# Foundation Models for Time Series Analysis: A Tutorial and Survey

Yuxuan Liang
The Hong Kong University of Science
and Technology (Guangzhou)
yuxliang@outlook.com

Haomin Wen
Beijing Jiao Tong University &
HKUST(GZ)
wenhaomin@bjtu.edu.cn

Yuqi Nie
Princeton University
ynie@princeton.edu

Yushan Jiang
University of Connecticut
yushan.jiang@uconn.edu

Ming Jin
Monash University
ming.jin@monash.edu

Dongjin Song
University of Connecticut
dongjin.song@uconn.edu

Shirui Pan
Griffith University
s.pan@griffith.edu.au

Qingsong Wen
Squirrel AI, USA
qingsongedu@gmail.com

## ABSTRACT

Time series analysis stands as a focal point within the data mining community, serving as a cornerstone for extracting valuable insights crucial to a myriad of real-world applications. Recent advances in Foundation Models (FMs) have fundamentally reshaped the paradigm of model design for time series analysis, boosting various downstream tasks in practice. These innovative approaches often leverage pre-trained or fine-tuned FMs to harness generalized knowledge tailored for time series analysis. This survey aims to furnish a comprehensive and up-to-date overview of FMs for time series analysis. While prior surveys have predominantly focused on either application or pipeline aspects of FMs in time series analysis, they have often lacked an in-depth understanding of the underlying mechanisms that elucidate why and how FMs benefit time series analysis. To address this gap, our survey adopts a methodology-centric classification, delineating various pivotal elements of time-series FMs, including model architectures, pre-training techniques, adaptation methods, and data modalities. Overall, this survey serves to consolidate the latest advancements in FMs pertinent to time series analysis, accentuating their theoretical underpinnings, recent strides in development, and avenues for future exploration.

## CCS CONCEPTS

• **Information systems → Spatial-temporal systems**.

## KEYWORDS

Time series, foundation model, deep learning

**Figure 1: Roadmaps of representative TSFMs.**

## 1 INTRODUCTION

Time series data are characterized by their sequential order and temporal dependencies, encapsulating valuable information about the dynamics of diverse systems and processes [45, 86, 113]. Various time series data (e.g., stock price, traffic flow, electricity) present unique challenges and opportunities for computational analysis, each requiring tailored approaches to effectively capture their inherent properties. The analysis and understanding of time series data is an important piece of data mining, facilitating crucial insights and decisions in many domains [46, 94], including finance [19, 108], healthcare [52, 62], cloud computing [116], environments [14, 70], energy [123], and urban computing [79, 88].

In recent years, the advancements of deep learning, especially the transformer-based models [85], have revolutionized the field of time series analysis [95]. Following the pioneering work [54], there has been a surge in research interest exploring the application of

**Table 1: Comparison between our survey and related surveys. (Abbr: Taxonomy means the main taxonomy used in the survey. Standard means standard time series, Spatial means spatial time series, Others include trajectory and event).**

| Survey | Taxonomy | Standard | Spatial | Others |
|---|---|---|---|---|
| Jin et al. [48] | Data | ✔ | ✔ | ✘ |
| Jiang et al. [45] | Pipeline | ✔ | ✘ | ✔ |
| Zhang et al. [114] | Pipeline | ✔ | ✘ | ✘ |
| Miller et al. [66] | Pipeline | ✔ | ✘ | ✘ |
| (Ours) | **Methodology** | ✔ | ✔ | ✔ |

transformers for time series analysis [101, 118, 119]. The motivation behind deep learning and transformers lies in their ability to automatically learn comprehensive representations from raw data, thus capturing complex nonlinear relationships and temporal dependencies without the need for manual feature engineering. Such capability leads to significant performance improvements compared with traditional statistical methods across numerous time series applications. Foundation models (FMs), such as large language models (LLMs) in natural language processing (NLP) [117] and advanced models in computer vision (CV) [2], have emerged as powerful paradigms capable of achieving state-of-the-art performances in their respective fields. The success of these FMs can be attributed to their ability to *leverage vast amounts of data to cultivate general-purpose representations*, subsequently fine-tuning them, or even deploying them directly in a zero-shot manner to excel across a diverse spectrum of downstream tasks. This approach not only economizes on the need for task-specific model development but also encapsulates a broad understanding of the world, endowing these models with exceptional versatility and efficiency [6, 49].

Inspired by the remarkable achievements of FMs in broad domains like CV and NLP, the concept of Time Series Foundation Models (TSFMs) has garnered attention as a promising direction for time series analysis. TSFMs aim to harness the power of the foundation model paradigm to develop generalized models proficient in understanding and forecasting time series data spanning diverse domains. By capitalizing on large-scale time series datasets, TSFMs hold the promise of attaining superior performance on a spectrum of time series tasks, offering a unified framework that can accelerate research and application developments in this field.

Despite the promising prospects and rapid development of TSFMs, a systematic analysis of TSFMs from a methodological standpoint has been notably absent in prior literature. Existing studies, as depicted in Table 1, have concentrated on either the data perspective [48] or the pipeline perspective [45] of TSFMs. To bridge this gap, this survey aims to provide a comprehensive methodological analysis of foundation models for learning a variety of time series. This examination will center on scrutinizing their *model architectures*, *pre-training techniques*, *adaptation methods*, and *data modalities*. Through this endeavor, we seek to illuminate an overall picture of core elements in TSFMs, thereby enhancing comprehension regarding the rationale behind their efficacy and the mechanisms driving their substantial potential in time series analysis.

In contrast to previous surveys, this manuscript incorporates the most extensive array of time series data types (see Table 1),

spatial time series, as well as other types such as the trajectory and event. We further summarize the developmental roadmap of current TSFMs in Figure 1, in order to foster further innovations and understanding in the dynamic and ever-evolving landscape of TSFMs. In short, our major contributions lie in three aspects:

- **Comprehensive and up-to-date survey.** We offer a comprehensive and up-to-date survey on foundation models for a wide spectrum of time series, encompassing standard time series, spatial time series, and other types (i.e., trajectories and events).
- **Novel methodology-centric taxonomy.** We introduce a novel taxonomy that offers a thorough analysis from a methodological standpoint on TSFMs with the first shot, enabling a full understanding of the mechanism on why and how FMs can achieve admirable performance in time series data.
- **Future research oppotunities.** We discuss and highlight future avenues for enhancing time series analysis using foundation models, urging researchers to delve deeper into this area.

## 2 BACKGROUND

**Foundation Models.** Foundation models (FMs), also known as large pre-trained models, are a class of deep models that are pre-trained on vast amounts of data, thus equipped with a wide range of general knowledge and patterns. To this end, these models serve as a versatile starting point for various tasks across different domains. Specifically, FMs can be fine-tuned or adapted to specific tasks with relatively small amounts of task-specific data, showcasing remarkable flexibility and efficiency. In CV, FMs such as text-prompted model CLIP [73] and visual-prompted model SAM [51] have propelled advancements in image recognition, object detection, and more. In NLP, FMs such as BERT [25] and GPT-3 [8] have revolutionized text understanding and generation tasks. Inspired by the great success of FMs in the above domains, this survey delves into the utilization of these models in the realm of time series analysis.

Concretely, we investigate TSFMs from a methodology perspective: the components of foundation models encompass the data modality, architecture, pre-training, and adaptation technicals: 1) data modality refers to the type of data used for model training, from single modality such as time series, text, images, and audio to multimodality; 2) architecture refers to which deep neural network is adopted as the backbone of FM, with Transformers [85, 95] being a popular choice for their ability to handle sequential data effectively; 3) Pre-training involves how to train the model on large, diverse datasets to gain a broad understanding of the data, using supervised or self-supervised learning; 4) Adaptation, such as fine-tuning or few-shot learning, is employed to accommodate the pre-trained FMs to specific tasks. This comprehensive framework of FMs, spanning from data modality to adaptation, facilitates the understanding of using them in time series analysis.

**Categories of Time Series.** A time series is commonly described as an ordered sequence of data points. Figure 2 illustrates various types of time series discussed in this survey, including standard time series, spatial time series, trajectories, and events. Note that trajectories and events can be regarded as time series since each data point is associated with a specific timestamp (and location), allowing for analysis using time series techniques such as anomaly detection. These time series are formulated as follows.
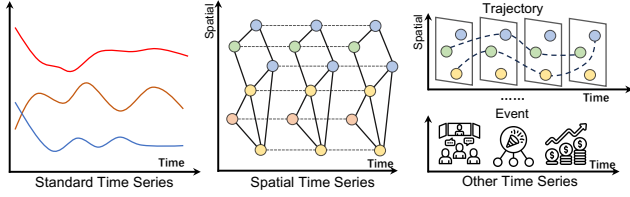
**Figure 2: Illustration of various types of time series.**

*Definition 1 (Standard Time Series).* The standard time series is defined as a sequence of $T$ data points ordered by time. It can be denoted by $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T\} \in \mathbb{R}^{T \times D}$, where $\mathbf{x}_t \in \mathbb{R}^D$ is the data point at time step $t$, and $D$ is the dimension of each data points. When $D = 1$, $\mathbf{X}$ is referred to as a univariate time series, while $D > 1$, $\mathbf{X}$ is a multivariate time series.

