SIFT: ENHANCING TIME SERIES FOUNDATION MODELS VIA SEMANTIC INVARIANCE AND STRUCTURAL FIDELITY FINE-TUNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time Series Foundation Models (TSFMs) have achieved remarkable zero-shot performance through extensive pre-training on massive time series datasets. Nevertheless, due to the low-dimensional properties and diverse structural patterns of time series data, performing naive fine-tuning on TSFMs often leads to overfitting and falling into the mean-prediction trap. To address these challenges, we propose **SIFT**, a robust adaptation method that enhances time series foundation models by preserving **S**emantic **I**nvariance and structural **F**idelity throughout the fine-**T**uning process. We employ semantic-invariant adversarial augmentation, which utilizes semantic spectrum decomposition to partition the semantic space and then generates perturbations within the non-core semantic subspace to bolster the model's robustness against these perturbations, mitigating overfitting. We implement a component-based structural fidelity enhancement, which facilitates component-wise mixup and imposes a reconstruction objective to improve the model's ability to preserve structural fidelity, alleviating the mean-prediction trap. Extensive experiments on representative TSFMs covering 10 real-world datasets demonstrate that **SIFT** can significantly enhance the performance of TSFMs.

1 INTRODUCTION

Recently, time series foundation models (TSFMs) have emerged as a promising paradigm for time series analysis (Qiu et al., 2025; Liu et al., 2026; Li et al., 2025a). By pre-training on massive time series datasets, TSFMs acquire a broad understanding of temporal patterns, enabling them to deliver impressive zero-shot forecasting performance (Woo et al., 2024; Goswami et al., 2024; Liu et al., 2024a). However, real-world time series frequently exhibit subtle temporal characteristics and domain-dependent variations that are difficult to fully capture during pre-training. As a result, such zero-shot forecasting often offers suboptimal results in practical applications. To unlock the full potential of TSFMs, a dedicated fine-tuning method is essential to adapt the model for the nuanced and intricate temporal patterns inherent in specialized downstream scenarios.

Currently, a broad range of parameter-efficient fine-tuning (PEFT) methods have been proposed, including LoRA (Hu et al., 2022), AdaLoRA (Zhang et al., 2023), Prefix Tuning (Li & Liang, 2021), and Linear Probing (Kumar et al., 2022). These methods are primarily developed in NLP and CV and have been widely adapted to diverse downstream tasks. More recently, TSFM-specific approaches such as MSFT (Qiao et al., 2025) have been introduced to better accommodate time series by explicitly integrating multi-scale(Wang et al., 2023) modeling into the fine-tuning process. However, existing methods still struggle to preserve semantic consistency and fine-grained structural information in time series data during fine-tuning. Several critical challenges therefore remain unresolved.

Challenge 1: TSFMs are prone to overfitting when adapting to low-dimensional time series. TSFMs are typically pre-trained with large-scale architectures to learn transferable representations across diverse time-series domains. However, downstream tasks often involve time series exhibiting intrinsic low-dimensional structures (Huang et al., 2025; Yu et al., 2025), where core semantics, such as long-term trends and seasonal patterns, are concentrated in a sparse subspace. This mismatch between the high-capacity pre-trained model and the low-dimensional downstream data increases the risk of overfitting to spurious patterns or non-transferable variations during fine-tuning. Such

