, *MTL_finetuned* and *Combine_trained* on the Geolife, Interact-VIC, and Interact-MTL datasets (Table 5.2). The distributions of ReFGeM feature values across all datasets and models are visualized in Appendix A.2 Figures A.5–A.12. The relative JS scores correspond well with the visualized feature distributions, confirming the consistency between quantitative and visual assessments.

Table 5.2: ReFGeM feature comparison using Jensen–Shannon (JS) score among *Geolife_trained*, *Interact_trained*, *Interact_finetuned*, *MTL_trained*, *MTL_finetuned*, and *combine_trained* models on the Geolife, Interact-VIC, and Montreal datasets. Lower JS score is better; best score for each feature and dataset is highlighted in **bold**, second best is underlined.

Features	Geolife Dataset						Interact-VIC Dataset						Montreal Dataset					
	Geo-Tr.	VIC-Tr.	VIC-FT.	MTL-Tr.	MTL-FT.	Comb-Tr.	Geo-Tr.	VIC-Tr.	VIC-FT.	MTL-Tr.	MTL-FT.	Comb-Tr.	Geo-Tr.	VIC-Tr.	VIC-FT.	MTL-Tr.	MTL-FT.	Comb-Tr.
C1	0.2820	0.3004	0.2848	0.2659	0.3077	**0.1971**	0.3795	0.2398	0.1668	0.2102	0.2034	**0.1350**	0.4607	0.3334	0.3193	**0.2261**	0.2482	0.3212
C2	0.4462	0.3673	0.3678	**0.3144**	0.3167	0.3733	0.3131	0.3558	0.2339	0.3059	**0.2203**	0.2257	**0.4218**	0.5114	0.4887	0.4851	0.4573	0.4637
C3	0.2329	0.2425	0.1502	0.2337	0.2014	**0.0991**	0.6267	0.2837	**0.1191**	0.2652	0.1560	0.1647	**0.2846**	0.5971	0.5234	0.5472	0.4709	0.4578
C4	0.2767	0.4867	0.3724	0.5340	0.4531	**0.1999**	0.2796	0.2715	0.1227	0.2923	0.1914	**0.1030**	0.1360	0.2825	0.2192	0.3365	0.3667	**0.1040**
C5	**0.1818**	0.3312	0.3871	0.2899	0.4266	0.2237	0.8264	**0.1555**	0.2396	0.1827	0.2574	0.2451	0.8326	0.4504	0.3640	0.3524	**0.2364**	0.3440
Buffer Area	0.2149	0.2281	0.2249	0.1982	0.2845	**0.1199**	0.5243	0.3255	**0.1780**	0.2716	0.1880	0.2321	0.6731	0.5978	0.5239	0.5273	0.4620	**0.4556**
Convex Hull	0.2014	0.2307	0.2152	0.2601	0.2950	**0.0991**	0.5094	0.2858	**0.1202**	0.2747	0.1765	0.2288	0.6444	0.5837	0.5182	0.5230	0.4731	**0.4509**
Fractal Dimension	0.2015	0.1912	0.2059	0.1675	0.2478	**0.1001**	0.4993	0.2846	0.1514	0.2284	**0.1280**	0.1986	0.6928	0.5977	0.5134	0.5088	**0.4310**	0.4700

Across all three cities, the *Combine_trained* model exhibits consistent performance, achieving the best or second-best JS score for most features. When it does not achieve the lowest score, the best-performing model is typically the one trained on the corresponding local dataset. On the Beijing (Geolife) dataset, the *Combine_trained* model achieves the lowest JS score on 6 out of 8 ReFGeM features and the second-lowest score on C5. Its average JS divergence is approximately 30% lower than that of the *Geolife_trained* model. On the Victoria dataset, fine-tuned models perform well on spatial-related features, including buffer area, convex hull, and fractal dimension. The *VIC_finetuned* model attains the lowest JS score on buffer area and convex hull, with the *MTL_finetuned* model ranking second. For fractal dimension, the *MTL_finetuned* model achieves the lowest score, followed by the *VIC_finetuned* model. Although the *MTL_finetuned* model is not trained on Victoria data, it outperforms models trained or fine-tuned on the Victoria dataset. This behavior can be attributed to similarities between the Montreal and Victoria training datasets. The ReFGeM feature distributions of these two datasets are closely aligned. First, both datasets are collected using Sensedoc devices, which record daily activities from waking up to going to bed at the same sampling frequency. Second, after truncating trajectories to 200 points (approximately 16 minutes), the space of feasible movement patterns becomes limited. These factors reduce cross-city differences and enable generalization. The Montreal-based models can outperform locally trained Victoria models due to differences in participant populations and study areas. Interact-VIC focuses on cyclists within the Greater Victoria area, whereas Interact-MTL includes adult participants without additional inclusion criteria and covers a broader region, including the Island of Montreal, Laval, Longueuil, Brossard, and Saint-Lambert. The larger geographic scope and participant diversity introduce more varied mobility patterns,

supporting transfer across cities. On the Montreal test set, the *Combine_trained* model achieves the lowest JS score on buffer area and convex hull.

Among the ReFGeM features, C1 - C4 exhibit smaller magnitudes. The values of C1 - C4 span several orders of magnitude, ranging from approximately 10^{-9} to 10^{1}, with a large proportion of values concentrated at small scales. In contrast, C5 ranges from 10^{2} to 10^{3}. C5 (dwell time) dominates the numerical scale of the entropy related ReFGeM features and is emphasized in the analysis. The lowest JS score for C5 is consistently achieved by models trained or fine-tuned on the local dataset, indicating that dwell time is city-specific. The divergence in C5 is larger between the Beijing and Canadian datasets than between the two Canadian datasets. The Geolife dataset exhibits a distinct dwell-time distribution because it is collected through participant-initiated sharing rather than passive full-day observation. This collection process leads to incomplete trajectories and alters dwell-time characteristics.

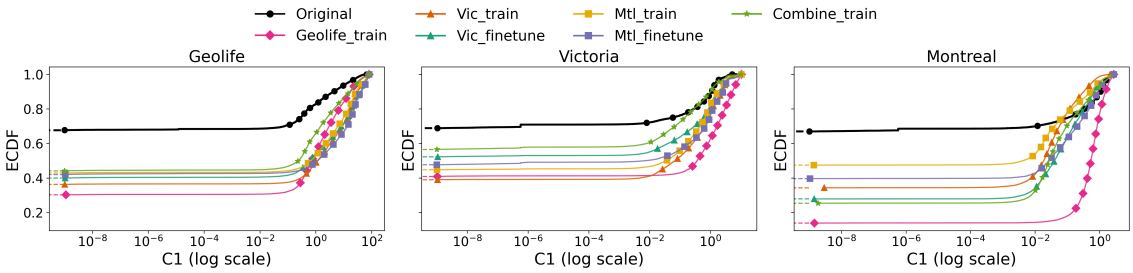
When comparing training strategies, fine-tuned models outperform models trained from scratch, indicating that exposure to diverse data improves performance. The results of the *Combine_trained* model further support this conclusion. Models trained only on Victoria data perform well in-domain but show limited transfer across cities, whereas Montreal-related models transfer to both Victoria and Beijing, and the Geolife-trained model transfers to Montreal.

The Empirical Cumulative Distribution Functions (ECDFs; Fig. 5.2) for each ReFGeM feature were computed on all datasets, comparing real and synthetic trajectories. The black line corresponds to the ECDF of the ReFGeM feature computed from the real dataset, while the colored lines represent ECDFs from synthetic trajectories generated by different models. Each row corresponds to a single ReFGeM feature, with three subplots showing the Geolife, Interact-VIC, and Interact-MTL datasets, respectively. The ECDFs indicate that the original ReFGeM features of Victoria and Montreal are quite similar, while the Geolife dataset differs from these two private datasets. C1 and C2 exhibit the largest gaps between original and synthetic trajectories. For C1 (inverse squared speed), all three original datasets show a high proportion of small values around 0.7, whereas synthetic datasets tend to generate fewer low C1 values, reducing the proportion to approximately 0.3–0.5. This indicates that high-speed trajectories are more common in the original datasets, while synthetic trajectories tend to have lower speeds. For C2 (squared dwell time), the original datasets show a low percentage of short dwell times (ECDF values below 10^{2} are around 0.2–0.3), with the majority of dwell times being longer. In the Geolife

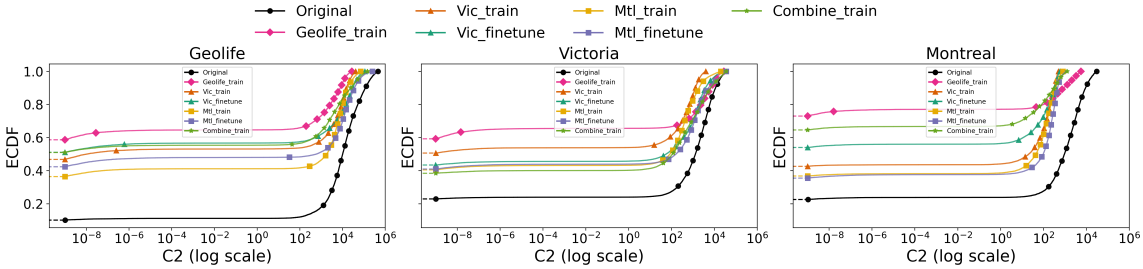
dataset, the ECDF sharply rises from 10^3 to 10^6, with most values above 10^3, whereas Victoria and Montreal exhibit a narrower range, mostly between 10^2 and 10^4. This highlights the differences between Geolife and the two Canadian datasets. Because C1 and C2 are squared quantities, discrepancies between original and synthetic datasets are further amplified. For C5 (dwell time), the *Geolife_trained* model (pink-purple with diamond marker) fails to accurately capture this feature when generating synthetic datasets for other cities. It tends to produce trajectories with excessively long dwell times, reflecting the characteristics of its own training data. For the remaining features describing spatial coverage and route complexity, synthetic datasets approximate the originals. However, the Montreal dataset appears harder to replicate, as its ECDF lines are farther from the best-performing synthetic datasets. This difference arises from the larger geographic and behavioral diversity in the Montreal dataset, which increases the variety of feasible trajectories despite similar overall feature distributions.

Overall, the JS divergence analysis and ECDF comparisons show consistent patterns. The *Combine_trained* model performs stably across all three cities. For C5 (dwell time), models trained on the corresponding local dataset achieve the lowest JS scores. The ECDFs indicate that replicating C1–C5, which quantify spatiotemporal movement patterns through combinations of speed and dwell time, is more challenging than replicating spatial coverage (buffer area and convex hull) or spatial complexity (fractal dimension). Short trajectories limit spatial coverage and route complexity, whereas C1–C5 are strongly influenced by local context, participant behavior, and transport patterns.