*Definition 2 (Spatial Time Series).* It refers to a sequence of data points with both temporal and spatial dimensions, which can be represented by $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_T\} \in \mathbb{R}^{N \times T \times D}$, where $\mathbf{X}_t \in \mathbb{R}^{N \times D}$ denotes the signals generated by $N$ sensors with each equipped with $D$ features. Besides, the $N$ sensors are usually associated with spatial correlations, according to which the spatial time series can be further divided into two subtypes: i) spatio-temporal graph, when the spatial correlation of those sensors is described by a graph $G$ with adjacent matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$; ii) spatio-temporal raster, when sensors are distributed uniformly as a grid in geographical space.

*Definition 3 (Trajectory).* A trajectory is a sequence of timestamped locations that describe the movements of an object in the geographical space. It can be formulated as $\mathcal{T} = \{(l_1, l_2, \cdots, l_T\} \in \mathbb{R}^{T \times 2}$, where $l_t$ means the object's location at time $t$, represented by the two-dimensional coordinates, i.e., latitude and longitude.

*Definition 4 (Event Sequence).* An event sequence is a temporally ordered set of events that describe the progression of actions or occurrences within a specific context. It can be formalized as $\mathcal{E} = \{(e_1, t_1), (e_2, t_2), \ldots, (e_n, t_n)\}$, where $e_i$ is an event described by a predicate-argument structure that captures the nature of the occurrence, and $t_i$ denotes the timestamp when $e_i$ occurs.

## 3 TAXONOMY

The proposed taxonomy is illustrated in Figure 3, and the related works can be found in Table 2. The proposed taxonomy unfolds a structured and comprehensive classification to enhance the understanding of foundation models on time series analysis. It is organized into four hierarchical levels, starting with the data category, followed by the model architecture, pre-training techniques, and finally, the application domain. Unlike previous taxonomies, ours distinguishes itself by delving deeper into the foundation models from the methodology perspective, with a keen focus on their architectural designs, pre-training, and adaptation techniques. This method-centric view is pivotal for researchers, providing valuable insights into the mechanisms of why and how foundation models show great potential for time series analysis.

Diving into the details of data categories, we classify the time series data into three distinct types: standard time series, spatial time series, and others, which encompass trajectory and event data. Standard time series data, characterized by their sequential order and temporal dependencies, form the backbone of traditional time

series analysis. Spatial time series data, on the other hand, introduce an additional layer of complexity by incorporating geographical or spatial information, making them crucial for applications in urban computing and environmental monitoring. Lastly, the "others" category, including trajectory and event data, represents diverse datasets where time plays a critical role, such as the movement of objects over time or the occurrence of specific events, offering a broadened perspective on time series analysis.

From the methodology perspective: i) regarding model architecture, the proposed taxonomy highlights three primary categories: Transformer-based, non-Transformer-based, and diffusion-based models. Transformer-based models leverage self-attention mechanisms to capture long-range dependencies within time series, offering significant advantages in handling sequential data. Non-transformer-based models, with their diverse architectures, cater to a wide range of time series tasks by efficiently processing temporal patterns. Diffusion-based models, a novel addition, employ stochastic processes to model the data generation process, presenting innovative solutions for time series analysis. ii) In terms of pre-training techniques, the proposed taxonomy divides them into fully-supervised and self-supervised methods, the latter of which includes contrastive, generative, and hybrid approaches. This classification shows how different FMs are trained with or without labels. iii) Adaptation strategies, such as zero-shot learning, prompt engineering, tokenization, and fine-tuning, further exemplify the versatility of FMs in customizing to specific time series applications.

## 4 DATA PERSPECTIVE

In this section, we explore advancements in TSFMs from various data perspectives: *standard time series*, *spatial time series*, and *others*. We further categorize our discussion within each subsection into *task-oriented* or *general-purpose* foundation models.

### 4.1 Standard Time Series

Standard time series possess diverse properties, including varying sampling rates and temporal patterns, which pose significant challenges in developing relevant FMs. These models aim to identify universal patterns within extensive time series data from varied sources, either to enhance specific tasks or for broad analysis.

Most of the existing attempts are in the category of task-oriented standard time series foundation models. They leverage single or multiple data modalities to craft robust models targeting particular time series tasks, typically forecasting or classification. For models involved only in a single (i.e., time series) modality, they may either be developed from scratch or on existing pre-trained models from other domains like large language or vision models [117].

In the first group, Lag-Llama [75] and TimeGPT-1 [35] represent pioneering efforts as forecasting foundation models. Both models undergo pre-training on a vast collection of time series data spanning multiple domains. Lag-Llama employs a decoder-only transformer architecture, utilizing lags as covariates, whereas TimeGPT-1 features an encoder-decoder structure with several transformer layers, facilitating efficient zero-shot forecasting. Another noteworthy contribution is TTMs [31], a recent endeavor in creating a domain-agnostic forecasting model built upon TSMixer [30], which itself is pre-trained on diverse time series datasets from various
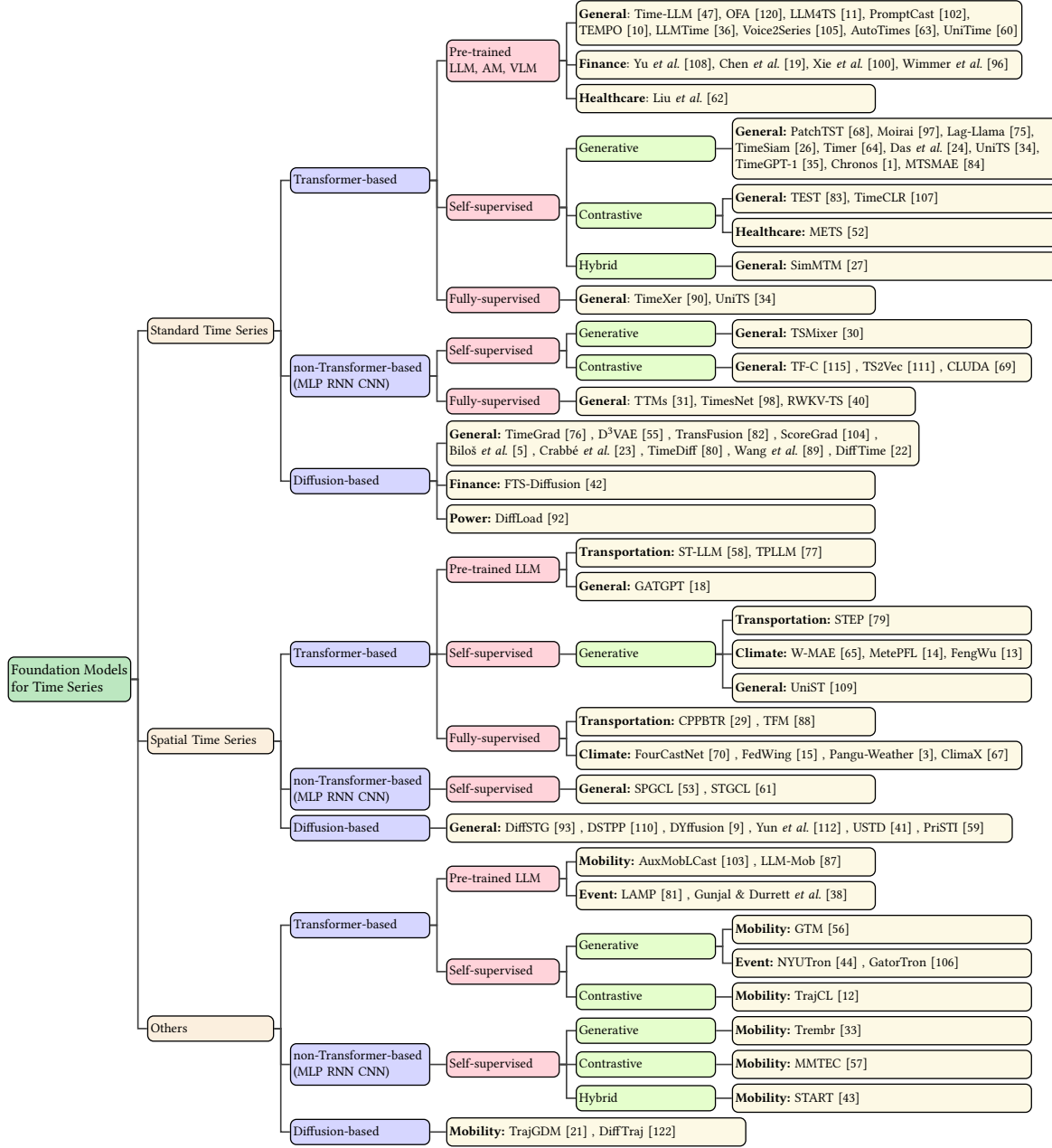
**Figure 3: A comprehensive taxonomy of TSFMs, categorized according to data and methodologies.**

domains. Echoing Lag-Llama's approach, TimesFM [24] emerges as a decoder-only model exhibiting strong zero-shot forecasting capabilities. Concurrently, Moirai [97] introduces an approach with its masked encoder-based universal forecasting transformer, coupled with a new pre-training dataset (LOTSA), containing 27 billion observations from nine distinct domains. Additionally, the exploration extends to diffusion models like TimeGrad [76] and TransFusion [82], which primarily focus on optimizing a variational bound on data likelihood, transforming white noise into meaningful samples of the target distribution.

Pre-training from scratch can be expensive, which has spurred the development of alternative approaches that leverage pre-trained models from other domains, such as large language, vision, and acoustic models. For instance, LLM4TS [11] and TEMPO [10] successfully perform time series forecasting across various datasets by fine-tuning GPT-2 [74] backbones, predicated on the notion that LLMs can be adapted to process non-linguistic datasets by activating their inherent capabilities. Similarly, Voice2Series [105] engages in the synchronization of time series and acoustic data to harness the classification prowess of an acoustic model for time series data.

