Geolocation Representation from Large Language Models are Generic Enhancers for Spatio-Temporal Learning

Junlin He[*], **Tong Nie**[*], **Wei Ma**[†]

The Hong Kong Polytechnic University
Hong Kong SAR, China
{junlinspeed.he, tong.nie}@connect.polyu.hk, wei.w.ma@polyu.edu.hk

Abstract

In the geospatial domain, universal representation models are significantly less prevalent than their extensive use in natural language processing and computer vision. This discrepancy arises primarily from the high costs associated with the input of existing representation models, which often require street views and mobility data. To address this, we develop a novel, training-free method that leverages large language models (LLMs) and auxiliary map data from OpenStreetMap to derive geolocation representations (LLMGeovec). LLMGeovec can represent the geographic semantics of city, country, and global scales, which acts as a generic enhancer for spatio-temporal learning. Specifically, by direct feature concatenation, we introduce a simple yet effective paradigm for enhancing multiple spatio-temporal tasks including geographic prediction (GP), long-term time series forecasting (LTSF), and graph-based spatio-temporal forecasting (GSTF). LLMGeovec can seamlessly integrate into a wide spectrum of spatio-temporal learning models, providing immediate enhancements. Experimental results demonstrate that LLMGeovec achieves global coverage and significantly boosts the performance of leading GP, LTSF, and GSTF models. Our codes are available at https://github.com/Umaruchain/LLMGeovec.

Introduction

Geolocation representation models encode geographical coordinates into latent embeddings with enriched geographic contextual information. Such embeddings ensure that similar representations reflect analogous sociodemographic attributes, activity patterns, and climatic characteristics across locations over the globe (Wang et al. 2022; Jean et al. 2019; Lee et al. 2021; Wang, Li, and Rajagopal 2020; Zhang et al. 2021, 2023a; Zhou et al. 2023b; Kim and Yoon 2022).

These geolocation representations are naturally suited to enhance spatio-temporal learning because they carry spatial contextual semantics. Previous research has focused only on how geolocation representations can be used for geographic prediction (GP). Specifically, GP tasks are trained on the attributes of some locations and used to predict the attributes of the remaining locations. These attributes include crime rate

(Li et al. 2022b; Kim and Yoon 2022), poverty rate (Jean et al. 2016; Chi et al. 2022; Marty and Duhaut 2024), public health (Yeh et al. 2021; Nilsen et al. 2021; Draidi Areed et al. 2022; Chang et al. 2022; Sheehan et al. 2019), and so on. However, these geolocation representations are not used for more complex tasks.

The reason why previous approaches do not further extend the complex applications is that they do not achieve global coverage and they have a heavy reliance on expensive input data such as street views, travel patterns, and traffic trajectories (Wang, Li, and Rajagopal 2020; Kim and Yoon 2022; Lee et al. 2021; Zhang et al. 2023a). Although studies have used free and globally available satellite imagery for geolocation representation, their effectiveness has been hampered by the low resolution and absence of important features such as activity patterns (Manvi et al. 2023; Robinson, Hohman, and Dilkina 2017; Head et al. 2017; Jean et al. 2019; Elmustafa et al. 2022; Xi et al. 2022; Ma et al. 2023; Sun 2024).

Our objective is to develop a generic and effective geolocation representation method that utilizes only readily accessible global data to improve more spatio-temporal learning tasks: GP, long-term time series forecasting (LTSF), and graph-based spatio-temporal forecasting (GSTF). The latter two are typical spatio-temporal datasets that, given the values of many nodes at historical moments, predict the future values of those nodes. They differ in that the former tends to deal with correlations between nodes by channel-mixing strategies while the latter aggregates spatial connections between nodes by graph neural networks (GNNs).

Recent advancements have demonstrated the extensive spatio-temporal and human-related knowledge embedded within large language models (LLMs). Some studies have transformed GP tasks as text generation tasks in LLMs (Manvi et al. 2023, 2024), and some have even found that LLMs learn linear representations of space and time across multiple scales (global, country, city) (Gurnee and Tegmark 2023). Inspired by these findings, we explore the potential of LLMs to generate effective geolocation representations.

In this paper, we introduce a novel, training-free method that uses LLM and OpenStreetMap auxiliary map data to derive geolocation representations (LLMGeovec). As illustrated in Fig. 1, our approach first extracts textual descriptions of the coordinates from OpenStreetMap, which provides a sufficient geographic context. LLMs process these descriptions,

---

[*]These authors contributed equally.
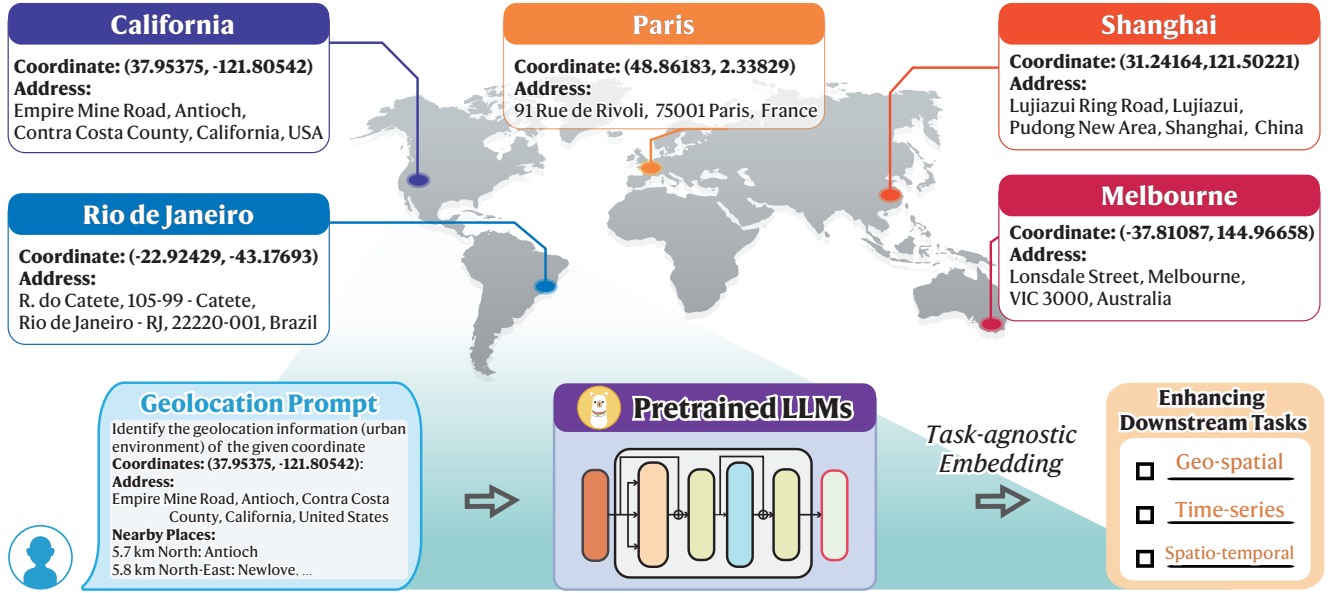
[†]Corresponding author.

Figure 1: Our geolocation representation method consists of two phases: firstly, generating geolocation prompts for coordinates from map data, and then generating representations for text descriptions from pre-trained LLMs. It achieves global coverage and generates representations that can be used in various spatio-temporal tasks.

and the final hidden states of individual tokens are averaged to form the LLMGeovec embedding for each coordinate.

LLMGeovec is not only the first geolocation representation model to achieve global coverage using LLMs, it also presents a simple yet effective paradigm for enhancing spatio-temporal learning with LLMs. In GP, LLMGeovec can be used either as a standalone representation or concatenated with existing representations generated using street view and human mobility data. In LTSF, different temporal patterns of various nodes need to be handled, and the challenge is how to not only identify the unique characteristics of nodes but also model the correlations between them (Nie et al. 2024a, 2022; Zeng et al. 2022). LLMGeovec can be concatenated to the temporal features of individual nodes, allowing models to distinguish between different nodes while modeling their geographical connections. In terms of GSTF, current approaches focus on capturing spatial dependencies by using GNNs to aggregate node features with the guidance of an adjacency matrix (Shao et al. 2022a,d,c; Wu et al. 2019a; Chen et al. 2020; Geng et al. 2019; Nie et al. 2023). Adding LLMGeovec as new node features before performing GNNs can provide more a priori knowledge about the spatial semantics of nodes.

Our experiments demonstrate that LLMGeovec is plug-and-play and improves the performance of various spatio-temporal learning. In GD, LLMGeovec alone achieves SOTA performance for tasks of all scales, and its performance is further enhanced when spliced with other features. In LTSF and GSTF, LLMGeovec enhances many of the latest temporal models and spatio-temporal graph neural networks (STGNNs). Notably, we find that LLMGeovec with a simple MLP can outperform many STGNNs, demonstrating that LLMGeovec already contains rich spatial correlations, and has the potential to replace heavy GNNs.

In summary, we present three major contributions:

- We propose LLMGeovec, a novel, training-free approach that leverages LLMs to generate semantically rich geolocation representations. By utilizing OpenStreetMap data, LLMGeovec functions as a universal and effective geolocation representation model.

- LLMGeovec achieves comprehensive global geographic coverage and offers a simple yet effective paradigm for enhancing spatio-temporal learning using LLMs, resulting in direct performance improvements.

- Extensive experimental analysis demonstrates that LLM-Geovec achieves global coverage and significantly boosts the performance of leading GD, LTSF, and GSTF models.

Related Work

Geolocation Representation Models

Geolocation representation models encode spatial coordinates into latent embeddings enriched with contextual geographic information. These embeddings ensure that similar representations reflect analogous social attributes and climatic characteristics across diverse locations (Wang et al. 2022; Jean et al. 2019; Lee et al. 2021; Wang, Li, and Rajagopal 2020; Zhang et al. 2021, 2023a; Zhou et al. 2023b; Kim and Yoon 2022).