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

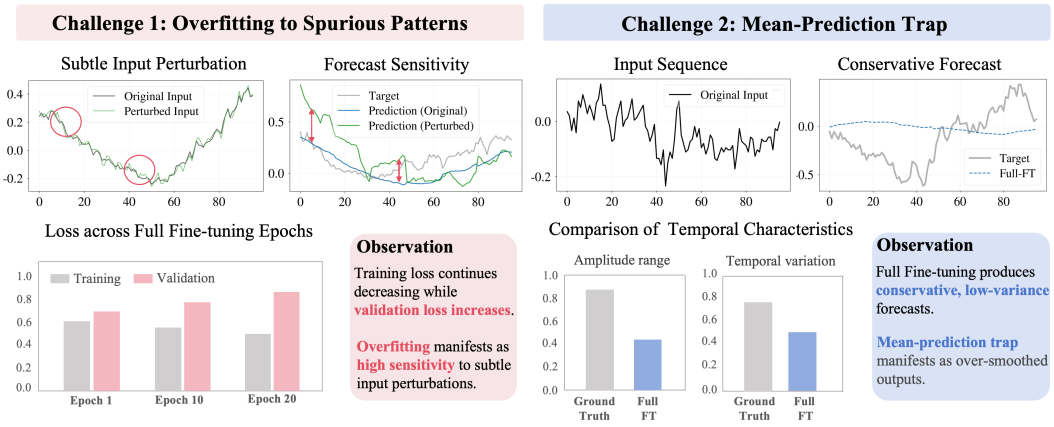


Figure 1: Two primary bottlenecks in TSFM fine-tuning. (1) Overfitting to intrinsic low-dimensional structures results in high sensitivity to subtle input perturbations. (2) Downstream scenarios that deviate significantly from pre-training patterns lead to over-smoothed outputs and structural absence.

overfitting often manifests as high sensitivity to input perturbations. As shown in Figure 1, even subtle perturbations in the input can induce disproportionately large deviations in the output space, indicating poor robustness to local fluctuations or minor distribution shifts introduced during fine-tuning. Consequently, this fragility hinders effective transfer of pre-trained temporal knowledge and semantic representations (Mukhoti et al., 2024) to downstream tasks (Chen et al., 2019).

Challenge 2: TSFMs are susceptible to the mean-prediction trap under domain divergence. While TSFMs internalize transferable temporal priors through massive pre-training, their generalization capability is often compromised when downstream datasets deviate substantially from those pre-training distributions. Modern TSFMs employ diverse forecasting objectives, including pointwise regression losses, likelihood-based distribution modeling, and quantile regression. Despite these different formulations, they primarily optimize numerical prediction accuracy or predictive distribution quality during fine-tuning. Under domain divergence, this lack of explicit structural preservation may lead the model to underrepresent fine-grained temporal dynamics, resulting in conservative, low-variance forecasts—a phenomenon referred to as the mean-prediction trap (Wan et al., 2026). As illustrated in Figure 1, this manifests as pronounced over-smoothing (Sun et al., 2025), suppressed peaks, and a tendency to regress toward the average trajectory. Such behavior undermines the structural fidelity of forecasts, limiting the model's ability to capture high-amplitude temporal variations when downstream patterns diverge from the pre-trained priors.

To address **Challenge 1**, we propose a **semantic-invariant adversarial augmentation** strategy to improve model robustness against perturbations that do not alter core semantics, reducing overfitting to spurious patterns and non-transferable variations. There are two key components in this strategy: (1) *semantic spectrum decomposition*, which partitions the semantic space into subspaces to identify the perturbation space for the adversarial attacks; and (2) *adversarial perturbation augmentation*, which first produces input perturbations that maximally expose the model's sensitivity while preserving the core semantics. During fine-tuning, we minimize the discrepancy between the model's predictions on original and perturbed inputs, compelling the model to prioritize core semantic features over non-core variations. This strategy enables the model to capture core semantic features effectively, while improving robustness by reducing sensitivity to input perturbations.

To address **Challenge 2**, we propose a **component-based structural fidelity enhancement** framework to mitigate the mean-prediction trap by guiding the model to reconstruct high-fidelity structures. To prevent the model from collapsing into conservative, low-variance forecasts, we adopt *structural component mixup* to synthesize structurally diverse hard samples, which first prioritizes high-volatility samples then decomposes these samples into latent trend and seasonal components to facilitate component-wise mixup. Furthermore, we impose *component-wise fidelity reconstruction objective* to provide fine-grained structural supervision, encouraging the model to preserve structural fidelity. Therefore, the model acquires the capacity to track volatile fluctuations, alleviating the over-smoothing and the mean-prediction phenomenon.