Another approach is presented by Wimmer *et al.* [96], who utilize vision-language models (VLMs) to predict market changes. Beyond fine-tuning existing models, a distinct methodology involves direct inference from LLMs for time series forecasting, showcasing commendable zero-shot performance. A notable example of this is LLMTime [36], which introduces various strategies for effectively tokenizing time series data and transforming discrete token distributions into flexible continuous value densities.

Beyond approaches that focus solely on a single data modality of time series, there have been initiatives towards developing multi-modal, task-oriented foundation models. A notable example is Time-LLM [47], which introduces a reprogramming framework to integrate textual and time series information, repurposing an existing LLM into time series forecasters without additional computational costs. In a similar vein, METS [52] employs a trainable ECG encoder alongside a frozen language model to process paired ECG and clinical reports. Further, there is emerging research on directly prompting LLMs for specific time series tasks. For instance, Prompt-Cast [102] converts numerical inputs and outputs into prompts, framing the forecasting task as a sentence-to-sentence conversion to leverage language models directly for forecasting. Other studies, such as one involving LLMs prompted with historical stock price data, company metadata, and past economic/financial news, aim to enhance stock return forecasting [108]. Another example combines GNNs with ChatGPT to predict stock movements [19], illustrating the diverse applications of these methodologies. Additional noteworthy efforts include [100] and [62].

Notably, recent efforts have been moved towards creating general-purpose, single-modality standard TSFMs. TS2Vec [111] is a pioneering effort by introducing a universal framework for representing time series via contrastive learning. SimMTM [27] explores cross-domain applications, where pre-trained models via masked time series modeling exhibit superior fine-tuning performance in forecasting and classification tasks. More recent works, such as Timer [64] and UniTS [34], further advance the field by facilitating general time series analysis through single, large-scale pre-trained models. Moreover, there is a growing interest in adapting pre-trained models, such as LLMs, for broad time series analysis. OFA [120] and TEST [83] exemplify this trend, though both approaches necessitate end-to-end fine-tuning for specific tasks.

## 4.2 Spatial Time Series

In complex real-world systems, time series data often display intricate spatial dependencies alongside temporal dynamics, manifesting in forms such as spatio-temporal *graphs* and *rasters*. Similar to the discussion in Sec. 4.1, research on spatial time series typically encompasses areas such as forecasting and classification. Unlike foundation models for standard time series, most existing research on spatial time series is still in its early stages, often characterized by being domain-specific, single-modality, and task-oriented. In the following, we categorize related works into two specific data modalities and discuss them in different subsections.

*4.2.1 Spatio-Temporal Graph.* Most foundation models for spatio-temporal graphs are task-oriented and only focused on graph data. In the transportation sector, TFM [88] utilizes graph structures and algorithms to analyze the behavior and interactions within

transportation systems, showing promising results in urban traffic forecasting. ST-LLM [58] combines spatio-temporal information with a partially frozen LLM to improve traffic predictions, while DiffSTG [93] applies denoising diffusion models to spatio-temporal graphs for probabilistic traffic forecasting. Efforts towards domain-agnostic models include STEP [79], which links spatio-temporal GNNs with a pre-trained transformer for enhanced forecasting by learning from extensive historical data. Similarly, STGCL [61] and SPGCL [53] explore the integration of contrastive learning into spatio-temporal graph forecasting, indicating its potential benefits. Research on general-purpose models for spatio-temporal graphs is limited. A notable example, USTD [41], introduces a unified model for both forecasting and kriging tasks, employing an uncertainty-aware diffusion approach to address diverse challenges effectively.

*4.2.2 Spatio-Temporal Raster.* Spatio-temporal raster refers to a data modality that captures and organizes spatial information over various time points in a grid-like format. This modality is primarily utilized in climate foundation models. For instance, FourCast-Net [70] is a global, data-driven weather forecasting model delivering accurate short to medium-range predictions worldwide. Similar models, such as FengWu [13] and W-MAE [65], follow suit. Notably, Pangu-Weather [4], which is trained on 39 years of global data, achieves superior deterministic forecasting outcomes across all evaluated variables compared to leading numerical weather prediction systems. On a different note, ClimaX [67] aims at general-purpose climate foundation models, pre-trained with diverse datasets covering various variables, spatio-temporal scopes, and physical contexts. It is designed for fine-tuning across a wide array of climate and weather-related tasks, such as forecasting, projection, and down-scaling, even for atmospheric variables and spatio-temporal scales not encountered during its pre-training phase. However, there is a scarce number of domain-agnostic models for spatio-temporal raster data. DYffusion [9], for example, capitalizes on the temporal dynamics inherent in raster data, integrating these dynamics directly with the model's diffusion steps to create a stochastic, time-conditioned interpolator and forecasting network.

## 4.3 Others

In addition to standard and spatial time series, various other types of data incorporate the temporal dimension, including trajectories, events, and clinical records. A majority of studies in this category focus on trajectory data. For Transformer-based models, AuxMobLCast [103] fine-tunes pre-trained LLMs through mobility prompting and auxiliary POI Category classification to forecast human mobility patterns, effectively bridging the gap between natural language processing and temporal sequence prediction. LLM-Mob [87] encodes human mobility data into structured prompts that instruct LLMs to consider both long-term and short-term behavioral patterns, along with time-specific context, to generate accurate and explainable predictions of future locations. For non-Transformer-based models, Trembr [33] leverages auto-encoding techniques to extract road network and temporal information embedded in trajectories effectively. While START [43] introduces a hybrid approach to trajectory embedding learning by combining masked language model [25] and SimCLR [17] to enhance its learning capability. More recently, GTM [56] separates trajectory features into three
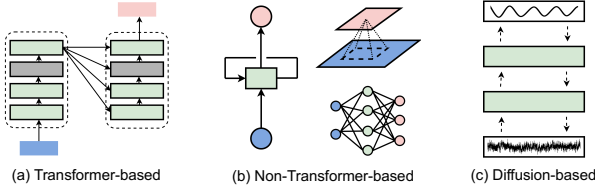
**Figure 4: Architectures of TSFMs.**

domains, which can be masked and generated independently to meet specific input and output requirements of a given task. Then, GTM is pre-trained by reconstructing densely sampled trajectories in an auto-regressive manner given re-sampled sparse counterparts. For the diffusion-based model, DiffTraj [122] reconstructs and synthesizes geographic trajectories from white noise through a conditioned reverse trajectory denoising process.

# 5 METHODOLOGY PERSPECTIVE

In this section, we dissect TSFMs from a methodology perspective, focusing on *architecture* and *pipeline* (including pre-train and adaptation) intricacies. This discussion aims to elucidate the intricate mechanisms driving these models' efficacy and adaptability.

## 5.1 Architecture

As shown in Figure 4, we first delve into the architecture of TSFMs, including *Transformer-based models*, *non-Transformer-based nodels* and *diffusion-based models*, focusing on the underlying mechanisms that shape their capabilities, as well as how they could be applied on various time series.

*5.1.1 Transformer-based Models.* The architecture of FMs has seen a significant convergence towards the Transformer [85], a model architecture first introduced for NLP tasks. The core innovation of the Transformer lies in its utilization of the attention mechanism, which allows the model to dynamically focus on different parts of the input data. The attention function can be succinctly described as $\texttt{Attention}(Q, K, V) = \texttt{Softmax}(QK^T/\sqrt{d_k})V$, where $Q$, $K$, and $V$ represent the queries, keys, and values matrices respectively, each with dimensions $T \times d_k$, and $d_k$ serves as a scaling factor to moderate the dot products' magnitude. It is evident from the formula that the attention mechanism has the capability to learn global, long-range dependencies in data. This distinguishes it from previous architectures, which were often limited by their local receptive fields or dependency windows. Besides, the Transformer's design is inherently friendly to parallelization, which allows for significant scalability, enabling the processing of large datasets and the construction of models with billions of parameters. Such scalability and efficiency in capturing intricate data patterns have led to the widespread adoption of the Transformer architecture beyond its initial application in natural language processing (NLP) [25] to fields including computer vision (CV), speech, video, time series (Table 2) and beyond.

The choice of foundation model framework remains debated in the realm of time series analysis, contrasting the trend towards decoder-only models in natural language processing. Notable works

in this area includes encoder-only [34, 68, 97], encoder-decoder [1, 26, 35], and decoder-only [24, 64, 75] models. *Ansari et al.* [1] analyze the applicability of the encoder-decoder framework to decoder-only models. *Liu et al.* [64] discuss that while the encoder-only model is favored in time series forecasting for its effectiveness on small datasets, the decoder-only architecture, with its strong generalization and capacity, could be preferred for large-scale time series models. The diversity in the architectural choices underscores the potential and necessity for further exploration within this field.

In terms of standard time series analysis, the Transformer architecture leverages its sequence modeling capabilities to capture temporal dynamics. This includes either repurposing pretrained LLMs for time series to leverage their preexisting sequence modeling strengths [102], or directly using Transformer as a base for TSFMs, training from scratch to achieve models best suited for the specifics of time series data [35]. Besides, various techniques have been innovated to augment the functionality of Transformer models in time series analysis comprehensively. A common practice in TSFMs segments time series into patches, which can effectively encapsulate local dynamics within input tokens [10, 11, 20, 24, 47, 68, 75, 97, 120]. Another critical design is the normalization layer, where reversible instance normalization [50] techniques, standardizing data through instance-specific mean and variance then reverting it at the output layer, have found extensive application across the above models. Moreover, specialized approaches such as multi-resolution analysis, exemplified by Moirai [97] through the employment of varying patch sizes, and decomposition strategies, as implemented by TEMPO [10] via the separation of complex interactions into the trend, seasonal, and residual components, have been shown to enhance model efficacy substantially.