Currently, there are three primary types of geolocation representation models: GNN-based models, image-based models, and natural language processing (NLP)-based models. GNN-based models construct graphs from correlations between locations, such as geographic distance, points of interest (POI), and human mobility patterns. These models generate node representations through message passing on the constructed graphs (Zhang et al. 2021; Kim and Yoon

2022; Zhang et al. 2023a; Zhou et al. 2023b). In contrast, Image-based models utilize street view or satellite imagery and employ contrastive learning to generate representations tied to specific coordinates (Jean et al. 2019; Wang, Li, and Rajagopal 2020; Liu et al. 2023c; Li et al. 2022a; Xi et al. 2022; Ma, Ni, and Chen 2024). Meanwhile, NLP-based models leverage textual representations to represent the textual descriptions associated with the corresponding locations.

However, GNN-based models are highly dependent on human mobility data, which restricts their applicability to urban environments where such records are available. This limitation also challenges the modeling of cities on a global scale. Image-based models face challenges as well; those relying solely on satellite imagery often lack critical human activity information (Xi et al. 2022; Manvi et al. 2023), while street view images can be costly and not universally accessible. NLP-based models show promise due to the abundance of geographically relevant textual data available online, which is typically free and globally accessible. However, existing NLP-based models, such as using Doc2vec to represent textual descriptions from Wikipedia, are inherently limited in their data source and model ability to fully capture the richness of geographic information (Sheehan et al. 2019).

To address these limitations, this paper introduces LLM-Geovec, an NLP-based geolocation representation model that extracts extensive spatio-temporal and human-related knowledge compressed in LLMs to represent locations effectively.

LLMs for GP

Recent advancements in LLMs have seen their application in various GP tasks. Utilizing pre-trained LLMs, researchers have addressed various challenges such as forecasting dementia patterns over time series data, predicting urban functionalities, and estimating socio-climatic variables (Mai et al. 2023; Manvi et al. 2024; Zhang et al. 2023b). Despite these applications, the efficacy of LLMs not specifically fine-tuned for geographic tasks remains suboptimal. Significant efforts have been directed towards customizing LLMs for geospatial analytics. For example, some researchers have fine-tuned LLMs in geoscience text corpora to enhance their performance in geographic question answering, summarization, and text classification tasks (Deng et al. 2024). Furthermore, more recent studies have constructed training sets derived from OpenStreetMap data and associated GP tasks, leading to improved performance by fine-tuning LLMs on training sets (Manvi et al. 2023).

However, fine-tuning LLMs is resource-intensive, often requiring substantial computational and data resources (Hu et al. 2021; Kaddour et al. 2023). Previous approaches mainly generate texts with LLMs to approximate GP, focusing more on relevance rather than precision (Lopez-Lira and Tang 2023). In contrast, our proposed LLMGeovec framework leverages pre-trained LLMs for direct geolocation representation. This approach facilitates the use of geographic knowledge of LLMs within various prediction models. Our experimental results confirm the robustness and utility of LLMGeovec in practical GP.

LLMs for LTSF and GSTF

LTSF and GSTF involve the analysis of spatio-temporal data, which encapsulate both the temporal dynamics of individual nodes and the spatial dependencies among them. Recent advancements have explored the integration of LLMs to leverage their sequence modeling capabilities and the spatio-temporal knowledge they encode. This is typically achieved by tokenizing time series and graph data, fine-tuning LLMs on these tokens, and subsequently employing customized prompts to improve forecast accuracy (Jiang et al. 2024; Li et al. 2024; Zhou et al. 2023a; Chang et al. 2024; Sun et al. 2023; Cao et al. 2023; Jin et al. 2023).

However, existing approaches predominantly utilize LLMs as direct predictors, necessitating substantial computational resources for fine-tuning and often falling short in embedding spatio-temporal knowledge into existing advanced forecasting models. Our work addresses this limitation by extracting geolocation representations from LLMs, enriching the LTSF and GSTF models with enhanced spatial correlation learning, leading to direct performance improvements.

Preliminaries

In this section, we introduce the definitions of geolocation representation learning, GP, LTSF and GSTF.

Geolocation Representation Learning. Given a set of nodes $P \in \mathbb{R}^{2 \times N}$, where N represents the number of nodes, and $P_i = (P_i^{\text{lon}}, P_i^{\text{lat}})$ denotes the longitude and latitude of the i-th node, the goal of geolocation representation learning is to construct an effective encoder f that transforms P into geographically informative representations $Z = f(P) \in \mathbb{R}^{M \times N}$ with M denoting the dimension of the representation.

GP. Given a set of nodes $P \in \mathbb{R}^{2 \times N}$, each node is associated with geographic attributes such as climatic indicators (e.g., average annual temperature, humidity) and social indicators (e.g., regional average educational attainment, average annual income, poverty rate, crime rate). For a given set of geographic attributes $A \in \mathbb{R}^{1 \times N}$, GP in the context of geolocation representation learning involves training a linear regressor with $Z_{[:K]} = f(P_{[:K]})$ to fit $A_{[:K]}$ using K training samples. The performance of the regressor on the test sets $Z_{[K:]} = f(P_{[K:]})$ and $A_{[K:]}$ is used to measure the quality of the encoder f and the representations Z.

LTSF. We consider a multivariate time series (MTS) $X \in \mathbb{R}^{H \times N}$, where N represents the number of nodes (variates) and H is the number of historical time slots. The objective is to predict future values $Y \in \mathbb{R}^{F \times N}$, with F as the number of future time slots. Each value of node i in time slot t is denoted by X_t^i, and their coordinates by $P \in \mathbb{R}^{2 \times N}$.

GSTF. Different from LTSF, GSTF constructs a weighted adjacency matrix $A \in \mathbb{R}^{N \times N}$, where $A_{ij} = 1/\text{dist}(P_i, P_j)$ and $\text{dist}(P_i, P_j)$ represents the spatial distance between node P_i and P_j. A graph G is then formed based on A. Unlike LTSF, GSTF leverages GNNs to aggregate the features of nodes X_t at t-th time slot, enhancing prediction by incorporating spatial relationships.

LLMGeovec: A Generic Enhancer for Spatio-Temporal Learning

As depicted in Fig. 1, the proposed LLMGeovec encapsulates two primary phases: prompt generation and text embedding via LLMs. Initially, we generate geographic descriptions based on specified coordinates with map data. These descriptions are then transformed into embeddings by LLMs. The obtained embeddings can be used for GP, LTSF, and GSTF.

Prompt Generation

Given a coordinate, we generate universal prompts, intentionally devoid of task-specific data, to enable the effective extraction of geographic knowledge of LLMs. As outlined in Fig. 1, the prompt structure incorporates:

- **Instruction**: guides LLMs in identifying essential geographic information linked to specific coordinates.
- **Address**: uses reverse-geocoding to detail the hierarchy of location, from local neighborhoods to national identifiers.
- **Nearby Places**: enumerates the ten nearest points of interest within a 100-kilometer radius, including their names, distances, directions and bearings.

Data sources include OpenStreetMap (Neis and Zipf 2012), with addresses derived through Nominatim's reverse geocoding (Serere, Resch, and Havas 2023) and nearby places via the Overpass API (Olbricht et al. 2011). This approach aligns with and extends previous studies (Manvi et al. 2023, 2024) by focusing on the extraction of generalized geographic information without specifying downstream tasks.

Text Embedding Using LLMs

With the geolocation prompts generated, we proceed to embed these textual descriptions using LLMs. Recent studies have explored enhancing text embeddings generated by LLMs, typically by modifying attention mechanisms or repeating prompts to circumvent the limitations of decoder-only models (BehnamGhader et al. 2024; Muennighoff 2022; Ma et al. 2024; Wang et al. 2023; Springer et al. 2024). Our structured prompts, particularly with crucial geographic context presented at the end of prompts, allow LLMs to generate sufficiently high-quality geolocation representations without repetition of prompts or modification of models. To be specific, we use the average word embeddings from the last layer of a pre-trained LLM as the text representation, ensuring our LLMGeovec method remains adaptable to the latest LLMs without training. In addition, by avoiding fine-tuning, our method preserves the intrinsic geographic knowledge within LLMs (Zhai et al. 2023; Lin et al. 2023).

Incorporating LLMGeovec into GP

Consistent with many previous studies, high-quality geolocation representations can be used for GP with the help of partial region labeling (Wang et al. 2022; Jean et al. 2019; Lee et al. 2021; Wang, Li, and Rajagopal 2020; Zhang et al. 2021, 2023a; Zhou et al. 2023b; Kim and Yoon 2022). This is a direct application of LLMGeovec. Specifically, we divide the locations into a training set and a test set, and use linear regression in the training set to map geolocation representation to location attributes. This is followed by testing in the test set. It is worth noting that since LLMGeovec achieves global coverage, it can be used in GPs of various scales (global, country, city) and can also be combined with other geolocation representations through feature concatenation.

Incorporating LLMGeovec into LTSF

In this section, we describe the integration of LLMGeovec with LTSF models. We start by outlining a general LTSF model, which typically consists of a token embedding layer E, an encoder C, and a predictor D (Chen et al. 2023; Li et al. 2023a; Liu et al. 2023b; Yi et al. 2024; Zhang, Guo, and Wang 2023). The embedding layer E projects the the t-th historical record $X_t \in \mathbb{R}^{1 \times N}$ into hidden temporal embeddings $S_t = E(X_t) \in \mathbb{R}^{d_t \times N}$, where d_t is the embedding dimension. Note that there can be a normalization operation in this embedder such as RevIN (Kim et al. 2021) to address the nonstationarity of time series. The encoder C then models the node-to-node and slot-to-slot relationships across H historical time slots, and the predictor D generates predictions $\hat{Y} \in \mathbb{R}^{F \times N}$ for the future F time slots. This process is formulated as follows:

$$S = \{S_0, \cdots, S_t, \cdots, S_H\}, \quad S_t = E(X_t), \qquad (1)$$

$$\text{Loss}_{\text{LTSF}} = \min ||D(C(S)) - Y||_F^2, \qquad (2)$$

where the model parameters are updated automatically through gradient descent. In practice, the encoder C can be instantiated by Transformer blocks, convolution, and MLPs, to model either channel dependencies or token correlations. Many state-of-the-art LTSF models follow this architectural template, such as TSMixer (Chen et al. 2023), RMLP (Li et al. 2023a), iTransformer (Liu et al. 2023b), and FreTS (Yi et al. 2024). For other Transformer-based models that employ token-wise embedding, such as models in (Wu et al. 2021; Zhou et al. 2021; Zhang, Wang, and Zhang 2024), we can adapt them with a simple inverting strategy (Liu et al. 2023b).