Overall, we propose SIFT, a fine-tuning method for time series foundation models that preserves semantic invariance and structural fidelity, enabling stable adaptation to diverse downstream time series forecasting tasks. Our primary contributions are summarized as follows:

- We propose a semantic-invariant adversarial augmentation strategy to reduce overfitting to spurious patterns and prioritize core semantics, ensuring robust fine-tuning of TSFMs.
- We develop a component-based structural fidelity enhancement to bolster model adaptability to distributional shifts, mitigating over-smoothing and the mean-prediction trap.
- Extensive experiments on 10 real-world downstream datasets using 4 representative pre-trained models demonstrate that our method outperforms existing baselines, achieving state-of-the-art performance in the fine-tuning of TSFMs for time series forecasting.

2 Related Work

Time Series Foundation Model. Recent years have witnessed a surge in time series foundation models (TSFMs). Existing architectures can be broadly categorized into three primary architectures: (1) Encoder-only paradigms: MOIRAI (Woo et al., 2024) unifies diverse multivariate inputs through sequence flattening, while MOMENT (Goswami et al., 2024) and UniTS (Gao et al., 2024) utilize masked reconstruction and task-specific tokenization to capture high-level features across varying domains. (2) Decoder-only architectures: Inspired by Large Language Models, frameworks like TimesFM (Das et al., 2024) and Timer (Liu et al., 2024b) treat forecasting as an autoregressive process. (3) Encoder-decoder and generative variants: This category focuses on integrating structural or domain-specific priors into the generative process. Notable examples include TTM (Ekambaram et al., 2024), which leverages multi-resolution sampling for cross-channel patterns, and ROSE (Wang et al., 2024), which emphasizes frequency-domain decomposition. Furthermore, Sundial (Liu et al., 2025) introduces a generative Transformer framework for modeling continuous temporal flows. However, while TSFMs have achieved remarkable breakthroughs in zero-shot scenarios, how to effectively perform fine-tuning on downstream tasks remains an under-explored yet critical problem.

Fine-tuning for TSFMs. Methods for fine-tuning TSFMs have largely inherited Parameter-Efficient Fine-Tuning (PEFT) techniques from NLP and CV, which can be broadly categorized into three main paradigms: (1) Selective fine-tuning, such as Linear Probing (Kumar et al., 2022), which freezes the backbone and updates only the task-specific output head; (2) Additive fine-tuning, notably Prefix Tuning (Li & Liang, 2021), which introduces trainable continuous vectors to hidden states to guide the model's representations; and (3) Reparameterization-based fine-tuning, including LoRA (Hu et al., 2022) and its variant AdaLoRA (Zhang et al., 2023), which adapt internal weights through low-rank decomposition and importance-based budget allocation. While efficient, these general-purpose strategies are often insufficient to handle the unique non-stationarity of temporal signals. Consequently, recent research has pivoted toward paradigms tailored specifically for the temporal domain, notably MSFT (Qiao et al., 2025). By adopting a causal perspective, MSFT models multiple temporal scales (Lester et al., 2021; Liu et al., 2022) to leverage the forecasting potential of backbones. Despite these advances, existing paradigms often still overlook the overfitting and over-smoothing phenomenon during practical downstream model adaptation.

3 Methodology

Problem Formulation. In multivariate time series forecasting, the goal is to predict future values $Y \in \mathbb{R}^{H \times C}$ based on historical observations $X \in \mathbb{R}^{L \times C}$, where L, H, and C denote the look-back length, forecasting horizon, and number of variates, respectively. In the context of fine-tuning, we leverage a time series foundation model $f(\cdot)$ with parameters θ, which has been pre-trained on large-scale source datasets to capture general temporal patterns. Given a target dataset $\mathcal{D} = \{(X, Y)\}$ from a specific downstream domain, our objective is to fine-tune θ on the target domain, optimizing the model to minimize forecasting error and achieve superior performance on downstream tasks.

