A Survey of Post-Training in Time Series Foundation Models

Anonymous authors
Paper under double-blind review

Abstract

Time series foundation models (TSFMs) have emerged as general-purpose models for time series analysis, but pretraining alone is often insufficient for reliable downstream deployment. Bridging this gap requires further intervention to handle domain shift, task heterogeneity, limited supervision, and computational constraints, which motivates post-training as a broad class of methods to adapt, augment, compose, calibrate, or specialize pretrained TSFMs for downstream tasks. In this survey, we organize TSFM post-training methods by their locus of intervention in the prediction pipeline, yielding five categories: parameter adaptation, context augmentation, model composition, output processing and uncertainty control, and compression and specialization. For each category, we summarize representative methods, trace their development, and discuss their limitations. We further identify future directions toward controlled adaptation, reliable context construction, uncertainty-aware model composition, calibrated output processing, and deployment-aware specialization. Overall, by providing a comprehensive view of the emerging TSFM post-training landscape, this survey aims to support future research to navigate the design space between a pretrained TSFM and its reliable downstream deployment.

1 Introduction

Time series data record how real-world systems evolve, capturing operationally critical quantities such as prices, demand, physiological measurements, energy load, weather variables, industrial sensor readings, and computing-resource usage. They arise across finance (Lai et al., 2018b; Sonkavde et al., 2023; Tsay, 2005, 2014), healthcare (Čepulionis & Lukoševičiūtė, 2016; Chui et al., 2017; Penfold & Zhang, 2013), economics (Nerlove et al., 2014), electricity consumption (Qin et al., 2023; Trindade, 2015; Zhu et al., 2023), weather forecasting (Chen et al., 2023; Max Planck Institute, 2021; Pathak et al., 2022), supply management and retail (Böse et al., 2017; Li et al., 2022), industrial processes, observability, and monitoring (Cohen et al., 2026; Liu et al., 2017), and cloud computing (Yang et al., 2023; Zhang et al., 2021). The typical tasks considered include forecasting, classification, anomaly detection, imputation, representation learning, and event-oriented analysis such as segmentation and change-point detection.

From specialized methods ... A plethora of methods have been developed in the past decades to solve these tasks, from mathematical methods (Chen & Tao, 2021; Sorjamaa et al., 2007) and statistical approaches such as AutoRegressive Integrated Moving Average (Box et al., 1974; Box & Jenkins, 1990; Stevenson, 2007, ARIMA) and seasonal naive (Hyndman & Athanasopoulos, 2021), to machine learning algorithms such as Support Vector Machines (Hearst et al., 1998, SVMs) and Gradient Boosting Machines (Friedman, 2001, GBMs). Before the recent shift toward foundation models, most deep learning approaches to time series were developed as task-specific architectures trained to solve a particular task (Casolaro et al., 2023), ranging from fully connected, recurrent and convolutional neural networks (Fan et al., 2019; Lai et al., 2018a; Oreshkin et al., 2020; Rangapuram et al., 2018; Salinas et al., 2020; Sen et al., 2019; Smyl, 2019) to transformer-based models (Ilbert et al., 2024; Liu et al., 2024b; Nie et al., 2023; Woo et al., 2024; Wu et al., 2021; Zhou et al., 2021, 2022). Much of this progress has come from training global models over large collections of time series, rather than fitting separate models to individual series. By exploiting shared structure across series, this cross-series learning paradigm has contributed to the success of top-performing methods in forecasting

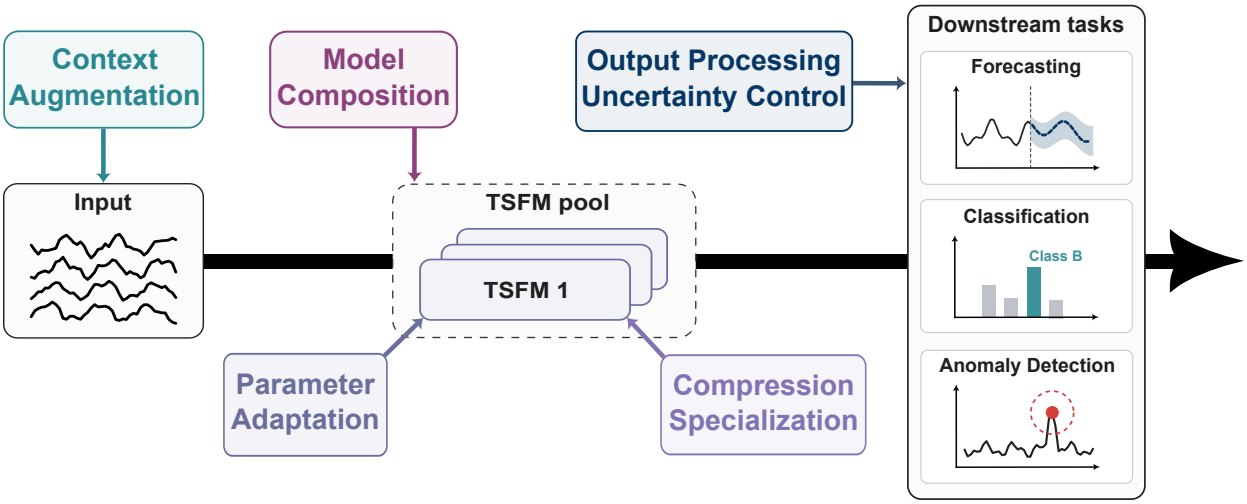

Figure 1: **Families of TSFMs post-training methods.** Many post-training methods transfer across tasks, domains, and pretrained architectures. As such, their main distinguishing feature is their locus of intervention in the prediction pipeline. This leads to the 5 families of post-training methods studied in this survey, which can act at the input context level, on the model parameters, or on the model outputs.

competitions such as M4 and M5 (Anderer & Li, 2022; Smyl, 2019), helping establish deep learning as a dominant paradigm for time series applications.

... **To foundation models.** Recently, the time series field has undergone the same paradigm shift as natural language processing (Devlin et al., 2019; Radford et al., 2018, 2019; Raffel et al., 2020; Touvron et al., 2023) and computer vision (Carion et al., 2026; Caron et al., 2021; Dosovitskiy et al., 2021; He et al., 2022; Kirillov et al., 2023; Oquab et al., 2024; Radford et al., 2021; Ravi et al., 2025; Siméoni et al., 2025), moving from models trained on task-specific data towards pretraining on a large-scale corpora of diverse time series. Early transfer-learning work in time series showed that a model trained on source collections with a given sampling frequency could generalize directly to unseen target series without target-side adaptation (Oreshkin et al., 2021). The field then moved from frequency-specific transfer learning toward foundation-style pretraining, in which large sequence models are trained across diverse time series to support broad zero-shot forecasting. We refer to this emerging class of pretrained forecasting models as *Time Series Foundation Models* (TSFMs). TimeGPT (Garza & Mergenthaler-Canseco, 2023) was an early example of this direction, showing how pretrained transformer architectures could be adapted to forecasting. We note that several specific challenges occur when developing TSFMs, especially when it comes to building pretraining databases. Indeed, time series data are inherently noisy and often require expert knowledge to ensure their quality. This limits the open-source availability of high-quality data equivalent to ImageNet (Deng et al., 2009) for computer vision and C4 (Dodge et al., 2021; Raffel et al., 2020) or FineWeb (Penedo et al., 2024) for LLMs. A lot of effort has been put into alleviating these issues, both from industry and academia. As such, many TSFMs' recipes were proposed over the years (Ansari et al., 2024, 2025; Arango et al., 2025; Cohen et al., 2026; Das et al., 2024; Ekambaram et al., 2024; Feofanov et al., 2025; Liu et al., 2026a, 2025d; Woo et al., 2024), with a noticeable convergence towards the transformer architecture (Vaswani et al., 2017). Methods to build pretraining databases were also proposed, including synthetic data (Dooley et al., 2023; Oreshkin et al., 2026; Taga et al., 2024; Xie et al., 2026a), and benchmarks and evaluations were proposed in tandem to measure the improvement of those models (Aksu et al., 2024; Bagnall et al., 2018; Cohen et al., 2026; Dau et al., 2019; Erickson et al., 2020; Shchur et al., 2023, 2026).

The advent of post-training. Despite the impressive capabilities of foundation models, pretraining is rarely the final step for practical deployment. For instance, base large language models (LLMs) acquire

general knowledge during pretraining (Hoffmann et al., 2022; Radford et al., 2019; Zhou et al., 2023) and become helpful chatbots after post-training (Lambert, 2026; Ouyang et al., 2022), where they can be given reasoning abilities and be aligned with human intent (Guo et al., 2025; Kamath et al., 2025; Touvron et al., 2023; Yang et al., 2025). In the context of TSFMs, post-training serves three main purposes. The first one has to do with task alignment. While pretraining enables learning general and reusable temporal knowledge, most downstream tasks have specific characteristics, such as the granularity, seasonality, feature correlations, the number of classes, or the forecasting horizon (Das et al., 2024; Ekambaram et al., 2024; Woo et al., 2024). Second, efficiently solving the task requires specialization, which can be achieved during post-training by adapting the model to the specific constraints at hand. Last but not least, TSFMs can benefit from strategies inspired by LLMs such as Retrieval-Augmented Memory (Lee et al., 2026a; Lewis et al., 2020; Liang et al., 2026b, RAG), or in-context learning (Brown et al., 2020; Faw et al., 2025, ICL). Post-training methods help boost the performance in these areas, improving the generalization to unseen distributions. While post-training turns pretrained LLMs into aligned chatbots, the post-training of TSFMs transforms them from universal temporal learners to specialized and reliable experts.

Prior surveys. Surveys are important to help organize the field and give an overview of the current trends, limitations, and open problems. Recent surveys on post-training methods provide valuable insights but focus on large language models (Kumar et al., 2025; Lambert, 2026; Tie et al., 2025) and vision-language models (Lambert, 2026; Tie et al., 2025). With the increase of TSFMs post-training approaches and given the important differences in terms of data, end-goals, and challenges, a specific treatment is needed for TSFMs. Many surveys on TSFMs focus on the task of forecasting (Casolaro et al., 2023), which is prevalent in most recent papers (Ansari et al., 2025; Cohen et al., 2026; Liu et al., 2026a; Woo et al., 2024), on the pretraining phase (Kottapalli et al., 2025; Liang et al., 2024), on the role of synthetic data (Liu et al., 2025c), or on the emergence of multimodal time series analysis (Jiang et al., 2025). Yet, a clear focus on post-training methods remains elusive. Close to our work is Liu et al. (2026b) that describes the evolution from the pretraining to post-training in TSFMs. However, the analysis remains limited to finetuning approaches, thus overlooking other important post-training components. In contrast, this survey provides a comprehensive overview of post-training approaches for TSFMs, categorizing them into the 5 families of Fig. 1, according to their locus of intervention in the prediction pipeline. We leverage this perspective to provide the structured taxonomy displayed in Fig. 2. Our tailored approach allows us to go beyond forecasting and to discuss the current trends, limitations, and open problems in the literature to help improve.

Contributions. This survey aims to give readers a clear understanding of the development and evolution of post-training in TSFMs, along with the limitations and areas of improvement. Our key contributions can be summarized as follows:

1. **Comprehensive overview.** We organized a survey of TSFM post-training methods, covering both peer-reviewed publications and preprints available across major publication venues and preprint platforms.
2. **Structured taxonomy.** We provide a taxonomy of TSFMs post-training methods that clarifies where the intervention occurs in the prediction pipeline, offering insights into the evolution and limitations of current techniques.
3. **Vision for future research.** To help foster the development of next-generation methods, we identify critical open problems and propose actionable directions for future work.

Overview. The survey is organized as follows. In Section 2, we recall some background notions on TSFMs models, how they are pretrained, and how they are post-trained. The taxonomy of post-training methods is introduced in Section 3 and organized according to where the post-training interventions occur on the prediction pipeline. In the following sections, we discuss in detail the different families of methods, their development, and current limitations: the parameter adaptation methods are discussed in Section 4, the context augmentation methods are discussed in Section 5, the strategies of model composition are introduced in Section 6, the approaches of output processing and uncertainty control are presented in Section 7 and, finally, the compression and specialization methods are presented in Section 8. Finally, in Section 9, we synthesize the main limitations of current methods and outline future research directions.

2 Background

Time Series Foundation Models. Let $\mathbf{x}_{1:T}(\mathbf{x}_1, \ldots, \mathbf{x}_T) \in \mathbb{R}^{T \times C}$ denote a time series of length T with C channels. Depending on the downstream task, the target output $\mathbf{y} \in \mathcal{Y}$ can be a future trajectory, a class label, an imputed sequence, an anomaly score, or a predictive distribution(Kottapalli et al., 2025; Liang et al., 2024).

A time series foundation model is a pretrained model

$$f_{\theta_0} : \mathcal{X} \to \mathcal{Y},$$

where θ_0 denotes parameters learned from large-scale time series corpora. Unlike task-specific models trained for a single dataset, horizon, or application, TSFMs are designed to transfer temporal representations (Feofanov et al., 2025) or predictive capabilities (Ansari et al., 2024) across diverse downstream tasks. This positions them as a general-purpose modeling paradigm for time series analysis.

Pretraining of TSFMs. The goal of TSFM pretraining is to learn broadly reusable temporal knowledge from a pretraining distribution, which may consist of real-world time series (Gao et al., 2024; Garza et al., 2024), synthetic time series(Moroshan et al., 2026; Xie et al., 2026a), or a mixture of both. Given a pretraining corpus $\mathcal{D}^{\text{pre}}$, the model parameters are learned by optimizing a generic pretraining objective:

$$\theta_0 = \arg\min_{\theta} \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{\text{pre}}} \left[\mathcal{L}_{\text{pre}}(f_\theta, \mathbf{x}, \mathbf{y}) \right].$$

The pretraining objective may be based on forecasting (Woo et al., 2024), reconstruction (Goswami et al., 2024), masked modeling (Zhang et al., 2025a), contrastive learning (Lin et al., 2024), or sequence-to-sequence prediction (Rasul et al., 2023). Through this process, the model is expected to capture common temporal patterns such as trends, seasonality, periodicity, local dynamics, cross-channel dependencies, and distributional uncertainty.

After pretraining, the obtained model f_{θ_0} can be used as a general-purpose backbone for different time series tasks. However, the pretraining distribution is usually not identical to the target deployment distribution. Target domains may differ in temporal resolution, noise level, covariates, channel structure, forecasting horizon, label space, or nonstationary behavior.

Post-training of TSFMs. Although TSFMs are pretrained for broad transfer, practical deployment often requires adaptation to a target domain, task, or operational constraint (Jiang et al., 2025). The most familiar mechanism is finetuning, where a pretrained model is updated on target-domain data (Arango et al., 2025; Qiao et al., 2026). However, TSFM adaptation after pretraining extends beyond parameter updates, also involving context construction, retrieval, model composition, output refinement, uncertainty calibration, and deployment-oriented compression. We refer to post-training as any procedure applied after pretraining to improve a TSFM on a target domain $\mathcal{D}^{\text{tar}}$. Given a pretrained model f_{θ_0}, a target input $\mathbf{x}_{1:T}$, and optional adaptation resources $\mathcal{R}$. Post-training can be written in a general form as:

$$\hat{\mathbf{y}} = \mathcal{A}\left(f_{\theta_0}, \mathbf{x}_{1:T}, \mathcal{R}\right),$$

where $\mathcal{A}$ denotes the post-training mechanism. The resource set $\mathcal{R}$ may include labeled target-domain samples, unlabeled streams, calibration data, retrieved examples, auxiliary covariates, external memory, model pools, or deployment constraints.