For an LTSF task, we collect the latitude and longitude of each node that generates the time series to construct the node set P. We then select an LLM (e.g., LLaMa3) and use our proposed LLMGeovec to generate the geolocation representation $Z \in \mathbb{R}^{M \times N}$. A two-layer MLP acts as an adapter for LLMGeovec, projecting Z into a low-dimensional space $Z' = \text{Adapter}(Z) \in \mathbb{R}^{d_s \times N}$ to align with the LTSF task. This process is described by the following equations:

$$Z' = \text{Adapter}(Z), \qquad (3)$$

$$S' = \{S'_0, \cdots, S'_t, \cdots, S'_H\}, \quad S'_t = \text{Concat}(E(X_t), Z'), \qquad (4)$$

$$\text{Loss}'_{\text{LTSF}} = \min ||D(C(S')) - Y||_F^2, \qquad (5)$$

where $S'_i \in \mathbb{R}^{(d_t + d_s) \times N}$ is formed by concatenating the series embeddings $E(X_i)$ with the geolocation representations Z' along the feature dimension. The parameters of both the adapter and the original components are updated automatically via gradient descent.

Incorporating LLMGeovec into GSTF

Previous spatio-temporal prediction models often employ GNNs to capture spatial relationships between nodes, aggregating them into node features, which are then input into the temporal modeling component sequentially or alternatively (Shao et al. 2022a,d,c; Wu et al. 2019a; Tang, He, and Zhao 2022; Tang et al. 2022; Zhang et al. 2025). We refer to the spatio-temporal model as STGNN. Given the graph G, the structural template is framed as follows:

$$S = \{S_0, \cdots, S_t, \cdots, S_H\}, \quad S_t = E(X_t),$$
$$\hat{S} = \text{STGNN}(S, G), \qquad (6)$$
$$\text{Loss}_{\text{GSTF}} = \min ||D(\hat{S}) - Y||_F^2.$$

where the embedder E projects the node signal to hidden states, and all node states are collected into the STGNN processor to generate the graph representation $\hat{S}$.

On top of them, we first concatenate LLMGeovec into node features (e.g., temporal readings), which are then processed by STGNN. This process is described by:

$$Z' = \text{Adapter}(Z), \qquad (7)$$
$$S' = \{S_0', \cdots, S_t', \cdots, S_H'\}, \quad S_t' = \text{Concat}(E(X_t), Z'), \qquad (8)$$
$$\hat{S}' = \text{STGNN}(S', G), \qquad (9)$$
$$\text{Loss}_{\text{GSTF}}' = \min ||D(\hat{S}') - Y||_F^2, \qquad (10)$$

where the parameters of both the adapter, the STGNN and D are updated automatically via gradient descent.

Note that this scheme is applicable for STGNNs with different types of processing methods. Specifically, different models have different instantiations of Eq. 9, e.g., spatio-temporal message passing mechanisms. Both the mainstream time-then-space and time-and-space STGNN family discussed in (Cini et al. 2023) can be seamlessly adopted simply by concatenating LLMGeovec into the input of each node.

Table 1: A multi-scale and multi-topic GP benchmark

Tasks	Source	Scale	Attribute	Training/Testing
Annual Air Temperature	Chelsa	Global	Climate	80k/20k
Annual Precipitation	Chelsa	Global	Climate	80k/20k
Monthly Climate Moisture	Chelsa	Global	Climate	40k/20k
Population Density	WorldPop	Global	Society	80k/20k
Nighttime Light Intensity	EOG	Global	Society	80k/20k
Human Modification Terrestrial	SEDAC	Global	Society	80k/20k
Global Gridded Relative Deprivation	SEDAC	Global	Society	80k/20k
Ratio of Built-up Area to Non-built Up Area	SEDAC	Global	Society	80k/20k
Child Dependency Ratio	SEDAC	Global	Society	80k/20k
Subnational Human Development	SEDAC	Global	Society	80k/20k
Infant Mortality Rates	SEDAC	Global	Society	80k/20k
Asset Index	DHS	Global	Society	20k/5k
Sanitation Index	DHS	Global	Society	20k/5k
Women BMI	DHS	Global	Society	40k/10k
Poverty Rate	DHS	Country	Society	5k/1k
Population Density	FaceBook	Country	Society	5k/1k
Women BMI	DHS	Country	Society	5k/1k
Population Density	NYC	City	Society	1k/424
Education Level	NYC	City	Society	1k/424
Income Level	NYC	City	Society	1k/424
Crime Rate	NYC	City	Society	1k/424

Table 2: Baseline methods of GP

Methods	Scales	Data Resources
Node2vec, GCN, GAT	City	Road Graph
ZE-Mob, MGFN, MV-PN	City	Road Graph, Mobility
HDGE, HUGAT, MVURE, HKGL	City	Check-in, PoI
Image Supervised Learning	Country	Street View
Object Counts	Country	Street View
Mapillarygcn	Country	Street View
Bert-whitening	All	Map
GTE-large	All	Map
GTE-qwen2 7B	All	Map
LLMGeovec	All	Map

Numeric Experiments

We study the effectiveness of LLMGeovec through extensive experiments. We first demonstrate that in geolocation representation models, LLMGeovec performs SoTA in GP tasks at all three scales: city, national, and global, and even outperforms end-to-end supervised training models. We examine two LLMs, LLaMa3 8B and Mistral 8x7B, both of which are able to produce high-quality geolocation representations with LLMGeovec. Moreover, LLMGeovec is seamlessly embedded into various LTSF and GSTF models and directly improves the model performance under various tasks. Notably, in the GSTF tasks, the use of a simple MLP and LLMGeovec outperforms many GNN-based approaches, and LLMGeovec shows great potential as an alternative to time-consuming GNNs. For a detailed description of the models and the datasets (GP, LTSF, GSTF), please refer to Appendix.

LLMGeovec for GP

To comprehensively validate the quality of LLMGeovec and its effectiveness in GP tasks, we constructed a multi-scale, multi-topic benchmark encompassing a range of scenarios from city-level poverty rates to global population density. Unlike many existing powerful baselines, our approach can generate high-quality geolocation representations for any location without expensive data or extensive training.

A Multi-scale and Multi-topic GP Benchmark. As illustrated in Table 1, at the global scale, we collect 14 GP tasks. These include three climate indicators such as Annual Air Temperature and 11 social indicators like Population Density and Human Modification (detailed descriptions please refer to Appendix). We use 100,000 locations with global coverage, generated by Manvi et al. (2024) (Africa: 19,855; Asia: 55,893; Europe: 6,825; North America: 8,440; South America: 5,189; Oceania: 2,049). In line with Manvi et al. (2023, 2024), each GP task is associated with a corresponding Geo-TIFF file. For each coordinate, the average value of 12 pixels surrounding the coordinate is taken as the value of the coordinate. Following the protocol of Kim and Yoon (2022), we perform five cross-validation using ridge linear regression implemented in Sklearn (Feurer et al. 2020; McDonald 2009), and the average of mean absolute error (MAE), root mean square error (RMSE) and R^2 are reported. On the country and city scales, we use existing benchmarks, including social indicators in India (Lee et al. 2021) and NYC (Zhou et al.

Table 3: Results of the LTSF benchmarks. We report the average forecast error of different models under different prediction lengths. Full results are presented in Appendix. Part of the baseline results are from (Nie et al. 2024a). IMP shows the average percentage of MSE improvement of LLMGeovec.

Models Metric	iTransformer MSE	MAE	w/ LLMGeovec MSE	MAE	TSMixer MSE	MAE	w/ LLMGeovec MSE	MAE	RMLP MSE	MAE	w/ LLMGeovec MSE	MAE	Informer MSE	MAE	w/ LLMGeovec MSE	MAE	IMP %
Global Wind	4.582	1.51	3.979	1.380	4.261	1.424	4.132	1.407	4.905	1.498	4.180	1.414	4.905	1.576	4.844	1.566	13.30%
Global Temp	13.079	2.653	11.945	2.601	12.035	2.480	11.441	2.398	13.447	2.558	12.525	2.480	18.370	3.209	18.639	3.234	5.19%
Solar Energy	0.233	0.262	0.206	0.265	0.255	0.294	0.219	0.289	0.261	0.313	0.235	0.286	0.264	0.308	0.263	0.313	11.59%
Demand-SH	0.331	0.298	0.322	0.297	0.355	0.332	0.336	0.305	0.345	0.326	0.318	0.286	0.896	0.618	0.779	0.666	2.47%
Air Quality	1.922	0.631	1.856	0.619	2.068	0.665	1.989	0.650	1.857	0.627	1.820	0.613	3.584	0.864	2.858	0.771	3.46%
Traffic-SD	0.136	0.225	0.106	0.201	0.116	0.212	0.105	0.197	0.205	0.296	0.168	0.264	0.199	0.298	0.152	0.254	22.01%

2023b). We also employ ridge linear regression in Sklearn and report MAE, RMSE, and R^2 on test sets.

Baselines. We compare the LLMGeovec generated by LLaMa3 8B (Touvron et al. 2023) and Mistral 8x7B (Jiang et al. 2023). As shown in Tab. 2, we also compare the text embedding generated by Bert-whitening, GTE-large, and GTE-qwen2 7B (Su et al. 2021; Li et al. 2023b). In the city and country scales, we additionally compare Image-based and GNN-based geolocation representation models.