For spatial time series, the attention mechanism is utilized to model both the spatial and temporal dependency. For instance, ST-LLM [58] employs a novel partially frozen attention strategy for traffic prediction, leveraging spatial-temporal embeddings to capture the intricate dynamics of traffic data across space and time. Conversely, other studies opt for independent modeling of spatial and temporal relationships. TFM [88] is a case in point, which employs attention mechanisms within a dynamic graph encoder for spatial modeling, integrating time encoding for temporal aspects, embodying principles of transformers in addressing traffic system's spatial-temporal dependencies. Besides simultaneously modeling spatial and temporal relationships, there exists an alternative approach that augments the Transformer model with additional spatial models or external spatial information to enhance its capabilities in the temporal modeling of time series. An example of this is STEP [79], which uses unsupervised pre-trained TransFormer blocks to model temporal relationship from long-term history time series, while applying a graph structure learner and spatio-temporal GNNs based on the representation of TransFormer blocks. Furthermore, the application of Transformer models extends to the domain of spatial-temporal prompt learning, as evidenced by initiatives such as MetePFL [14] and FedWing [15].

In addition to conventional time series data, the Transformer architecture has demonstrated efficacy across a diverse array of temporal datasets, such as trajectory and healthcare records, as summarized in Figure 3. This expansion highlights the Transformer's versatile capacity for temporal data analysis.

*5.1.2 Non-Transformer-based Models.* Excluding the widespread adoption of Transformers, a diverse array of traditional pre-training methods leveraged models such as Multi-Layer Perceptrons (MLPs) [31], Recurrent Neural Networks (RNNs) [33], and Convolutional Neural Networks (CNNs) [98] as the backbone for pre-training. These models, each with their unique strengths, are notable for their effectiveness in both conventional and spatial time series data.

MLPs and CNNs are both acclaimed for their capabilities in modeling spatial and temporal data effectively. CNN-based architectures, in particular, have garnered significant attention in self-supervised learning for general time series representation, with a notable emphasis on the usage of ResNet [27, 115] and dilated convolution layers [69, 111] as foundational backbones. Those approaches predominantly employ 1D convolutional operations. In contrast, TimesNet [98] introduces a novel perspective by converting 1D time series data into 2D tensors, facilitating the adaptive identification of multi-periodicity and the extraction of complex temporal variations through the use of a parameter-efficient inception block. MLP-based models, on the other hand, are lauded for their lightweight design, offering benefits in terms of reduced computational time and cost. TSMixer [30] and TTMs [31], as instances, both claiming superior efficiency in memory usage and processing speed while still delivering competitive performance.

RNNs have been acknowledged for their proficiency in temporal data modeling [33, 39]. Recently, there has been a resurgence of interest in RNN architectures, which poses a compelling challenge to the prevailing Transformer-based models. This trend is driven by the quest for models that are not only more resource-efficient but also adept at handling longer sequences through their inherent linear complexity. A notable embodiment is the RWKV-TS [40], which leverages the RWKV [72], an RNN-type foundation model architecture, demonstrating promising potential for general time series analysis. This emerging trend presents a valuable opportunity for time series research and applications.

*5.1.3 Diffusion-based Models.* Diffusion-based foundation models have gained prominence in CV [71, 78] and video [7] due to their proficiency in learning complex data distributions, yet their exploration in time series analysis remains nascent. These models function by gradually introducing and then reversing noise to data, effectively learning the generative process of original data through the reverse diffusion process. This unique mechanism equips diffusion models with great potential to serve as versatile foundation models capable of tackling prediction, imputation, and anomaly detection in time series.

In standard time series and other temporal data, diffusion models predict future states by capturing temporal dynamics, generating smooth transitions from current to potential future states [76, 92]. Applied to spatial time series, they extend this capability to model spatial correlations alongside temporal ones, providing insights into the interplay between space and time, particularly beneficial in fields like traffic forecasting [93].

## 5.2 Pipeline

In this part, we review TSFMs from the pipeline perspective, including diverse model acquisition and adaptation mechanisms.

*5.2.1 Pre-training.* Pre-training is an initial and crucial step for building TSFMs, since the knowledge learned in this phase enables the models to generalize across different contexts and quickly adapt to various downstream tasks with minimal adaptations. On the other hand, the diverse nature of pre-training data (*e.g.*, standard time series, spatial time series, and trajectories), as well as the way the data is used, lead to a wide spectrum of pre-training mechanisms when building and deploying foundation model. In this survey, we propose a new perspective mostly based on learning objectives in the pre-training phase, to categorize existing methods for TSFMs. These mechanisms include fully-supervised, self-supervised (generative, contrastive, hybrid of generative and contrastive), and others.

Fully-supervised pre-training refers to the strategy where the foundation model is initially trained on one or multiple large time series datasets with labels to capture the complex temporal dynamics and learn generalizable representations. TTMs [31] proposes a universal time series foundation model supervised framework that is able to handle the heterogeneity of multiple time series datasets and effectively build the forecasting capability during pre-training, via the design of multi-resolution enhancements (*e.g.*, adaptive patching, data augmentation via downsampling, *etc.*) Fully-supervised pre-training for TSFMs is particularly suited for scenarios where there is sufficient labeled historical data. Moreover, this pre-training technique is more frequently used in some domain-specific applications such as transportation [29, 88] and climate [3, 70], where the model can be directly tailored for downstream forecasting tasks with the ease of minimal adaptations.

We categorize the generative pre-training strategy as a general modeling of time series representations, including reconstruction and probabilistic modeling of time series inputs. In reconstruction-based pre-training, an effective learning objective is to recover the original input space via masked autoencoding strategies [14, 79]. In the probabilistic modeling methods, the latent representation space formed from temporal or spatial-temporal encoders is optimized toward an estimated density via maximizing log-likelihood, based on which the forecasts can be sampled [76, 93]. Moreover, it is also beneficial to leverage contrastive learning to enhance the robustness of pre-training time series foundation models. The key is to construct and utilize the self-supervision signals by generating informative positive pairs as well as filtering out unsuitable negative pairs when performing augmentation [61]. In addition to the aforementioned two self-supervised strategies, the efficacy of the hybrid variant has also been validated, where the pre-trained model on fewer time series data outperforms the supervised counterpart [43].

In general, self-supervised pre-training enables foundation models to exploit the vast amounts of unlabeled time series data, providing generic temporal knowledge that can be further fine-tuned for specific downstream tasks. Compared with fully-supervised pre-training, it provides a more generic and realistic solution for the acquisition of a time series foundation model.

Note that the above pre-training methods typically build the model from scratch and obtain the universal knowledge from data with the same modality (*i.e.*, time series). Nevertheless, recent advancements in time series research have heightened the usage of LLMs [10, 11, 19, 36, 38, 47, 58, 62, 63, 81, 87, 100, 102, 103, 108, 120], VLMs [96], and AMs [105] that are pre-trained from other data modalities (text sequence, image-text sequence, acoustic signals).

*5.2.2 Adaptation.* The adaptation phase tailors TSFM to specific tasks or datasets, enhancing its performance on those tasks by leveraging the learned generic temporal knowledge. We partition prior methods into four branches, including direct usage, fine-tuning, prompt engineering, and time series tokenization (see Figure 5).
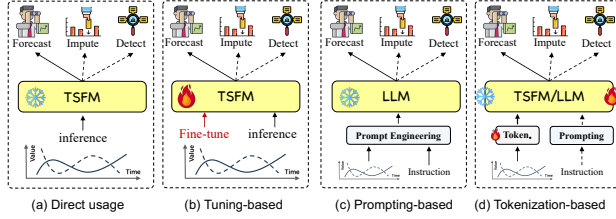


**Figure 5: Illustration of different adaptation techniques.**

Direct usage (also called zero-shot), means no further fine-tuning on the target datasets, suggesting the sufficient capability of a pre-trained model for downstream tasks. It can also indicate the homogeneity between the pre-trained dataset and target dataset, especially for some real-world applications where a foundation model is built to fulfill domain-specific tasks [4, 13].

Fine-tuning is a common strategy to adapt foundation models to target tasks. Based on the way the foundation model is used on the target dataset, there are three mainstream works: fine-tuning the whole model [65, 67, 70] or specific components (*e.g.*, training positional embeddings and layer normalization, while keeping feedforward and attention layers frozen when fine-tuning LLMs) [11, 120], to directly infer results, or integrate foundation models as part of the whole model [19, 52, 83, 103].

Prompt engineering is more specialized in LLM-based TSFMs. The prompt can be handcrafted with task-specific textual input and directly used to query the output for downstream prediction [87, 102] or intermediate embedding as feature enhancement [103]. Besides, the prompt can also be parameterized vectors and end-to-end learnable when optimizing the model on target datasets [10, 83]. In comparison to static prompts, the use of trainable prompts enhances the ability of LLMs to comprehend and match the context of given time series inputs. For example, TEMPO [10] constructs a trainable prompt pool with distinct key-value pairs, and retrieves the most representative prompt candidates with the highest similarity scores.

Time series tokenization aims to effectively represent the time series as embeddings, which is also more frequently adopted in transformer-based architectures [10, 47, 68]. Common tokenization techniques include reversible instance normalization [50] that mitigates distribution shift, patching with channel independence strategy that effectively and efficiently extracts the time series context [68], as well as the joint usage of time series decomposition to explicitly represent explainable components [10] for the ease of subsequent temporal modeling.