We propose SIFT, a fine-tuning method for time series foundation models that preserves semantic invariance and structural fidelity (Figure 2) : (1) The **semantic-invariant adversarial augmentation** strategy is designed to bolster model robustness and mitigate overfitting by enhancing the model's ability to capture core semantic features. A semantic spectrum decomposition is employed to partition

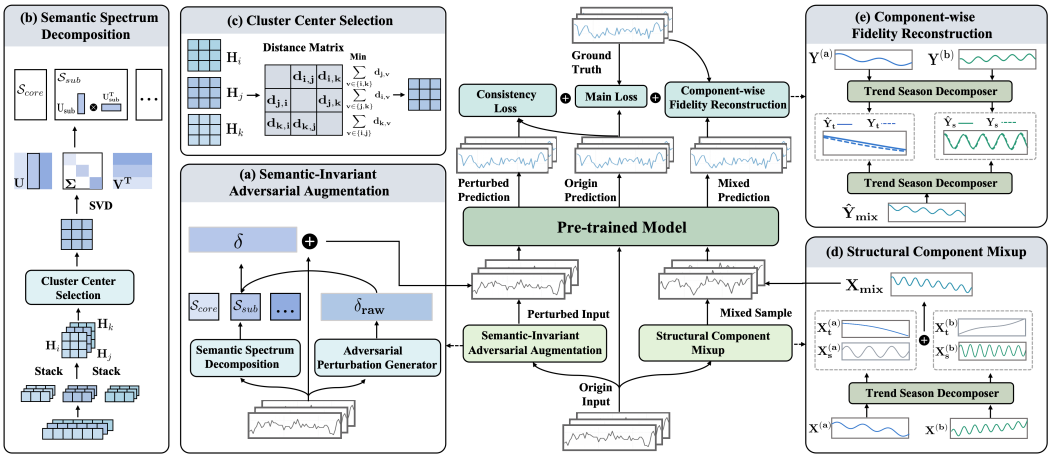


Figure 2: The framework of SIFT, comprising two main modules: semantic-invariant adversarial augmentation (SIAA) to mitigate overfitting and component-based structural fidelity enhancement (CSFE) to preserve temporal structure through component mixup and reconstruction.

the semantic space. Then, an adversarial perturbation generator produces perturbations to maximize TSFM's prediction divergence between the original and perturbed inputs, while a subspace projection using the predefined partitioning ensures these perturbations are confined to the subordinate semantic space to preserve semantic invariance. During fine-tuning, the discrepancy between the model's predictions for original and perturbed samples is minimized. (2) The **component-based structural fidelity enhancement** is proposed to mitigate the mean-prediction trap by compelling the model to reconstruct structural details. We first identify training samples exhibiting high temporal volatility. These samples perform trend-seasonal decomposition, followed by a structural component mixup that recombines their decomposed components to synthesize structurally diverse hard examples. Finally, we use a component-wise fidelity reconstruction loss to align the decomposed trend and seasonal components of the outputs with those of the ground-truth, preserving the sharp structural fidelity.

3.1 Semantic-Invariant Adversarial Augmentation

To enhance the robustness of TSFMs during fine-tuning and mitigate overfitting, we propose a semantic-invariant adversarial augmentation strategy. The key idea is to generate perturbations within a semantically constrained subspace, preserving core semantic features while amplifying non-core variations to expose the model's sensitivity. During fine-tuning, we enforce prediction consistency under such subspace-constrained perturbations, encouraging the model to focus on the stable core semantics rather than spurious patterns or non-transferable variations.

Semantic Spectrum Decomposition. Instead of introducing perturbations directly in the raw signal space that risks degrading essential semantic structures, we first perform a semantic decomposition to disentangle the underlying structured components of the input. For simplicity, we describe this process using a single time series channel. Given a time series input $X \in \mathbb{R}^L$, we define a trajectory embedding operator Ψ that maps 1D time series X into a trajectory matrix $\mathbf{H}$ (Rukhin, 2002):

$$\mathbf{H} = \Psi(X) = \begin{bmatrix} X(1) & X(1+S) & \cdots & X(1+(K-1)S) \\ X(2) & X(2+S) & \cdots & X(2+(K-1)S) \\ \vdots & \vdots & \ddots & \vdots \\ X(W) & X(W+S) & \cdots & X(W+(K-1)S) \end{bmatrix} \in \mathbb{R}^{W \times K} \tag{1}$$

where W and S denote the window width and stride, respectively, and the column dimension K is given by $K = \lfloor \frac{L-W}{S} \rfloor + 1$. This transformation converts the original temporal signal into a matrix representation that explicitly exposes local temporal dependencies and structural patterns.