Table 4: Performance of GP (global)

Tasks	LLMGeovec (LLaMa 3 8B) MAE	RMSE	R²	LLMGeovec (Mistral 8 x 7B) MAE	RMSE	R²	Bert-whitening (Bert base) MAE	RMSE	R²	GTE-large MAE	RMSE	R²	GTE-qwen2 7B MAE	RMSE	R²
Annual Air Temperature	9.90	14.32	0.95	11.05	16.03	0.94	24.03	32.73	0.76	22.23	30.15	0.80	14.05	19.99	0.91
Annual Precipitation	2016.60	3021.76	0.86	2176.82	3245.12	0.83	3717.73	5293.82	0.56	3519.57	5012.55	0.61	2604.86	3931.23	0.76
Monthly Climate Moisture	1302.14	2021.21	0.55	1345.82	2097.32	0.52	1644.56	2715.44	0.19	1619.65	2610.16	0.25	1394.10	2288.39	0.43
Population Density	695.10	1020.11	0.88	759.81	1115.38	0.82	1342.98	2185.67	0.30	1266.82	1986.15	0.42	896.52	1417.02	0.70
Nighttime Light Intensity	3.55	4.58	0.97	3.79	4.89	0.96	8.55	11.37	0.81	7.63	9.99	0.85	4.77	6.15	0.94
Human Modification Terrestrial	0.07	0.09	0.78	0.07	0.09	0.75	0.12	0.15	0.39	0.11	0.14	0.47	0.08	0.11	0.68
Global Gridded Relative Deprivation	6.56	8.98	0.85	6.70	9.15	0.84	10.43	13.60	0.65	9.83	12.95	0.68	8.13	10.86	0.78
Ratio of Built-up Area to Non-built Up Area	8.44	11.07	0.78	8.72	11.41	0.77	13.04	16.40	0.52	12.51	15.81	0.56	10.48	13.41	0.68
Child Dependency Ratio	5.84	8.29	0.86	5.88	8.34	0.88	9.50	13.12	0.64	9.11	12.47	0.68	7.17	10.09	0.79
Subnational Human Development	5.79	8.20	0.89	5.82	8.22	0.89	9.81	13.30	0.70	9.15	12.36	0.75	7.10	9.95	0.83
Infant Mortality Rates	3.98	6.06	0.93	4.02	6.14	0.93	7.42	10.76	0.77	7.24	10.20	0.80	4.97	7.50	0.89
Asset Index	0.02	0.03	0.93	0.02	0.03	0.92	0.06	0.08	0.53	0.05	0.07	0.62	0.04	0.06	0.78
Sanitation Index	0.09	0.12	0.95	0.10	0.13	0.93	0.23	0.30	0.67	0.20	0.26	0.75	0.15	0.20	0.85
Women BMI	0.76	1.01	0.95	0.83	1.12	0.94	1.82	2.38	0.77	1.55	2.04	0.83	1.16	1.57	0.90

Table 5: Performance of GP (country)

Methods	R² Poverty Rate	Population Density	Women BMI
Image Supervised Learning	0.51	0.85	0.52
Object Counts	0.52	0.81	0.53
Mapillarygcn	0.53	0.89	0.56
Bert-whitening (Bert base)	0.51	0.77	0.45
LLMGeovec (Mistral 8 x 7B)	**0.66**	**0.96**	**0.65**
LLMGeovec (LLaMa 3 8B)	**0.66**	**0.96**	**0.65**

Performances of GP. As illustrated in Table 4 5, and 6, the LLMGeovec family significantly outperforms Bert-whitening in generating geolocation representations from the same textual descriptions. This highlights the superiority of LLMs in leveraging a vast Internet corpus for pre-training. Additionally, the results show that LLaMa3 8B generally performs better than Mistral 8x7B. We attribute this to the multilingual corpus used in the pre-training of LLaMa3 8B, which enhances its understanding of geographic knowledge across different regions. At a global scale, LLMGeovec (LLaMa3 8B) demonstrates robust performance, with R^2 values exceeding 0.75 for all tasks except for Monthly Climate Moisture, which has an R2 of 0.55. Most tasks achieve R^2 values above 0.90. At the national scale, LLMGeovec (LLaMa3 8B) im-

proves the R^2 scores across various tasks by 0.07 to 0.10, compared to the SOTA model (MapillaryGCN), which employs end-to-end supervised training using Street View data and GNNs. At the city scale, LLMGeovec outperforms or is comparable to baselines that utilize extensive human activity data and sophisticated graph learning techniques for all tasks except Crime Rate. Notably, concatenating the geolocation representations generated by LLMGeovec with those from existing methods (e.g., HKGL) substantially boosts the performance of downstream tasks.

Table 6: Performance of GP (city)

Methods	Poverty Rate MAE	RMSE	R²	Education Level MAE	RMSE	R²	Income Level MAE	RMSE	R²	Crime Rate MAE	RMSE	R²
Node2vec	0.45	0.66	0.078	0.09	0.12	0.68	0.23	0.30	0.51	0.50	0.64	0.43
GCN	0.45	0.66	0.07	0.09	0.12	0.68	0.21	0.28	0.56	0.44	0.58	0.53
GAT	0.44	0.64	0.12	0.09	0.12	0.66	0.23	0.29	0.52	0.48	0.61	0.47
ZE-Mob	0.47	0.68	0.018	0.13	0.16	0.42	0.29	0.37	0.24	0.60	0.75	0.20
MGFN	0.45	0.66	0.07	0.11	0.14	0.56	0.25	0.33	0.41	0.50	0.64	0.42
MV-PN	0.46	0.66	0.11	0.14	0.18	0.26	0.32	0.40	0.12	0.63	0.77	0.16
HDGE	0.45	0.66	0.07	0.10	0.14	0.58	0.24	0.32	0.45	0.53	0.67	0.37
HUGAT	0.47	0.67	0.04	0.13	0.16	0.38	0.30	0.38	0.22	0.53	0.67	0.37
MVURE	0.45	0.65	0.12	0.10	0.12	0.66	0.23	0.29	0.52	0.47	0.61	0.48
HKGL	0.42	0.62	0.20	0.08	0.11	0.73	0.20	0.27	0.60	0.40	0.51	0.64
Bert-whitening (Bert base)	0.47	0.68	0.02	0.11	0.15	0.48	0.27	0.35	0.33	0.59	0.75	0.20
LLMGeovec (LLaMa 3 8B)	0.43	0.64	0.13	0.08	0.11	0.74	0.20	0.26	0.62	0.51	0.66	0.39
HKGL w/ LLMGeovec (LLaMa 3 8B)	**0.39**	**0.58**	**0.28**	**0.08**	**0.10**	**0.76**	**0.19**	**0.25**	**0.652**	**0.39**	**0.50**	**0.64**

Performance of Prompt Variants. To examine the effect of the various parts of the prompts on the quality of LLM-Geovec, we try several variants of the prompts and test them on GP tasks in NYC. As shown in Tab. 7, Address, which is the latitude and longitude counterpart of the location hierarchy, from local neighborhoods to national identifiers, is the most important part of motivating geographic knowledge in LLM. As for the K nearest places, we discover that when K is too small, it can not provide effective geographic information; when K is too large, the neighboring nodes will be too convergent, which leads to poor results in downstream tasks.

Table 7: Performance of Prompt Variants

Prompt	Poverty Rate MAE	RMSE	R²	Education Level MAE	RMSE	R²	Income Level MAE	RMSE	R²	Crime Rate MAE	RMSE	R²
Instruction + Address + Top 10 NearbyPlaces	0.13	0.74	0.62	0.39	0.17	0.75	0.61	0.36	0.12	0.65	0.53	0.34
Instruction + Address + Top 5 NearbyPlaces	0.17	0.75	0.61	0.36	0.12	0.65	0.53	0.34	0.06	0.61	0.48	0.32
Instruction + Address + Top 1 NearbyPlaces	0.12	0.65	0.53	0.34	0.06	0.61	0.48	0.32	-0.01	0.42	0.29	0.09
Instruction + Address	0.06	0.61	0.48	0.32	-0.01	0.42	0.29	0.09	0.13	0.74	0.62	0.39
Instruction	-0.01	0.42	0.29	0.09	0.13	0.74	0.62	0.39	0.17	0.75	0.61	0.36

Performance of Regions with Sparse POIs. For the NearbyPlaces in the prompt, we use the node field in the OSM

data, and when there are POIs (e.g., bar, restaurant) near the coordinates, NearbyPlaces will be a variety of such buildings or stores. If there are no POIs near the coordinates, NearbyPlaces will be the names of nearby streets. We show that LLMGeovec performs consistently in regions with extensive POI coverage (e.g., North America) and those with little coverage (e.g., Africa). Specifically, we utilize the continent boundaries to extract nodes belonging to North America and Africa separately and perform the five-fold cross-validation within the continents, the results of which are shown in Tab 8. where LLMGeovec's results are similar in both regions.

Table 8: Performance of Regions with Sparse POIs.

Region	Population Density	Night Light Density	Annual Air Temperature
Africa	0.90	0.94	0.88
North America	0.82	0.97	0.92

LLMGeovec for LTSF

In the previous section, we discussed how to seamlessly embed LLMGeovec into existing LTSF models. Next, we will conduct detailed experiments to verify the effectiveness of LLMGeovec using popular LTSF benchmarks and various models. Due to limited computational resources, we choose LLMGeovec (LLaMa3 8B) with the best performance in the GP to be added to the various models.

Datasets and models. Following the settings of Zhang et al. (2024a); Wu et al. (2022); Shao et al. (2022a), we select five LTSF datasets from a wide range of domains, including Solar Energy, Global Wind, Global Temperature, Traffic flow, Delivery demand, and air quality. Several representative LTSF models are selected, including both Transformer-based and MLP-based methods. They are iTransformer (Liu et al. 2023b), TSMixer (Chen et al. 2023), RMLP (Li et al. 2023a), and Informer (Zhou et al. 2021).