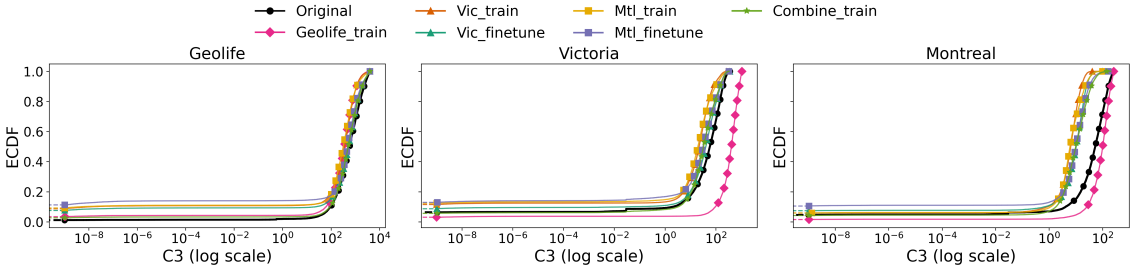
(a) C1



(b) C2



(c) C3



(d) C4

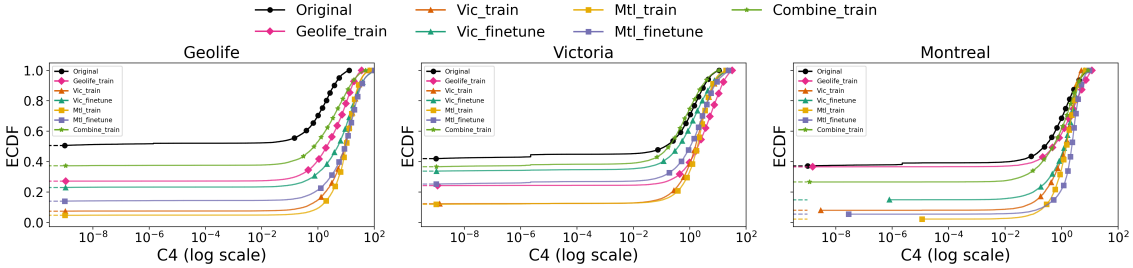
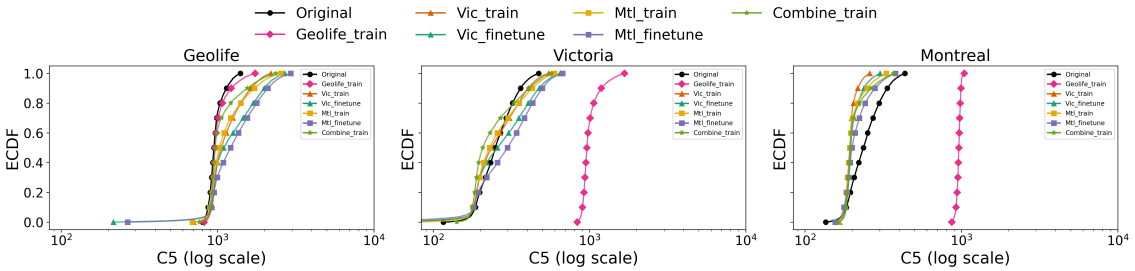
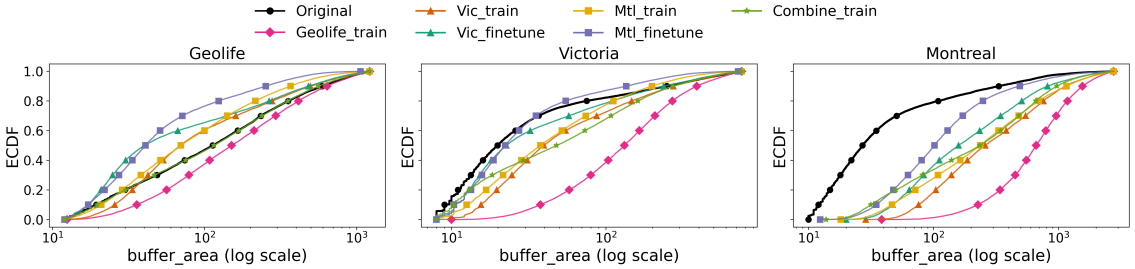


Figure 5.2: ECDF plots for each feature (Part 1). Comparing real vs synthetic trajectories across datasets. (Continued on next page)

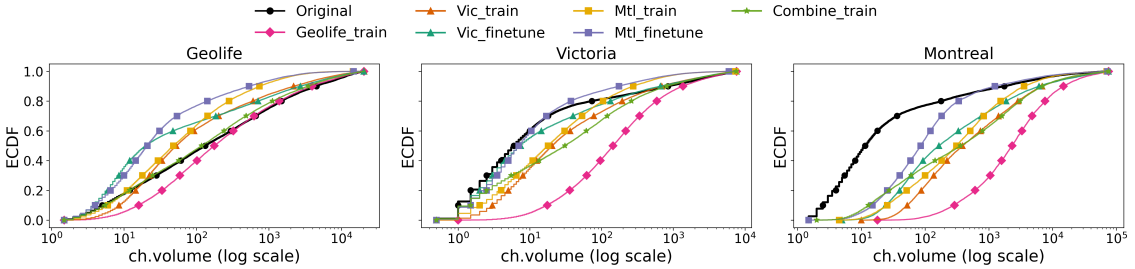
(e) C5



(f) Buffer area



(g) Convex hull



(h) Fractal dimension

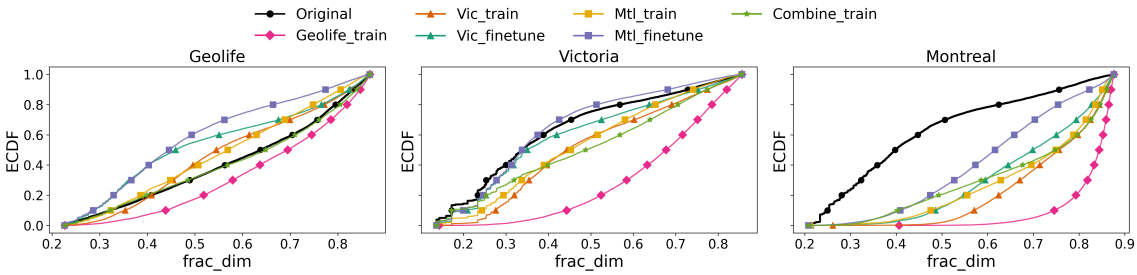


Figure 5.2: (Continued) ECDF plots for each feature.

Chapter 6

Ancillary Work

This chapter presents a set of exploratory experiments conducted at an early stage of the study. The experiments are limited in scope and focus on a subset of datasets and model configurations, including the GeoLife dataset and the Victoria dataset, as well as the original GeoLife-trained model *Geolife_trained* and a VIC-fine-tuned model initialized from the GeoLife-pretrained checkpoint and further trained on the Interact-VIC dataset *VIC_finetuned*. Due to the preliminary nature of these experiments and the limited performance gains observed, this line of investigation was not pursued further in the main body of the thesis. The results provide useful empirical insights into trajectory segmentation and road network integration, and are therefore included here as ancillary work.

6.1 Dwell Journey Detection

Dwells and journeys constitute a complete trajectory. Dwells occur when individuals are stationary, such as waiting at a bus stop or staying in a coffee shop, whereas journeys represent movements between locations, such as walking from the office to the bus station or taking a stroll in a park. Analysis of the estimated entropy coefficients indicates that dwell-related behavior dominates, as the marginal dwell time (C5) has a much larger magnitude than velocity (C4). This suggests that splitting trajectories into dwell and journey segments is necessary and potentially beneficial. Trajectory segmentation is commonly performed using movement parameters, defined as statistical properties of the movement process computed at each trajectory point. Representative parameters include velocity, speed, heading, acceleration, turning an-

gle, angular range, displacement, straightness index, sinuosity, tortuosity, and related locally derived measures [21, 20]. In this work, trajectories are segmented into dwell and journey intervals using a displacement-based criterion. Displacement is computed over a 10-point sliding window. Points are classified as part of a journey if the displacement within the window exceeds 30 meters; otherwise, they are classified as dwell. The 30-meter threshold is chosen because a 10-point window corresponds to 50 seconds, which would approximately correspond to 30 meters of displacement under a slow walking speed ($\approx 0.6m/s$), exceeding the expected range of GPS wandering noise of 0.3-1.6m [30]. Trip segments shorter than five points are discarded as noise. Additional analysis of the segmented trajectories is provided in Appendix A.1.

6.2 Road Network Integration

During trajectory generation, some synthetic trajectories may exhibit impossible paths—for example, passing through buildings, lakes, or other non-traversable areas. This issue arises because the generative models do not explicitly incorporate road network constraints. To address this limitation, road network integration is applied to all journey segments within each trajectory. Road network data are obtained using the Python library `OSMnx`, which is used to download complete driving, cycling, and walking networks from OpenStreetMap for both Beijing and Victoria. For each trajectory's trip segment, I downsample the location points to ensure that the start and end points are included, along with every fifth point in between. I then compute the shortest path that passes through all these waypoints in sequence, using Dijkstra's algorithm between each pair of consecutive waypoints. From the returned route, I extract the full geometric shape of the path by retrieving the polylines of each road edge traversed. The total route distance is then calculated as the sum of the individual edge lengths. Finally, I uniformly upsample along the entire path to match the original trip's length (in number of points), and concatenate the reconstructed trip segments with their corresponding dwell segments to form the complete trajectory.

A valid route is not always guaranteed to be found during this process. The success rate depends on the density and connectivity of the road network. The Beijing road graph contains over 350,000 nodes and 900,000 edges offering a high likelihood of finding valid routes between arbitrary points. The Victoria road graph has only about 8,500 nodes and 24,000 edges, providing fewer connectivity options. For trips where no valid route can be found, I retain the original synthetic trajectories generated by

the model.

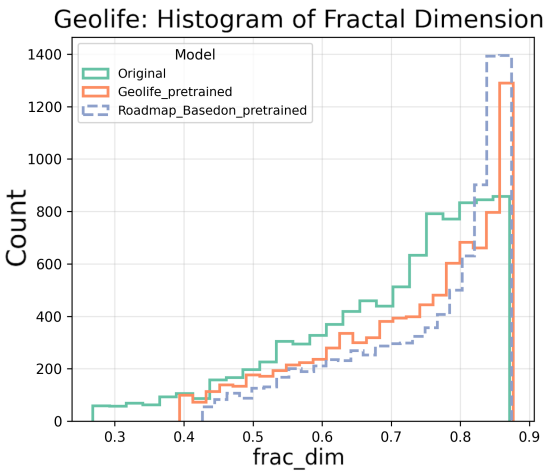
6.3 Road Network Results Analysis

After examining the trip and dwell distributions for all datasets (Appendix Fig. A.1), I selected the best synthetic dataset for each city—*Geolife_pretrained* synthetic Geolife dataset and *Interact_finetuned* synthetic Interact dataset—for road network integration. Since my focus is on mapping trip segments to the nearest road network, the feature most directly affected is the fractal dimension. My goal is to assess whether road network integration can bring the synthetic distribution closer to the real dataset, reducing the JS divergence.