Post-training is broader than conventional finetuning. It may update all or part of the model parameters (Qiao et al., 2026), add lightweight trainable modules (Arango et al., 2025), modify the input context (Qiu et al., 2026), retrieve relevant examples (Ning et al., 2026), combine multiple pretrained models (Shi et al., 2025), refine the output (Bolton et al., 2026), calibrate uncertainty (Chinta et al., 2026), or compress the model for efficient deployment (Li et al., 2026b). In this sense, post-training is not a single technique, but a family of adaptation strategies that improve the usefulness of pretrained TSFMs under target domain and deployment requirements. This survey studies TSFM post-training methods based on the locus of intervention in the prediction pipeline. The overall taxonomy is provided in the following section.

3 Taxonomy of TSFMs Post-Training Methods

We organize existing TSFM post-training methods by their locus of intervention in the inference pipeline, rather than by downstream task, model backbone, or application domain. Given a pretrained time series foundation model $f_{\theta_0} : \mathcal{X} \to \mathcal{Y}$, we define post-training as any procedure that improves its performance, robustness, calibration, or deployment efficiency on a target domain $\mathcal{D}_{\text{tar}}$ by modifying one of these components. Since many of these methods transfer across tasks, domains, and pretrained backbones, what principally distinguishes them is where they intervene. This motivates a taxonomy of five categories, described in Fig. 1: modifications in the model parameters, input context, model composition, output space, or computational structure. The overall taxonomy of the methods is provided in Fig. 2.

A. Parameter adaptation. Parameter adaptation intervenes directly in the model parameters of a pretrained TSFM. It either updates some or all of the original model parameters, as in full-parameter finetuning, continual pretraining, and test-time adaptation, or introduces additional trainable components while keeping most of the backbone frozen, as in parameter-efficient finetuning. This can be formally expressed as the following parameter update:

$$\theta^\star = \theta_0 + \Delta\theta,$$

where $\Delta\theta$ denotes the post-training update, ranging from full-model parameter changes to a small set of new trainable parameters.

B. Context augmentation. Context augmentation intervenes the input side of a TSFM with retrieved examples, external memory, support samples, or in-context augmentation. Instead of primarily changing the backbone parameters, these methods construct an additional context $\mathcal{C}_x$ for each input series:

$$\hat{\mathbf{y}}_{1:H} = f_{\theta_0}\big(\mathbf{x}_{1:T}, \mathcal{C}_x\big).$$

We further categorize the methods into retrieval augmentation, memory augmentation, and in-context augmentation.

C. Model composition. Model composition treats TSFM post-training as a model allocation problem, combining multiple pretrained or adapted models rather than relying on a single one. Given a pool of pretrained or adapted models: $\mathcal{M} = \{f_{\theta_1}, \ldots, f_{\theta_K}\}$, a router r_ϕ assigns each target input to the models based on the input series and, optionally, model metadata such as their training domains, supported frequencies, context lengths, horizons, validation performance, or uncertainty estimates. A common soft-routing formulation is

$$\boldsymbol{\alpha} = r_\phi(\mathbf{x}_{1:T}, \mathcal{M}) \in \Delta^{K-1}, \qquad \hat{\mathbf{y}}_{1:H} = \sum_{k=1}^{K} \alpha_k f_{\theta_k}(\mathbf{x}_{1:T}),$$

where $\Delta^{K-1} = \{\boldsymbol{\alpha} \in \mathbb{R}_+^K : \sum_{k=1}^K \alpha_k = 1\}$ is the probability simplex over the K models. The router parameters ϕ may be learned from validation or target-domain data, estimated from historical model performance, induced by retrieval or similarity scores, or specified by metadata-based rules using properties such as domain, frequency, horizon, context length, or uncertainty. This category includes model selection, ensembling, fusion, routing, and mixture-of-experts-style composition.

D. Output processing and uncertainty control. Output processing and uncertainty control intervene on the initial output of a pretrained TSFM $\tilde{\mathbf{y}}_{1:H} = f_{\theta_0}(\mathbf{x}_{1:T})$. An output-level post-processor g_ψ then maps the raw output, possibly together with the input context, to a refined prediction, calibrated interval, probabilistic sample, or anomaly score:

$$\hat{\mathbf{y}}_{1:H} = g_\psi\left(\tilde{\mathbf{y}}_{1:H}, \mathbf{x}_{1:T}\right).$$

The parameters $\psi \in \Psi$ can be learned from target-domain residuals, estimated on a calibration set, updated online from recent forecast errors, or set by distributional, conformal. Depending on the objective, ψ can include correction-model weights, residual quantiles, conformal thresholds, variance or scale parameters,

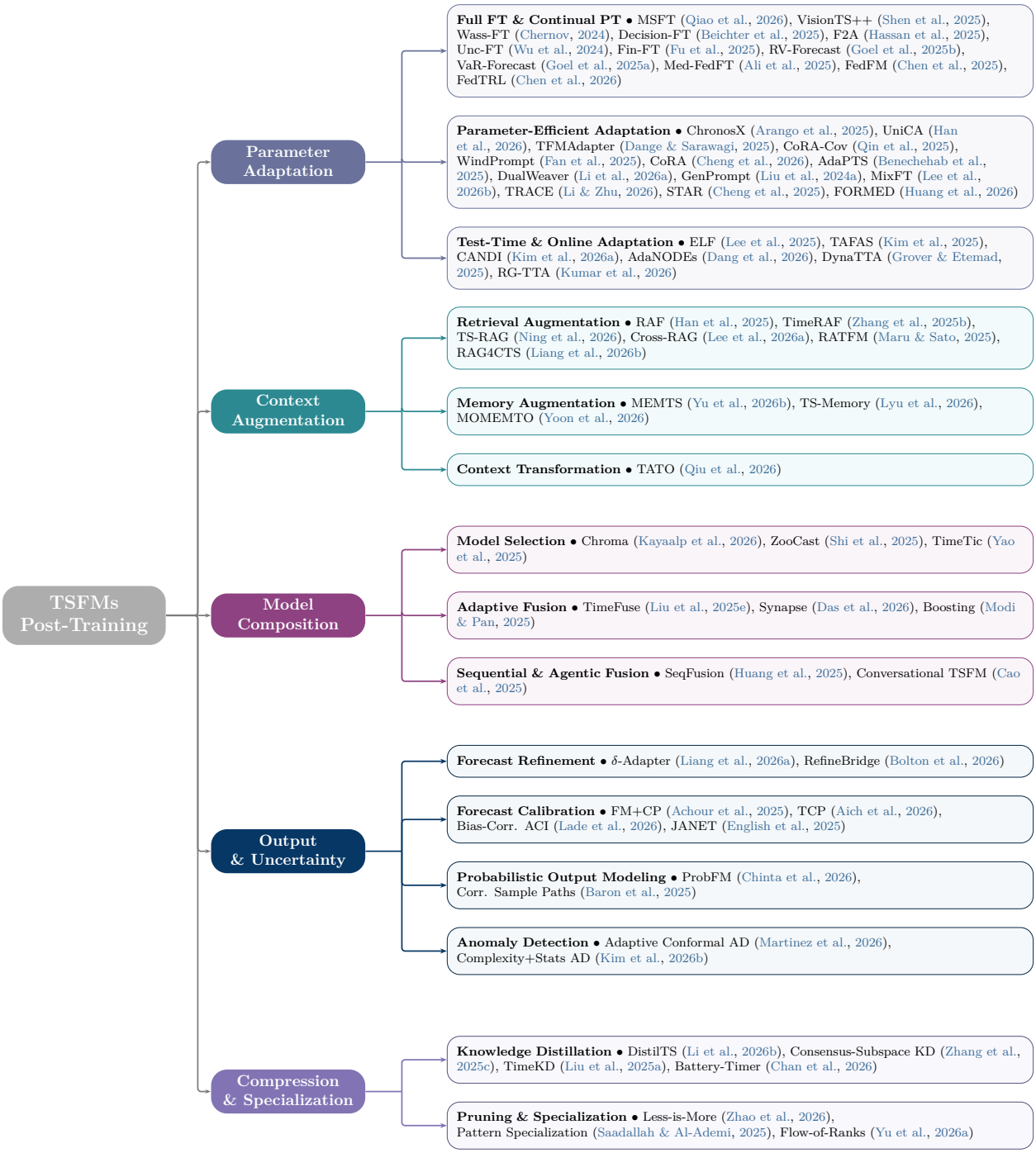

Figure 2: Taxonomy of post-training methods for time series foundation models. Methods are grouped by the locus of intervention: parameter space, input context, model composition, output processing, and computational structure. Each method is assigned to its primary subcategory to avoid duplication.

recalibration maps, or anomaly-score normalization constants. Thus, this category encompasses output refinement, conformal calibration, probabilistic recalibration, residual correction, and anomaly-score processing techniques.

E. Compression and specialization. Compression and specialization operate during the deployment of pretrained TSFM. The goal is to convert a broadly pretrained model f_{θ_0} into a cheaper, faster, or more targeted variant $f_{\bar{\theta}}$ for a specific deployment regime, such that:

$$\mathrm{Cost}(f_{\bar{\theta}}) < \mathrm{Cost}(f_{\theta_0}),$$

while maintaining competitive accuracy, calibration, and robustness on the target domain. Here, Cost can quantify parameter count, memory footprint, inference latency, context-length complexity, energy consumption, or serving cost. The specialized parameters $\bar{\theta}$ can be obtained through techniques such as distillation, pruning, quantization, low-rank approximation, architecture replacement, or targeted adaptation to a restricted set of domains, horizons, or tasks. This category includes both model compression for efficient deployment and specialization for constrained operational settings.

4 Parameter Adaptation Methods

Parameter adaptation modifies the pretrained model parameters or attaches new trainable components to a frozen backbone. We organize the methods into three subcategories according to the locus and timing of adaptation: full-parameter finetuning and continual pretraining, parameter-efficient finetuning, and test-time adaptation.

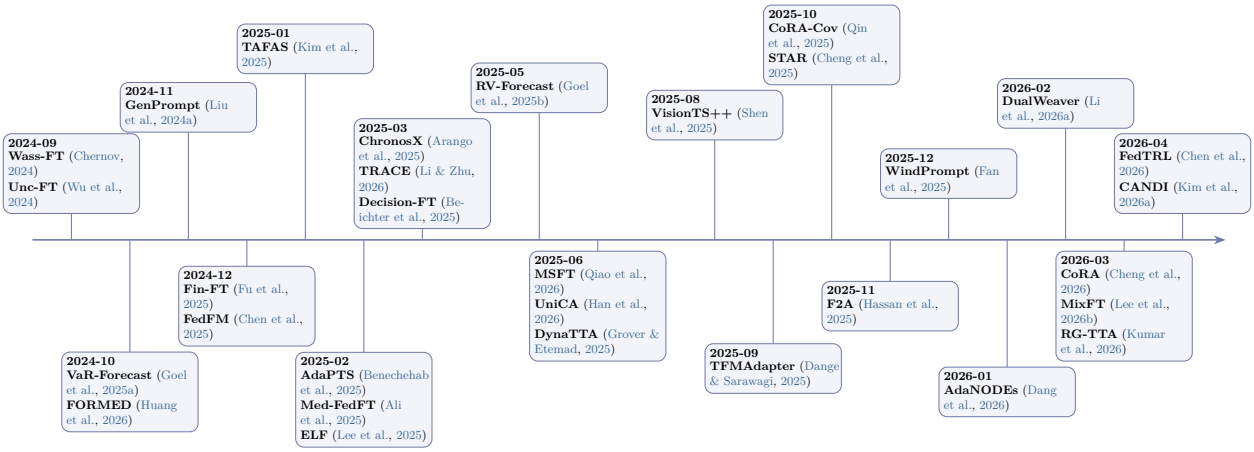

Figure 3: Timeline of parameter adaptation methods for post-training of TSFMs.

4.1 Full-parameter finetuning and Continual Pretraining

Full-parameter finetuning and continual pretraining adapt a pretrained TSFM by updating a large portion of its parameters. Full-parameter finetuning typically uses supervised target-domain objectives, whereas continual pretraining continues self-supervised or forecasting-oriented training on additional temporal data. Both approaches are more expensive than lightweight adaptation methods, but offer greater capacity for domain specialization. Methods in this category differ primarily in the adaptation objective, the data used for updating the model, and whether the goal is task-specific specialization or broader domain adaptation.

Multi-scale finetuning methods (Qiao et al., 2026) adapt TSFMs by exposing the model to supervision at multiple temporal resolutions, which is useful when short-term local patterns, medium-term seasonality, and long-term trends coexist. Other approaches tailor the finetuning objective to the target task. Wasserstein finetuning uses a distributional loss to improve predictive distribution matching (Chernov, 2024), while decision-focused finetuning optimizes forecasts with respect to downstream decision quality (Beichter et al., 2025). Task-specific finetuning has also been studied in anomaly-related settings, including Forecast2Anomaly (Hassan et al., 2025) and uncertainty-aware finetuning (Wu et al., 2024).

Full-parameter finetuning is also widely used for domain-specific adaptation. In finance, large time series models are finetuned or continually pretrained for market forecasting, realized volatility prediction, and value-at-risk estimation, where the target distribution is highly nonstationary and can differ substantially from general-purpose pretraining data (Fu et al., 2025; Goel et al., 2025a,b). When target-domain data cannot be centralized, federated post-training adapts TSFMs across decentralized clients without sharing raw data. Federated medical and heterogeneous time-series studies combine local updates with global aggregation to address privacy constraints, client heterogeneity, and cross-domain variation (Ali et al., 2025; Chen et al., 2025, 2026). Cross-modal continual pretraining methods, such as VisionTS++ (Shen et al., 2025), further show that backbones pretrained in other modalities can be adapted to time series by continuing training on large-scale temporal data.

Overall, full-parameter finetuning and continual pretraining offer the highest-capacity form of TSFM adaptation, allowing pretrained models to be reshaped for new domains, tasks, and data regimes. This flexibility is particularly valuable when the target distribution differs substantially from the pretraining data or when the adaptation objective departs from generic forecasting accuracy. At the same time, these methods are computationally expensive, data-dependent, and vulnerable to overfitting, catastrophic forgetting, and proliferation of domain-specific model copies. These trade-offs motivate parameter-efficient and modular post-training strategies that seek similar specialization with lower adaptation and deployment cost.

4.2 Parameter-Efficient Adaptation

Parameter-efficient adaptation provides a lower-cost alternative to full-parameter finetuning by specializing a pretrained TSFM while keeping most of the backbone fixed. These methods optimize only a small subset of existing parameters or introduce lightweight trainable components, reducing adaptation cost, memory overhead, and the need to maintain separate full model copies for different domains, tasks, or deployment settings. Such methods are particularly attractive when full-parameter finetuning is too expensive or when only limited target-domain data are available.

A major branch focuses on covariate and exogenous variable adaptation. Methods such as ChronosXs (Arango et al., 2025), UniCA (Han et al., 2026), TFMAdapter (Dange & Sarawagi, 2025), and covariate-aware CoRA (Qin et al., 2025) introduce trainable modules that inject past or future covariates into frozen TSFMs, enabling univariate or target-only models to exploit external information such as weather, events, categorical variables, and other side signals. The same principle is also applied in application-specific prompt tuning, where exogenous meteorological variables are encoded as time series prompts for wind power forecasting (Fan et al., 2025).