Hyperparameters Settings. We adapt the suggested hyperparameters in Time-Series-Library benchmark (Wang et al. 2024) for all model.

Performances of LTSF. As shown in Table 3, LLMGeovec can consistently improve the original performances of different models in almost all scenarios. This effect is noticeable in datasets related to both natural processes and human activities, which demonstrates the generality of LLM-based geolocation representation.

Performances Comparisions using different geolocation embeddings. In this section, we compare the effects of two geolocation representations on LTSF models, and for reference, we also add learnable embeddings (Shao et al. 2022b) (e.g., STID) with the same feature dimensions as LLMGeovec. For a fair comparison, all three methods use the same Adapter and model parameters. As shown in Tab. 9, LLMGeovec, which contains richer spatial semantics, achieves the greatest improvement.

LLMGeovec for GSTF

Finally, we evaluate the effectiveness of LLMGeovec in GSTF tasks and models. As with LTSF, we choose LLaMa3 8B to generate LLMGeovec.

Table 9: Different geolocation embeddings on LTSF.

Global Wind	LLMGeovec (LLaMa 3 8B)		Bert-whitening (Bert Base)		STID	
	MSE	MAE	MSE	MAE	MSE	MAE
iTransformer	**3.560**	1.300	3.563	1.301	3.575	**1.299**
TSMixer	**3.524**	**1.292**	3.758	1.316	3.68	1.303
RMLP	**3.221**	**1.237**	3.563	1.309	3.591	1.310

Datasets and models. We select the large-scale LargeST traffic flow benchmark (Liu et al. 2023a) and the LaDe demand dataset (Wu et al. 2023) for evaluations. Several competitive baselines that are widely adopted in related work are considered, including DCRNN (Li et al. 2017), STGCN (Yu, Yin, and Zhu 2017), ASTGCN (Guo et al. 2019), AGCRN (Bai et al. 2020), GWNET (Wu et al. 2019b), MTGNN (Wu et al. 2020b), and STID (Shao et al. 2022b).

Hyperparameters Settings. We adopt the suggested hyperparameters in LargeST (Liu et al. 2023a) for all models.

Table 10: Experimental results of GSTF in LargeST dataset.

Models	SD		GLA		GBA		IMP
	MAE	RMSE	MAE	RMSE	MAE	RMSE	(%)
HA	60.78	87.39	59.58	86.19	56.43	79.81	–
DCRNN	25.23	39.17	22.73	35.65	22.35	35.26	8.32%
+LLMGeovec	**18.70**	**31.36**	**21.43**	**34.76**	**21.69**	**34.37**	
STGCN	20.10	34.60	22.48	38.55	23.14	37.90	3.51%
+LLMGeovec	**19.83**	**33.21**	**22.03**	**37.45**	**22.43**	**36.51**	
ASTGCN	25.13	39.88	28.44	44.13	26.15	40.25	7.98%
+LLMGeovec	**23.89**	**38.08**	**23.74**	**38.27**	**23.24**	**37.78**	
AGCRN	18.45	34.40	20.61	36.23	20.55	_33.91_	0.60%
+LLMGeovec	**18.21**	**33.82**	_19.88_	**35.96**	_19.77_	34.12	
GWNET	19.38	31.88	21.23	33.68	20.84	34.58	3.92%
+LLMGeovec	_18.03_	**30.06**	**20.29**	_32.62_	**20.66**	**33.58**	
MTGNN	23.69	36.83	23.47	37.68	23.73	36.01	8.09%
+LLMGeovec	**19.03**	**31.17**	**21.76**	**34.58**	**22.55**	**35.77**	
MLP	27.84	43.92	29.12	45.76	29.15	45.64	26.53%
+LLMGeovec	**19.00**	_30.03_	**21.07**	**34.56**	**21.42**	**34.92**	

Performances in GSTF. Tab. 10 reports the evaluation results on the LargeST benchmark. Mainstream GNN-based models can benefit from the incorporation of LLMGeovec. This clearly shows that LLMGeovec is able to complement the spatial relationships captured by GNNs with the rich geographic knowledge of LLMs. Surprisingly, the vanilla MLP model equipped with LLMGeovec can achieve comparable performance to the GNN counterparts. This suggests that LLMGeovec can even be used as an alternative to GNNs to provide geographic correlation for temporal models.

Overhead of LLMGeovec. For the GTSF and LTSF tasks, we concatenate the d-dimensional LLMGeovec to the original feature vector while reducing the dimensionality of the original feature vector by d. This way, the overall dimensionality remains unchanged. The only additional parameters introduced are the LLM Adapter (A two-layer MLP that uses Leaky_Relu in the middle), which is used to reduce the input dimensionality of the LLMGeovec. As shown in 11, our model introduces only an acceptable number of additional parameters and maintains computational efficiency comparable

to the original model.

Table 11: Ovearhead of LLMGeovec.

Model	Parameters	Running Speed	GPU Memory
STGCN w/o LLMGeovec	3482K	3.62 s/it	3.5 GB
STGCN w/ LLMGeovec	3351K	3.58 s/it	3.5 GB

Performances in zero-shot scenarios. In addition to enhancing various models in full training scenarios for GSTF, LLMGeovec also has the potential for enhancing zero-shot transfer. We compare the performance of the learnable node embedding (STID) introduced by Shao et al. (2022a) and LLMGeovec in zero-shot scenarios. As a reference, we also test the transferability of the baseline GWNET and MLP. The LaDe data is adopted for this experiment. It is evident from Fig. 2 that when models are transferred to a new region in a zero-shot scenario, the learnable embedding harms the performance of MLP significantly because the embedding has adapted to the source data with specific patterns. In contrast, universal LLMGeovec can be generalized to other regions without any adjustment. This indicates that LLMGeovec features intrinsic geolocation knowledge that is generalizable for different regions. In addition, more advanced techniques such as test-time adaptation in graphs can be applied to achieve better few-shot performance (Sun et al. 2024).
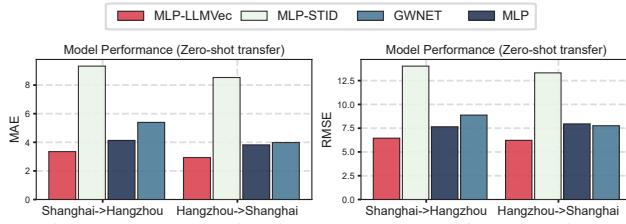

Figure 2: Comparsion in zero-shot transfer scenarios.

Conclusion and Future Work

The acquisition of universal geolocation representations to improve downstream tasks has been a long-standing pursuit. This paper presents our first attempt to utilize recent advanced LLMs to extract such representations. By the merit of the geospatial knowledge within LLMs, the extracted embedding from the pre-activated layer achieves global coverage and serves as a generic enhancer for spatio-temporal learning. We demonstrate the effectiveness of embedding in various tasks, including GP, LTSF, and GSTF. Empirical results indicate that LLMGeovec can improve the performances of various models simply by incorporating it into the input (i.e., feature concatenation). In future work, we are interested to see if larger LLMs (e.g., LLaMa3 70B) can further improve the quality of LLMGeovec. We can adopt LLMGeovec in more challenging spatio-temporal learning tasks, such as spatio-temporal imputation (Nie et al. 2024b; Yuan et al. 2022) and traffic flow generation (Wu et al. 2020a). It is also interesting to explore the possibility of integrating LLMGeovec into pre-trained foundational models for unified spatio-temporal learning (Jin et al. 2023; Yuan et al. 2024; Zhang et al. 2024b).

References

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33: 17804–17815.

BehnamGhader, P.; Adlakha, V.; Mosbach, M.; Bahdanau, D.; Chapados, N.; and Reddy, S. 2024. Llm2vec: Large language models are secretly powerful text encoders. *arXiv preprint arXiv:2404.05961*.

Cao, D.; Jia, F.; Arik, S. O.; Pfister, T.; Zheng, Y.; Ye, W.; and Liu, Y. 2023. Tempo: Prompt-based generative pre-trained transformer for time series forecasting. *arXiv preprint arXiv:2310.04948*.

Chang, C.; Wang, W.-Y.; Peng, W.-C.; and Chen, T.-F. 2024. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. *arXiv preprint arXiv:2308.08469*.

Chang, T.; Hu, Y.; Taylor, D.; and Quigley, B. M. 2022. The role of alcohol outlet visits derived from mobile phone location data in enhancing domestic violence prediction at the neighborhood level. *Health & Place*, 73: 102736.

Chen, S.-A.; Li, C.-L.; Yoder, N.; Arik, S. O.; and Pfister, T. 2023. Tsmixer: An all-mlp architecture for time series forecasting. *arXiv preprint arXiv:2303.06053*.

Chen, W.; Chen, L.; Xie, Y.; Cao, W.; Gao, Y.; and Feng, X. 2020. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 3529–3536.

Chi, G.; Fang, H.; Chatterjee, S.; and Blumenstock, J. E. 2022. Microestimates of wealth for all low-and middle-income countries. *Proceedings of the National Academy of Sciences*, 119(3): e2113658119.

Cini, A.; Marisca, I.; Zambon, D.; and Alippi, C. 2023. Taming Local Effects in Graph-based Spatiotemporal Forecasting. *arXiv preprint arXiv:2302.04071*.

Deng, C.; Zhang, T.; He, Z.; Chen, Q.; Shi, Y.; Xu, Y.; Fu, L.; Zhang, W.; Wang, X.; Zhou, C.; et al. 2024. K2: A foundation language model for geoscience knowledge understanding and utilization. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, 161–170.

Draidi Areed, W.; Price, A.; Arnett, K.; and Mengersen, K. 2022. Spatial statistical machine learning models to assess the relationship between development vulnerabilities and educational factors in children in Queensland, Australia. *BMC Public Health*, 22(1): 2232.