For the Geolife dataset, the JS divergence of the fractal dimension increases after road network integration, from 0.1619 to 0.1856, indicating a marginally lower similarity to real trajectories. In contrast, for the Interact-VIC dataset, the JS divergence decreases from 0.2667 to 0.2635. In both cases, I conclude that post-hoc map matching is insufficient. Further research on integrating geographic context during training and post-processing is required.

Visual inspection of histograms and KDE plots for fractal dimension (Appendix Fig. A.2) shows that the overall effect of road network integration is limited, primarily because trip segments constitute a minority of trajectory points, while dwell segments dominate. To better understand the impact, I filtered trajectories that were modified by the road network and re-examined their fractal dimension distributions (Fig. 6.1). In Beijing, I observed that the modified trajectories sometimes exhibit even higher fractal dimensions than the original synthetic data, likely due to the dense and intricate road network, which introduces additional twists and complexity. Conversely, in Victoria, road network integration can bring fractal dimension closer to those of the real trajectories. The difference in outcomes between the two cities also reflects the relative success of the route-finding process. During the road matching process, Victoria experienced more cases where valid paths could not be found. Beijing's dense road network allowed for nearly all trips to be successfully matched, sometimes resulting in tortuous trajectories.

(a) Geolife



(b) Interact-VIC

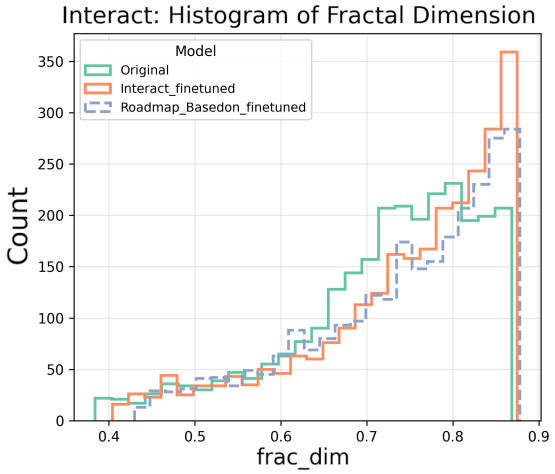


Figure 6.1: Histogram of fractal dimension computed on the Geolife and Interact datasets, comparing real trajectories with synthetic trajectories generated by different models. **(a)** Geolife dataset , **(b)** Interact dataset. Only trajectories containing modified trips are shown.

These findings suggest that road network integration can improve realism in certain contexts, but its application requires caution. In cities with complex networks, modifications can unintentionally increase route complexity. With appropriate constraints, integrating road network information remains a promising avenue for enhancing trajectory synthesis.

Chapter 7

Discussion and Summary

7.1 Discussion

The ultimate goal is to generate synthetic trajectories that can preserve the statistical characteristics of real human mobility across all ReFGeM features—ideally driving the Jensen–Shannon (JS) divergence toward zero. Currently, even the best-performing model maintains an average JS score close to 0.2, indicating that meaningful discrepancies remain between synthetic and real data. A closer examination of feature-level results shows that the multi-scale entropy coefficients (C1–C5) remain particularly challenging to replicate. The bi-modal distributions observed for these entropy rate constants(Appendix A.2, Fig. A.5–A.9) suggest that future models could benefit from explicitly distinguishing between dwell and trip states. Although exposing model with more city's data consistently narrows the distributional gap, noticeable discrepancies persist, implying that accurately modeling the temporal transitions between dwell and trip phases—core elements of human mobility dynamics—remains unresolved.

The cross-city transfer results show that the stable performance of the *Combine_trained* model reflects the benefit of training on multi-city data. Across all evaluation settings, the combined model consistently achieves the best or second-best performance, indicating that a sufficiently large and diverse training corpus can serve as a general foundation for synthetic mobility generation. These results suggest that, with adequate data coverage, it is plausible to generate realistic city-level synthetic trajectories—not to predict individual movements, but to reproduce aggregate mobility characteristics for a given city. Models trained or fine-tuned exclusively on the Victoria dataset exhibit limited transferability and perform well only on Victoria.

Models trained on Montreal or Geolife data show stronger cross-city generalization. This asymmetry indicates that cross-city transferability depends on both dataset diversity and geographic complexity. The Victoria dataset covers a smaller study area and a more homogeneous participant group, resulting in simpler mobility structures. By comparison, both the Montreal and Beijing datasets include broader geographic regions and more diverse participant populations, providing richer spatiotemporal patterns for the models to learn. The Geolife dataset is collected through participant-initiated uploads of individual trips [80], which introduces selection bias and incomplete daily coverage. In contrast, the Interact-VIC and Interact-MTL datasets are collected using wearable sensors that record mobility continuously throughout the day. This difference in data collection protocols is reflected in the observed gaps in features such as C5 (dwell time) when transferring models between Geolife and the Canadian cities. These results show that cross-city synthetic mobility generation is influenced by the consistency of data collection protocols as well as the diversity of populations and urban contexts represented in the training data. Training on the combined dataset reduces these discrepancies across cities, indicating that aggregation across heterogeneous sources provides a more stable basis for city-level synthetic mobility generation.

Road network integration produces mixed outcomes across regions. In Victoria, the integration reduced JS divergence, while in Beijing it led to more complex trajectories with higher fractal dimensions. This divergence reflects both the differences in road network density and connectivity and the success rate of route-finding: Victoria's sparse network led to many unmatched trips, limiting changes, whereas Beijing's dense road graph often generated overly intricate routes. While road constraints influence trajectory geometry, the improvements from post hoc integration remain limited. This suggests that road network information should be incorporated into the model training process rather than applied as a post-processing step.

This study demonstrates the value of this feature-based evaluation framework through the ReFGeM approach for assessing the quality of synthetic human mobility data. The framework not only quantifies how closely generated trajectories approximate real-world mobility patterns but also highlights specific aspects where existing models fall short, such as their inability to capture transitions between dwell and trip states. These versatile features cover multiple aspects of trajectory behavior, providing not only a measure of quality but also directional guidance for synthetic trajectory generation improvement. These findings suggest that ReFGeM can serve

as a practical feedback mechanism to guide model design and optimization, potentially functioning as a feedback metric in reinforcement learning to support iterative improvement toward more realistic mobility synthesis. The results also indicate that substantial work remains before synthetic trajectories become feasible: future models must generate trajectories which are at a minimum sufficiently credible to be considered valid under ReFGeM evaluation. ReFGeM operates at an aggregate level, and the generation protocol samples conditional information from learned distributions rather than deriving it directly from the test dataset. This design enables evaluation without exposing individual-level trajectories, supporting group-level analysis while preserving privacy. Together, these properties make ReFGeM a practical and privacy-aware framework for evaluating and guiding synthetic human mobility models.

Several limitations remain that point to promising directions for future work. My choice of baseline models was constrained by the goal of maintaining consistency with [9]. Although more advanced trajectory generative models exist, many do not offer public available pretrained implementations, limiting the ability to include them in a controlled comparison. As additional pretrained models are released, the proposed evaluation schema can be applied. The models I evaluated were restricted to trajectories of only 200 points at 5-second intervals ($\approx$ 16 minutes), which does not represent common trips like commuting behavior. While the 5-second interval was adopted for consistency with prior implementations, it may not be optimal across datasets. The sparsity and short duration contribute to underestimation of fractal dimensions and incomplete behavioral patterns. In future work, extending trajectory length beyond the current published state of the art and prioritizing datasets with consistent, high-frequency sampling may yield more realistic outcomes. Current models do not distinguish between dwell and trip phases, which constrains their ability to capture the transition dynamics between movement states. Future research could incorporate explicit representations of these two modes through hierarchical or hybrid generative architectures to better reflect real-world temporal structures and improve entropy-related feature quality. Embedding structural priors such as road network complexity or urban morphology descriptors could also help mitigate the observed asymmetry in cross-city transfer. Assembling multi-city training corpora with diverse mobility patterns may move the field toward a universal, city-agnostic trajectory generator.

7.2 Summary

This work presented a comprehensive evaluation of synthetic human mobility trajectories using the ReFGeM feature framework, demonstrating its utility for systematically assessing the realism and quality of generative models. I benchmarked several existing GAN-based and diffusion-based models, highlighting their varying abilities to reproduce the multi-dimensional feature distributions observed in real-world mobility data. Building upon this baseline, I investigated the effects of different training strategies, including training from scratch, fine-tuning on high-quality local datasets, and training on combined multi-city data, and examined its cross-city generalization capability. Through these experiments, I established a systematic understanding of how model architecture, data quality, and spatial context influence the divergence between synthetic and real trajectories, and demonstrate the utility of feature-based analysis of synthetic trajectories. Despite a remaining average JS divergence of around 0.2, the ReFGeM framework revealed where and why existing models diverge from true human mobility, offering interpretable diagnostics that complement conventional trajectory-level evaluations.

Beyond the specific results, this study contributes a feature-based evaluation paradigm that bridges generative modeling and mobility analysis. By capturing statistical, geometric, and temporal characteristics in a unified representation, ReFGeM offers a necessary foundation for future research on trajectory synthesis, model validation, and data-driven urban mobility understanding. Looking forward, integrating dwell–trip state modeling, geographic conditioning, and longer temporal contexts may help move synthetic trajectory generation toward greater behavioral plausibility and generalizability across cities. Advancing the ability to generate realistic trajectories also contributes to a deeper computational understanding of human mobility, forming the basis for agent-based simulations that can support urban design, transportation planning, and policy evaluation across both real and virtual environments.