Another branch adapts univariate TSFMs to multivariate forecasting. Correlation-aware CoRA introduces lightweight adapters to model time-varying and time-invariant inter-channel dependencies (Cheng et al., 2026). AdaPTS (Benechehab et al., 2025) and DualWeaver (Li et al., 2026a) extend univariate backbones to probabilistic or multivariate forecasting by adding stochastic feature space adapters, surrogate series mechanisms, or feature fusion modules. Generalized prompt tuning follows a related principle in healthcare time series, using trainable prompts to integrate cross-channel information while keeping the backbone frozen (Liu et al., 2024a).

Beyond extending model capabilities in time series forecasting, parameter-efficient adaptation can also address data heterogeneity and task diversity. MixFT (Lee et al., 2026b) provides a data-centric strategy by partitioning heterogeneous time series into more homogeneous groups and adapting lightweight modules for each group. On the task side, TRACE studies general PEFT strategies across both forecasting and anomaly detection (Li & Zhu, 2026). STAR introduces state adapters for multivariate anomaly detection, where the meaning of an observation depends on the underlying system state (Cheng et al., 2025). FORMED repurposes a frozen TSFM backbone for medical time-series classification through lightweight task-specific adaptation (Huang et al., 2026).

Overall, parameter-efficient adaptation provides a modular way to extend frozen TSFMs to new covariates, multivariate structures, domains, and tasks without retraining the full backbone. Its main advantage is scalability: different target settings can be supported by lightweight modules while sharing the same pretrained

model. However, this modularity also shifts the design burden to the adapter architecture, placement, and training objective. The resulting performance depends on whether the frozen backbone contains reusable temporal representations and whether the added modules have sufficient capacity to capture exogenous effects, cross-channel dependencies, and task-specific structure. These trade-offs make parameter-efficient adaptation a practical but still underdeveloped direction for scalable TSFM post-training.

4.3 Test-Time and Online Adaptation

Test-Time and Online Adaptation updates a model component during deployment, allowing the model to respond to nonstationarity, regime shifts, and streaming feedback. ELF trains lightweight forecasters and weighting modules online while keeping the TSFM backbone fixed (Lee et al., 2025). Some methods focus on lightweight online updating. TAFAS studies nonstationary forecasting with test-time adaptation and uses gated calibration to preserve useful source-model information while adapting to the target stream (Kim et al., 2025). For anomaly detection, CANDI emphasizes curated adaptation, selecting reliable test-time samples before updating the model to avoid learning from anomalous or contaminated observations (Kim et al., 2026a). Other methods further refine the adaptation dynamics. AdaNODEs models continuous drift by Neural ODEs for source-free test-time adaptation (Dang et al., 2026), while DynaTTA (Grover & Etemad, 2025) and RG-TTA (Kumar et al., 2026) use shift or regime guided controllers to determine when and how strongly the model should adapt.

Taken together, these methods highlight the central trade-off in test-time and online adaptation: TSFMs must remain flexible enough to track nonstationarity, while avoiding unstable updates caused by noise, transient fluctuations, or anomalous observations. Recent work therefore moves beyond unconstrained online updating toward controlled adaptation, using lightweight modules, calibration gates, curated samples, drift models, or regime detectors to decide when and how strongly to adapt. This makes test-time adaptation a promising but delicate post-training strategy, whose success depends on balancing responsiveness to target-domain change with preservation of the pretrained model's general knowledge.

4.4 Development and Limitations

Parameter adaptation methods have evolved from direct finetuning toward increasingly lightweight and deployment-aware strategies as shown in Fig. 3. Full-parameter finetuning and continual pretraining offer strong domain specialization but demand substantial computation and target-domain data. Parameter-efficient methods reduce this cost by optimizing only a small number of parameters on top of frozen backbones. Test-time and online adaptation further shift the adaptation process into deployment, trading offline training cost for the ability to respond to nonstationary streams.

Despite their effectiveness, parameter adaptation methods also face several limitations. Full-parameter finetuning can be computationally expensive and may cause catastrophic forgetting, particularly when the target domain is scarce or noisy. Parameter-efficient methods reduce cost but introduce architectural decisions, such as where to insert adapters, how to encode covariates, and how to model cross-channel dependence, which remain largely task-specific and lack principled selection criteria. Test-time adaptation can be unstable when the test stream contains anomalies, delayed labels, abrupt regime changes, or weak self-supervised signals. More broadly, all parameter adaptation methods must navigate between adaptation capability and overfitting. Updating too aggressively risks fitting noise in limited target data, while updating too conservatively may fail to bridge the distribution gap from pretraining.

5 Context Augmentation Methods

Context augmentation constitutes an input-level post-training paradigm for TSFMs, which improves TSFMs by providing additional information alongside the target input, without modifying model parameters. The central question is how to construct useful contexts for each target input. We identify three subcategories by the structure of the context: retrieval augmentation, memory augmentation, and context transformation.

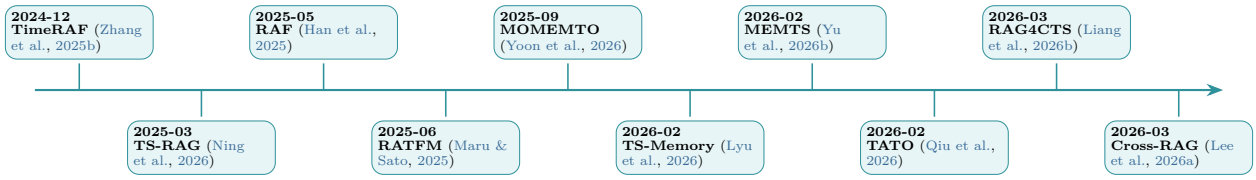

Figure 4: Timeline of context augmentation methods for post-training of TSFMs.

5.1 Retrieval Augmentation

Retrieval augmentation methods retrieve related historical series or segments and inject them into the forecasting or detection process. RAF (Han et al., 2025) provides an early formulation of retrieval-augmented forecasting for TSFMs. It retrieves related time series examples and incorporates them into the forecasting context, showing that retrieval can improve zero-shot forecasting. TimeRAF (Zhang et al., 2025b) further develops this idea by constructing customized time series knowledge bases for specific forecasting tasks. It uses a learnable retriever to extract useful information and introduces channel prompting to integrate the retrieved knowledge along the channel dimension. TS-RAG (Ning et al., 2026) also retrieves semantically relevant time series segments from a dedicated knowledge database, but focuses on fusing the retrieved patterns with the TSFM representation through a learnable mixture-of-experts augmentation module. Cross-RAG (Lee et al., 2026a) observes that a fixed set of retrieved samples may contain irrelevant information. It therefore introduces query–retrieval cross attention to model input-level relevance between the target series and the retrieved examples. This allows the model to selectively attend to useful retrieved samples. RATFM (Maru & Sato, 2025) extends retrieval augmentation from forecasting to anomaly detection. It retrieves normal examples from the target domain and uses them as test-time adaptation signals, enabling pretrained TSFMs to approach in-domain performance without domain-dependent parameter updates. RAG4CTS (Liang et al., 2026b) studies retrieval augmentation for covariate time series in industrial predictive maintenance.

Taken together, these methods show that retrieval augmentation provides a non-parametric route to TSFM post-training: instead of updating the backbone, the model is supplied with relevant historical examples, normal patterns, or task-specific temporal knowledge at inference time. The central design questions are what to retrieve, how to define relevance, and how to fuse retrieved information with the target series. Recent work therefore moves from simple retrieval of related examples toward learned retrievers, customized knowledge bases, cross-attention relevance modeling, and mixture-of-experts fusion. This makes retrieval augmentation especially attractive when target-domain data are available but parameter updates are costly or risky. However, the approach remains sensitive to retrieval quality, since irrelevant or distributionally mismatched examples may introduce noise; robust retrieval, filtering, and selective fusion are therefore key open challenges.

5.2 Memory Augmentation

Memory augmentation replaces explicit retrieval with memory modules that store domain knowledge or normal patterns in a compact form. MEMTS (Yu et al., 2026b) addresses the latency and scalability limitations of retrieval-augmented adaptation. The method introduces a Knowledge Persistence Module that internalizes domain-specific temporal dynamics, such as recurring seasonal patterns and trends, into learnable latent prototypes. TS-Memory (Lyu et al., 2026) first builds a kNN teacher that produces confidence quantile targets from the retrieved futures, then distills the retrieval correction into a lightweight memory adapter. In inference, the memory adapter fuses with the backbone prediction. Memory augmentation has also been applied to anomaly detection. MOMEMTO (Yoon et al., 2026) introduces a patch-based memory module that stores representative normal patterns to solve the problem that high-capacity reconstruction models may over-generalize and reconstruct unseen anomalies too well. The memory gate constrains reconstruction through normal prototypes and improves anomaly detection, especially in few-shot and multi-domain settings.

Overall, memory augmentation provides a scalable alternative to retrieval-based post-training by compressing useful temporal knowledge into learned prototypes, adapters, or memory banks. This makes memory augmentation attractive when retrieval is too costly or when compact domain specialization is needed. Its effectiveness, however, depends on how the memory is constructed and maintained: prototypes must capture stable target-domain structure while avoiding noise, contamination, and overfitting to narrow regimes. Robust memory update rules, uncertainty-aware prototype selection, and mechanisms for forgetting outdated patterns remain important open directions.

5.3 Context Transformation

Context transformation adapts a frozen TSFM by transforming the target input before prediction. Given an input series $\mathbf{x}_{1:T}$, these methods learn or search for a transformation τ_η such that

$$\mathbf{z}_{1:T'} = \tau_\eta(\mathbf{x}_{1:T}), \qquad \hat{\mathbf{y}}_{1:H} = f_{\theta_0}(\mathbf{z}_{1:T'}).$$

The transformation may change how the context is sliced, normalized, denoised, corrected, or represented. TATO (Qiu et al., 2026) is a representative example. Related input-side methods, such as the input nudging component of δ-Adapter (Liang et al., 2026a), also suggest that small transformations at the model interface can improve forecasting without modifying the deployed backbone.

Context transformation is most useful when the pretrained backbone must remain fixed, but the target input can be adapted to better match the model interface. The main open challenge is to design transformations that correct distributional mismatch while preserving task-relevant temporal information.

5.4 Development and Limitations

Context augmentation methods have evolved from injecting external evidence into the prediction process to reshaping the input representation itself, as summarized in Fig. 4. Retrieval-augmented methods retrieve related historical examples from external knowledge bases and incorporate them into forecasting or detection. Memory augmentation compresses such external knowledge into compact modules, reducing the need for explicit retrieval at inference time. Context transformation goes one step further by adapting the input representation directly, without adding external examples or modifying the pretrained backbone.

Despite this progression, context augmentation methods face several limitations. First, their effectiveness depends strongly on the quality of the augmented context. Irrelevant retrieved examples can introduce noise and degrade predictions, especially in sparse or poorly matched target domains. Second, retrieval-based methods introduce memory, indexing, and latency overheads that may be prohibitive in real-time settings. Memory-based methods mitigate these costs, but may over-compress domain knowledge or struggle to update efficiently under distribution shift. More broadly, context augmentation must balance flexibility with reliability: the added or transformed context should expose useful target-domain structure without amplifying spurious correlations, stale patterns, or anomalous observations.

6 Model Composition Methods

Model composition methods constitute a model-level post-training paradigm for TSFMs. This line of work maintains a pool of pretrained, post-trained, or statistically enhanced forecasters and learns how to allocate predictive responsibility among them for a given target time series. The central question is determining which models to select and how to weight their contributions. We identify three subcategories by how the allocation is performed: model selection, adaptive fusion, and sequential fusion.

6.1 Model Selection

Model selection methods select one or a small subset of models from a candidate pool for a given target series, dataset, or task. Chroma (Kayaalp et al., 2026) investigates whether a set of smaller pretrained forecasting models can replace a single large TSFM. It shows that ensembling or model selection over a model

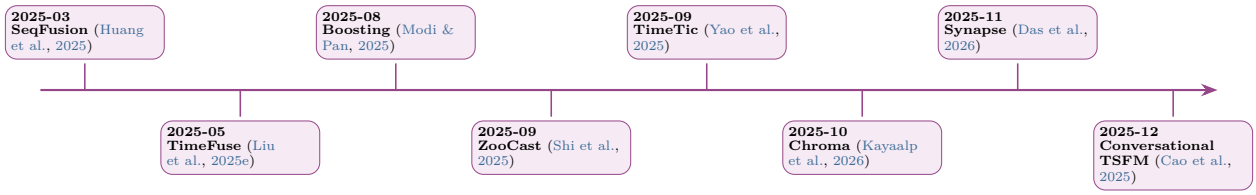

Figure 5: Timeline of model composition methods for post-training of TSFMs.

bank, which consists of diverse specialists post-trained from a shared base model, can achieve competitive forecasting performance with substantially fewer activated parameters. ZooCast (Shi et al., 2025) embeds models in a unified representation space and selects suitable forecasters through similarity matching. This design is attractive when new TSFMs are continuously released, since the model zoo can be expanded without retraining the entire selection system. TimeTic (Yao et al., 2025) extends model selection from zero-shot performance prediction to transferability estimation. Given observed model–dataset–performance relationships on source datasets, it predicts how a TSFM will perform after finetuning on an unseen target dataset, selecting the model expected to yield the strongest downstream adaptation.

Model selection is most useful when the candidate pool contains genuinely complementary models, but its reliability depends on estimating model suitability from limited target-domain evidence. The main open challenge is to develop selection criteria that remain robust under distribution shift and evolving model zoos.

6.2 Adaptive Fusion

Adaptive fusion methods combine the predictions of multiple models using input-dependent or context-dependent weights. TimeFuse (Liu et al., 2025e) extracts meta features from the input time series and trains a learnable fusor to predict sample-level fusion weights over models. Synapse (Das et al., 2026) focuses on the complementary expertise of TSFMs. It dynamically assigns predictive weights according to context performance and constructs a robust forecast distribution by adaptively sampling from the output quantiles of constituent models. Modi & Pan (2025) revisit classical ensemble and statistical methods for TSFM, improving forecasting robustness through bagging, stacking, residual modeling, prediction intervals, and iterative error feedback.

Unlike model selection, which commits to one or a few forecasters, adaptive fusion keeps multiple models active and combines them through input- or context-dependent weights. This can better exploit complementary model behavior and provide smoother adaptation across regimes, but it also increases inference cost and may become unstable when the weighting signal is noisy or distributionally shifted. A central open question is how to make this additional flexibility translate into consistent gains over model selection while keeping inference costs under control.

6.3 Sequential Fusion

Sequential fusion treats model composition as a multi-step decision process. SeqFusion (Huang et al., 2025) selects suitable pretrained models according to the temporal characteristics of the target series, invokes them sequentially, and fuses their predictions. Conversational TSFM (Cao et al., 2025) pushes this direction further toward agentic orchestration. It positions an LLM as a judge that evaluates, explains and coordinates an ensemble of TSFMs. The LLM is trained to associate ensemble weights with interpretable temporal reasoning and then refines its decisions through multi-turn interaction.

Sequential fusion extends model composition from one-shot routing to staged orchestration, allowing later decisions to depend on intermediate forecasts, diagnostics, or model explanations. This can improve flexibility and interpretability, but it also introduces additional latency, error propagation, and dependence on the reliability of the controller. A key open question is when multi-step orchestration provides measurable gains over simpler selection or fusion strategies, especially under realistic inference-cost constraints.