Elmustafa, A.; Rozi, E.; He, Y.; Mai, G.; Ermon, S.; Burke, M.; and Lobell, D. 2022. Understanding economic development in rural Africa using satellite imagery, building footprints and deep models. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, 1–4.

Feurer, M.; Eggensperger, K.; Falkner, S.; Lindauer, M.; and Hutter, F. 2020. Auto-sklearn 2.0: The next generation. *arXiv preprint arXiv:2007.04074*, 24: 8.

Geng, X.; Li, Y.; Wang, L.; Zhang, L.; Yang, Q.; Ye, J.; and Liu, Y. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 3656–3663.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 922–929.

Gurnee, W.; and Tegmark, M. 2023. Language models represent space and time. *arXiv preprint arXiv:2310.02207*.

Head, A.; Manguin, M.; Tran, N.; and Blumenstock, J. E. 2017. Can human development be measured with satellite imagery? *Ictd*, 17: 16–19.

Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; and Chen, W. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Jean, N.; Burke, M.; Xie, M.; Davis, W. M.; Lobell, D. B.; and Ermon, S. 2016. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301): 790–794.

Jean, N.; Wang, S.; Samar, A.; Azzari, G.; Lobell, D.; and Ermon, S. 2019. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3967–3974.

Jiang, A. Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D. S.; Casas, D. d. l.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; et al. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825*.

Jiang, Y.; Pan, Z.; Zhang, X.; Garg, S.; Schneider, A.; Nevmyvaka, Y.; and Song, D. 2024. Empowering time series analysis with large language models: A survey. *arXiv preprint arXiv:2402.03182*.

Jin, M.; Wang, S.; Ma, L.; Chu, Z.; Zhang, J. Y.; Shi, X.; Chen, P.-Y.; Liang, Y.; Li, Y.-F.; Pan, S.; et al. 2023. Time-llm: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728*.

Kaddour, J.; Harris, J.; Mozes, M.; Bradley, H.; Raileanu, R.; and McHardy, R. 2023. Challenges and applications of large language models. *arXiv preprint arXiv:2307.10169*.

Kim, N.; and Yoon, Y. 2022. Effective urban region representation learning using heterogeneous urban graph attention network (HUGAT). *arXiv preprint arXiv:2202.09021*.

Kim, T.; Kim, J.; Tae, Y.; Park, C.; Choi, J.-H.; and Choo, J. 2021. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.

Lee, J.; Grosz, D.; Uzkent, B.; Zeng, S.; Burke, M.; Lobell, D.; and Ermon, S. 2021. Predicting livelihood indicators from community-generated street-level imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 268–276.

Li, T.; Xin, S.; Xi, Y.; Tarkoma, S.; Hui, P.; and Li, Y. 2022a. Predicting multi-level socioeconomic indicators from structural urban imagery. In *Proceedings of the 31st ACM international conference on information & knowledge management*, 3282–3291.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Li, Z.; Huang, C.; Xia, L.; Xu, Y.; and Pei, J. 2022b. Spatial-temporal hypergraph self-supervised learning for crime prediction. In *2022 IEEE 38th international conference on data engineering (ICDE)*, 2984–2996. IEEE.

Li, Z.; Qi, S.; Li, Y.; and Xu, Z. 2023a. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv preprint arXiv:2305.10721*.

Li, Z.; Xia, L.; Tang, J.; Xu, Y.; Shi, L.; Xia, L.; Yin, D.; and Huang, C. 2024. Urbangpt: Spatio-temporal large language models. *arXiv preprint arXiv:2403.00813*.

Li, Z.; Zhang, X.; Zhang, Y.; Long, D.; Xie, P.; and Zhang, M. 2023b. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*.

Lin, Y.; Tan, L.; Lin, H.; Zheng, Z.; Pi, R.; Zhang, J.; Diao, S.; Wang, H.; Zhao, H.; Yao, Y.; et al. 2023. Speciality vs generality: An empirical study on catastrophic forgetting in fine-tuning foundation models. *arXiv preprint arXiv:2309.06256*.

Liu, X.; Xia, Y.; Liang, Y.; Hu, J.; Wang, Y.; Bai, L.; Huang, C.; Liu, Z.; Hooi, B.; and Zimmermann, R. 2023a. LargeST: A Benchmark Dataset for Large-Scale Traffic Forecasting. *arXiv preprint arXiv:2306.08259*.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023b. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*.

Liu, Y.; Zhang, X.; Ding, J.; Xi, Y.; and Li, Y. 2023c. Knowledge-infused contrastive learning for urban imagery-based socioeconomic prediction. In *Proceedings of the ACM Web Conference 2023*, 4150–4160.

Lopez-Lira, A.; and Tang, Y. 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619*.

Ma, X.; Ma, M.; Hu, C.; Song, Z.; Zhao, Z.; Feng, T.; and Zhang, W. 2023. Log-can: local-global class-aware network for semantic segmentation of remote sensing images. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Ma, X.; Ni, Z.; and Chen, X. 2024. TinyViM: Frequency Decoupling for Tiny Hybrid Vision Mamba. arXiv:2411.17473.

Ma, X.; Wang, L.; Yang, N.; Wei, F.; and Lin, J. 2024. Fine-tuning llama for multi-stage text retrieval. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2421–2425.

Mai, G.; Huang, W.; Sun, J.; Song, S.; Mishra, D.; Liu, N.; Gao, S.; Liu, T.; Cong, G.; Hu, Y.; et al. 2023. On the opportunities and challenges of foundation models for geospatial artificial intelligence. *arXiv preprint arXiv:2304.06798*.

Manvi, R.; Khanna, S.; Burke, M.; Lobell, D.; and Ermon, S. 2024. Large language models are geographically biased. *arXiv preprint arXiv:2402.02680*.

Manvi, R.; Khanna, S.; Mai, G.; Burke, M.; Lobell, D.; and Ermon, S. 2023. Geollm: Extracting geospatial knowledge from large language models. *arXiv preprint arXiv:2310.06213*.

Marty, R.; and Duhaut, A. 2024. Global poverty estimation using private and public sector big data sources. *Scientific Reports*, 14(1): 3160.

McDonald, G. C. 2009. Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1): 93–100.

Muennighoff, N. 2022. Sgpt: Gpt sentence embeddings for semantic search. *arXiv preprint arXiv:2202.08904*.

Neis, P.; and Zipf, A. 2012. Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 1(2): 146–165.

Nie, T.; Mei, Y.; Qin, G.; Sun, J.; and Ma, W. 2024a. Channel-Aware Low-Rank Adaptation in Time Series Forecasting. *arXiv preprint arXiv:2407.17246*.

Nie, T.; Qin, G.; Ma, W.; Mei, Y.; and Sun, J. 2024b. ImputeFormer: Low Rankness-Induced Transformers for Generalizable Spatiotemporal Imputation. *arXiv preprint arXiv:2312.01728*.

Nie, T.; Qin, G.; Wang, Y.; and Sun, J. 2023. Correlating sparse sensing for large-scale traffic speed estimation: A Laplacian-enhanced low-rank tensor kriging approach. *Transportation research part C: emerging technologies*, 152: 104190.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. *arXiv preprint arXiv:2211.14730*.

Nilsen, K.; Tejedor-Garavito, N.; Leasure, D. R.; Utazi, C. E.; Ruktanonchai, C. W.; Wigley, A. S.; Dooley, C. A.; Matthews, Z.; and Tatem, A. J. 2021. A review of geospatial methods for population estimation and their use in constructing reproductive, maternal, newborn, child and adolescent health service indicators. *BMC health services research*, 21: 1–10.

Olbricht, R.; et al. 2011. Overpass API. *Anwenderkonferenz für Freie und Open Source Software für Geoinformationssysteme*.

Robinson, C.; Hohman, F.; and Dilkina, B. 2017. A deep learning approach for population estimation from satellite imagery. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, 47–54.

Serere, H. N.; Resch, B.; and Havas, C. R. 2023. Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection. *Plos one*, 18(3): e0282942.

Shao, Z.; Zhang, Z.; Wang, F.; Wei, W.; and Xu, Y. 2022a. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4454–4458.

Shao, Z.; Zhang, Z.; Wang, F.; Wei, W.; and Xu, Y. 2022b. Spatial-temporal identity: A simple yet effective baseline for multivariate time series forecasting. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4454–4458.

Shao, Z.; Zhang, Z.; Wang, F.; and Xu, Y. 2022c. Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 1567–1577.

Shao, Z.; Zhang, Z.; Wei, W.; Wang, F.; Xu, Y.; Cao, X.; and Jensen, C. S. 2022d. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. *arXiv preprint arXiv:2206.09112*.

Sheehan, E.; Meng, C.; Tan, M.; Uzkent, B.; Jean, N.; Burke, M.; Lobell, D.; and Ermon, S. 2019. Predicting economic development using geolocated wikipedia articles. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2698–2706.

Springer, J. M.; Kotha, S.; Fried, D.; Neubig, G.; and Raghunathan, A. 2024. Repetition improves language model embeddings. *arXiv preprint arXiv:2402.15449*.

Su, J.; Cao, J.; Liu, W.; and Ou, Y. 2021. Whitening sentence representations for better semantics and faster retrieval. *arXiv preprint arXiv:2103.15316*.

Sun, C.; Li, Y.; Li, H.; and Hong, S. 2023. TEST: Text prototype aligned embedding to activate LLM's ability for time series. *arXiv preprint arXiv:2308.08241*.

Sun, H. 2024. Ultra-High Resolution Segmentation via Boundary-Enhanced Patch-Merging Transformer. arXiv:2412.10181.

Sun, H.; Xu, L.; Jin, S.; Luo, P.; Qian, C.; and Liu, W. 2024. PROGRAM: PROtotype GRAph Model based Pseudo-Label Learning for Test-Time Adaptation. In *The Twelfth International Conference on Learning Representations*.