Appendix A

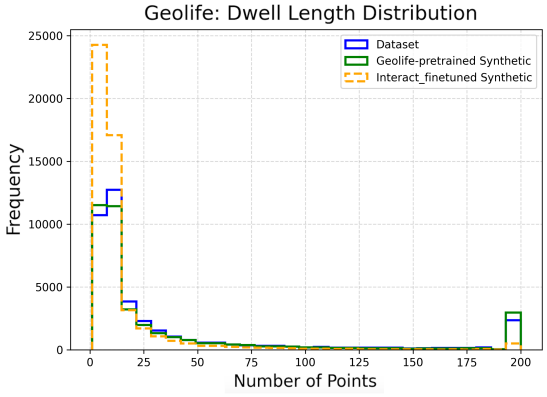
Additional Information

A.1 Supplementary Analysis of Auxiliary Work

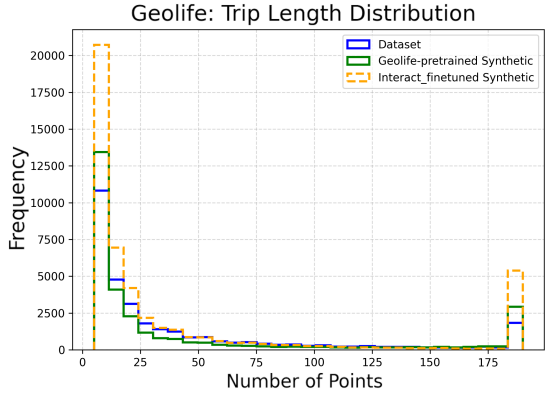
This section presents the visualization results from the ancillary work stage: including separate distributions of trip and dwell segment lengths, comparisons of fractal dimension distributions before and after road network integration, and examples of trajectory alignment with the Beijing road network.

Figure A.1 shows the distributions of trip and dwell segment lengths (in number of points) after trip–dwell detection. The first row corresponds to Geolife and the second to Interact-VIC, with each subfigure comparing the original data, *Geolife_pretrained*, and *Interact_finetuned* synthetic trajectories. Distributions showing greater deviation are indicated by dashed lines. For Geolife, the *Interact_finetuned* model generates many short trips and dwells, suggesting over-segmentation of trajectories into brief movement and stationary phases. In contrast, Interact-VIC real data contain longer dwells dominating entire trajectories, while the synthetic versions underestimate trip frequency and favor extended stationary periods.

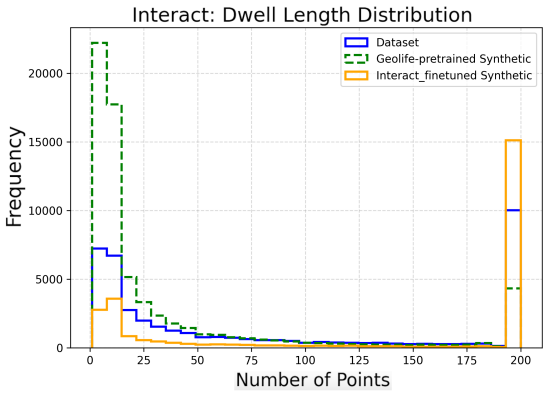
(a) Geolife: Length Distribution of Dwells



(b) Geolife: Length Distribution of Journeys



(c) Interact-VIC: Length Distribution of Dwells



(d) Interact-VIC: Length Distribution of Journeys

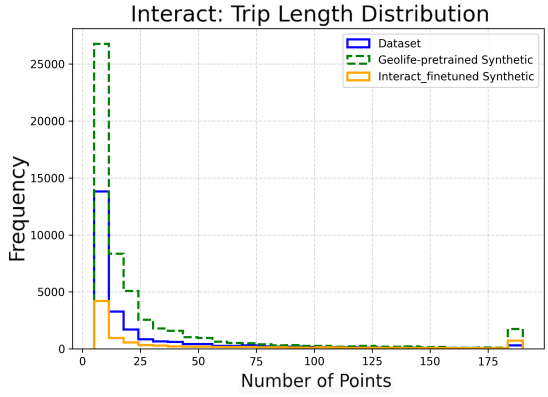
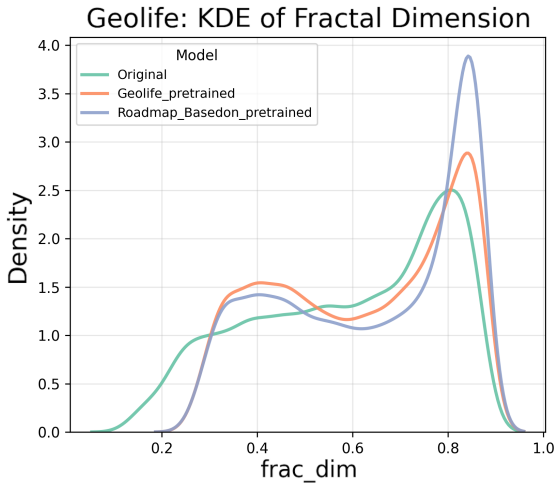


Figure A.1: Trip length and dwell-time distributions across Geolife and Interact datasets (real and synthetic). Subfigures **(a)**,**(c)** show dwells length(number of points) distributions; **(b)**,**(d)** show journeys length (number of points)distributions for each dataset respectively.

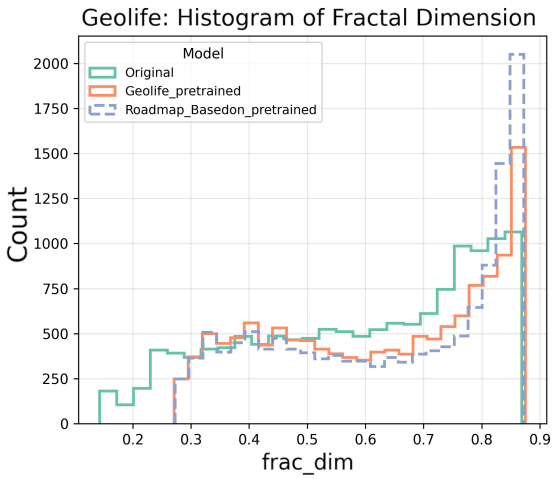
Figure A.2 shows the distributions of fractal dimensions computed from real and synthetic trajectories for the Geolife and Interact-VIC datasets, before and after integrating road network information. For each dataset, the model that best captures its characteristics is selected: *Geolife_pretrained* for Geolife and *Interact_finetuned* for Interact. In the Geolife dataset, the integration of road network information tends to affect trajectories with higher fractal dimensions, as shown by the dashed distribution skewed around the 0.8–0.9 range. This suggests that road alignment introduces more complex path shapes. For Interact-VIC, the pre- and post-integration curves

nearly overlap, showing minimal difference, due to the low success rate of aligning trajectories to the road network.

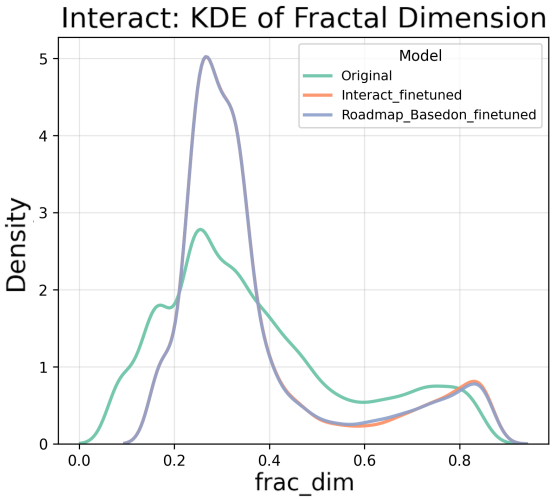
(a) Geolife:Kernel Density Estimation of Fractal Dimension



(b) Geolife: Histogram of Fractal Dimension



(c) Interact-VIC: Kernel Density Estimation of Fractal Dimension



(d) Interact-VIC:Histogram of Fractal Dimension

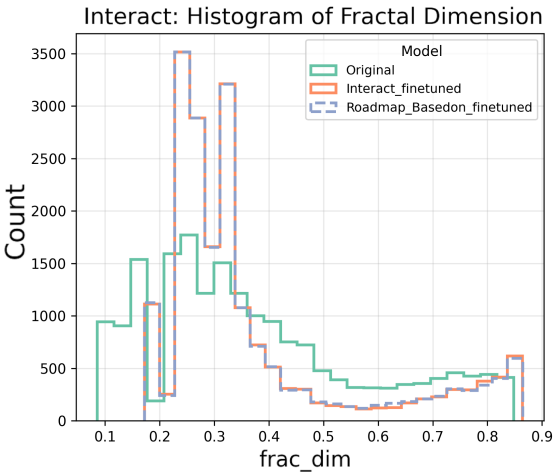


Figure A.2: KDE(Kernel Density Estimation) and histogram of fractal dimension computed on the Geolife and Interact datasets, comparing real trajectories with synthetic trajectories generated by different models. **(a)**–**(b)** correspond to the Geolife dataset (KDE and histogram, respectively), and **(c)**–**(d)** correspond to the Interact dataset (KDE and histogram, respectively).

Figure A.3 shows the KDE and histogram of fractal dimensions for the Geolife dataset, highlighting the performance of the *Interact_finetuned* model with Beijing road network integration to evaluate cross-city transferability. After applying road network integration, the number of synthetic trajectories with fractal dimensions in the 0.4–0.6 range decreases, bringing the distribution closer to the real dataset, while trajectories with fractal dimensions above 0.6 increase, better aligning the synthetic data with the patterns observed in actual trajectories.

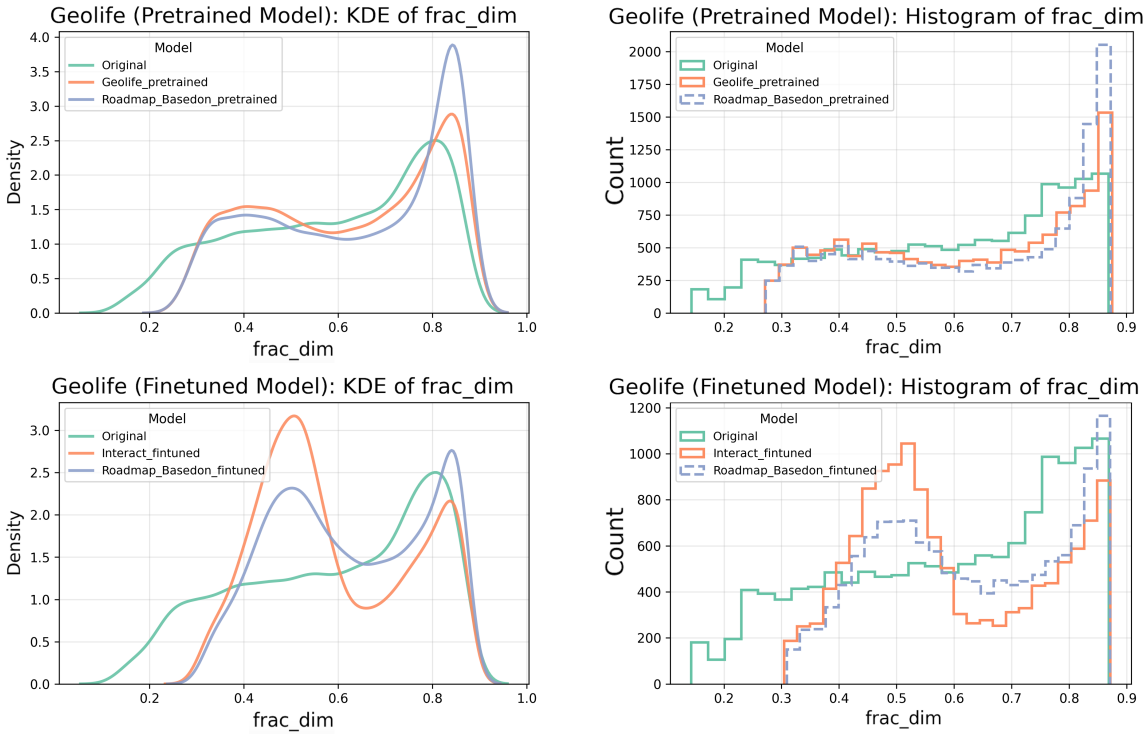


Figure A.3: KDE and histogram of fractal dimension computed on the Geolife dataset, comparing real Geolife trajectories with synthetic trajectories generated by different models. **(a)**–**(b)** correspond to the Geolife dataset with synthetic trajectories generated by the Geolife-pretrained model and its Roadmap post-processing variant (KDE and histogram, respectively), while **(c)**–**(d)** correspond to the Geolife dataset with synthetic trajectories generated by the Interact-finetuned model and its Roadmap post-processing variant (KDE and histogram, respectively).