6.4 Development and Limitations

Model Composition has evolved from a simple combination to structured orchestration, as shown in Fig. 5. Early approaches apply classical strategies such as averaging and bagging. Subsequent methods introduce input-conditional fusion, sequential selection pipelines, and transferability-based model selection. Recent agentic methods further add iterative reasoning, enabling an LLM to select, explain, and refine model combinations through multi-turn interaction.

Despite the advances, model composition remains at an early stage. First, many routers rely on shallow signals, such as handcrafted meta features, representation distances, historical validation performance, or learned ensemble weights, which may not capture causal dynamics and regime changes. Second, computational cost is not yet standardized. Future studies would benefit from standardized reporting of post-training cost, routing cost, activated parameter count, memory, and latency. Third, interpretability remains limited. Although agentic methods attempt to explain routing decisions, most current methods do not verify whether the learned weights reflect to meaningful temporal properties rather than spurious benchmark correlations. Finally, the current literature is dominated by forecasting, while routing across anomaly detection, classification, and imputation tasks remains underexplored.

7 Output Processing and Uncertainty Control Methods

Output processing provides an output-level paradigm for TSFM post-training. Rather than modifying the pretrained backbone or its input context, these methods operate after a model has produced an initial forecast, predictive distribution, reconstruction, or anomaly score. The central question is how to improve the accuracy, calibration, reliability, or decision utility of these outputs through post-hoc refinement. We organize this category into four subfamilies according to the object being refined and the strategy used: forecast refinement, forecast calibration, probabilistic output modeling, and anomaly detection.

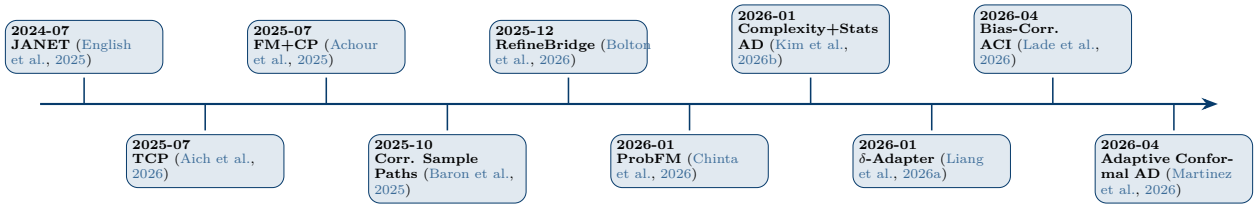

Figure 6: Timeline of output processing and uncertainty control methods for post-training of TSFMs.

7.1 Forecast Refinement

Forecast refinement methods directly adjust the initial forecast produced by a TSFM. Given an initial prediction $\tilde{\mathbf{y}}_{1:H} = f_{\theta_0}(\mathbf{x}_{1:T})$, these methods learn a post-processing map

$$\hat{\mathbf{y}}_{1:H} = \tilde{\mathbf{y}}_{1:H} + g_\psi\left(\mathbf{x}_{1:T}, \tilde{\mathbf{y}}_{1:H}\right),$$

where g_ψ estimates a residual correction that compensates for systematic errors, biases, or distributional mismatch in the raw forecast. The Forecast After the Forecast (Liang et al., 2026a) proposes δ-Adapter, a lightweight post-processing framework that combines input nudging, output residual correction, and uncertainty calibration. RefineBridge (Bolton et al., 2026) treats TSFM forecasts as generative priors and uses a context-conditioned Schrödinger Bridge refinement module to transport the initial forecast toward the target distribution.

Forecast refinement is most valuable when TSFM errors contain stable, target-specific structure such as horizon-dependent bias, scale mismatch, or recurring residual patterns. The main challenge is to estimate the residual map g_ψ from limited target data while detecting cases where the raw TSFM forecast should be left unchanged.

7.2 Forecast Calibration

Forecast calibration methods transform TSFM forecasts into calibrated prediction intervals or joint prediction regions. The primary goal is to provide uncertainty sets with desired coverage properties. Foundation models for time series forecasting with conformal prediction (Achour et al., 2025) investigates how zero-shot TSFMs can benefit conformal prediction, since more target domain samples can be reserved for calibration rather than model training. Temporal Conformal Prediction (Aich et al., 2026) combines quantile forecasting with rolling split-conformal calibration and a Robbins–Monro update to adapt intervals under nonstationarity. Bias-Corrected Adaptive Conformal Inference (Lade et al., 2026) further observes that standard adaptive conformal inference can only adjust the interval width, but not shift the interval center. It therefore estimates online forecast bias and re-centers multi-horizon prediction intervals. JANET (English et al., 2025) generalizes conformal prediction from marginal intervals to joint prediction regions for univariate and multivariate time series.

In this setting, calibration acts as a lightweight reliability layer on top of TSFM forecasts, correcting coverage without retraining the backbone. The main challenge is to maintain coverage under temporal dependence, nonstationarity, and multi-horizon error correlation while keeping prediction sets sharp. This makes calibration-data efficiency, online bias correction, and joint-region construction central directions for TSFM uncertainty control.

7.3 Probabilistic Output Modeling

Probabilistic output modeling methods construct, parameterize or refine the predictive distribution of a TSFM. Unlike conformal calibration that focuses on coverage guarantees, these methods aim to model distributional structure, uncertainty decomposition, or temporal dependence among future horizons. ProbFM (Chinta et al., 2026) introduces a probabilistic TSFM framework based on Deep Evidential Regression. It outputs distributional evidence and decomposes predictive uncertainty into epistemic and aleatoric components. Efficiently Generating Correlated Sample Paths (Baron et al., 2025) addresses a different limitation of multi-step TSFMs: many models provide marginal distributions for each future horizon but ignore the joint dependence across horizons. The method introduces a copula-based post-processing layer to generate correlated sample paths from existing multi-step TSFM outputs in a single forward pass.

Probabilistic output modeling complements calibration by targeting the shape and dependence structure of the predictive distribution, not only its coverage. Its main benefit is richer uncertainty information, including epistemic–aleatoric decomposition and coherent multi-horizon sample paths. The key challenge is to estimate marginal distributions, cross-horizon dependence, and uncertainty decompositions from limited target-domain evidence.

7.4 Anomaly Detection

Anomaly detection methods convert TSFM outputs into a reliable anomaly detector (scores or p-values). Adaptive Conformal Anomaly Detection (Martinez et al., 2026) uses predictions from pretrained foundation models to construct adaptive conformal bounds. The resulting anomaly score is directly interpretable as a false alarm rate or p-value. Complexity and Statistics Guided Anomaly Detection (Kim et al., 2026b) studies reconstruction-based anomaly detection with TSFMs. It identifies overgeneralization and overstationarization as two key failure modes, and improves anomaly scores by incorporating high-frequency complexity measures and restoring statistical features.

Anomaly detection benefits from TSFM outputs when forecast errors, reconstruction residuals, or calibrated bounds can be converted into interpretable evidence of abnormal behavior. The main challenge is to separate true anomalies from distribution shift, model misspecification, and benign high-frequency variation. Future work should focus on score construction, false-alarm calibration, and adaptation rules that remain reliable when anomalies are rare, labels are limited, and normal behavior changes over time.

7.5 Development and Limitations

Output processing methods have evolved from simple forecast correction toward richer probabilistic and calibration-aware post-processing as shown in Fig. 6. One line focuses on forecast refinement, where lightweight modules such as δ-Adapter and RefineBridge correct the initial forecast without changing the deployed backbone. A second line focuses on calibrated forecasting sets, including conformalized prediction intervals, adaptive multi-horizon intervals, bias-corrected conformal inference, and joint prediction regions. More recent work models predictive distribution more explicitly, either by decomposing uncertainty or by generating correlated sample paths across horizons.

Despite their practicality, output processing methods have their specific limitations. First, they are constrained by the information contained in the original model output. A post-processor can correct systematic errors, but it cannot recover signals that the backbone has never represented. Second, calibration methods often rely on representative calibration data, and their guarantees may weaken under severe distribution shift or rapidly changing regimes. Third, probabilistic output methods must balance statistical fidelity with computational efficiency, especially when modeling long-horizon joint dependence.

8 Compression and Specialization Methods

Model compression and specialization constitute an efficiency-driven paradigm to adapt TSFMs for deployment in resource-constrained settings or specialized domains. The works usually operate at the architectural and parameter levels, typically following the initial training phase or during the adaptation to a downstream task. The central question is how to reduce the computational footprint and latency of a TSFM while maintaining, or even enhancing, the predictive performance on a specific set of tasks. We identify two subcategories by how the model is compressed or specialized: knowledge distillation, pruning, and specialization.

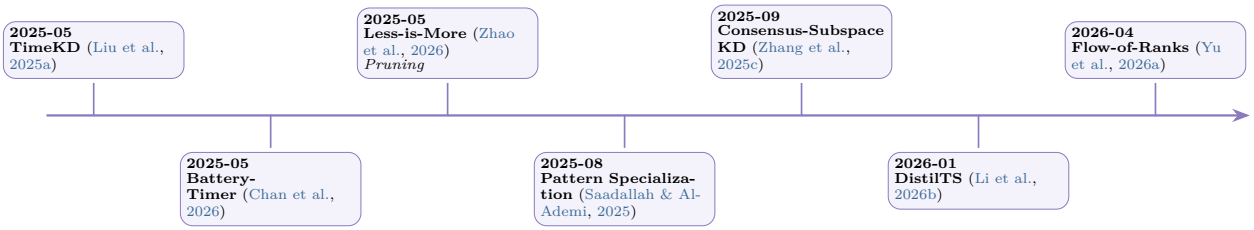

Figure 7: Timeline of compression and specialization methods for post-training of TSFMs.

8.1 Knowledge Distillation

Knowledge distillation (Hinton et al., 2015) compresses the knowledge of a large teacher model into a smaller student while preserving the teacher's generalization ability. Several recent works adapt this paradigm to the specific challenges of TSFMs. DistilTS (Li et al., 2026b) proposes a distillation framework with a horizon-specific objective to mitigate the difficulty discrepancy between short-horizon and long-horizon forecasting, along with a temporal alignment module to bridge the architectural differences between teacher and student models. Zhang et al. (2025c) is inspired by the observation that high-level embeddings converge across different model scales and formulates distillation as a consensus subspace optimization task. More specifically, the approach extracts scale-invariant low-rank subspaces using singular value decomposition of embeddings. The obtained consensus projection matrices are used to finetune student models, with scalable uncertainty mechanisms to improve generalization to unseen datasets. TimeKD (Liu et al., 2025a) distills the knowledge from an LLM for multivariate time series forecasting. It generates high-quality representations of a cross-modal teacher using ground-truth prompts and calibrated language models. The student is then obtained via privileged knowledge distillation, where the teacher LLM is used for reconstruction instead of prediction.

Battery-Timer (Chan et al., 2026) proposes to first finetune TSFMs with low-rank adaptation (Hu et al., 2022, LoRa) on battery degradation data and then apply knowledge distillation to obtain compact experts. This leads to effective and specialized models with reduced inference costs.

Knowledge distillation is attractive when a large TSFM provides strong generalization but is too costly for deployment. Its main benefit is to transfer forecasting behavior, representations, or domain-specific expertise into smaller students with lower inference cost. The main challenge is to decide what should be distilled: point forecasts, horizon-wise errors, latent representations, uncertainty estimates, or task-specific reasoning. Future work should clarify which distillation targets best preserve zero-shot generalization, calibration, and long-horizon behavior under constrained model capacity.

8.2 Pruning and Specialization

Pruning and specialization methods produce efficient, targeted models by removing redundant components or narrowing a general-purpose TSFM to specific patterns. Zhao et al. (2026) propose a structured pruning method to remove task-irrelevant parameters. The "prune-then-finetune" strategy ensures an efficient finetuning strategy to obtain specialized models with a smaller parameter space, leading to better performance than directly finetuning the original model. Saadallah & Al-Ademi (2025) propose to extract specific patterns during finetuning based on a clustering approach, which yields a specialized model that remains efficient in non-stationary environments. In Yu et al. (2026a), the authors study the rank structure of time series transformers, finding that data is concentrated in low-rank subspaces which allow an efficient compression of attention layers. Building on the notion of *flow-of-ranks*, which described the mechanism by which nonlinearity increases the rank across the transformer depth, the authors propose a novel compression method. These results on Chronos (Ansari et al., 2024) showcase a reduction of 65% of inference time and 81% of memory footprint while maintaining the performance of the original model.

Unlike distillation, which transfers behavior from a teacher to a smaller student, pruning and specialization reduce or reshape the TSFM itself. This can preserve more of the original model while lowering inference and memory costs, but it risks removing capacity needed for transfer, calibration, or rare regimes. Future work should evaluate pruning against distillation under matched compute budgets and stress-test compressed models across horizons, domains, and regime shifts.

8.3 Development and Limitations

Compression and specialization methods have developed along two main directions, as shown in Fig. 7. The first one is largely inspired by knowledge distillation with methods using a teacher-student framework, such as DistilTS or TimeKD. The goal is to train compact and smaller student models to achieve the behavior of larger teachers. The other direction works at the architectural level and aims to retain only the components relevant to a specific task. This can be done by pruning the model's weights (Yu et al., 2026a; Zhao et al., 2026) or by identifying specific patterns to tune during the adaptation (Saadallah & Al-Ademi, 2025). It is worth noting that the two directions are complementary. Combining specialization with distillation can help develop efficient and compact experts at deployment time (Chan et al., 2026). Despite promising results, several challenges remain. The existing literature focuses almost exclusively on forecasting. It remains unclear how these strategies perform in other common time-series tasks like classification, anomaly detection, or imputation. Even within forecasting, the choice of distillation objective is nontrivial. Whether to align intermediate representations or final predictions, and how to handle the varying difficulty across forecast horizons, remain open questions. Evaluation practices also need strengthening. Current work typically reports only parameter counts, while practitioners require inference latency, FLOPs, and memory overhead to make informed deployment decisions. Standardized efficiency benchmarks would enable more meaningful comparisons across methods.

9 Future Directions

The preceding sections show that TSFM post-training is no longer synonymous with finetuning: it includes interventions on model parameters, input contexts, model composition, output distributions, and deployed architectures. We now discuss open challenges that cut across these categories and identify future directions for making post-trained TSFMs more robust, efficient, calibrated, and deployable.

9.1 Advancing Parameter Adaptation Methods

Parameter adaptation is the most direct form of TSFM post-training: it improves downstream behavior by modifying the pretrained model itself or by adding trainable components to a frozen backbone. Future work should shift from unconstrained adaptation toward controlled adaptation, where the system decides which parameters to update, how far the adapted model may move from the pretrained backbone, and when target-domain evidence is sufficient to justify adaptation. This is crucial, because parameter updates can yield strong specialization, but they also alter internal representations and may cause forgetting, instability, or overfitting under limited and nonstationary target data.