In addition to the main branches of adapting TSFMs, it is also worth noting that some fine-tuning strategies take real-world constraints into account. For example, the fine-tuning is performed in a privacy-preserved manner [14, 16].

## 5.3 Modality

During the pre-training/adaptation of TSFMs, prior methods involve single or multiple data modalities, where standard time series, trajectory, raster, and text can be treated as different forms with unique domain perspectives. In this part, we review the data modalities that are used in existing TSFMs across different domains.

*5.3.1 Single-modality.* A majority of current TSFMs are constructed and tailored on the basis of single-modal data. Compared with multi-modal methods, the single-modal time series modeling strategy gains the advantages of inherent simplicity and bypasses the challenges of handling modality gaps, yet frequently demonstrates excellent empirical results across a wide range of real-world applications, such as traffic [58, 79] and climate forecasting [13, 67].

*5.3.2 Multi-modality.* However, the single-modal methods may not encapsulate the full picture for several challenging downstream tasks in finance [19, 108] and healthcare domains [52, 62]. To cope with this issue, there have been initiatives towards developing multimodal, task-oriented FMs, where additional information provides useful information to enhance the model capability. In *Chen et al.*, an external ChatGPT is queried to construct an evolving graph structure representing companies, based on the analysis of news headlines at specific time steps. As such, the inferred graph and stock prices are fed into the time series model (that uses GNN and LSTM for information propagation) to generate stock price movement predictions. Another example in healthcare also demonstrated the effectiveness of multi-modal medical context modeling, which aligns the embedding of ECG (Electrocardiogram) and corresponding medical text reports under a self-supervised contrastive learning framework and performs ECG classification. In general multi-modal time series analysis, similar cross-modality alignment strategies (*e.g.*, contrastive learning [83], reprogramming [47], token-wise prompting [63]) are adopted, where the multi-modal inputs are often the textual description of datasets and pre-training word embedding from LLMs. As a notable example, Time-LLM [47] introduces a reprogramming framework that aligns the language knowledge from pre-trained word embedding and time series information via linear projection and multi-head attention, where the handcrafted dataset descriptions are also used to quired text token embeddings as prompts, which further enhances the embedding space and informs the LLM to comprehend the task contexts. As such, utilizing multi-modal data facilitates the repurpose of the existing LLM into time series forecasters without additional computational costs.

## 6 CONCLUSION

The rapid development of FMs has revolutionized the research fields in different domains. In this survey, we provide a comprehensive and updated review of FMs specifically designed for time series analysis. A novel taxonomy is proposed from a methodology-centric perspective by classifying FMs based on key components including model architecture, pre-training technique, adaptation technique, and data modality. Our survey facilitates understanding the underlying mechanism of applying the FMs to time series. Furthermore, we believe that the consolidation of the latest advancements, as well as the potential future direction (see Appendix), can inspire more innovative works within the field of time series analysis.

## ACKNOWLEDGEMENTS

## REFERENCES

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Syndar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. 2024. Chronos: Learning the Language of Time Series. *arXiv preprint arXiv:2403.07815* (2024).

[2] Muhammad Awais, Muzammal Naseer, Salman Khan, Rao Muhammad Anwer, Hisham Cholakkal, Mubarak Shah, Ming-Hsuan Yang, and Fahad Shahbaz Khan. 2023. Foundational Models Defining a New Era in Vision: A Survey and Outlook. *arXiv preprint arXiv:2307.13721* (2023).

[3] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* (2023), 1–6.

[4] Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619, 7970 (2023), 533–538.

[5] Marin Biloš, Kashif Rasul, Anderson Schneider, Yuriy Nevmyvaka, and Stephan Günnemann. 2022. Modeling temporal data as continuous functions with process diffusion. *arXiv preprint arXiv:2211.02590* (2022).

[6] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[7] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. 2024. Video generation models as world simulators. (2024). https://openai.com/research/video-generation-models-as-world-simulators

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[9] Salva Rühling Cachay, Bo Zhao, Hailey James, and Rose Yu. 2023. DYffusion: A Dynamics-informed Diffusion Model for Spatiotemporal Forecasting. *arXiv preprint arXiv:2306.01984* (2023).

[10] Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. 2023. TEMPO: Prompt-based Generative Pre-trained Transformer for Time Series Forecasting. *arXiv preprint arXiv:2310.04948* (2023).

[11] Ching Chang, Wen-Chih Peng, and Tien-Fu Chen. 2023. LLM4TS: Two-Stage Fine-Tuning for Time-Series Forecasting with Pre-Trained LLMs. *arXiv preprint arXiv:2308.08469* (2023).

[12] Yanchuan Chang, Jianzhong Qi, Yuxuan Liang, and Egemen Tanin. 2023. Contrastive Trajectory Similarity Learning with Dual-Feature Attention. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 2933–2945.

[13] Kang Chen, Tao Han, Junchao Gong, Lei Bai, Fenghua Ling, Jing-Jia Luo, Xi Chen, Leiming Ma, Tianning Zhang, Rui Su, et al. 2023. FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead. *arXiv preprint arXiv:2304.02948* (2023).

[14] Shengchao Chen, Guodong Long, Tao Shen, and Jing Jiang. 2023. Prompt Federated Learning for Weather Forecasting: Toward Foundation Models on Meteorological Data. In *International Joint Conference on Artificial Intelligence*.

[15] Shengchao Chen, Guodong Long, Tao Shen, Tianyi Zhou, and Jing Jiang. 2023. Spatial-temporal Prompt Learning for Federated Weather Forecasting. *arXiv preprint arXiv:2305.14244* (2023).

[16] Shengchao Chen, Guodong Long, Tao Shen, Tianyi Zhou, and Jing Jiang. 2023. Spatial-temporal Prompt Learning for Federated Weather Forecasting. arXiv:2305.14244 [cs.LG]

[17] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, Vol. 119. 1597–1607.

[18] Yakun Chen, Xianzhi Wang, and Guandong Xu. 2023. Gatgpt: A pre-trained large language model with graph attention network for spatiotemporal imputation. *arXiv preprint arXiv:2311.14332* (2023).

[19] Zihan Chen, Lei Nico Zheng, Cheng Lu, Jialu Yuan, and Di Zhu. 2023. ChatGPT Informed Graph Neural Network for Stock Movement Prediction. *arXiv preprint arXiv:2306.03763* (2023).

[20] Jinguo Cheng, Chunwei Yang, Wanlin Cai, Yuxuan Liang, and Yuankai Wu. 2024. NuwaTS: Mending Every Incomplete Time Series. *arXiv preprint arXiv:2405.15317* (2024).

[21] Chen Chu, Hengcai Zhang, Peixiao Wang, and Feng Lu. 2024. Simulating human mobility with a trajectory generation framework based on diffusion model. *International Journal of Geographical Information Science* (2024), 1–32.

[22] Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, and Svitlana Vyetrenko. 2024. On the constrained time-series generation problem. *Advances in Neural Information Processing Systems* 36 (2024).

[23] Jonathan Crabbé, Nicolas Huynh, Jan Stanczuk, and Mihaela van der Schaar. 2024. Time Series Diffusion in the Frequency Domain. *arXiv preprint arXiv:2402.05933* (2024).

[24] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. 2023. A decoder-only foundation model for time-series forecasting. *arXiv preprint arXiv:2310.10688* (2023).

[25] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*. 4171–4186.

[26] Jiaxiang Dong, Haixu Wu, Yuxuan Wang, Yunzhong Qiu, Li Zhang, Jianmin Wang, and Mingsheng Long. 2024. TimeSiam: A Pre-Training Framework for Siamese Time-Series Modeling. *arXiv preprint arXiv:2402.02475* (2024).

[27] Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, and Mingsheng Long. 2023. SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling. *Advances in Neural Information Processing Systems* (2023).

[28] Yuntao Du, Jindong Wang, Wenjie Feng, Sinno Pan, Tao Qin, Renjun Xu, and Chongjun Wang. 2021. Adarnn: Adaptive learning and forecasting of time series. In *Proceedings of the 30th ACM international conference on information & knowledge management*. 402–411.

[29] Wenying Duan, Liu Jiang, Ning Wang, and Hong Rao. 2019. Pre-Trained Bidirectional Temporal Representation for Crowd Flows Prediction in Regular Region. *IEEE Access* 7 (2019), 143855–143865.

[30] Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2023. TSMixer: Lightweight MLP-Mixer Model for Multivariate Time Series Forecasting. *arXiv preprint arXiv:2306.09364* (2023).

[31] Vijay Ekambaram, Arindam Jati, Nam H Nguyen, Pankaj Dayama, Chandra Reddy, Wesley M Gifford, and Jayant Kalagnanam. 2024. TTMs: Fast Multilevel Tiny Time Mixers for Improved Zero-shot and Few-shot Forecasting of Multivariate Time Series. *arXiv preprint arXiv:2401.03955* (2024).

[32] Cheng Feng, Long Huang, and Denis Krompass. 2024. Only the Curve Shape Matters: Training Foundation Models for Zero-Shot Multivariate Time Series Forecasting through Next Curve Shape Prediction. arXiv:2402.07570 [cs.LG]

[33] Tao-Yang Fu and Wang-Chien Lee. 2020. Trembr: Exploring Road Networks for Trajectory Representation Learning. *ACM Trans. Intell. Syst. Technol.* 11, 1 (2020), 10:1–10:25.

[34] Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. 2024. UniTS: Building a Unified Time Series Model. *arXiv preprint arXiv:2403.00131* (2024).