Figure A.4 presents examples of trajectories after Beijing road network integration. Blue dots indicate the original synthetic waypoints, while red dots show the mapped trajectories. In Figure A.4a, mapping introduces zig-zags to align with the nearest roads. Figure A.4b illustrates a short, simple trajectory being mapped onto complex roads, including overpasses and ring roads, resulting in a higher fractal dimension.

(a) (b)

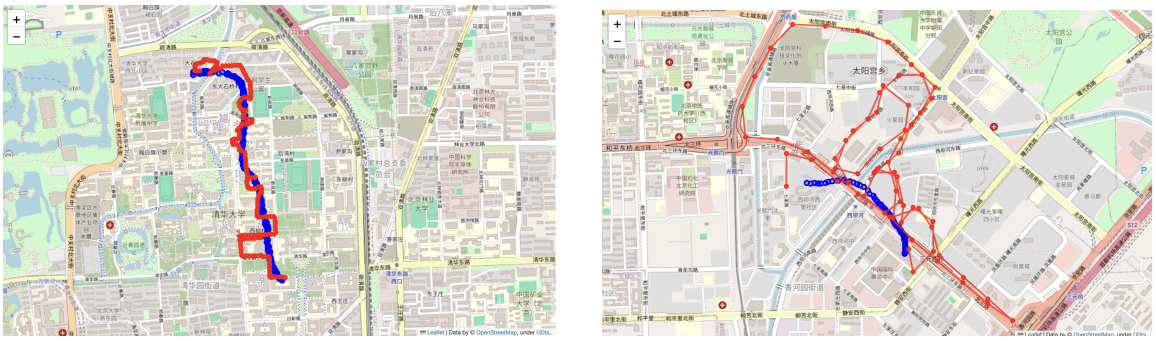


Figure A.4: Examples of trajectories after Beijing road network mapping. Blue dots are original synthetic waypoints, and red dots are the mapped trajectories.

A.2 ReFGeM Feature Distribution Plots

Figures A.5–A.12 show the distributions of ReFGeM features for synthetic trajectories generated by all model variants trained or fine-tuned on the Interact datasets. Features include multi-scale entropy coefficients (C1–C5), buffer area, convex hull, and fractal dimension. All features except fractal dimension use log–log scales; fractal dimension is linear. Subfigures are organized in three rows (Geolife top, Interact-VIC middle, Interact-MTL bottom) and seven columns (original data, *Geolife_trained* synthetic, *VIC_trained* synthetic, *VIC_finetuned* synthetic,*MTL_trained* synthetic, *MTL_finetuned* synthetic), *Combine_trained*. Synthetic trajectories capture some real-data patterns, but discrepancies remain. C1–C4 show bi-modal distributions reflecting dwell and trip states; in C1–C4, the relative height of low-value bars and small scale indicate dominance of C5, representing dwell time. The *Combine_trained* model reproduces spatial and geometric features effectively, particularly for the Geolife synthetic dataset. For example, the buffer area and convex hull of the Geolife test set follow an approximately log-uniform distribution; the *Geolife_trained* model fails to

capture this distribution accurately, whereas the *Combine_ trained* model reproduces it well.

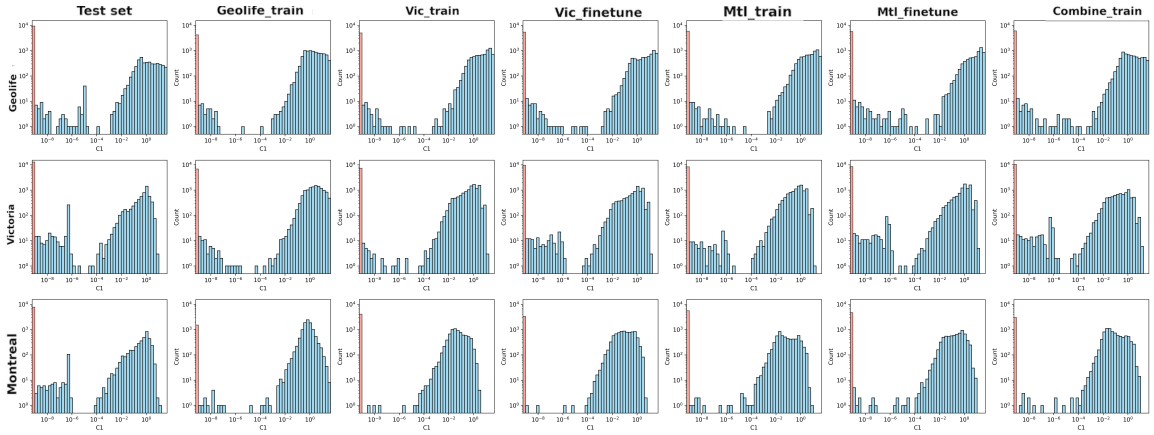


Figure A.5: ReFGeM feature distribution: C1. Values outside the 2nd–98th percentiles are removed as outliers.

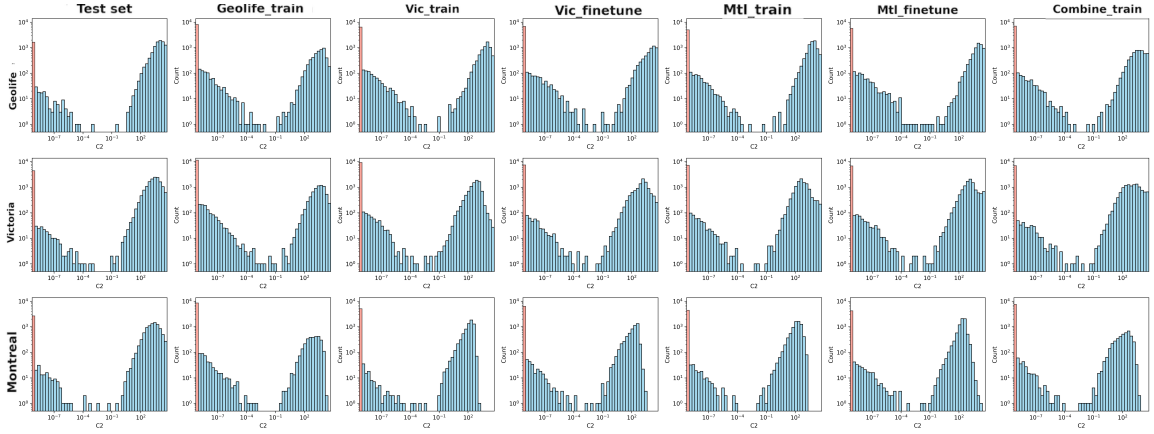


Figure A.6: ReFGeM feature distribution: C2. Values outside the 2nd–98th percentiles are removed as outliers.

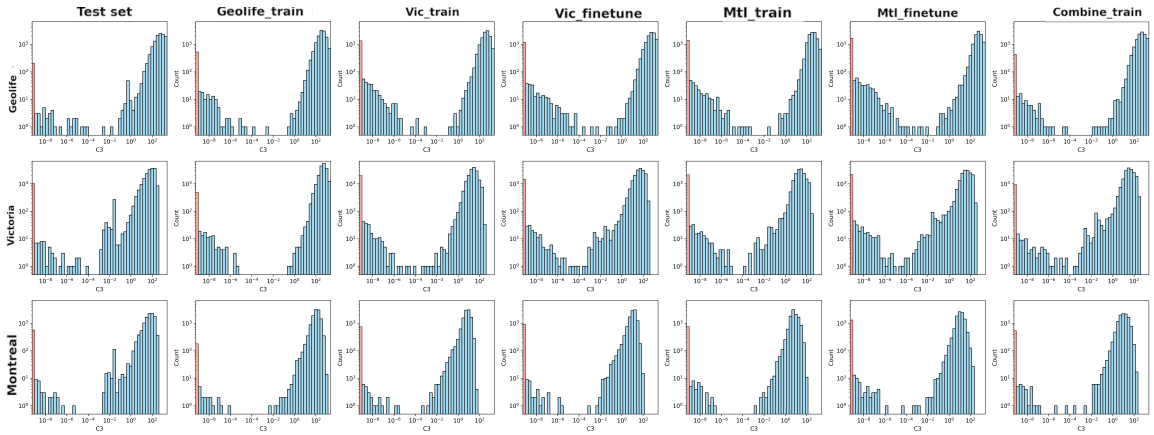


Figure A.7: ReFGeM feature distribution: C3. Values outside the 2nd–98th percentiles are removed as outliers.

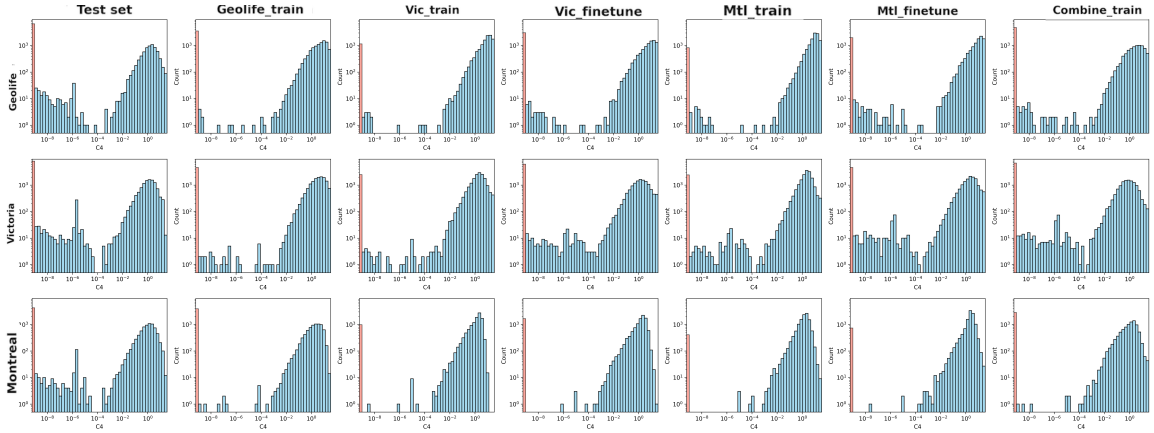


Figure A.8: ReFGeM feature distribution: C4. Values outside the 2nd–98th percentiles are removed as outliers.