For **full fine-tuning and continual pretraining**, the evidence is mixed in an informative way. Fine-tuning can yield large gains: dataset-wise fine-tuning of Chronos-T5 Small makes it the best-performing model on Benchmark II, surpassing larger zero-shot Chronos variants and task-specific baselines (Ansari et al., 2024). However, gains are not guaranteed. Process-model forecasting experiments find that LoRA and full fine-tuning can improve over zero-shot TSFMs, but the improvements are dataset-dependent, sometimes small, and may disappear on smaller or more complex datasets (Yu et al., 2025b). Future work should therefore make full-model updates conditional on evidence that lighter adaptation is insufficient. This requires diagnostics that compare full fine-tuning against PEFT, prediction-head adaptation, calibration, retrieval, and in-context adaptation under matched data and compute budgets. Recent results support this direction: FourierFT adapts Chronos-Tiny with only 2,400 trainable parameters (Gupta et al., 2024), TRACE reduces forecasting-head parameters by more than 70% while using gated LoRA-module selection (Li & Zhu, 2026), and in-context fine-tuning uses continued pretraining to teach a TSFM to exploit target examples at inference time (Faw et al., 2025). Finally, adaptation objectives should reflect deployment value: decision-focused fine-tuning of Moirai improves average total daily costs by 9.45% over prediction-focused fine-tuning in feeder optimization (Beichter et al., 2025).

For **parameter-efficient adaptation**, future work should move from task-specific adapter design toward unified and adaptive lightweight adaptation. Existing methods introduce adapters, LoRA modules, or task heads to adapt frozen TSFMs. However, these modules are often designed for specific settings. Future methods should develop general adapter interfaces that can jointly handle cross-channel dependence and task adapters.

For **test-time and online adaptation**, future work should move from continuous updating toward selective and safe adaptation. Existing methods adapt model components during deployment to handle nonstationarity and streaming feedback. However, updating on every test sample can be unstable when the stream contains anomalous signals. Future methods should therefore decide when adaptation should be triggered, which samples are reliable for updating, and how strongly the model should adapt. Safety mechanisms such as update rejection, rollback, and uncertainty filtering will be important for deploying TSFMs in real streaming environments.

9.2 Advancing Context Augmentation Methods

This family of methods improves a frozen or lightly adapted TSFM by constructing an auxiliary context for the target series. Future work should move beyond simply adding more context and focus on reliable context construction. In particular, an effective context-augmentation system should determine what information is useful, when external context is harmful, how contextual knowledge can be stored and updated efficiently, and how the input itself should be transformed to match the distribution expected by the pretrained model.

For **retrieval augmentation**, future work should move from similarity retrieval toward utility-aware retrieval. Existing methods retrieve related series or historical segments and inject them into the forecasting or detection process. However, temporal similarity does not always imply downstream usefulness. Two series may share similar shapes, seasonality, or frequency patterns while having different future dynamics. Therefore, retrievers should be trained or evaluated according to their contribution to the final task loss rather than only their similarity. Moreover, future methods should explicitly handle harmful retrievals by estimating retrieval confidence, filtering irrelevant series or abstaining when no reliable context is available. Finally, retrieval-augmented TSFMs should be evaluated under realistic deployment costs, including retrieval latency, memory footprint, and end-to-end inference time.

For **memory augmentation**, future work should focus on dynamic, compact, and confidence-aware memory. Memory-based methods reduce the latency and scalability issues of explicit retrieval by storing domain knowledge, temporal prototypes, or normal patterns in a compact form. The main challenge is the trade-off between memory efficiency and information fidelity. A memory module should be small enough for fast inference, but fidelious enough to preserve rare events. Future methods should therefore study how to update memory under distribution shift while avoiding collapse from anomalous, noisy, or low-confidence observations. This is especially important for anomaly detection, where storing abnormal samples as normal prototypes can degrade the detector. A promising direction is to equip memory modules with confidence gates, so that the model can decide how strongly to rely on memory for each input. In privacy-sensitive domains such as healthcare, another important direction is privacy-preserving memory, where raw historical series are replaced by synthetic summaries or memory representations.

For **context transformation**, future work should treat adaptation as learned context engineering. Future methods should learn how to normalize, rescale, denoise, decompose, or re-represent the input series so that it better matches the pretraining distribution of the backbone. At the same time, such transformations should be constrained and uncertainty-aware, because overly aggressive transformations may remove informative signals. A further direction is to integrate context transformation with retrieval and memory: transformed inputs may improve retrieval quality, while retrieved examples or memory prototypes may guide how the input should be transformed.

9.3 Advancing Model Composition Methods

This family of methods improves TSFMs by maintaining a pool of models and learning how to select or weight them. Future work should focus on reliable model allocation. In particular, an effective model composition system should determine which models are useful for a given input, when multiple models should be combined, and whether the routing decision is supported by meaningful temporal properties. This requires model composition to become uncertainty-aware and interpretable.

For **model selection**, future work should move from a zero-shot model selection toward adaptation-aware selection. Existing methods select one model or a small subset of models from a candidate pool according to the target series, dataset, or task. However, the model with the best zero-shot performance is not necessarily the model that will perform best after post-training. Therefore, future selectors should estimate not only immediate performance, but also the expected gain after fine-tuning or calibration. Another important direction is dynamic model zoo maintenance. Since new TSFMs are continuously released, future systems should be able to add new models, characterize their strengths, and remove redundant models without retraining the whole router.

For **adaptive fusion**, future work should move from sample-level weighted averaging toward uncertainty-aware fusion. Existing methods combine predictions from multiple TSFMs using weights. However, a single weight for each model may be too coarse, because different models may be useful for different horizons, channels, and quantile levels. Future fusion methods should therefore learn horizon-wise, channel-wise weights. They should also combine predictive distributions rather than only point forecasts, so that uncertainty information from different models can be preserved and calibrated.

For **sequential and agentic fusion**, interpretability should also become more rigorous. Natural language explanations are not sufficient unless they are grounded in measurable temporal properties, such as seasonality, volatility, regime shifts or historical expert reliability.

9.4 Advancing Output Processing and Uncertainty Control Methods

This family of methods improves TSFMs after an initial prediction. Future work should focus on turning raw TSFM outputs into a reliable output for deployment. In particular, a reliable output processing system should determine which part of the initial output should be corrected, and how output signals can be converted into actionable evidence for downstream decisions.

For **forecast refinement**, future work should move from generic residual correction toward structured error correction. Existing methods usually refine the initial forecast by learning a map on top of the frozen backbone output. However, forecasting errors may arise from different sources, including horizon bias, channel mismatch, seasonal misalignment, or long-term trend errors. Future refinement methods should therefore identify and correct these error components separately.

For **forecast calibration**, future work should move from marginal interval calibration toward calibration under temporal dependence and distribution shift. Since time series are sequential and often nonstationary, coverage estimated on a fixed calibration set may fail after a distribution shift. Future methods should adapt prediction sets over time and evaluate coverage across horizons, channels, and high-risk periods.

For **probabilistic output modeling**, future work should move from marginal predictive distributions toward coherent trajectory distributions. Future methods should preserve temporal and cross-channel dependence while remaining scalable for long horizons and high-dimensional multivariate series.

For **anomaly detection**, future work should move toward calibrated anomaly evidence. Existing methods convert forecasts or reconstructions into anomaly scores, but raw reconstruction or forecasting errors are often difficult to interpret across datasets. Future methods should produce anomaly evidence with operational meaning, such as p-values or event-level abnormality measures.

9.5 Advancing Compression and Specialization Methods

This family of methods improves TSFMs by reducing their computational cost or by specializing them for a target domain. Future work should move from simple model size reduction toward deployment-aware specialization. In particular, a reliable compression and specialization system should determine what temporal knowledge should be preserved, which computations can be safely removed, and how efficiency gains translate into practical improvements in latency, memory usage, and robustness under target-domain shift.

For **knowledge distillation**, future work should move from output imitation toward structure- and uncertainty-aware distillation. Existing methods usually train a compact student model to imitate a larger TSFM. However, matching point forecasts alone may discard cross-horizon dependence or predictive uncertainty. Future distillation methods should therefore preserve their temporal representations, trajectory distributions, and calibrated uncertainty.

For **pruning and structural specialization**, future work should move from static parameter removal toward adaptive and task-conditioned computation. Time series tasks often differ substantially in horizon, frequency, and noise level. A fixed pruned structure may therefore be efficient for one setting but suboptimal for another. Future methods should develop task-conditioned pruning. Such methods should also analyze whether the removed components correspond to redundant computation or to temporal knowledge that is only useful under other tasks.

9.6 Toward Agentic and Multimodal TSFMs

Beyond the five post-training families discussed above, an important future direction is extending TSFMs post-training toward agentic and multimodal settings. Current TSFMs are effective in modeling numerical

time series. while LLMs offer complementary strengths in semantic reasoning, planning, and tool use. Integrating other foundation models into TSFMs post-training and enhancing them remains an open challenge.

In agentic settings, an LLM can orchestrate a complete pipeline: selecting the appropriate TSFM, constructing the query, retrieving external context, interpreting TSFM outputs, and translating output into downstream actions. For example, TSFM-generated forecasts can improve LLM-based decision making in financial trading settings (Xie et al., 2026b). ChronoSteer (Wang et al., 2025) shows that textual events can be converted into revision instructions that steer TSFM predictions, and TS-Reasoner (Yu et al., 2025a) aligns TSFM representations with LLM reasoning for time series understanding tasks. These works suggest that future post-training may extend beyond adapting the TSFM itself to shaping outputs that are interpretable and actionable for LLM-based agents. Interfaces between LLMs and TSFMs, such as translating textual instructions into forecast revisions or aligning temporal representations with language-model reasoning, will likely require joint post-training. More broadly, post-training may need to extend beyond predictive accuracy, taking into account the composability and controllability of TSFM outputs within agentic pipelines.

In multimodal settings, real-world time series are usually accompanied by textual descriptions, event logs, clinical notes, visualizations, or expert annotations. However, most TSFMs are still applied and post-trained on numerical inputs alone (see (Jiang et al., 2025) for a survey on multimodal time series analysis.) Recent multimodal benchmarks confirm that textual context can provide complementary information (Liu et al., 2025b; Williams et al., 2025), for instance via question-answering (Xie et al., 2026c), and early work explores unified interfaces between time series and understanding over temporal data (Wang et al., 2024). Effective multimodal post-training will require objectives that guide TSFMs to leverage heterogeneous contexts that they were not pretrained on, as well as calibration strategies that account for the varying informativeness of multimodal inputs across domains and time. This leads to post-training methods that go beyond numerical adaptation toward cross-modal alignment and selective context integration.

10 Conclusion

Post-training is becoming a central step in turning time series foundation models from general-purpose pretrained backbones into reliable models for downstream use. In this survey, we reviewed this emerging area through a pipeline-based taxonomy covering parameter adaptation, context augmentation, model composition, output processing, uncertainty control, compression and specialization. Looking forward, the field should move from generic performance improvement toward reliable, uncertainty-aware, and deployment adaptation. We hope this survey provides a useful structure for understanding current progress and for guiding future research on robust and efficient time series foundation models.

References

Sami Achour, Yassine Bouher, Duong Nguyen, and Nicolas Chesneau. Foundation models for time series forecasting: Application in conformal prediction, 2025. URL https://arxiv.org/abs/2507.08858.

Agnideep Aich, Ashit Baran Aich, and Dipak C. Jain. Temporal conformal prediction (tcp): A distribution-free statistical and machine learning framework for adaptive risk forecasting, 2026. URL https://arxiv.org/abs/2507.05470.

Taha Aksu, Gerald Woo, Juncheng Liu, Xu Liu, Chenghao Liu, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. GIFT-eval: A benchmark for general time series forecasting model evaluation. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024. URL https://openreview.net/forum?id=Z2cMOOANFX.

Mahad Ali, Curtis Lisle, Patrick W. Moore, Tammer Barkouki, Brian J. Kirkwood, and Laura J. Brattain. Fine-tuning foundation models with federated learning for privacy preserving medical time series forecasting, 2025. URL https://arxiv.org/abs/2502.09744.

Matthias Anderer and Feng Li. Hierarchical forecasting with a top-down alignment of independent-level forecasts. *International Journal of Forecasting*, 38(4):1405–1414, 2022. Special Issue: M5 competition.

Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Bernie Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=gerNCVqqtR. Expert Certification.

Abdul Fatir Ansari, Oleksandr Shchur, Jaris Küken, Andreas Auer, Boran Han, Pedro Mercado, Syama Sundar Rangapuram, Huibin Shen, Lorenzo Stella, Xiyuan Zhang, Mononito Goswami, Shubham Kapoor, Danielle C. Maddix, Pablo Guerron, Tony Hu, Junming Yin, Nick Erickson, Prateek Mutalik Desai, Hao Wang, Huzefa Rangwala, George Karypis, Yuyang Wang, and Michael Bohlke-Schneider. Chronos-2: From univariate to universal forecasting, 2025. URL https://arxiv.org/abs/2510.15821.

Sebastian Pineda Arango, Pedro Mercado, Shubham Kapoor, Abdul Fatir Ansari, Lorenzo Stella, Huibin Shen, Hugo Henri Joseph Senetaire, Ali Caner Turkmen, Oleksandr Shchur, Danielle C. Maddix, Michael Bohlke-Schneider, Bernie Wang, and Syama Sundar Rangapuram. Chronosx: Adapting pretrained time series models with exogenous variables. In Yingzhen Li, Stephan Mandt, Shipra Agrawal, and Emtiyaz Khan (eds.), *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pp. 2242–2250. PMLR, 03–05 May 2025. URL https://proceedings.mlr.press/v258/arango25a.html.

Anthony Bagnall, Hoang Anh Dau, Jason Lines, Michael Flynn, James Large, Aaron Bostrom, Paul Southam, and Eamonn Keogh. The uea multivariate time series classification archive, 2018, 2018. URL https://arxiv.org/abs/1811.00075.

Ethan Baron, Boris N. Oreshkin, Ruijun Ma, Hanyu Zhang, Kari Torkkola, Michael W. Mahoney, Andrew Gordon Wilson, and Tatiana Konstantinova. Efficiently generating correlated sample paths from multi-step time series foundation models. In *Recent Advances in Time Series Foundation Models Have We Reached the 'BERT Moment'?*, 2025. URL https://openreview.net/forum?id=ZFwFB4MEu2.

Maximilian Beichter, Nils Friederich, Janik Pinter, Dorina Werling, Kaleb Phipps, Sebastian Beichter, Oliver Neumann, Ralf Mikut, Veit Hagenmeyer, and Benedikt Heidrich. Decision-focused fine-tuning of time series foundation models for dispatchable feeder optimization. *Energy and AI*, 21:100533, September 2025. ISSN 2666-5468. doi: 10.1016/j.egyai.2025.100533. URL http://dx.doi.org/10.1016/j.egyai.2025.100533.

Abdelhakim Benechehab, Vasilii Feofanov, Giuseppe Paolo, Albert Thomas, Maurizio Filippone, and Balázs Kégl. AdaPTS: Adapting univariate foundation models to probabilistic multivariate time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=yeICCRy3lE.

Anthony Bolton, Wuyang Zhou, Zehua Chen, Giorgos Iacovides, and Danilo Mandic. Refinebridge: Generative bridge models improve financial forecasting by foundation models. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1796–1800, 2026. doi: 10.1109/ICASSP55912.2026.11463629.

Joos-Hendrik Böse, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Dustin Lange, David Salinas, Sebastian Schelter, Matthias Seeger, and Yuyang Wang. Probabilistic demand forecasting at scale. *Proc. VLDB Endow.*, 10(12):1694–1705, August 2017. ISSN 2150-8097. doi: 10.14778/3137765.3137775. URL https://doi.org/10.14778/3137765.3137775.