[35] Azul Garza and Max Mergenthaler-Canseco. 2023. TimeGPT-1. *arXiv preprint arXiv:2310.03589* (2023).

[36] Nate Gruver, Marc Finzi, Shikai Qiu, and Andrew Gordon Wilson. 2023. Large Language Models Are Zero-Shot Time Series Forecasters. *Advances in Neural Information Processing Systems* (2023).

[37] Albert Gu and Tri Dao. 2023. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv:2312.00752 [cs.LG]

[38] Anisha Gunjal and Greg Durrett. 2023. Drafting Event Schemas using Language Models. *arXiv preprint arXiv:2305.14847* (2023).

[39] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. 2021. Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting* 37, 1 (2021), 388–427.

[40] Haowen Hou and F Richard Yu. 2024. RWKV-TS: Beyond Traditional Recurrent Neural Network for Time Series Tasks. *arXiv preprint arXiv:2401.09093* (2024).

[41] Junfeng Hu, Xu Liu, Zhencheng Fan, Yuxuan Liang, and Roger Zimmermann. 2023. Towards Unifying Diffusion Models for Probabilistic Spatio-Temporal Graph Learning. arXiv:2310.17360 [cs.LG]

[42] Hongbin Huang, Minghua Chen, and Xiao Qiao. 2024. Generative Learning for Financial Time Series with Irregular and Scale-Invariant Patterns. In *The Twelfth International Conference on Learning Representations*. https://openreview.net/forum?id=CdjnzWsQax

[43] Jiawei Jiang, Dayan Pan, Houxing Ren, Xiaohan Jiang, Chao Li, and Jingyuan Wang. 2022. Self-supervised Trajectory Representation Learning with Temporal Regularities and Travel Semantics. *CoRR* abs/2211.09510 (2022).

[44] Lavender Yao Jiang, Xujin Chris Liu, Nima Pour Nejatian, Mustafa Nasir-Moin, Duo Wang, Anas Abidin, Kevin Eaton, Howard Antony Riina, Ilya Laufer, Paawan Punjabi, et al. 2023. Health system-scale language models are all-purpose prediction engines. *Nature* (2023), 1–6.

[45] Yushan Jiang, Zijie Pan, Xikun Zhang, Sahil Garg, Anderson Schneider, Yuriy Nevmyvaka, and Dongjin Song. 2024. Empowering Time Series Analysis with Large Language Models: A Survey. *arXiv preprint arXiv:2402.03182* (2024).

[46] Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I Webb, Irwin King, and Shirui Pan. 2023. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly

detection. *arXiv preprint arXiv:2307.03759* (2023).

[47] Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, et al. 2023. Time-LLM: Time series forecasting by reprogramming large language models. *arXiv preprint arXiv:2310.01728* (2023).

[48] Ming Jin, Qingsong Wen, Yuxuan Liang, Chaoli Zhang, Siqiao Xue, Xue Wang, James Zhang, Yi Wang, Haifeng Chen, Xiaoli Li, et al. 2023. Large models for time series and spatio-temporal data: A survey and outlook. *arXiv preprint arXiv:2310.10196* (2023).

[49] Ming Jin, Yifan Zhang, Wei Chen, Kexin Zhang, Yuxuan Liang, Bin Yang, Jindong Wang, Shirui Pan, and Qingsong Wen. 2024. Position: What Can Large Language Models Tell Us about Time Series Analysis. In *International Conference on Machine Learning (ICML'24)*.

[50] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. 2021. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.

[51] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026.

[52] Jun Li, Che Liu, Sibo Cheng, Rossella Arcucci, and Shenda Hong. 2023. Frozen Language Model Helps ECG Zero-Shot Learning. In *Medical Imaging with Deep Learning*.

[53] Rongfan Li, Ting Zhong, Xinke Jiang, Goce Trajcevski, Jin Wu, and Fan Zhou. 2022. Mining spatio-temporal relations via self-paced graph contrastive learning. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 936–944.

[54] Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. 2019. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems* 32 (2019).

[55] Yan Li, Xinjiang Lu, Yaqing Wang, and Dejing Dou. 2022. Generative time series forecasting with diffusion, denoise, and disentanglement. *Advances in Neural Information Processing Systems* 35 (2022), 23009–23022.

[56] Yan Lin, Jilin Hu, Shengnan Guo, Bin Yang, Christian S. Jensen, Youfang Lin, and Huaiyu Wan. 2024. GTM: General Trajectory Modeling with Auto-regressive Generation of Feature Domains. arXiv:2402.07232 [cs.LG]

[57] Yan Lin, Huaiyu Wan, Shengnan Guo, Jilin Hu, Christian S. Jensen, and Youfang Lin. 2023. Pre-training General Trajectory Embeddings with Maximum Multiview Entropy Coding. arXiv:2207.14539 [cs.CV]

[58] Chenxi Liu, Sun Yang, Qianxiong Xu, Zhishuai Li, Cheng Long, Ziyue Li, and Rui Zhao. 2024. Spatial-temporal large language model for traffic prediction. *arXiv preprint arXiv:2401.10134* (2024).

[59] Mingzhe Liu, Han Huang, Hao Feng, Leilei Sun, Bowen Du, and Yanjie Fu. 2023. PriSTI: A Conditional Diffusion Framework for Spatiotemporal Imputation. *arXiv preprint arXiv:2302.09746* (2023).

[60] Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. 2024. UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting. arXiv:2310.09751 [cs.LG]

[61] Xu Liu, Yuxuan Liang, Chao Huang, Yu Zheng, Bryan Hooi, and Roger Zimmermann. 2022. When do contrastive learning signals help spatio-temporal graph forecasting?. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. 1–12.

[62] Xin Liu, Daniel McDuff, Geza Kovacs, Isaac Galatzer-Levy, Jacob Sunshine, Jiening Zhan, Ming-Zher Poh, Shun Liao, Paolo Di Achille, and Shwetak Patel. 2023. Large Language Models are Few-Shot Health Learners. *arXiv preprint arXiv:2305.15525* (2023).

[63] Yong Liu, Guo Qin, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. 2024. AutoTimes: Autoregressive Time Series Forecasters via Large Language Models. *arXiv preprint arXiv:2402.02370* (2024).

[64] Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. 2024. Timer: Transformers for Time Series Analysis at Scale. *arXiv preprint arXiv:2402.02368* (2024).

[65] Xin Man, Chenghong Zhang, Changyu Li, and Jie Shao. 2023. W-MAE: Pretrained weather model with masked autoencoder for multi-variable weather forecasting. *arXiv preprint arXiv:2304.08754* (2023).

[66] John A. Miller, Mohammed Aldosari, Farah Saeed, Nasid Habib Barna, Subas Rana, I. Budak Arpinar, and Ninghao Liu. 2024. A Survey of Deep Learning and Foundation Models for Time Series Forecasting. arXiv:2401.13912 [cs.LG]

[67] Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. 2023. ClimaX: A foundation model for weather and climate. *International Conference on Machine Learning* (2023).

[68] Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730* (2022).

[69] Yilmazcan Ozyurt, Stefan Feuerriegel, and Ce Zhang. 2022. Contrastive learning for unsupervised domain adaptation of time series. *arXiv preprint arXiv:2206.06243* (2022).

[70] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, et al. 2022. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214* (2022).

[71] William Peebles and Saining Xie. 2023. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4195–4205.

[72] Bo Peng, Eric Alcaide, Quentin Anthony, Alon Albalak, Samuel Arcadinho, Huanqi Cao, Xin Cheng, Michael Chung, Matteo Grella, Kranthi Kiran GV, et al. 2023. Rwkv: Reinventing rnns for the transformer era. *arXiv preprint arXiv:2305.13048* (2023).

[73] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[74] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[75] Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Arian Khorasani, George Adamopoulos, Rishika Bhagwatkar, Marin Biloš, Hena Ghonia, Nadhir Vincent Hassen, Anderson Schneider, et al. 2023. Lag-llama: Towards foundation models for time series forecasting. *arXiv preprint arXiv:2310.08278* (2023).

[76] Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. 2021. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*. PMLR, 8857–8868.

[77] Yilong Ren, Yue Chen, Shuai Liu, Boyue Wang, Haiyang Yu, and Zhiyong Cui. 2024. TPLLM: A Traffic Prediction Framework Based on Pretrained Large Language Models. *arXiv preprint arXiv:2403.02221* (2024).

[78] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[79] Zezhi Shao, Zhao Zhang, Fei Wang, and Yongjun Xu. 2022. 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.

[80] Lifeng Shen and James Kwok. 2023. Non-autoregressive Conditional Diffusion Models for Time Series Prediction. *arXiv preprint arXiv:2306.05043* (2023).

[81] Xiaoming Shi, Siqiao Xue, Kangrui Wang, Fan Zhou, James Y Zhang, Jun Zhou, Chenhao Tan, and Hongyuan Mei. 2023. Language Models Can Improve Event Prediction by Few-Shot Abductive Reasoning. In *Advances in Neural Information Processing Systems*.

[82] Md Fahim Sikder, Resmi Ramachandranpillai, and Fredrik Heintz. 2023. Transfusion: generating long, high fidelity time series using diffusion models with transformers. *arXiv preprint arXiv:2307.12667* (2023).

[83] Chenxi Sun, Yaliang Li, Hongyan Li, and Shenda Hong. 2023. TEST: Text Prototype Aligned Embedding to Activate LLM's Ability for Time Series. *arXiv preprint arXiv:2308.08241* (2023).