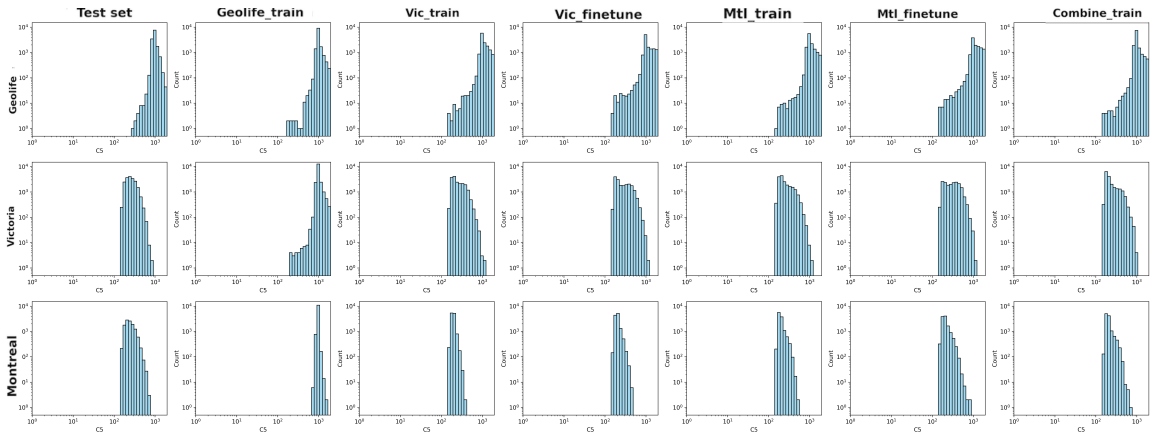


Figure A.9: ReFGeM feature distribution: C5. Values outside the 2nd–98th percentiles are removed as outliers.

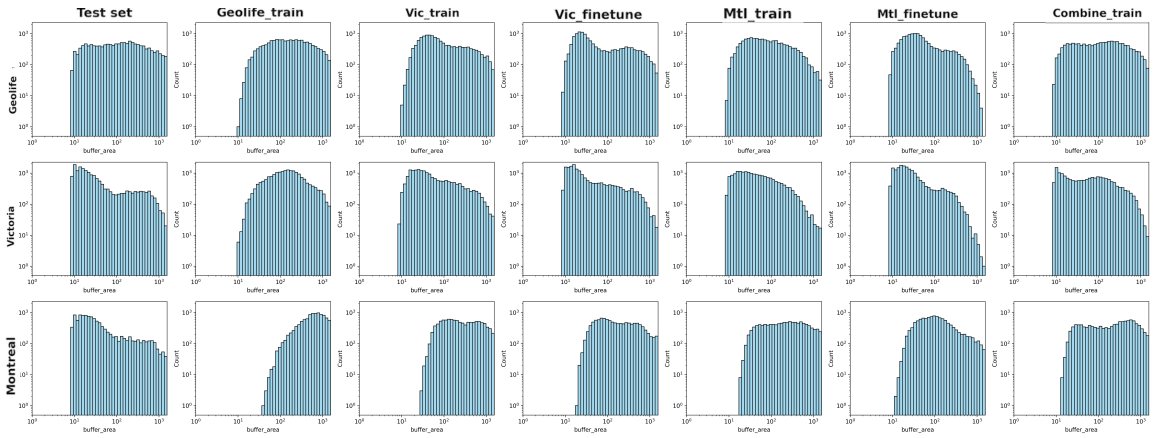


Figure A.10: ReFGeM feature distribution: Buffer Area. Values outside the 2nd–98th percentiles are removed as outliers.

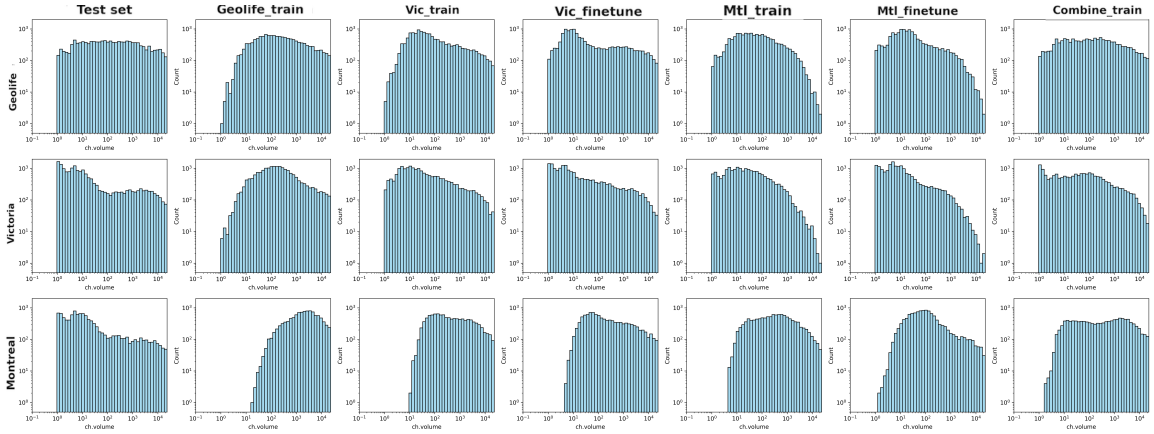


Figure A.11: ReFGeM feature distribution: Convex Hull Volume. Values outside the 2nd–98th percentiles are removed as outliers.

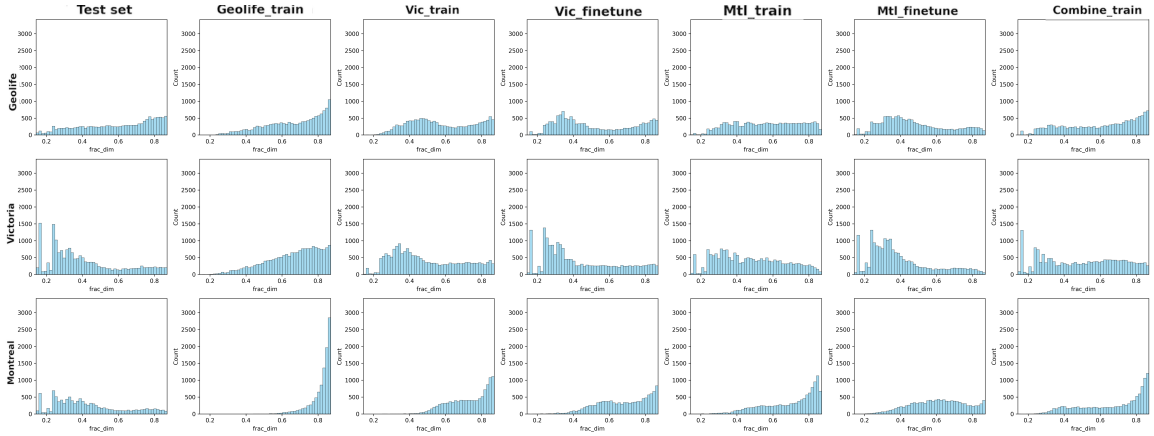


Figure A.12: ReFGeM feature distribution: Fractal Dimension. Values outside the 2nd–98th percentiles are removed as outliers.

A.3 Model Parameters

Table A.1 lists the parameters of the *Noise-TrajGAN* model, following the implementation in [9]. Table A.2 summarizes the parameters of *GeoTrajGAN* and its WGAN-LP variant, which was tested to mitigate mode collapse, also based on [9]. Table A.3 details the configurations of all *DiffTraj* models and their variants, including the unconditional DiffTraj trained on Geolife, the 8-stat DiffTraj trained from scratch on Geolife(*Geolife_trained*), Interact-VIC(*VIC_trained*), Interact-MTL(*MTL_trained*), a model trained on the combined Geolife, Interact-VIC, and Interact-MTL datasets (*Combine_trained*) , and the fine-tuned version on Interact-VIC (*VIC_finetuned*) and Interact-MTL (*MTL_finetuned*). These implementations follow [10].

Table A.1: Model Parameters (Noise_TrajGAN)

Parameter	**Value**
Model Name	Noise_TrajGAN_GEOLIFE
Features	[latlon, hour, day]
Latent Dim	100
Noise Dim	100
Gradient Penalty	True
Lipschitz Penalty	True
Use Regularizer	True
WGAN Mode	True
Differential Privacy	False
	Embedding Layers
latlon	Linear($2 \rightarrow 64$) + ReLU
hour	Linear($24 \rightarrow 24$) + ReLU
day	Linear($7 \rightarrow 7$) + ReLU
Feature Fusion	Linear($95 \rightarrow 100$) + ReLU
	Generator
Fusion Layer	Linear($100 \rightarrow 100$) + ReLU
LSTM	LSTM(100, 100, batch_first=True)
Output (latlon)	Linear($100 \rightarrow 2$) + Tanh
Output (hour)	Linear($100 \rightarrow 24$) + Softmax
Output (day)	Linear($100 \rightarrow 7$) + Softmax
Loss Function	TrajLoss(BCELoss)
Optimizer	AdamW(lr=0.001, betas=(0.5,0.999), weight_decay=0.01)
	Discriminator
Embedding Layers	Same as Generator
Feature Fusion	Linear($95 \rightarrow 100$) + ReLU
LSTM	LSTM(100, 100, batch_first=True)
Output Layer	Linear($100 \rightarrow 1$)
Loss Function	BCEWithLogitsLoss()
Optimizer	AdamW(lr=0.001, betas=(0.5,0.999), weight_decay=0.01)

Table A.2: Configurations of GeoTrajGAN and GeoTrajGAN-WGAN-LP Parameters

Parameter	**GeoTrajGAN**	**GeoTrajGAN-WGAN-LP**
Generator		
n_dim	2	2
Uses PointNet	True	True
Code Size	256	256
Normalization	NormOption.LAYER	NormOption.LAYER
Activation	ReLU()	ReLU()
Mode	linear	linear
Sequential Mode	True	True
Use LSTM	True	True
LSTM Latent Dim	64	64
Bidirectional LSTM	True	True
Merge Mode	sum	sum
Optimizer	AdamW(lr=1e-05, betas=(0.5,0.999))	AdamW(lr=1e-05, betas=(0.5,0.999))
Loss Function	BCEWithLogitsLoss()	WGAN
Discriminator		
n_dim	2	2
Uses STN	True	True
Code Size	256	256
Normalization	NormOption.LAYER	NormOption.LAYER
Activation	ReLU()	ReLU()
Mode	linear	linear
Sequential Mode	True	True
LSTM at Start	False	False
LSTM Latent Dim	64	64
Bidirectional LSTM	True	True
Merge Mode	sum	sum
Optimizer	AdamW(lr=1e-05, betas=(0.5,0.999))	AdamW(lr=1e-05, betas=(0.5,0.999))
Trajectory Critic	True	True
Trajectory STN	False	False
Shared PointNet	False	False
Loss Function	BCEWithLogitsLoss()	WGAN-LP
lambda_gp (GP Multiplier)	Not applicable	10

Table A.3: Full configuration of Set of DiffTraj models on Geolife and Finetuned on Interact datasets.