G. E. P. Box, G. M. Jenkins, and J. F. MacGregor. Some Recent Advances in Forecasting and Control. *Journal of the Royal Statistical Society Series C*, 23(2):158–179, June 1974. doi: 10.2307/2346997. URL https://ideas.repec.org/a/bla/jorssc/v23y1974i2p158-179.html.

George Edward Pelham Box and Gwilym Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc., USA, 1990. ISBN 0816211043.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss,

Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 1877–1901. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf.

Defu Cao, Michael Gee, Jinbo Liu, Hengxuan Wang, Wei Yang, Rui Wang, and Yan Liu. Conversational time series foundation models: Towards explainable and effective forecasting, 2025. URL https://arxiv.org/abs/2512.16022.

Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris Coll-Vinent, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, RISHI HAZRA, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollar, Nikhila Ravi, Kate Saenko, Pengchuan Zhang, and Christoph Feichtenhofer. SAM 3: Segment anything with concepts. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=r35clVtGzw.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021.

Angelo Casolaro, Vincenzo Capone, Gennaro Iannuzzo, and Francesco Camastra. Deep learning for time series forecasting: Advances and open problems. *Information*, 14(11), 2023. ISSN 2078-2489. doi: 10.3390/info14110598. URL https://www.mdpi.com/2078-2489/14/11/598.

Paulius Čepulionis and Kristina Lukoševičiūtė. Electrocardiogram time series forecasting and optimization using ant colony optimization algorithm. *Mathematical Models in Engineering*, 2(1):69–77, Jun 2016. ISSN 2351-5279. URL https://www.extrica.com/article/17229.

Joey Chan, Zhen Chen, and Ershun Pan. Foundation models knowledge distillation for battery capacity degradation forecast. *Reliability Engineering & System Safety*, 270:112139, June 2026. ISSN 0951-8320. doi: 10.1016/j.ress.2025.112139. URL http://dx.doi.org/10.1016/j.ress.2025.112139.

Renyi Chen and Molei Tao. Data-driven prediction of general hamiltonian dynamics via learning exactly-symplectic maps. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 1717–1727. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/chen21r.html.

Shengchao Chen, Guodong Long, Tao Shen, and Jing Jiang. Prompt federated learning for weather forecasting: Toward foundation models on meteorological data. In Edith Elkind (ed.), *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pp. 3532–3540. International Joint Conferences on Artificial Intelligence Organization, 8 2023. doi: 10.24963/ijcai.2023/393. URL https://doi.org/10.24963/ijcai.2023/393. Main Track.

Shengchao Chen, Guodong Long, Jing Jiang, and Chengqi Zhang. Federated foundation models on heterogeneous time series. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. ISBN 978-1-57735-897-8. doi: 10.1609/aaai.v39i15.33739. URL https://doi.org/10.1609/aaai.v39i15.33739.

Shengchao Chen, Guodong Long, Dikai Liu, and Jing Jiang. Bi-level heterogeneous learning for time series foundation models: A federated learning approach, 2026. URL https://arxiv.org/abs/2604.06727.

Hanyin Cheng, Ruitong Zhang, Yuning Lu, Peng Chen, Meng Wang, Yang Shu, Bin Yang, and Chenjuan Guo. Star: Boosting time series foundation models for anomaly detection through state-aware adapter, 2025. URL https://arxiv.org/abs/2510.16014.

Hanyin Cheng, Xingjian Wu, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. CoRA: Boosting time series foundation models for multivariate forecasting through correlation-aware adapter. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=JRlNrcTllN.

Andrei Chernov. Fine-tuning a time series foundation model with wasserstein loss. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024. URL https://openreview.net/forum?id=cVy5qBFmcr.

Arundeep Chinta, Lucas Vinh Tran, and Jay Katukuri. Probfm: Probabilistic time series foundation model with uncertainty decomposition, 2026. URL https://arxiv.org/abs/2601.10591.

Kwok Tai Chui, Wadee Alhalabi, Sally Siu Hang Pang, Patricia Ordonez de Pablos, Ryan Wen Liu, and Mingbo Zhao. Disease diagnosis in smart healthcare: Innovation, technologies and applications. *Sustainability*, 9(12):2309, 2017.

Ben Cohen, Emaad Khwaja, Youssef Doubli, Salahidine Lemaachi, Chris Lettieri, Charles Masson, Hugo Miccinilli, Elise Ramé, Qiqi Ren, Afshin Rostamizadeh, Jean Ogier du Terrail, Anna-Monica Toon, Kan Wang, Stephan Xie, Zongzhe Xu, Viktoriya Zhukova, David Asker, Ameet Talwalkar, and Othmane Abou-Amal. This time is different: An observability perspective on time series foundation models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=1jDAYXfcS2.

Ting Dang, Soumyajit Chatterjee, Hong Jia, Yu Wu, Flora Salim, and Fahim Kawsar. Adanodes: Test time adaptation for time series forecasting using neural odes. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2771–2775, 2026. doi: 10.1109/ICASSP55912.2026.11463542.

Afrin Dange and Sunita Sarawagi. Tfmadapter: Lightweight instance-level adaptation of foundation models for forecasting with covariates. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, CIKM '25, pp. 498–507, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400720406. doi: 10.1145/3746252.3761272. URL https://doi.org/10.1145/3746252.3761272.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=jn2iTJas6h.

Sarkar Snigdha Sarathi Das, Palash Goyal, Mihir Parmar, Yiwen Song, Long Le, Lesly Miculicich, Jinsung Yoon, Rui Zhang, Hamid Palangi, and Tomas Pfister. Synapse: Adaptive arbitration of complementary expertise in time series foundational models. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=j3HqbsCwt1.

{Hoang Anh} Dau, Anthony Bagnall, Kaveh Kamgar, {Chin-Chia Michael} Yeh, Yan Zhu, Shaghayegh Gharghabi, {Chotirat Ann} Ratanamahatan, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, November 2019. ISSN 2329-9266. doi: 10.1109/JAS.2019.000000.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 248–255. IEEE, 2009. URL https://ieeexplore.ieee.org/abstract/document/5206848/.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423/.

Jesse Dodge, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner. Documenting large webtext corpora: A case study on the colossal clean crawled corpus. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 1286–1305,

Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.98. URL https://aclanthology.org/2021.emnlp-main.98/.

Samuel Dooley, Gurnoor Singh Khurana, Chirag Mohapatra, Siddartha Venkat Naidu, and Colin White. ForecastPFN: Synthetically-trained zero-shot forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=tScBQRNgjk.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Vijay Ekambaram, Arindam Jati, Pankaj Dayama, Sumanta Mukherjee, Nam H Nguyen, Wesley M. Gifford, Chandra Reddy, and Jayant Kalagnanam. Tiny time mixers (TTMs): Fast pre-trained models for enhanced zero/few-shot forecasting of multivariate time series. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=3O5YCEWETq.

Eshant English, Eliot Wong-Toi, Matteo Fontana, Stephan Mandt, Padhraic Smyth, and Christoph Lippert. Janet: Joint adaptive prediction-region estimation for time-series. *Machine Learning*, 114(8): 177, Jun 2025. ISSN 1573-0565. doi: 10.1007/s10994-025-06812-2. URL https://doi.org/10.1007/s10994-025-06812-2.

Nick Erickson, Jonas Mueller, Alexander Shirkov, Hang Zhang, Pedro Larroy, Mu Li, and Alexander Smola. Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505*, 2020.

Chenyou Fan, Yuze Zhang, Yi Pan, Xiaoyue Li, Chi Zhang, Rong Yuan, Di Wu, Wensheng Wang, Jian Pei, and Heng Huang. Multi-horizon time series forecasting with temporal attention learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 2527–2535, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362016. doi: 10.1145/3292500.3330662. URL https://doi.org/10.1145/3292500.3330662.

Yuwei Fan, Tao Song, Chenlong Feng, Chao Liu, and Dongxiang Jiang. Wind power prediction using foundation large time series models enhanced by time series prompt in exogenous and tuning forms. *Applied Energy*, 400:126535, 2025. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2025.126535. URL https://www.sciencedirect.com/science/article/pii/S0306261925012656.

Matthew Faw, Rajat Sen, Yichen Zhou, and Abhimanyu Das. In-context fine-tuning for time-series foundation models. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=uxzgGLWPj2.

Vasilii Feofanov, Songkang Wen, Marius Alonso, Romain Ilbert, Hongbo Guo, Malik Tiomoko, Lujia Pan, Jianfeng Zhang, and Ievgen Redko. Mantis: Lightweight calibrated foundation model for user-friendly time series classification, 2025. URL https://arxiv.org/abs/2502.15637.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, October 2001.

Xinghong Fu, Masanori Hirano, and Kentaro Imajo. Financial fine-tuning a large time series model. In *2025 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CiFer)*, pp. 1–9, 2025. doi: 10.1109/CiFer64978.2025.10975735.

Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. UniTS: A unified multi-task time series model. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=nB0dYBptWW.

Azul Garza and Max Mergenthaler-Canseco. Timegpt-1. *arXiv preprint arXiv:2310.03589*, 2023.

Azul Garza, Cristian Challu, and Max Mergenthaler-Canseco. Timegpt-1, 2024. URL https://arxiv.org/abs/2310.03589.

Anubha Goel, Puneet Pasricha, and Juho Kanniainen. Time-series foundation ai model for value-at-risk forecasting, 2025a. URL https://arxiv.org/abs/2410.11773.

Anubha Goel, Puneet Pasricha, Martin Magris, and Juho Kanniainen. Foundation time-series ai model for realized volatility forecasting, 2025b. URL https://arxiv.org/abs/2505.11163.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=FVvf69a5rx.

Shivam Grover and Ali Etemad. Shift-aware test time adaptation and benchmarking for time-series forecasting. In *Second Workshop on Test-Time Adaptation: Putting Updates to the Test! at ICML 2025*, 2025. URL https://openreview.net/forum?id=a399SmgWGl.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, Xiaokang Zhang, Xingkai Yu, Yu Wu, Z. F. Wu, Zhibin Gou, Zhihong Shao, Zhuoshu Li, Ziyi Gao, Aixin Liu, Bing Xue, Bingxuan Wang, Bochao Wu, Bei Feng, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chong Ruan, Damai Dai, Deli Chen, Dongjie Ji, Erhang Li, Fangyun Lin, Fucong Dai, Fuli Luo, Guangbo Hao, Guanting Chen, Guowei Li, H. Zhang, Hanwei Xu, Honghui Ding, Huazuo Gao, Hui Qu, Hui Li, Jianzhong Guo, Jiashi Li, Jingchang Chen, Jingyang Yuan, Jinhao Tu, Junjie Qiu, Junlong Li, J. L. Cai, Jiaqi Ni, Jian Liang, Jin Chen, Kai Dong, Kai Hu, Kaichao You, Kaige Gao, Kang Guan, Kexin Huang, Kuai Yu, Lean Wang, Lecong Zhang, Liang Zhao, Litong Wang, Liyue Zhang, Lei Xu, Leyi Xia, Mingchuan Zhang, Minghua Zhang, Minghui Tang, Mingxu Zhou, Meng Li, Miaojun Wang, Mingming Li, Ning Tian, Panpan Huang, Peng Zhang, Qiancheng Wang, Qinyu Chen, Qiushi Du, Ruiqi Ge, Ruisong Zhang, Ruizhe Pan, Runji Wang, R. J. Chen, R. L. Jin, Ruyi Chen, Shanghao Lu, Shangyan Zhou, Shanhuang Chen, Shengfeng Ye, Shiyu Wang, Shuiping Yu, Shunfeng Zhou, Shuting Pan, S. S. Li, Shuang Zhou, Shaoqing Wu, Tao Yun, Tian Pei, Tianyu Sun, T. Wang, Wangding Zeng, Wen Liu, Wenfeng Liang, Wenjun Gao, Wenqin Yu, Wentao Zhang, W. L. Xiao, Wei An, Xiaodong Liu, Xiaohan Wang, Xiaokang Chen, Xiaotao Nie, Xin Cheng, Xin Liu, Xin Xie, Xingchao Liu, Xinyu Yang, Xinyuan Li, Xuecheng Su, Xuheng Lin, X. Q. Li, Xiangyue Jin, Xiaojin Shen, Xiaosha Chen, Xiaowen Sun, Xiaoxiang Wang, Xinnan Song, Xinyi Zhou, Xianzu Wang, Xinxia Shan, Y. K. Li, Y. Q. Wang, Y. X. Wei, Yang Zhang, Yanhong Xu, Yao Li, Yao Zhao, Yaofeng Sun, Yaohui Wang, Yi Yu, Yichao Zhang, Yifan Shi, Yiliang Xiong, Ying He, Yishi Piao, Yisong Wang, Yixuan Tan, Yiyang Ma, Yiyuan Liu, Yongqiang Guo, Yuan Ou, Yuduan Wang, Yue Gong, Yuheng Zou, Yujia He, Yunfan Xiong, Yuxiang Luo, Yuxiang You, Yuxuan Liu, Yuyang Zhou, Y. X. Zhu, Yanping Huang, Yaohui Li, Yi Zheng, Yuchen Zhu, Yunxian Ma, Ying Tang, Yukun Zha, Yuting Yan, Z. Z. Ren, Zehui Ren, Zhangli Sha, Zhe Fu, Zhean Xu, Zhenda Xie, Zhengyan Zhang, Zhewen Hao, Zhicheng Ma, Zhigang Yan, Zhiyu Wu, Zihui Gu, Zijia Zhu, Zijun Liu, Zilin Li, Ziwei Xie, Ziyang Song, Zizheng Pan, Zhen Huang, Zhipeng Xu, Zhongyu Zhang, and Zhen Zhang. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645:633–638, Sep 2025. ISSN 1476-4687. doi: 10.1038/s41586-025-09422-z. URL https://doi.org/10.1038/s41586-025-09422-z.

Divij Gupta, Anubhav Bhatti, and Surajsinh Parmar. Beyond LoRA: Exploring efficient fine-tuning techniques for time series foundational models. *arXiv preprint arXiv:2409.11302*, 2024. doi: 10.48550/arXiv.2409.11302. URL https://arxiv.org/abs/2409.11302.

Lu Han, Yu Liu, Lan Li, Qiwen Deng, Jian Jiang, Yinbo sun, Zhe Yu, Binfeng Wang, Xingyu Lu, Lintao Ma, Han-Jia Ye, and De-Chuan Zhan. UniCA: Unified covariate adaptation for time series foundation model. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=I8q4MZb4OP.

Sungwon Han, Seungeon Lee, Meeyoung Cha, Sercan O Arik, and Jinsung Yoon. Retrieval augmented time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=GUDnecJdJU.

Atif Hassan, Tarun Kumar, Ashish Mishra, Sergey Serebryakov, Satish Kumar Mopur, Phanidhar Koganti, Murthy Chelankuri, Ramanagopal Vogety, Suparna Bhattacharya, and Martin Foltin. Forecast2anomaly (f2a): Adapting multivariate time series foundation models for anomaly prediction, 2025. URL https://arxiv.org/abs/2511.03149.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15979–15988, 2022. doi: 10.1109/CVPR52688.2022.01553.

Marti A. Hearst, Susan T. Dumais, Edgar Osuna, John Platt, and Bernhard Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, 1998.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network, 2015. URL https://arxiv.org/abs/1503.02531.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Oriol Vinyals, Jack W. Rae, and Laurent Sifre. Training compute-optimal large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Nan Huang, Haishuai Wang, Zihuai He, Marinka Zitnik, and Xiang Zhang. Repurposing foundation model for generalizable medical time series classification. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=wNEzRYiyZM.