[84] Peiwang Tang and Xianchao Zhang. 2022. MTSMAE: Masked Autoencoders for Multivariate Time-Series Forecasting. In *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 982–989.

[85] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*. 5998–6008.

[86] Jun Wang, Wenjie Du, Wei Cao, Keli Zhang, Wenjia Wang, Yuxuan Liang, and Qingsong Wen. 2024. Deep Learning for Multivariate Time Series Imputation: A Survey. *arXiv preprint arXiv:2402.04059* (2024).

[87] Xinglei Wang, Meng Fang, Zichao Zeng, and Tao Cheng. 2023. Where Would I Go Next? Large Language Models as Human Mobility Predictors. arXiv:2308.15197 [cs.AI]

[88] Xuhong Wang, Ding Wang, Liang Chen, and Yilun Lin. 2023. Building Transportation Foundation Model via Generative Graph Transformer. *arXiv preprint arXiv:2305.14826* (2023).

[89] Xu Wang, Hongbo Zhang, Pengkun Wang, Yudong Zhang, Binwu Wang, Zhengyang Zhou, and Yang Wang. 2023. An observed value consistent diffusion model for imputing missing values in multivariate time series. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2409–2418.

[90] Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Yunzhong Qiu, Haoran Zhang, Jianmin Wang, and Mingsheng Long. 2024. TimeXer: Empowering Transformers for Time Series Forecasting with Exogenous Variables. *arXiv preprint arXiv:2402.19072* (2024).

[91] Zepu Wang, Yuqi Nie, Peng Sun, Nam H Nguyen, John Mulvey, and H Vincent Poor. 2023. St-mlp: A cascaded spatio-temporal linear framework with channel-independence strategy for traffic forecasting. *arXiv preprint arXiv:2308.07496* (2023).

[92] Zhixian Wang, Qingsong Wen, Chaoli Zhang, Liang Sun, and Yi Wang. 2023. DiffLoad: uncertainty quantification in load forecasting with diffusion model. *arXiv preprint arXiv:2306.01001* (2023).

[93] Haomin Wen, Youfang Lin, Yutong Xia, Huaiyu Wan, Qingsong Wen, Roger Zimmermann, and Yuxuan Liang. 2023. DiffSTG: Probabilistic spatio-temporal graph forecasting with denoising diffusion models. In *the 31st ACM International Conference on Advances in Geographic Information Systems*. 1–12.

[94] Qingsong Wen, Linxiao Yang, Tian Zhou, and Liang Sun. 2022. Robust time series analysis and applications: An industrial perspective. In *28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4836–4837.

[95] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. 2023. Transformers in time series: A survey. In *International Joint Conference on Artificial Intelligence(IJCAI)*. 6778–6786.

[96] Christopher Wimmer and Navid Rekabsaz. 2023. Leveraging vision-language models for granular market change prediction. *arXiv preprint arXiv:2301.10166* (2023).

[97] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. 2024. Unified training of universal time series forecasting transformers. *arXiv preprint arXiv:2402.02592* (2024).

[98] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The eleventh international conference on learning representations*.

[99] Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, and Roger Zimmermann. 2024. Deciphering spatio-temporal graph forecasting: A causal lens and treatment. *Advances in Neural Information Processing Systems* 36 (2024).

[100] Qianqian Xie, Weiguang Han, Yanzhao Lai, Min Peng, and Jimin Huang. 2023. The Wall Street Neophyte: A Zero-Shot Analysis of ChatGPT Over MultiModal Stock Movement Prediction Challenges. *arXiv preprint arXiv:2304.05351* (2023).

[101] Jiehui Xu, Haixu Wu, Jianmin Wang, and Mingsheng Long. 2021. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy. In *International Conference on Learning Representations*.

[102] Hao Xue and Flora D Salim. 2022. PromptCast: A New Prompt-based Learning Paradigm for Time Series Forecasting. *arXiv preprint arXiv:2210.08964* (2022).

[103] Hao Xue, Bhanu Prakash Voutharoja, and Flora D Salim. 2022. Leveraging language foundation models for human mobility forecasting. In *the 30th International Conference on Advances in Geographic Information Systems*. 1–9.

[104] Tijin Yan, Hongwei Zhang, Tong Zhou, Yufeng Zhan, and Yuanqing Xia. 2021. ScoreGrad: Multivariate probabilistic time series forecasting with continuous energy-based generative models. *arXiv preprint arXiv:2106.10121* (2021).

[105] Chao-Han Huck Yang, Yun-Yun Tsai, and Pin-Yu Chen. 2021. Voice2series: Reprogramming acoustic models for time series classification. In *International conference on machine learning*. PMLR, 11808–11819.

[106] Xi Yang, Aokun Chen, Nima PourNejatian, Hoo Chang Shin, Kaleb E Smith, Christopher Parisien, Colin Compas, Cheryl Martin, Anthony B Costa, Mona G Flores, et al. 2022. A large language model for electronic health records. *NPJ Digital Medicine* 5, 1 (2022), 194.

[107] Chin-Chia Michael Yeh, Xin Dai, Huiyuan Chen, Yan Zheng, Yujie Fan, Audrey Der, Vivian Lai, Zhongfang Zhuang, Junpeng Wang, Liang Wang, et al. [n. d.]. Toward a foundation model for time series data. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*.

[108] Xinli Yu, Zheng Chen, Yuan Ling, Shujing Dong, Zongyi Liu, and Yanbin Lu. 2023. Temporal Data Meets LLM–Explainable Financial Time Series Forecasting. *arXiv preprint arXiv:2306.11025* (2023).

[109] Yuan Yuan, Jingtao Ding, Jie Feng, Depeng Jin, and Yong Li. 2024. UniST: A Prompt-Empowered Universal Model for Urban Spatio-Temporal Prediction. arXiv:2402.11838 [cs.LG]

[110] Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. 2023. Spatio-temporal Diffusion Point Processes. *arXiv preprint arXiv:2305.12403* (2023).

[111] Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. 2022. Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8980–8987.

[112] Taeyoung Yun, Haewon Jung, and Jiwoo Son. 2023. Imputation as Inpainting: Diffusion models for SpatioTemporal Data Imputation. (2023).

[113] Kexin Zhang, Qingsong Wen, Chaoli Zhang, Rongyao Cai, Ming Jin, Yong Liu, James Zhang, Yuxuan Liang, Guansong Pang, Dongjin Song, et al. 2023. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *arXiv preprint arXiv:2306.10125* (2023).

[114] Xiyuan Zhang, Ranak Roy Chowdhury, Rajesh K. Gupta, and Jingbo Shang. 2024. Large Language Models for Time Series: A Survey. arXiv:2402.01801 [cs.LG]

[115] Xiang Zhang, Ziyuan Zhao, Theodoros Tsiligkaridis, and Marinka Zitnik. [n. d.]. Self-supervised contrastive pre-training for time series via time-frequency consistency. *Advances in Neural Information Processing Systems* 35 ([n. d.]).

[116] Yingying Zhang, Zhengxiong Guan, Huajie Qian, Leili Xu, Hengbo Liu, Qingsong Wen, Liang Sun, Junwei Jiang, Lunting Fan, and Min Ke. 2021. CloudRCA: A root cause analysis framework for cloud computing platforms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*.

[117] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023).

[118] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35. 11106–11115.

[119] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*. 27268–27286.

[120] Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. 2023. One Fits All: Power General Time Series Analysis by Pretrained LM. *Advances in Neural Information Processing Systems* (2023).

[121] Zhengyang Zhou, Qihe Huang, Kuo Yang, Kun Wang, Xu Wang, Yudong Zhang, Yuxuan Liang, and Yang Wang. 2023. Maintaining the Status Quo: Capturing Invariant Relations for OOD Spatiotemporal Learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*. 3603–3614.

[122] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. 2024. Difftraj: Generating gps trajectory with diffusion probabilistic model. *Advances in Neural Information Processing Systems* 36 (2024).

[123] Zhaoyang Zhu, Weiqi Chen, Rui Xia, Tian Zhou, Peisong Niu, Bingqing Peng, Wenwei Wang, Hengbo Liu, Ziqing Ma, Xinyue Gu, et al. 2023. Energy forecasting with robust, flexible, and explainable machine learning algorithms. *AI Magazine* 44, 4 (2023), 377–393.

## A APPENDIX

In this section, we discuss the future research directions and opportunities of TSFMs from the methodology perspective.

**Incooporating Multi-modalities.** As illustrated in this survey, a majority of current foundation models for time series are developed based on a single modality. However, many real-world dynamic systems are coupled with various modalities (time series, text, even image data). It would be a promising direction to leverage various modalities along with the time series in TSFM to learn more comprehensive and generalized knowledge, therefore significantly boosting the performance of different downstream tasks.

**Exploring more Efficient Architectures.** Currently, the Transformer serves as the dominant architecture for building the foundation model. Though promising, Transformer-based foundation models have quadratic scaling with respect to the sequence length due to their self-attention mechanism. This makes them computationally expensive and memory-intensive for processing long sequences. Therefore, it is an interesting avenue for future study to explore more efficient FM backbone architectures, such as state-space models Mamba [37].

**Developing more Effective Pipelines.** Time series data has unique properties such as temporal distribution shift [28, 91, 121] (i.e., the data distribution will evolve over time) and causality (i.e., casual relationship can exist between different points in the time series) [99]. Therefore, it would be another existing as well as challenging future direction to develop TSFMs that can well address the temporal distribution shift or have a powerful Interpretability for downstream tasks.