Tang, Y.; He, J.; and Zhao, Z. 2022. Activity-aware human mobility prediction with hierarchical graph attention recurrent network. *arXiv preprint arXiv:2210.07765*.

Tang, Y.; Qu, A.; Chow, A. H.; Lam, W. H.; Wong, S. C.; and Ma, W. 2022. Domain adversarial spatial-temporal network: A transferable framework for short-term traffic forecasting across cities. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 1905–1915.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Wang, L.; Yang, N.; Huang, X.; Yang, L.; Majumder, R.; and Wei, F. 2023. Improving text embeddings with large language models. *arXiv preprint arXiv:2401.00368*.

Wang, Y.; Albrecht, C. M.; Braham, N. A. A.; Mou, L.; and Zhu, X. X. 2022. Self-supervised learning in remote sensing: A review. *IEEE Geoscience and Remote Sensing Magazine*, 10(4): 213–247.

Wang, Y.; Wu, H.; Dong, J.; Liu, Y.; Long, M.; and Wang, J. 2024. Deep Time Series Models: A Comprehensive Survey and Benchmark. *arXiv preprint arXiv:2407.13278*.

Wang, Z.; Li, H.; and Rajagopal, R. 2020. Urban2vec: Incorporating street view imagery and pois for multi-modal urban neighborhood embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 1013–1020.

Wu, C.; Chen, L.; Wang, G.; Chai, S.; Jiang, H.; Peng, J.; and Hong, Z. 2020a. Spatiotemporal scenario generation of traffic flow based on lstm-gan. *IEEE Access*, 8: 186191–186198.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.

Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in Neural Information Processing Systems*, 34: 22419–22430.

Wu, L.; Wen, H.; Hu, H.; Mao, X.; Xia, Y.; Shan, E.; Zhen, J.; Lou, J.; Liang, Y.; Yang, L.; et al. 2023. Lade: The first comprehensive last-mile delivery dataset from industry. *arXiv preprint arXiv:2306.10675*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020b. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, 753–763.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019a. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019b. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*.

Xi, Y.; Li, T.; Wang, H.; Li, Y.; Tarkoma, S.; and Hui, P. 2022. Beyond the first law of geography: Learning representations of satellite imagery by leveraging point-of-interests. In *Proceedings of the ACM Web Conference 2022*, 3308–3316.

Yeh, C.; Meng, C.; Wang, S.; Driscoll, A.; Rozi, E.; Liu, P.; Lee, J.; Burke, M.; Lobell, D. B.; and Ermon, S. 2021. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; and Niu, Z. 2024. Frequency-domain MLPs are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36.

Yu, B.; Yin, H.; and Zhu, Z. 2017. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.

Yuan, Y.; Ding, J.; Feng, J.; Jin, D.; and Li, Y. 2024. UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. *arXiv preprint arXiv:2402.11838*.

Yuan, Y.; Zhang, Y.; Wang, B.; Peng, Y.; Hu, Y.; and Yin, B. 2022. STGAN: Spatio-temporal generative adversarial network for traffic data imputation. *IEEE Transactions on Big Data*, 9(1): 200–211.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2022. Are transformers effective for time series forecasting? *arXiv preprint arXiv:2205.13504*.

Zhai, Y.; Tong, S.; Li, X.; Cai, M.; Qu, Q.; Lee, Y. J.; and Ma, Y. 2023. Investigating the catastrophic forgetting in multimodal large language models. *arXiv preprint arXiv:2309.10313*.

Zhang, F.; Guo, T.; and Wang, H. 2023. DFNet: Decomposition fusion model for long sequence time-series forecasting. *Knowledge-Based Systems*, 277: 110794.

Zhang, F.; Wang, M.; Zhang, W.; and Wang, H. 2025. THATSN: Temporal hierarchical aggregation tree structure network for long-term time-series forecasting. *Information Sciences*, 692: 121659.

Zhang, J.; He, Y.; Chen, W.; Kuang, L.-D.; and Zheng, B. 2024a. CorrFormer: Context-aware tracking with cross-correlation and transformer. *Computers and Electrical Engineering*, 114: 109075.

Zhang, M.; Li, T.; Li, Y.; and Hui, P. 2021. Multi-view joint graph representation learning for urban region embedding. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, 4431–4437.

Zhang, Q.; Huang, C.; Xia, L.; Wang, Z.; Yiu, S. M.; and Han, R. 2023a. Spatial-temporal graph learning with adversarial contrastive adaptation. In *International Conference on Machine Learning*, 41151–41163. PMLR.

Zhang, W.; Wang, H.; and Zhang, F. 2024. Skip-Timeformer: Skip-Time Interaction Transformer for Long Sequence Time-Series Forecasting. In *International joint conference on artificial intelligence*, 5499–5507.

Zhang, Y.; Wei, C.; Wu, S.; He, Z.; and Yu, W. 2023b. GeoGPT: understanding and processing geospatial tasks through an autonomous GPT. *arXiv preprint arXiv:2307.07930*.

Zhang, Z.; Sun, Y.; Wang, Z.; Nie, Y.; Ma, X.; Sun, P.; and Li, R. 2024b. Large language models for mobility in transportation systems: A survey on forecasting tasks. *arXiv preprint arXiv:2405.02357*.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.

Zhou, T.; Niu, P.; Sun, L.; Jin, R.; et al. 2023a. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36: 43322–43355.

Zhou, Z.; Liu, Y.; Ding, J.; Jin, D.; and Li, Y. 2023b. Hierarchical knowledge graph learning enabled socioeconomic indicator prediction in location-based social network. In *Proceedings of the ACM Web Conference 2023*, 122–132.

Appendices

Descriptions of datasets and models in GP

Datasets We utilized a variety of global-scale datasets to inform our models. The datasets are listed below with their key characteristics:

- **Annual Air Temperature:** Mean annual daily mean air temperature data.
  `CHELSA_bio1_1981-2010_V.2.1.tif`

- **Annual Precipitation:** Mean annual accumulated precipitation amount.
  `CHELSA_bio12_1981-2010_V.2.1.tif`

- **Monthly Climate Moisture:** Average monthly climate moisture index.
  `CHELSA_cmi_mean_1981-2010_V.2.1.tif`

- **Population Density:** GeoLLM aggregates WorldPop population data at 1km resolution, employing importance sampling by population size.
  `ppp_2020_1km_Aggregated.tif`

- **Nighttime Light Intensity:** Satellite images capturing nighttime luminosity from VIIRS with a 500-meter resolution.
  `VNL_npp_2023_global_vcmslcfg_v2_`
  `c202402081600.cvg.dat.tif`

- **Human Modification Terrestrial:** A cumulative metric of human modification of terrestrial lands at a 1-km resolution, modeled from 13 anthropogenic stressors.
  `lulc-human-modification-terrestrial-`
  `systems_geographic.tif`

- **Global Gridded Relative Deprivation:** Index of multidimensional deprivation and poverty, ranging from 0 (lowest) to 100 (highest), at 30 arc-second (1 km) resolution.
  `povmap-grdi-v1.tif`

- **Other Indicators:** Additional indicators include the Ratio of Built-up Area to Non-built Up Area, Child Dependency Ratio, and Subnational Human Development Index.
  `povmap-grdi-v1_BUILT.tif, povmap-grdi-`
  `v1_CDR_CopyRaster.tif, povmap-grdi-v1_`
  `SHDI.tif`

- **Infant Mortality Rates:** Subnational Infant Mortality Rate estimates for 234 countries and territories.
  `povmap_global_subnational_infant_`
  `mortality_rates_v2_01.tif`

- **SustainBench Indicators:** Asset index, sanitation index, and women's BMI collected from Demographic and Health Surveys across 48 countries.
  `dhs_asset_index.tif, dhs_sanitation_`
  `index.tif, dhs_women_bmi.tif`

For country-scale and city-scale datasets, we used the datasets provided by MapillaryGCN and HKGL, respectively. The MapillaryGCN dataset records poverty rate, population density, and women's BMI across 6,000 communities in India, as detailed in the original article. The HKGL dataset captures population density, education level, income level, and crime rate across 1,500 census tracts in NYC.

Models We evaluate several baseline models, as detailed below:

- **Image Supervised Learning**: Trains a ResNet34 pretrained on ImageNet1k to predict cluster-specific indicators from street view images.

- **Object Counts**: Utilizes object detection outputs from street view images, followed by MLPs for indicator prediction.

- **MapilaryGCN**: Combines street view images and object detection results with a GCN for feature aggregation and prediction.

- **Node2Vec**: Employs random walks to learn node embeddings using skip-gram models.

- **GCN**: Aggregates information from neighboring nodes for embedding learning.

- **GAT**: Utilizes attention mechanisms to differentially weight information from neighboring nodes during aggregation.

- **ZE-Mob**: Leverages the co-occurrence of origin-destination locations to learn location embeddings from mobility flow data.

- **MGFN**: Fuses mobility graphs with similar patterns, then learns location embeddings using a multi-level attention mechanism.

- **MV-PN**: Constructs multi-view POI-POI networks per location and learns embeddings through an encoder-decoder framework.

- **HDGE**: Jointly learns location embeddings from both spatial and flow graphs.

- **HUGAT**: Defines meta-paths to capture semantics in LBSN, applying a heterogeneous graph attention network for embedding learning.

- **MVURE**: Models various location correlations using different graphs, with a joint learning module for location embeddings.

- **HKGL**: Implements a hierarchical KG learning model, using LBKG for global knowledge distillation and sub-KGs for domain-specific knowledge capture.

- **Bert-whitening**: In Bert, sentence representations are obtained by averaging the vectors of individual words, on the basis of which the isotropy of sentence representations is enhanced by whitening operations.

Datasets and Models in LTSF

Datasets We evaluate the proposed models on a diverse set of datasets, each characterized by distinct temporal patterns:

- **Global Wind, Global Temp:** This dataset, provided by Corrformer, originates from the National Centers for Environmental Information (NCEI). It encompasses hourly averaged wind speeds and temperatures from 3,850 global stations, spanning from January 1, 2019, to December 31, 2020.