Ting-Ji Huang, Xu-Yang Chen, and Han-Jia Ye. Seqfusion: Sequential fusion of pre-trained models for zero-shot time-series forecasting, 2025. URL https://arxiv.org/abs/2503.02836.

Rob Hyndman and G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 3rd edition, 2021.

Romain Ilbert, Ambroise Odonnat, Vasilii Feofanov, Aladin Virmaux, Giuseppe Paolo, Themis Palpanas, and Ievgen Redko. SAMformer: Unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=8kLzL5QBh2.

Yushan Jiang, Kanghui Ning, Zijie Pan, Xuyang Shen, Jingchao Ni, Wenchao Yu, Anderson Schneider, Haifeng Chen, Yuriy Nevmyvaka, and Dongjin Song. Multi-modal time series analysis: A tutorial and survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 6043–6053, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3736567. URL https://doi.org/10.1145/3711896.3736567.

Aishwarya Kamath, Johan Ferret, Shreya Pathak, Nino Vieillard, Ramona Merhej, Sarah Perrin, Tatiana Matejovicova, Alexandre Ramé, Morgane Rivière, Louis Rouillard, Thomas Mesnard, Geoffrey Cideron, Jean bastien Grill, Sabela Ramos, Edouard Yvinec, Michelle Casbon, Etienne Pot, Ivo Penchev, Gaël Liu, Francesco Visin, Kathleen Kenealy, Lucas Beyer, Xiaohai Zhai, Anton Tsitsulin, Robert Busa-Fekete, Alex Feng, Noveen Sachdeva, Benjamin Coleman, Yi Gao, Basil Mustafa, Iain Barr, Emilio Parisotto, David Tian, Matan Eyal, Colin Cherry, Jan-Thorsten Peter, Danila Sinopalnikov, Surya Bhupatiraju, Rishabh Agarwal, Mehran Kazemi, Dan Malkin, Ravin Kumar, David Vilar, Idan Brusilovsky, Jiaming Luo, Andreas Steiner, Abe Friesen, Abhanshu Sharma, Abheesht Sharma, Adi Mayrav Gilady, Adrian Goedeckemeyer, Alaa Saade, Alex Feng, Alexander Kolesnikov, Alexei Bendebury, Alvin Abdagic, Amit Vadi, András György, André Susano Pinto, Anil Das, Ankur Bapna, Antoine Miech, Antoine Yang, Antonia Paterson, Ashish Shenoy, Ayan Chakrabarti, Bilal Piot, Bo Wu, Bobak Shahriari, Bryce Petrini, Charlie Chen, Charline Le Lan, Christopher A. Choquette-Choo, CJ Carey, Cormac Brick, Daniel Deutsch, Danielle Eisenbud, Dee Cattle, Derek Cheng, Dimitris Paparas, Divyashree Shivakumar Sreepathihalli, Doug Reid, Dustin Tran, Dustin Zelle, Eric Noland, Erwin Huizenga, Eugene Kharitonov, Frederick Liu, Gagik Amirkhanyan, Glenn Cameron, Hadi Hashemi, Hanna Klimczak-Plucińska, Harman Singh, Harsh Mehta, Harshal Tushar Lehri, Hussein Hazimeh, Ian Ballantyne, Idan Szpektor, Ivan Nardini, Jean Pouget-Abadie, Jetha Chan, Joe Stanton, John Wieting, Jonathan Lai, Jordi Orbay, Joseph Fernandez, Josh Newlan, Ju yeong Ji, Jyotinder Singh, Kat Black, Kathy Yu, Kevin Hui, Kiran Vodrahalli, Klaus Greff, Linhai Qiu, Marcella Valentine, Marina Coelho, Marvin Ritter, Matt Hoffman, Matthew Watson, Mayank Chaturvedi, Michael Moynihan, Min Ma, Nabila Babar, Natasha Noy, Nathan Byrd, Nick Roy, Nikola Momchev, Nilay Chauhan, Noveen Sachdeva, Oskar Bunyan, Pankil Botarda, Paul Caron, Paul Kishan Rubenstein, Phil Culliton, Philipp Schmid, Pier Giuseppe Sessa, Pingmei Xu, Piotr Stanczyk, Pouya

Tafti, Rakesh Shivanna, Renjie Wu, Renke Pan, Reza Rokni, Rob Willoughby, Rohith Vallu, Ryan Mullins, Sammy Jerome, Sara Smoot, Sertan Girgin, Shariq Iqbal, Shashir Reddy, Shruti Sheth, Siim Põder, Sijal Bhatnagar, Sindhu Raghuram Panyam, Sivan Eiger, Susan Zhang, Tianqi Liu, Trevor Yacovone, Tyler Liechty, Uday Kalra, Utku Evci, Vedant Misra, Vincent Roseberry, Vlad Feinberg, Vlad Kolesnikov, Woohyun Han, Woosuk Kwon, Xi Chen, Yinlam Chow, Yuvein Zhu, Zichuan Wei, Zoltan Egyed, Victor Cotruta, Minh Giang, Phoebe Kirk, Anand Rao, Kat Black, Nabila Babar, Jessica Lo, Erica Moreira, Luiz Gustavo Martins, Omar Sanseviero, Lucas Gonzalez, Zach Gleicher, Tris Warkentin, Vahab Mirrokni, Evan Senter, Eli Collins, Joelle Barral, Zoubin Ghahramani, Raia Hadsell, Yossi Matias, D. Sculley, Slav Petrov, Noah Fiedel, Noam Shazeer, Oriol Vinyals, Jeff Dean, Demis Hassabis, Koray Kavukcuoglu, Clement Farabet, Elena Buchatskaya, Jean-Baptiste Alayrac, Rohan Anil, Dmitry, Lepikhin, Sebastian Borgeaud, Olivier Bachem, Armand Joulin, Alek Andreev, Cassidy Hardin, Robert Dadashi, and Léonard Hussenot. Gemma 3 technical report, 2025. URL https://arxiv.org/abs/2503.19786.

Mert Kayaalp, Ali Caner Turkmen, Oleksandr Shchur, Pedro Mercado, Abdul Fatir Ansari, Michael Bohlke-Schneider, and Bernie Wang. Test-time efficient pretrained model portfolios for time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=iqUMjxfDNH.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. ISBN 978-1-57735-897-8. doi: 10.1609/aaai.v39i17.33965. URL https://doi.org/10.1609/aaai.v39i17.33965.

HyunGi Kim, Jisoo Mok, Hyungyu Lee, Juhyeon Shin, and Sungroh Yoon. Candi: Curated test-time adaptation for multivariate time-series anomaly detection under distribution shift. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15018–15026. Association for the Advancement of Artificial Intelligence, 2026a. doi: 10.1609/aaai.v40i17.38524. URL https://doi.org/10.1609/aaai.v40i17.38524.

Jongwon Kim, Samuel Yoon, Young Myoung Ko, Yerin Kim, Sung Il Kim, and JAEUNG TAE. Complexity- and statistics-guided anomaly detection in time series foundation models. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=rBt9aW3Mx7.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023. URL https://arxiv.org/abs/2304.02643.

Siva Rama Krishna Kottapalli, Karthik Hubli, Sandeep Chandrashekhara, Garima Jain, Sunayana Hubli, Gayathri Botla, and Ramesh Doddaiah. Foundation models for time series: A survey, 2025. URL https://arxiv.org/abs/2504.04011.

Indar Kumar, Akanksha Tiwari, Sai Krishna Jasti, and Ankit Hemant Lade. Rg-tta: Regime-guided meta-control for test-time adaptation in streaming time series, 2026. URL https://arxiv.org/abs/2603.27814.

Komal Kumar, Tajamul Ashraf, Omkar Thawakar, Rao Muhammad Anwer, Hisham Cholakkal, Mubarak Shah, Ming-Hsuan Yang, Phillip H. S. Torr, Fahad Shahbaz Khan, and Salman Khan. Llm post-training: A deep dive into reasoning large language models, 2025. URL https://arxiv.org/abs/2502.21321.

Ankit Lade, Sai Krishna J., and Indar Kumar. Bias-corrected adaptive conformal inference for multi-horizon time series forecasting, 2026. URL https://arxiv.org/abs/2604.13253.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pp. 95–104, New York, NY, USA, 2018a. Association for Computing Machinery. ISBN 9781450356572. doi: 10.1145/3209978.3210006. URL https://doi.org/10.1145/3209978.3210006.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *Association for Computing Machinery*, SIGIR '18, pp. 95–104,

New York, NY, USA, 2018b. ISBN 9781450356572. doi: 10.1145/3209978.3210006. URL https://doi.org/10.1145/3209978.3210006.

Nathan Lambert. Reinforcement learning from human feedback, 2026. URL https://arxiv.org/abs/2504.12501.

Seunghan Lee, Jaehoon Lee, Jun Seo, Sungdong Yoo, Minjae Kim, Tae Yoon Lim, Dongwan Kang, Hwanil Choi, SoonYoung Lee, and Wonbin Ahn. Cross-rag: Zero-shot retrieval-augmented time series forecasting via cross-attention, 2026a. URL https://arxiv.org/abs/2603.14709.

Thomas L Lee, William Toner, Martin Asenov, Artjom Joosen, and Rajkarn Singh. Lightweight online adaption for time series foundation model forecasts. In *Second Workshop on Test-Time Adaptation: Putting Updates to the Test! at ICML 2025*, 2025. URL https://openreview.net/forum?id=ohJDcNM33o.

Thomas L. Lee, Edoardo M. Ponti, and Amos Storkey. Adapting time series foundation models through data mixtures, 2026b. URL https://arxiv.org/abs/2603.02840.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.

Jinpeng Li, Zhongyi Pei, Huaze Xue, Bojian Zheng, Chen Wang, and Jianmin Wang. Dualweaver: Synergistic feature weaving surrogates for multivariate forecasting with univariate time series foundation models, 2026a. URL https://arxiv.org/abs/2602.22066.

Na Li et al. From demand forecasting to inventory ordering decisions for red blood cells through integrating machine learning, statistical modeling, and inventory optimization. *Transfusion*, 62(1):87–99, 2022.

Yuqi Li, Kuiye Ding, Chuanguang Yang, Szu-Yu Chen, and Yingli Tian. Distilling time series foundation models for efficient forecasting, 2026b. URL https://arxiv.org/abs/2601.12785.

Yuze Li and Wei Zhu. TRACE: Time series parameter efficient fine-tuning. *Neurocomputing*, 664:132098, February 2026. ISSN 0925-2312. doi: 10.1016/j.neucom.2025.132098. URL http://dx.doi.org/10.1016/j.neucom.2025.132098.

Daojun Liang, Qi Li, Yinglong Wang, Jing Chen, Hu Zhang, Xiaoxiao Cui, Qizheng Wang, and Shuo Li. The forecast after the forecast: A post-processing shift in time series. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=syfWdclGE1.

Kenny Ye Liang, Zhongyi Pei, Huan Zhang, Yuhui Liu, Shaoxu Song, and Jianmin Wang. Retrieval-augmented generation with covariate time series, 2026b. URL https://arxiv.org/abs/2603.04951.

Yuxuan Liang, Haomin Wen, Yuqi Nie, Yushan Jiang, Ming Jin, Dongjin Song, Shirui Pan, and Qingsong Wen. Foundation models for time series analysis: A tutorial and survey. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 6555–6565. ACM, August 2024. doi: 10.1145/3637528.3671451. URL http://dx.doi.org/10.1145/3637528.3671451.

Chenguo Lin, Xumeng Wen, Wei Cao, Congrui Huang, Jiang Bian, Stephen Lin, and Zhirong Wu. Nutime: Numerically multi-scaled embedding for large- scale time-series pretraining. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=TwiSBZ0p9u.

Chenghao Liu, Taha Aksu, Juncheng Liu, Xu Liu, Hanshu Yan, Quang Pham, Silvio Savarese, Doyen Sahoo, Caiming Xiong, and Junnan Li. Moirai 2.0: When less is more for time series forecasting, 2026a. URL https://arxiv.org/abs/2511.11698.

Chenxi Liu, Hao Miao, Qianxiong Xu, Shaowen Zhou, Cheng Long, Yan Zhao, Ziyue Li, and Rui Zhao. Efficient multivariate time series forecasting via calibrated language models with privileged knowledge distillation. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pp. 3165–3178, 2025a. doi: 10.1109/ICDE65448.2025.00237.

Haoxin Liu, Shangqing Xu, Zhiyuan Zhao, Lingkai Kong, Harshavardhan Kamarthi, Aditya B. Sasanur, Megha Sharma, Jiaming Cui, Qingsong Wen, Chao Zhang, and B. Aditya Prakash. Time-mmd: Multi-domain multimodal dataset for time series analysis, 2025b. URL https://arxiv.org/abs/2406.08627.

Mingzhu Liu, Angela H. Chen, and George H. Chen. Generalized prompt tuning: Adapting frozen univariate time series foundation models for multivariate healthcare time series, 2024a. URL https://arxiv.org/abs/2411.12824.

Ruonan Liu, Guotao Meng, Boyuan Yang, Chuang Sun, and Xuefeng Chen. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Transactions on Industrial Informatics*, 13(3):1310–1320, June 2017. ISSN 1551-3203. doi: 10.1109/TII.2016.2645238. Publisher Copyright: © 2016 IEEE.

Xu Liu, Taha Aksu, Juncheng Liu, Qingsong Wen, Yuxuan Liang, Caiming Xiong, Silvio Savarese, Doyen Sahoo, Junnan Li, and Chenghao Liu. Empowering time series analysis with synthetic data: A survey and outlook in the era of foundation models, 2025c. URL https://arxiv.org/abs/2503.11411.

Xu Liu, Juncheng Liu, Gerald Woo, Taha Aksu, Yuxuan Liang, Roger Zimmermann, Chenghao Liu, Junnan Li, Silvio Savarese, Caiming Xiong, and Doyen Sahoo. Moirai-moe: Empowering time series foundation models with sparse mixture of experts. In *Forty-second International Conference on Machine Learning*, 2025d. URL https://openreview.net/forum?id=SrEOUSyJcR.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b. URL https://openreview.net/forum?id=JePfAI8fah.

Zhen Liu, Boyuan Li, Hao Huang, Yanru Sun, Yucheng Wang, Min Wu, and Qianli Ma. From pre-training to post-training: A survey on time series foundation models. *TechRxiv*, 2026(0211), 2026b. doi: 10.36227/techrxiv.176978429.90235801/v2. URL https://www.techrxiv.org/doi/abs/10.36227/techrxiv.176978429.90235801/v2.

Zhining Liu, Ze Yang, Xiao Lin, Ruizhong Qiu, Tianxin Wei, Yada Zhu, Hendrik Hamann, Jingrui He, and Hanghang Tong. Breaking silos: Adaptive model fusion unlocks better time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025e. URL https://openreview.net/forum?id=sdFRCRk1pP.

Sisuo Lyu, Siru Zhong, Tiegang Chen, Weilin Ruan, Qingxiang Liu, Taiqiang Lv, Qingsong Wen, Raymond Chi-Wing Wong, and Yuxuan Liang. Ts-memory: Plug-and-play memory for time series foundation models, 2026. URL https://arxiv.org/abs/2602.11550.