**Protecting Privacy.** Protecting privacy is an essential concern when training foundation models on diverse sources and modalities of data, which raises potential risks of exposing sensitive information. As such, one future direction is the development of robust privacy-preserving techniques for training the TSFM from multi-source datasets, as well as keeping the utility of the trained FMs. This may include the advancement of federated learning approaches, where models can be trained across multiple decentralized devices or servers without exchanging raw data.

**Table 2: Summary of representative foundation models tailored for time series data modeling. N/A means not not applicable**

| | Category | Method | Modaility | Pre-training | Adaptation | Domain | Year |
|---|---|---|---|---|---|---|---|
| **Standard Time Series** | Transformer-based | Time-LLM [47] | Multi-Modality | Pretrained LLM | Prompt-engineering & Tokenization | General | 2023 |
| | | OFA [120] | Single-Modality | Pretrained LLM | Fine-tuning | General | 2023 |
| | | PromptCast [102] | Multi-Modality | Pretrained LLM | Prompt-engineering | General | 2022 |
| | | TEST [83] | Multi-Modality | Contrastive | Prompt-engineering & Tokenization | General | 2023 |
| | | LLM4TS [11] | Single-Modality | Pretrained LLM | Fine-tuning | General | 2023 |
| | | TEMPO [10] | Single-Modality | Pretrained LLM | Fine-tuning & Prompt-engineering | General | 2023 |
| | | LLMTime [36] | Single-Modality | Pretrained LLM | Zero-shot | General | 2023 |
| | | Yu *et al.* [108] | Multi-Modality | Pretrained LLM | Zero-shot | Finance | 2023 |
| | | Chen *et al.* [19] | Multi-Modality | Pretrained LLM | Zero-shot | Finance | 2023 |
| | | Xie *et al.* [100] | Multi-Modality | Pretrained LLM | Zero-shot | Finance | 2023 |
| | | Wimmer *et al.* [96] | Single-Modality | Pretrained VLM | Fine-tuning | Finance | 2023 |
| | | Liu *et al.* [62] | Multi-Modality | Pretrained LLM | Prompt-engineering | Healthcare | 2023 |
| | | METS [52] | Multi-Modality | Contrastive | Zero-shot | Healthcare | 2023 |
| | | Voice2Series [105] | Single-Modality | Pretrained AM | Tokenization | General | 2021 |
| | | PatchTST [68] | Single-Modality | Generative | Fine-tuning | General | 2022 |
| | | Moirai [97] | Single-Modality | Generative | Fine-tuning | General | 2024 |
| | | Timer [64] | Single-Modality | Generative | Fine-tuning | General | 2024 |
| | | TimeSiam [26] | Single-Modality | Generative | Fine-tuning | General | 2024 |
| | | TimeXer [90] | Single-Modality | Fully-supervised | - | General | 2024 |
| | | AutoTimes [63] | Multi-Modality | Pretrained LLM | Prompt-engineering & Tokenization | General | 2024 |
| | | Lag-Llama [75] | Single-Modality | Generative | Zero-shot & Fine-tuning | General | 2023 |
| | | Das *et al.* [24] | Single-Modality | Generative | Zero-shot | General | 2023 |
| | | TimeGPT-1 [35] | Single-Modality | Generative | Zero-shot & Fine-tuning | General | 2023 |
| | | UniTS [34] | Single-Modality | Fully-supervised & Generative | Zero-shot & Prompt-engineering | General | 2024 |
| | | Chronos [1] | Single-Modality | Generative | Zero-shot | General | 2024 |
| | | SimMTM [27] | Single-Modality | Hybrid | Fine-tuning | General | 2023 |
| | | MTSMAE [84] | Single-Modality | Generative | Fine-tuning | General | 2022 |
| | | TimeCLR [107] | Single-Modality | Contrastive | Fine-tuning | General | 2023 |
| | | UniTime [60] | Single-Modality | Pretrained LLM | Fine-tuning | General | 2024 |
| | | GTT [32] | Single-Modality | Fully-supervised | Zero-shot | General | 2024 |
| | Non-Transformer-based | TF-C [115] | Single-Modality | Contrastive | Fine-tuning | General | 2022 |
| | | CLUDA [69] | Single-Modality | Contrastive | Fine-tuning | General | 2022 |
| | | TS2Vec [111] | Single-Modality | Contrastive | Fine-tuning | General | 2021 |
| | | TimesNet [98] | Single-Modality | Fully-supervised | - | General | 2022 |
| | | TSMixer [30] | Single-Modality | Generative | Fine-tuning | General | 2023 |
| | | TTMs [31] | Single-Modality | Fully-supervised | Zero-shot & Fine-tuning & Prompt-engineering | General | 2024 |
| | | RWKV-TS [40] | Single-Modality | Fully-supervised | - | General | 2024 |
| | Diffsuion-based | TimeGrad [76] | Single-Modality | Generative | N/A | General | 2021 |
| | | D$^3$VAE [55] | Single-Modality | Generative | N/A | General | 2022 |
| | | TransFusion [82] | Single-Modality | Generative | N/A | General | 2023 |
| | | ScoreGrad [104] | Single-Modality | Generative | N/A | General | 2021 |
| | | Biloš et al. [5] | Single-Modality | Generative | N/A | General | 2022 |
| | | Crabbé et al. [23] | Single-Modality | Generative | N/A | General | 2024 |
| | | TimeDiff [80] | Single-Modality | Generative | N/A | General | 2023 |
| | | Wang et al. [89] | Single-Modality | Generative | N/A | General | 2023 |
| | | DiffTime [22] | Single-Modality | Generative | N/A | General | 2024 |
| | | FTS-Diffusion [42] | Single-Modality | Generative | N/A | Finance | 2023 |
| | | DiffLoad [92] | Single-Modality | Generative | N/A | Power | 2023 |
| **Spatial Time Series** | Transformer-based | CPPBTR [29] | Single-Modality | Fully-supervised | Fine-tuning | Transportation | 2019 |
| | | TFM [88] (graph TF) | Single-Modality | Fully-supervised | - | Transportation | 2023 |
| | | STEP [79] | Single-Modality | Generative | Fine-tuning | Transportation | 2022 |
| | | ST-LLM [58] | Single-Modality | Pretrained LLM | Fine-tuning | Transportation | 2024 |
| | | FourCastNet [70] | Single-Modality | Fully-supervised | Fine-tuning | Climate | 2022 |
| | | MetePFL [14] | Single-Modality | Generative | Federated Learning & Prompt-Learning | Climate | 2023 |
| | | FedWing [15] | Single-Modality | Fully-supervised | Federated Learning & Prompt-Learning | Climate | 2023 |
| | | ClimaX [67] | Single-Modality | Fully-supervised | Fine-tuning | Climate | 2023 |
| | | FengWu [13] | Single-Modality | Generative | Zero-shot | Climate | 2023 |
| | | Pangu-Weather [3] | Single-Modality | Fully-supervised | Zero-shot | Climate | 2023 |
| | | W-MAE [65] | Single-Modality | Generative | Fine-tuning | Climate | 2023 |
| | | GATGPT [18] | Single-Modality | Pretrained LLM | Fine-tuning | General | 2023 |
| | | TPLLM [77] | Single-Modality | Pretrained LLM | Fine-tuning | Transportation | 2024 |
| | | UniST [109] | Single-Modality | Generative | Fine-tuning | General | 2024 |
| | Non-Tranformer-based | SPGCL [53] | Single-Modality | Contrastive | - | General | 2022 |
| | | STGCL [61] | Single-Modality | Contrastive | Fine-tuning | General | 2021 |
| | Diffusion-based | DiffSTG [93] | Single-Modality | Generative | N/A | General | 2023 |
| | | DSTPP [110] | Single-Modality | Generative | N/A | General | 2023 |
| | | DYffusion [9] | Single-Modality | Generative | N/A | General | 2023 |
| | | Yun et al. [112] | Singe-Modality | Generative | N/A | General | 2023 |
| | | USTD [41] | Single-Modality | Generative | N/A | General | 2023 |
| | | PriSTI [59] | Single-Modality | Generative | N/A | General | 2023 |
| **Others** | Transformer-based | AuxMobLCast [103] | Single-Modality | Pretrained LLM | Fine-tuning & Prompt-engineering | Mobility | 2022 |
| | | LLM-Mob [87] | Single-Modality | Pretrained LLM | Prompt-engineering | Mobility | 2023 |
| | | TrajCL [12] | Single-Modality | Contrastive | Fine-tuning | Mobility | 2023 |
| | | GTM [56] | Single-Modality | Generative | Zero-shot | Mobility | 2024 |
| | | LAMP [81] | Singe-Modality | Pretrained LLM | Prompt-engineering | Event | 2023 |
| | | Gunjal & Durrett [38] | Singe-Modality | Pretrained LLM | Prompt-engineering | Event | 2023 |
| | | NYUTron [44] | Singe-Modality | Generative | Fine-tuning | Event(Healthcare) | 2023 |
| | | GatorTron [106] | Singe-Modality | Generative | Fine-tuning | Event(Healthcare ) | 2022 |
| | Non-Transformer-based | MMTEC [57] | Single-Modality | Contrastive | Fine-tuning | Mobility | 2022 |
| | | START [43] | Single-Modality | Hybrid | Fine-tuning | Mobility | 2022 |
| | | Trembr [33] | Single-Modality | Generative | Fine-tuning | Mobility | 2020 |
| | Diffusion-based | TrajGDM [21] | Singe-Modality | Generative | N/A | Mobility | 2024 |
| | | DiffTraj [122] | Singe-Modality | Generative | N/A | Mobility | 2024 |