Category / Parameter	**Uncond_DiffTraj_GEOLIFE**	**GEOLIFE/VIC/MTL/Combine_Trained**	**VIC/MTL_Finetuned**
General Settings			
Dataset	Geolife	Geolife/Interact-VIC/Interact-MTL/Combine	Interact-VIC/Interact-MTL
Model type	DiffTraj (unconditional)	DiffTraj (conditional)	DiffTraj (conditional)
Batch size	1024	1024	1024
K-fold	5	5	5
Seed	42	42	42
GPU	0	0	0
Gradient clip norm	1.0	1.0	1.0
EMA	False	False	False
Model Architecture			
Input channels	2	2	2
Base channels (ch)	128	128	128
Channel multiplier (ch_mult)	(1, 2, 2, 2)	(1, 2, 2, 2)	(1, 2, 2, 2)
Residual blocks	2	2	2
Dropout	0.1	0.1	0.1
Attention resolutions	(16)	(16)	(16)
Attribute dimension	8	8	8
Output channels	2	2	2
Guidance scale	3	3	3
Diffusion Process			
Beta schedule	Linear	Linear	Linear
β_{start}	0.0001	0.0001	0.0001
β_{end}	0.05	0.05	0.05
Diffusion timesteps	500	500	500
Generation timesteps	100	100	100
η	0.0	0.0	0.0
Optimizer			
Optimizer type	AdamW	AdamW	AdamW
Learning rate	**2e-4**	**2e-4**	**2e-5**
Betas	(0.9, 0.999)	(0.9, 0.999)	(0.9, 0.999)
Weight decay	None	None	None
Conditional Information			
Conditional	False	True	True
Embedding type	—	Integrated FC	Integrated FC
Embedding dimension	—	512	512
Hidden dimension	—	256	256
Noise type	—	Gaussian	Gaussian
Noise schedule	—	Enabled	Enabled
Clipping type	—	Norm clipping	Norm clipping
Max norm	—	1.0	1.0
Epsilon (ϵ)	—	1.0	1.0
Delta (δ)	—	0.0	0.0
Location	—	Hidden	Hidden
Source	—	DiffTraj	DiffTraj

A.4 GeoTrajGAN Mode Collapse Experiment

To address the mode collapse observed in GeoTrajGAN, I replaced the original Binary Cross-Entropy (BCE) loss with a WGAN-LP (Wasserstein GAN with Lipschitz penalty) loss [49]. However, this modification did not improve model performance; in fact, the generated points became more concentrated near boundaries and overall more scattered. Additional trials and hyperparameter tuning might alleviate this issue, but exploring these solutions is beyond the scope of this work.

Figure A.13 illustrates a comparison of outputs between the original GeoTrajGAN and the WGAN-LP variant, alongside the real Geolife trajectories.

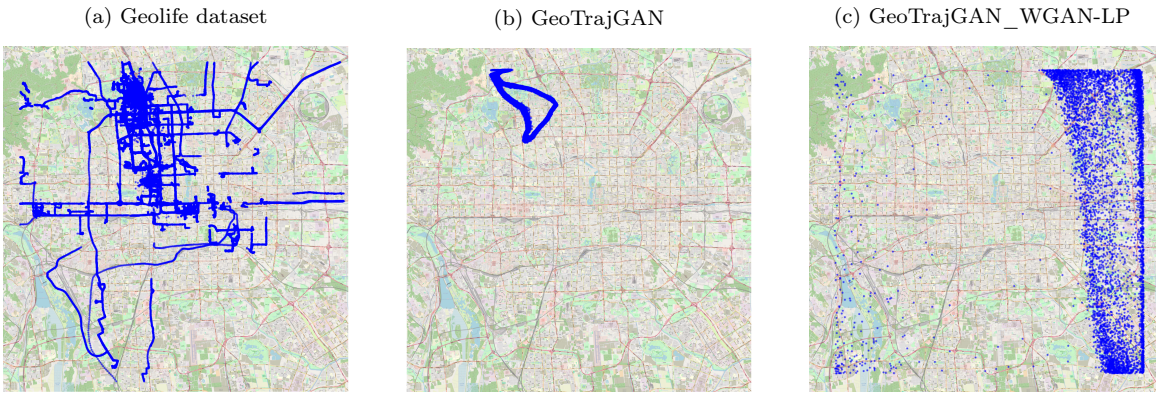


Figure A.13: Random samples of 5,000 trajectories from the Geolife dataset and synthetic trajectories generated by GeoTrajGAN and its WGAN-LP variant. The background displays the road network of Beijing. **(a)** Real Geolife trajectories, **(b)** GeoTrajGAN outputs, **(c)** GeoTrajGAN_WGAN-LP outputs.

Bibliography

[1] Acm sigspatial cup 2025: Human mobility prediction challenge. https://sigspatial2025.sigspatial.org/giscup/index.html. Accessed: 2025-09-30.

[2] Mobysens sensedoc. http://mobysens.com/en/. Accessed: 2025-12-12.

[3] Hossein Amiri, Richard Yang, and Andreas Züfle. Geolife+: Large-scale simulated trajectory datasets calibrated to the geolife dataset. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*, GeoSim '24, page 25–28, New York, NY, USA, 2024. Association for Computing Machinery.

[4] Pichamon Anantasech and Chotirat Ann Ratanamahatana. Enhanced weighted dynamic time warping for time series classification. In *Third International Congress on Information and Communication Technology: ICICT 2018, London*, pages 655–664. Springer, 2018.

[5] Hugo Barbosa-Filho, Marc Barthelemy, Gourab Ghoshal, Charlotte R. James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J. Ramasco, Filippo Simini, and Marcello Tomasini. Human Mobility: Models and Applications. *Physics Reports*, 734:1–74, March 2018. arXiv:1710.00004 [physics].

[6] Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD Workshop*, 1994.

[7] Lorenzo Bracciale, Marco Bonola, Pierpaolo Loreti, Giuseppe Bianchi, Raul Amici, and Antonello Rabuffi. Crawdad roma/taxi, 2022.

[8] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, January 2006. Publisher: Nature Publishing Group.

[9] Erik Buchholz, Sharif Abuadbba, Shuo Wang, Surya Nepal, and Salil Kanhere. Sok: Can trajectory generation combine privacy and utility? 07 2024.

[10] Erik Buchholz, Natasha Fernandes, David D. Nguyen, Alsharif Abuadbba, Shuo Wang, Surya Nepal, and Salil S. Kanhere. What is the cost of differential privacy for deep learning-based trajectory generation?, June 2025. 21 pages.

[11] Yanguang Chen. Fractal modeling and fractal dimension description of urban morphology. *Entropy*, 22(9):961, 2020.

[12] Yile Chen, Yicheng Tao, and Gao Cong. Enhancing large language models for mobility analytics with semantic location tokenization. *arXiv preprint arXiv:2506.11109*, 2025. Explicitly addresses the quantization error and vocabulary issues in coordinate tokenization.

[13] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[14] Chen Chu, Hengcai Zhang, Peixiao Wang, and Feng Lu. Simulating human mobility with a trajectory generation framework based on diffusion model. *International Journal of Geographical Information Science*, 38(5):847–878, 2024.

[15] Giuliano Cornacchia and Luca Pappalardo. A Mechanistic Data-Driven Approach to Synthesize Human Mobility Considering the Spatial, Temporal, and Social Dimensions Together. *ISPRS International Journal of Geo-Information*, 10(9):599, September 2021.

[16] Chris Culnane, A/Benjamin I. P. Rubinstein, and A/Vanessa Teague. Stop the Open Data Bus, We Want to Get Off, August 2019. arXiv:1908.05004 [cs].

[17] Teddy Cunningham, Konstantin Klemmer, Hongkai Wen, and Hakan Ferhatosmanoglu. Geopointgan: Synthetic spatial data with local label differential privacy. *arXiv preprint arXiv:2205.08886*, 2022.

[18] Yves-Alexandre de Montjoye, Sébastien Gambs, Vincent Blondel, Geoffrey Canright, Nicolas de Cordes, Sébastien Deletaille, Kenth Engø-Monsen, Manuel Garcia-Herranz, Jake Kendall, Cameron Kerry, Gautier Krings, Emmanuel

Letouzé, Miguel Luengo-Oroz, Nuria Oliver, Luc Rocher, Alex Rutherford, Zbigniew Smoreda, Jessica Steele, Erik Wetter, Alex "Sandy" Pentland, and Linus Bengtsson. On the privacy-conscientious use of mobile phone data. *Scientific Data*, 5(1):180286, December 2018. Publisher: Nature Publishing Group.

[19] Daniel Dias and LHMK Costa. Crawdad dataset coppe-ufrj/riobuses (v. 2018-03-19). *Downloaded from*, 2018.

[20] Somayeh Dodge, Patrick Laube, and Robert Weibel. Movement similarity assessment using symbolic representation of trajectories. *International Journal of Geographical Information Science*, 26(9):1563–1588, 2012.

[21] Somayeh Dodge, Robert Weibel, and Ehsan Forootan. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, 33(6):419–434, 2009.

[22] Yingling Fan and Asad J Khattak. Urban form, individual spatial footprints, and travel: Examination of space-use behavior. *Transportation Research Record*, 2082(1):98–106, 2008.

[23] Farzan Farnia and David Tse. A convex duality framework for gans. *arXiv preprint arXiv:1810.11740*, 2018.

[24] Jie Feng, Yong Li, Chao Zhang, Funing Sun, Fanchao Meng, Ang Guo, and Depeng Jin. DeepMove: Predicting Human Mobility with Attentional Recurrent Networks. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, pages 1459–1468, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[25] Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, page 3426–3433, New York, NY, USA, 2020. Association for Computing Machinery.

[26] A Stewart Fotheringham and David WS Wong. The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*, 23(7):1025–1044, 1991.