- **Solar Energy:** Consists of electricity generation data from 137 solar stations in Alabama, recorded at 15-minute intervals.

- **Demand-SH:** A delivery demand dataset from Shanghai, provided by LaDe, comprising 96,000 trajectories over a 6-month period.
- **Air Quality:** Includes air quality measurements from 437 cities across China.
- **Traffic-SD:** Contains traffic flow data at 716 nodes in San Diego, recorded every 5 minutes from January 1, 2017, to December 31, 2021.

Models We benchmark the performance of the following models:

- **iTransformer:** Applies the attention mechanism and a feed-forward network to inverted dimensions. Specifically, time points of individual series are embedded into variate tokens, which are utilized by the attention mechanism to capture multivariate correlations. Concurrently, the feed-forward network is employed on each variate token to learn non-linear representations.
- **TSMixer:** Based on mixing operations across both time and feature dimensions, TSMixer efficiently extracts relevant information.
- **RMLP:** Utilizes RevIN (reversible normalization) and Channel Independence (CI) to enhance overall forecasting performance.
- **Informer:** Investigates sparsity in the self-attention mechanism and proposes an efficient Transformer architecture tailored for LTSF.
- **STID**: Identifies nodes and time slots using learnable embedding.

Descriptions of datasets and models in GSTF

Datasets We utilize the following datasets:

- **LargeST-SD, GLD, GBA:** These datasets are provided by LargeST and include traffic flow data spanning five years from three cities, covering approximately 8,600 nodes.
- **Delivery Demand-ShangHai, Hangzhou:** Provided by LaDe, these datasets contain detailed package information, such as location and time requirements. They also include event logs documenting courier activities, such as task acceptance and completion.

Models Our research employs several models:

- **DCRNN:** This model leverages bidirectional random walks on graphs to capture spatial dependencies and uses an encoder-decoder architecture with scheduled sampling to address temporal dependencies.
- **STGCN:** STGCN employs a fully convolutional structure to address graph-based problems, enabling faster training and reduced model complexity.
- **ASTGCN:** Integrating gated recurrent units with adaptive graph convolutional networks, ASTGCN dynamically learns graph structures while capturing spatial and local temporal relationships.
- **AGCRN:** AGCRN introduces two adaptive modules that enhance GCN capabilities: a Node Adaptive Parameter

Learning (NAPL) module for node-specific pattern learning and a Data Adaptive Graph Generation (DAGG) module for inferring automatic inter-dependencies among traffic series.
- **GWNET:** GWNET develops a novel adaptive dependency matrix, learning through node embedding to capture the hidden spatial dependencies in data.
- **MTGNN:** This model features a graph learning module for extracting uni-directed variable relations, a mix-hop propagation layer, and a dilated inception layer, enhancing both spatial and temporal dependency capture within the time series.

Full results of LTSF

Full results of LTSF are shown in Table 12. We can see that LLMGeovec leads to direct improvement in various LSTF models and datasets with different prediction lengths.

Table 12: Full Results of the LTSF benchmarks. We report the forecast error of different models under different prediction lengths. IMP shows the average percentage of MSE improvement of LLMGeovec.

Models	iTransformer		w/ LLMGeovec		TSMixer		w/ LLMGeovec		RMLP		w/ LLMGeovec		Informer		w/ LLMGeovec		IMP
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	%
Global Wind 24	3.812	1.440	**3.222**	**1.237**	3.583	1.313	**3.524**	**1.292**	3.873	1.356	**3.562**	**1.300**	4.708	1.532	**4.683**	**1.524**	15.47%
48	4.441	1.456	**3.851**	**1.354**	4.191	1.414	**4.105**	**1.391**	4.521	1.476	**4.160**	**1.411**	4.842	1.558	**4.749**	**1.537**	13.29%
96	4.904	1.553	**4.298**	**1.435**	4.670	1.484	**4.416**	**1.479**	4.997	1.559	**4.434**	**1.457**	5.081	1.623	**4.960**	**1.603**	12.36%
168	5.171	1.602	**4.546**	**1.493**	4.598	1.486	**4.484**	**1.466**	5.260	1.607	**4.564**	**1.488**	4.987	**1.591**	4.983	1.600	12.08%
Avg	4.582	1.513	**3.979**	**1.380**	4.261	1.424	**4.132**	**1.407**	4.905	1.498	**4.180**	**1.414**	4.905	1.576	**4.844**	**1.566**	13.30%
Global Temp 24	9.000	2.048	**8.808**	**2.040**	7.564	1.945	**6.927**	**1.855**	8.555	1.997	**8.000**	**1.931**	**15.200**	**2.893**	16.226	3.014	2.13%
48	12.447	2.481	**11.346**	**2.383**	10.406	2.324	**10.290**	**2.304**	11.752	2.392	**11.086**	**2.332**	16.494	3.038	**16.181**	**2.997**	8.85%
96	16.295	2.916	**15.623**	**2.869**	13.738	2.696	**12.790**	**2.632**	15.293	2.782	**14.263**	**2.703**	**19.333**	3.311	19.334	3.309	4.12%
168	19.076	3.169	**18.003**	**3.114**	16.433	2.955	**15.768**	**2.900**	18.187	3.061	**16.752**	**2.954**	**22.453**	**3.596**	22.814	3.625	5.63%
Avg	13.079	2.653	**11.945**	**2.601**	12.035	2.480	**11.441**	**2.398**	13.447	2.558	**12.525**	**2.480**	18.370	3.209	**18.639**	3.234	5.19%
Solar Energy 96	0.203	0.237	**0.180**	**0.213**	0.222	0.281	**0.199**	**0.281**	0.233	0.296	**0.213**	**0.271**	0.236	**0.279**	0.227	0.289	11.33%
192	0.233	**0.261**	0.217	0.289	0.261	**0.301**	0.229	0.308	0.260	0.316	**0.239**	**0.291**	0.227	0.287	**0.303**	0.332	6.87%
336	0.248	**0.273**	0.216	0.291	0.271	0.299	**0.206**	**0.270**	0.276	0.323	**0.244**	**0.292**	0.262	0.310	0.254	**0.309**	12.90%
720	0.249	**0.275**	**0.211**	0.277	0.267	**0.293**	0.243	0.306	0.273	0.316	**0.244**	**0.288**	0.329	0.355	**0.259**	0.323	15.26%
Avg	0.233	0.262	**0.206**	**0.265**	0.255	0.294	**0.219**	**0.289**	0.261	0.313	**0.235**	**0.286**	0.264	0.308	**0.263**	0.313	11.59%
Demand-SH 48	0.238	0.256	**0.233**	**0.253**	0.259	0.282	**0.246**	**0.259**	0.244	0.271	**0.227**	**0.242**	0.474	0.401	**0.385**	**0.359**	2.10%
96	0.291	0.282	**0.285**	**0.281**	0.315	0.311	**0.301**	**0.290**	0.302	0.306	**0.279**	**0.268**	0.588	0.479	**0.489**	**0.418**	2.06%
168	0.360	0.314	**0.351**	**0.313**	0.383	0.348	**0.362**	**0.319**	0.373	0.343	**0.345**	**0.299**	1.028	0.686	**0.537**	**0.461**	2.50%
360	0.434	0.340	**0.420**	**0.339**	0.461	0.386	**0.433**	**0.350**	0.459	0.384	**0.420**	**0.332**	1.475	0.904	**0.704**	**0.525**	3.23%
Avg	0.331	0.298	**0.322**	**0.297**	0.355	0.332	**0.336**	**0.305**	0.345	0.326	**0.318**	**0.286**	0.896	**0.618**	0.779	0.666	2.47%
Air Quality 6	1.155	0.467	**1.126**	**0.465**	1.289	0.507	**1.259**	**0.505**	1.235	0.495	**1.158**	**0.468**	3.542	0.817	**2.880**	**0.713**	2.51%
12	1.672	0.593	**1.589**	**0.567**	**1.775**	0.606	1.787	**0.602**	1.629	0.583	**1.610**	**0.576**	3.409	0.807	**3.087**	**0.756**	4.97%
24	2.155	0.683	**2.081**	**0.673**	2.333	0.727	**2.196**	**0.701**	2.048	0.671	**2.028**	**0.664**	4.859	0.955	**3.236**	**0.792**	3.44%
48	2.707	0.781	**2.628**	**0.770**	2.875	0.819	**2.713**	**0.791**	2.517	0.757	**2.483**	**0.744**	3.524	0.878	**3.228**	**0.824**	2.92%
Avg	1.922	0.631	**1.856**	**0.619**	2.068	0.665	**1.989**	**0.650**	1.857	0.627	**1.820**	**0.613**	3.584	0.864	**2.858**	**0.771**	3.46%
Traffic-SD 96	0.104	0.195	**0.080**	**0.176**	0.090	0.183	**0.086**	**0.173**	0.175	0.264	**0.140**	**0.237**	0.167	0.267	**0.124**	**0.225**	23.08%
192	0.139	0.229	**0.104**	**0.202**	0.111	0.207	**0.098**	**0.191**	0.217	0.305	**0.174**	**0.270**	0.207	0.304	**0.140**	**0.239**	25.18%
336	0.167	0.252	**0.128**	**0.221**	0.138	0.234	**0.120**	**0.215**	0.229	0.320	**0.189**	**0.286**	0.224	0.321	**0.167**	**0.266**	23.35%
720	0.134	0.223	**0.112**	**0.205**	0.124	0.225	**0.115**	**0.209**	0.201	0.293	**0.168**	**0.264**	0.199	0.298	**0.176**	**0.285**	16.42%
Avg	0.136	0.225	**0.106**	**0.201**	0.116	0.212	**0.105**	**0.197**	0.205	0.296	**0.168**	**0.264**	0.199	0.298	**0.152**	**0.254**	22.01%