Natalia Martinez, Fearghal O'Donncha, Wesley M. Gifford, Nianjun Zhou, Dhaval C Patel, and Roman Vaculin. Adaptive conformal anomaly detection with time series foundation models for signal monitoring. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=7uFbs68MSI.

Chihiro Maru and Shoetsu Sato. Ratfm: Retrieval-augmented time series foundation model for anomaly detection, 2025. URL https://arxiv.org/abs/2506.02081.

Max Planck Institute. Weather dataset, 2021. URL https://www.bgc-jena.mpg.de/wetter/.

Dhruv D. Modi and Rong Pan. Enhancing transformer-based foundation models for time series forecasting via bagging, boosting and statistical ensembles, 2025. URL https://arxiv.org/abs/2508.16641.

Vladyslav Moroshan, Julien Siems, Arber Zela, Timur Carstensen, and Frank Hutter. Tempopfn: Synthetic pre-training of linear rnns for zero-shot time series forecasting, 2026. URL https://arxiv.org/abs/2510.25502.

Marc Nerlove, David M. Grether, and Jose L. Carvalho. *Analysis of Economic Time Series: A Synthesis*. Academic Press, 2014.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Kanghui Ning, Zijie Pan, Yu Liu, Yushan Jiang, James Y. Zhang, Kashif Rasul, Anderson Schneider, Lintao Ma, Yuriy Nevmyvaka, and Dongjin Song. TS-RAG: Retrieval-augmented generation based time series foundation models are stronger zero-shot forecaster. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=PymOnHw4Ty.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt. Featured Certification.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In *8th International Conference on Learning Representations, ICLR 2020*, 2020. URL https://openreview.net/forum?id=r1ecqn4YwB.

Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. Meta-learning framework with applications to zero-shot time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 9242–9250, 2021.

Boris N. Oreshkin, Mayank Jauhari, Ravi Kiran Selvam, Malcolm Wolff, Wenhao Pan, Shankar Ramasubramanian, Kin G. Olivares, Tatiana Konstantinova, Andres Potapczynski, Mengfei Cao, Dmitry Efimov, Michael W. Mahoney, and Andrew G. Wilson. Zero-shot forecasting by simulation alone. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2601.00970.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Guilherme Penedo, Hynek Kydlíček, Loubna Ben allal, Anton Lozhkov, Margaret Mitchell, Colin Raffel, Leandro Von Werra, and Thomas Wolf. The fineweb datasets: Decanting the web for the finest text data at scale. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. URL https://openreview.net/forum?id=n6SCkn2QaG.

Robert B. Penfold and Fang Zhang. Use of interrupted time series analysis in evaluating health care quality improvements. *Academic Pediatrics*, 13(6):S38–S44, 2013.

Zhongzheng Qiao, Chenghao Liu, Yiming Zhang, Ming Jin, Quang Pham, Qingsong Wen, Ponnuthurai Nagaratnam Suganthan, Xudong Jiang, and Savitha Ramasamy. Multi-scale finetuning for encoder-based time series foundation models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=OPOBV0zXu7.

Dalin Qin, Chenxi Wang, Qingsong Wen, Weiqi Chen, Liang Sun, and Yi Wang. Personalized federated darts for electricity load forecasting of individual buildings. *IEEE Transactions on Smart Grid*, 14(6): 4888–4901, 2023. doi: 10.1109/TSG.2023.3253855.

Guo Qin, Zhi Chen, Yong Liu, Zhiyuan Shi, Haixuan Liu, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Cora: Covariate-aware adaptation of time series foundation models, 2025. URL https://arxiv.org/abs/2510.12681.

Yunzhong Qiu, Zhiyao Cen, Zhongyi Pei, Chen Wang, and Jianmin Wang. Adapt data to model: Adaptive transformation optimization for domain-shared time series foundation models. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=uTK1SNgi1N.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training, 2018. URL https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners, 2019. URL https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020. URL http://jmlr.org/papers/v21/20-074.html.

Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/5cf68969fb67aa6082363a6d4e6468e2-Paper.pdf.

Kashif Rasul, Arjun Ashok, Andrew Robert Williams, Arian Khorasani, George Adamopoulos, Rishika Bhagwatkar, Marin Biloš, Hena Ghonia, Nadhir Hassen, Anderson Schneider, Sahil Garg, Alexandre Drouin, Nicolas Chapados, Yuriy Nevmyvaka, and Irina Rish. Lag-llama: Towards foundation models for time series forecasting. In *R0-FoMo:Robustness of Few-shot and Zero-shot Learning in Large Foundation Models*, 2023. URL https://openreview.net/forum?id=jYluzCLFDM.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollar, and Christoph Feichtenhofer. SAM 2: Segment anything in images and videos. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Ha6RTeWMd0.

Amal Saadallah and Abdulaziz Al-Ademi. Adaptive fine-tuning via pattern specialization for deep time series forecasting. In *Advanced Analytics and Learning on Temporal Data: 10th ECML PKDD Workshop, AALTD 2025, Porto, Portugal, September 19, 2025, Revised Selected Papers*, pp. 103–119, Berlin, Heidelberg, 2025. Springer-Verlag. ISBN 978-3-032-15534-4. doi: 10.1007/978-3-032-15535-1_7. URL https://doi.org/10.1007/978-3-032-15535-1_7.

David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. Deepar: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020. ISSN 0169-2070. doi: https://doi.org/10.1016/j.ijforecast.2019.07.001. URL https://www.sciencedirect.com/science/article/pii/S0169207019301888.

Rajat Sen, Hsiang-Fu Yu, and Inderjit Dhillon. Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2019. Curran Associates Inc.

Oleksandr Shchur, Caner Turkmen, Nick Erickson, Huibin Shen, Alexander Shirkov, Tony Hu, and Yuyang Wang. AutoGluon-TimeSeries: AutoML for probabilistic time series forecasting. In *International Conference on Automated Machine Learning*, 2023.

Oleksandr Shchur, Abdul Fatir Ansari, Caner Turkmen, Lorenzo Stella, Nick Erickson, Pablo Guerron, Michael Bohlke-Schneider, and Yuyang Wang. fev-bench: A realistic benchmark for time series forecasting, 2026. URL https://arxiv.org/abs/2509.26468.

Lefei Shen, Mouxiang Chen, Xu Liu, Han Fu, Xiaoxue Ren, Jianling Sun, Zhuo Li, and Chenghao Liu. Visionts++: Cross-modal time series foundation model with continual pre-trained vision backbones, 2025. URL https://arxiv.org/abs/2508.04379.

Hao-Nan Shi, Ting-Ji Huang, Lu Han, De-Chuan Zhan, and Han-Jia Ye. One-embedding-fits-all: Efficient zero-shot time series forecasting by a model zoo, 2025. URL https://arxiv.org/abs/2509.04208.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. Dinov3, 2025. URL https://arxiv.org/abs/2508.10104.

Slawek Smyl. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *International Journal of Forecasting*, 07 2019.

Gaurang Sonkavde, Deepak Sudhakar Dharrao, Anupkumar M. Bongale, Sarika T. Deokate, Deepak Doreswamy, and Subraya Krishna Bhat. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. *International Journal of Financial Studies*, 11(3), 2023. ISSN 2227-7072. doi: 10.3390/ijfs11030094. URL https://www.mdpi.com/2227-7072/11/3/94.

Antti Sorjamaa, Jin Hao, Nima Reyhani, Yongnan Ji, and Amaury Lendasse. Methodology for long-term prediction of time series. *Neurocomputing*, 70(16):2861–2869, 2007. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2006.06.015. URL https://www.sciencedirect.com/science/article/pii/S0925231207001610. Neural Network Applications in Electrical Engineering Selected papers from the 3rd International Work-Conference on Artificial Neural Networks (IWANN 2005).

Simon Stevenson. A comparison of the forecasting ability of arima models. *Journal of Property Investment & Finance*, 25(3):223–240, 2007. doi: 10.1108/14635780710746902.

Ege Onur Taga, Muhammed Emrullah Ildiz, and Samet Oymak. TimePFN: Effective multivariate time series forecasting with synthetic data. In *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024. URL https://openreview.net/forum?id=A9iqHtj3dk.

Guiyao Tie, Zeli Zhao, Dingjie Song, Fuyang Wei, Rong Zhou, Yurou Dai, Wen Yin, Zhejian Yang, Jiangyue Yan, Yao Su, Zhenhan Dai, Yifeng Xie, Yihan Cao, Lichao Sun, Pan Zhou, Lifang He, Hechang Chen, Yu Zhang, Qingsong Wen, Tianming Liu, Neil Zhenqiang Gong, Jiliang Tang, Caiming Xiong, Heng Ji, Philip S. Yu, and Jianfeng Gao. A survey on post-training of large language models, 2025. URL https://arxiv.org/abs/2503.06072.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. Llama: Open and efficient foundation language models, 2023. URL https://arxiv.org/abs/2302.13971.

Artur Trindade. Electricity Load Diagrams 2011-2014. UCI Machine Learning Repository, 2015. DOI: https://doi.org/10.24432/C58C86.

Ruey S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, 2005.

Ruey S. Tsay. *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley Series in Probability and Statistics. Wiley, 2014.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. Chattime: A unified multimodal time series foundation model bridging numerical and textual data, 2024. URL https://arxiv.org/abs/2412.11376.

Chengsen Wang, Qi Qi, Zhongwen Rao, Lujia Pan, Jingyu Wang, and Jianxin Liao. Chronosteer: Bridging large language model and time series foundation model via synthetic data, 2025. URL https://arxiv.org/abs/2505.10083.

Andrew Robert Williams, Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Jithendaraa Subramanian, Roland Riachi, James Requeima, Alexandre Lacoste, Irina Rish, Nicolas Chapados, and Alexandre Drouin.

Context is key: A benchmark for forecasting with essential textual information. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=ih2WuBT1Fn.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=Yd8eHMY1wz.

Dongxia Wu, Yilin Shen, and Hongxia Jin. Uncertainty-aware fine-tuning on time series foundation model for anomaly detection, 2024. URL https://openreview.net/forum?id=W1wlE4bPqP.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=J4gRj6d5Qm.

Shifeng Xie, Vasilii Feofanov, Ambroise Odonnat, Lei Zan, Marius Alonso, Jianfeng Zhang, Themis Palpanas, Lujia Pan, Keli Zhang, and Ievgen Redko. Cauker: Classification time series foundation models can be pretrained on synthetic data, 2026a. URL https://arxiv.org/abs/2508.02879.

Shifeng Xie, Ziwei Li, Themis Palpanas, Peilin Zhao, and Chenghao Liu. Time series foundation models improve LLM decisions: A case study in stock trading. In *1st ICLR Workshop on Time Series in the Age of Large Models*, 2026b. URL https://openreview.net/forum?id=XrGg8w03tJ.

Stephan Xie, Ben Cohen, Mononito Goswami, Junhong Shen, Emaad Khwaja, Chenghao Liu, David Asker, Othmane Abou-Amal, and Ameet Talwalkar. ARFBench: Benchmarking multimodal time series reasoning for software incident response. In *1st ICLR Workshop on Time Series in the Age of Large Models*, 2026c. URL https://openreview.net/forum?id=C4AXJvsgT6.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jing Zhou, Jingren Zhou, Junyang Lin, Kai Dang, Keqin Bao, Kexin Yang, Le Yu, Lianghao Deng, Mei Li, Mingfeng Xue, Mingze Li, Pei Zhang, Peng Wang, Qin Zhu, Rui Men, Ruize Gao, Shixuan Liu, Shuang Luo, Tianhao Li, Tianyi Tang, Wenbiao Yin, Xingzhang Ren, Xinyu Wang, Xinyu Zhang, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yinger Zhang, Yu Wan, Yuqiong Liu, Zekun Wang, Zeyu Cui, Zhenru Zhang, Zhipeng Zhou, and Zihan Qiu. Qwen3 technical report, 2025. URL https://arxiv.org/abs/2505.09388.

Yiyuan Yang, Chaoli Zhang, Tian Zhou, Qingsong Wen, and Liang Sun. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3033–3045, 2023.

Qingren Yao, Ming Jin, Chengqi Zhang, Chao-Han Huck Yang, Jun Qi, and Shirui Pan. Estimating time series foundation model transferability via in-context learning, 2025. URL https://arxiv.org/abs/2509.23695.

Samuel Yoon, Jongwon Kim, Juyoung Ha, and Young Myoung Ko. Momemto: Patch-based memory gate model in time series foundation model, 2026. URL https://arxiv.org/abs/2509.18751.

Annan Yu, Danielle C. Maddix, Boran Han, Xiyuan Zhang, Abdul Fatir Ansari, Oleksandr Shchur, Christos Faloutsos, Andrew Gordon Wilson, Michael W. Mahoney, and Bernie Wang. Understanding transformers for time series: Rank structure, flow-of-ranks, and compressibility. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=axR2KZwaD3.

Fangxu Yu, Hongyu Zhao, and Tianyi Zhou. Ts-reasoner: Aligning time series foundation models with llm reasoning, 2025a. URL https://arxiv.org/abs/2510.03519.

Xiaoyun Yu, Li fan, Xiangfei Qiu, Nanqing Dong, Yonggui Huang, Honggang Qi, Geguang Pu, Wanli Ouyang, Xi Chen, and Jilin Hu. Memts: Internalizing domain knowledge via parameterized memory for retrieval-free domain adaptation of time series foundation models, 2026b. URL https://arxiv.org/abs/2602.13783.

Yongbo Yu, Jari Peeperkorn, Johannes De Smedt, and Jochen De Weerdt. Time series foundation models for process model forecasting. *arXiv preprint arXiv:2512.07624*, 2025b. URL https://arxiv.org/abs/2512.07624.

Haoran Zhang, Yong Liu, Yunzhong Qiu, Haixuan Liu, Zhongyi Pei, Jianmin Wang, and Mingsheng Long. Timesbert: A bert-style foundation model for time series understanding, 2025a. URL https://arxiv.org/abs/2502.21245.

Huanyu Zhang, Chang Xu, Yi-Fan Zhang, Zhang Zhang, Liang Wang, and Jiang Bian. Timeraf: Retrieval-augmented foundation model for zero-shot time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 37(9):5654–5665, 2025b. doi: 10.1109/TKDE.2025.3579137.

Yingying Zhang, Zhengxiong Guan, Huajie Qian, Leili Xu, Hengbo Liu, Qingsong Wen, Liang Sun, Junwei Jiang, Lunting Fan, and Min Ke. CloudRCA: A root cause analysis framework for cloud computing platforms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 4373–4382, 2021.

Zexing Zhang, Qingxin Zhao, and Huimin Lu. Efficient compression of time-series foundation models via consensus subspace distillation, 2025c. URL https://openreview.net/forum?id=zOgvgw5sia.

Lifan Zhao, Yanyan Shen, Zhaoyang Liu, Xue Wang, and Jiaji Deng. Less is more: Unlocking specialization of time series foundation models via structured pruning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2026. URL https://openreview.net/forum?id=jy4bBsr1Jc.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srini Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, LILI YU, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, and Omer Levy. LIMA: Less is more for alignment. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=KBMOKmX2he.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Zhaoyang Zhu, Weiqi Chen, Rui Xia, Tian Zhou, Peisong Niu, Bingqing Peng, Wenwei Wang, Hengbo Liu, Ziqing Ma, Qingsong Wen, and Liang Sun. eforecaster: unifying electricity forecasting with robust, flexible, and explainable machine learning algorithms. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i13.26853. URL https://doi.org/10.1609/aaai.v37i13.26853.