[27] Marta C González, César A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[28] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Y. Bengio. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3, 06 2014.

[29] Sihui Guo, Tao Pei, Shuyun Xie, Ci Song, Jie Chen, Yaxi Liu, Hua Shu, Xi Wang, and Ling Yin. Fractal dimension of job-housing flows: A comparison between beijing and shenzhen. *Cities*, 112:103120, 2021.

[30] Veton Hamza, Bojan Stopar, Sterle Oskar, and Polona Prešeren. Observations and positioning quality of low-cost gnss receivers: a review. *GPS Solutions*, 28, 06 2024.

[31] Amin Hosseinpoor Milaghardan, Rahim Ali Abbaspour, and Christophe Claramunt. A geometric framework for detection of critical points in a trajectory using convex hulls. *ISPRS International Journal of Geo-Information*, 7(1):14, 2018.

[32] Jhoan Keider Hoyos-Osorio and Luis Gonzalo Sanchez-Giraldo. The representation jensen-shannon divergence. *arXiv preprint arXiv:2305.16446*, 2023.

[33] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. González. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, September 2016. Publisher: Proceedings of the National Academy of Sciences.

[34] James Jordon, Jinsung Yoon, and Mihaela Van Der Schaar. Pate-gan: Generating synthetic data with differential privacy guarantees. In *International Conference on Learning Representations (ICLR)*, 2019.

[35] Eamonn J Keogh and Michael J Pazzani. Scaling up dynamic time warping for datamining applications. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 285–289, 2000.

[36] Yan Kestens, Meghan Winters, Daniel Fuller, Scott Bell, Janelle Berscheid, Ruben Brondeel, Michael Cantinotti, Geetanjali Datta, Lise Gauvin, Margot Gough, Karen Laberee, Paul Lewis, Sébastien Lord, Hui Luan, Heather McKay,

Catherine Morency, Nazeem Muhajarine, Trisalyn Nelson, Callista Ottoni, and Wasfi Rania. Interact: A comprehensive approach to assess urban form interventions through natural experiments. *BMC Public Health*, 19, 01 2019.

[37] S. Kullback. *Information Theory and Statistics*. John Wiley & Sons, 1951.

[38] Xiao Liang, Jichang Zhao, Li Dong, and Ke Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3(1):2983, October 2013. Publisher: Nature Publishing Group.

[39] J. Lin. Divergence measures based on the shannon entropy. In *IEEE Transactions on Information Theory*, volume 37, pages 145–151, 1991.

[40] Xi Liu, Hanzhou Chen, and Clio Andris. trajgans: Using generative adversarial networks for geo-privacy protection of trajectory data (vision paper). In *Location privacy and security workshop*, pages 1–7, 2018.

[41] Chris Ma, D.K.Y. Yau, Nung Kwan Aaron Yip, and Nageswara Rao. Privacy vulnerability of published anonymous mobility traces. *Networking, IEEE/ACM Transactions on*, 21:720–733, 06 2013.

[42] Benoit B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, 1983.

[43] Àlex Miranda-Pascual, Patricia Guerra-Balboa, Javier Parra-Arnau, Jordi Forné, and Thorsten Strufe. Sok: Differentially private publication of trajectory data. *Proceedings on Privacy Enhancing Technologies*, 2023.

[44] Yves-Alexandre Montjoye, Michel Verleysen, and Vincent Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3:1376, 03 2013.

[45] Kun Ouyang, Reza Shokri, David Rosenblum, and Wenzhuo Yang. A non-parametric generative model for human trajectories. pages 3812–3817, 07 2018.

[46] Mia A. Papas et al. Measuring environmental exposures in people's activity space. *International Journal of Environmental Research and Public Health*, 2020.

[47] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in

human mobility. *Nature Communications*, 6(1):8166, September 2015. Publisher: Nature Publishing Group.

[48] Tuhin Paul, Kevin G. Stanley, and Nathaniel D. Osgood. Multiscale entropy rate analysis of complex mobile agents. *Royal Society Open Science*, 5(10):180488, October 2018.

[49] Henning Petzka, Asja Fischer, and Denis Lukovnikov. On the regularization of wasserstein gans. 09 2017.

[50] Natalia Ponomareva, Hussein Hazimeh, Alex Kurakin, Zheng Xu, Carson Denison, H. McMahan, Sergei Vassilvitskii, Steve Chien, and Abhradeep Thakurta. How to dp-fy ml: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77:1113–1201, 07 2023.

[51] Daniel G.C. Rainham et al. Are buffers around home representative of physical activity spaces? *International Journal of Behavioral Nutrition and Physical Activity*, 2013.

[52] Jinmeng Rao, Song Gao, Yuhao Kang, and Qunying Huang. LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection, June 2020. arXiv:2006.10521 [cs].

[53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. volume 9351, pages 234–241, 10 2015.

[54] Ludger Rüschendorf. The wasserstein distance and approximation theorems. *Probability Theory and Related Fields*, 70(1):117–129, 1985.

[55] Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. Socio-spatial properties of online location-based social networks. In *Proceedings of the Fifth International AAAI Conference on Web and Social Media (ICWSM)*, pages 329–336, 2011.

[56] Christian M. Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C. González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, July 2013. Publisher: Royal Society.

[57] Stefan Schönfelder and Kay W Axhausen. Activity spaces: measures of social exclusion? *Transport policy*, 10(4):273–286, 2003.

[58] Toru Shimizu, Kota Tsubouchi, and Takahiro Yabe. Geo-bleu: similarity measure for geospatial sequences. pages 1–4, 11 2022.

[59] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015.

[60] Aivin V. Solatorio. GeoFormer: Predicting Human Mobility using Generative Pre-trained Transformer (GPT). In *Proceedings of the 1st International Workshop on the Human Mobility Prediction Challenge*, pages 11–15, November 2023. arXiv:2311.05092 [cs].

[61] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modeling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, October 2010. arXiv:1010.0436 [cond-mat].

[62] Chris Stanford, Suman Adari, Xishun Liao, Brian He, Qinhua Jiang, Chenchen Kuai, Jiaqi Ma, Emmanuel Tung, Yinlong Qian, Lingyi Zhao, Zihao Zhou, Zeeshan Rasheed, and Khurram Shafique. Numosim: A synthetic mobility dataset with anomaly detection benchmarks. pages 68–78, 11 2024.

[63] Han Su, Shuncheng Liu, Bolong Zheng, Xiaofang Zhou, and Kai Zheng. A survey of trajectory distance measures and performance evaluation. *The VLDB Journal*, 29:1–30, 01 2020.

[64] Thomas M. Sutter, Imant Daunhawer, and Julia E. Vogt. Multimodal generative learning utilizing jensen–shannon divergence. In *Advances in Neural Information Processing Systems*, 2020.

[65] Jameson L Toole, Serdar Colak, Bradley Sturt, Lauren P Alexander, Alexandre Evsukoff, and Marta C González. The path most traveled: Travel demand estimation using big data. *Transportation Research Part C: Emerging Technologies*, 58:162–177, 2015.

[66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. 06 2017.

[67] Michail Vlachos, George Kollios, and Dimitrios Gunopulos. Discovering similar multidimensional trajectories. In *Proceedings 18th international conference on data engineering*, pages 673–684. IEEE, 2002.

[68] Jiawei Wang, Renhe Jiang, Chuang Yang, Zengqing Wu, Makoto Onizuka, Ryosuke Shibasaki, Noboru Koshizuka, and Chuan Xiao. Large Language Models as Urban Residents: An LLM Agent Framework for Personal Mobility Generation. 2024. Publisher: arXiv Version Number: 3.

[69] Xingrui Wang, Xinyu Liu, Ziteng Lu, and Hanfang Yang. Large scale gps trajectory generation using map based on two stage gan. *Journal of Data Science*, pages 126–141, 02 2021.

[70] Jonathan A. Ward, Andrew J. Evans, and Nicolas S. Malleson. Dynamic calibration of agent-based models using data assimilation. *Royal Society Open Science*, 3(4):150703, April 2016. Publisher: Royal Society.

[71] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, Kaoru Sezaki, Esteban Moro, and Alex Pentland. YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories. *Scientific Data*, 11(1):397, April 2024. Publisher: Nature Publishing Group.

[72] Dingqi Yang, Daqing Zhang, Vincent W Zheng, and Zhiyong Yu. Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 45(1):129–142, 2014.

[73] Jing Yuan, Yu Zheng, Chengyang Zhang, Wenlei Xie, Xing Xie, Guangzhong Sun, and Yan Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, page 99–108, New York, NY, USA, 2010. Association for Computing Machinery.

[74] Shannon N. Zenk et al. Destinations that older adults experience within their gps activity spaces. *International Journal of Environmental Research and Public Health*, 2015.

[75] Jing Zhang, Qihan Huang, Yirui Huang, Qian Ding, and Pei-Wei Tsai. Dp-trajgan: A privacy-aware trajectory generation model with differential privacy. *Future Gener. Comput. Syst.*, 142(C):25–40, May 2023.

[76] Rui Zhang, Kevin Stanley, Daniel Fuller, and Scott Bell. Differentiating population spatial behavior using representative features of geospatial mobility (refgem). *ACM Transactions on Spatial Algorithms and Systems*, 6:1–25, 02 2020.

[77] Yuheng Zhang, Yuan Yuan, Jingtao Ding, Jian Yuan, and Yong Li. Noise Matters: Diffusion Model-based Urban Mobility Generation with Collaborative Noise Priors, February 2025. arXiv:2412.05000 [cs].

[78] Xingyu Zhao, Xiao Zhang, Bohan Zhang, Jianpeng Qi, Junyu Dong, and Yanwei Yu. Ma2traj: Diffusion network with multi-attribute aggregation for trajectory generation. *Geoinformatica*, 29(3):789–810, May 2025.

[79] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 312–321, New York, NY, USA, 2008. Association for Computing Machinery.

[80] Yu Zheng, Xing Xie, and Wei-Ying Ma. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33:32–39, 06 2010.

[81] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 791–800, New York, NY, USA, 2009. Association for Computing Machinery.

[82] Lu Zhong, Lei Dong, Qi R. Wang, Chaoming Song, and Jianxi Gao. Universal expansion of human mobility across urban scales. *Nature Cities*, 2(7):603–607, 2025.

[83] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. Difftraj: Generating gps trajectory with diffusion probabilistic model. *Advances in Neural Information Processing Systems*, 36:65168–65188, 2023.

[84] Yuanshao Zhu, James Jianqiao Yu, Xiangyu Zhao, Qidong Liu, Ye Yongchao, Wei Chen, Zijian Zhang, Xuetao Wei, and Yuxuan Liang. Controltraj: Controllable trajectory generation with topology-constrained diffusion model. pages 4676–4687, 08 2024.