Inspired by prior spectral analyses (Chen et al., 2019), we perform singular value decomposition (SVD) on the trajectory matrix to characterize its internal spectral structure:

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{2}$$

where $\mathbf{U} \in \mathbb{R}^{W\times d}$ and $\mathbf{V} \in \mathbb{R}^{K\times d}$ are orthonormal matrices containing the left and right singular vectors, $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_d)$ with $\sigma_1 \geq \cdots \geq \sigma_d \geq 0$, and $d = \min(W, K)$. To obtain a semantic partition of the spectrum, we define two rank thresholds, $r_1 = \lfloor d\cdot\rho_1 \rfloor$ and $r_2 = \lfloor d\cdot\rho_2 \rfloor$, where scaling factors $0 < \rho_1 < \rho_2 \leq 1$. These thresholds divide the left singular vectors (indexed by $i = 1, \ldots, d$) into three groups, $\mathbf{U}_{\text{core}} = [\mathbf{u}_1, \ldots, \mathbf{u}_{r_1}]$, $\mathbf{U}_{\text{sub}} = [\mathbf{u}_{r_1+1}, \ldots, \mathbf{u}_{r_2}]$, and $\mathbf{U}_{\text{noise}} = [\mathbf{u}_{r_2+1}, \ldots, \mathbf{u}_d]$, spanning the core subspace ($\mathcal{S}_{\text{core}}$), the subordinate subspace ($\mathcal{S}_{\text{sub}}$), and the noise subspace ($\mathcal{S}_{\text{noise}}$).

According to the *Eckart-Young-Mirsky theorem* (Eckart & Young, 1936), the optimal rank-r_1 approximation of $\mathbf{H}$, which we denote as $\mathbf{H}_{\text{core}}$, is given by truncating the SVD:

$$\mathbf{H}_{\text{core}} = \arg\min_{\text{rank}(\mathbf{H}')=r_1} \|\mathbf{H} - \mathbf{H}'\|_F = \sum_{i=1}^{r_1} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \tag{3}$$

with a minimum Frobenius norm error of $\sqrt{\sum_{i=r_1+1}^{d} \sigma_i^2}$. This indicates that the subspace $\mathcal{S}_{\text{core}}$ captures the dominant semantic structures, which typically correspond to long-term trends and seasonal patterns. The noise boundary r_2 is inspired by the *Marchenko-Pastur (MP) law* (Marčenko & Pastur, 1967), which provides an empirical approximation for separating signal and noise in spectral analysis. The singular values associated with noise are expected to concentrate below the threshold:

$$\lambda_+ = \hat{\sigma}_\epsilon \left(1 + \sqrt{\frac{W}{K}}\right), \tag{4}$$

where $\hat{\sigma}_\epsilon$ is estimated from the tail of the spectrum, and W and K denote the dimensions of the matrix $\mathbf{H}$. We characterize r_2 as the largest index satisfying $\sigma_{r_2} > \lambda_+$. The subspace spanned by the singular vectors with indices greater than r_2, denoted $\mathcal{S}_{\text{noise}}$, consists of stochastic fluctuations.

The singular components indexed between $r_1 + 1$ and r_2 span the subordinate semantic subspace $\mathcal{S}_{\text{sub}}$. This subspace captures components that are neither core temporal semantics encoded in $\mathcal{S}_{\text{core}}$ nor stochastic fluctuations contained in $\mathcal{S}_{\text{noise}}$. Such components are more likely to encode spurious patterns or non-transferable variations, making the model particularly prone to overfitting to them. We therefore use this subspace as the primary domain for adversarial perturbation. For multivariate time series, we first cluster channels according to their temporal similarity and use the cluster centers as representative sequences. We then perform spectral stratification on each cluster center.

Adversarial Perturbation Augmentation. To enhance the robustness of the TSFM during fine-tuning, we design a two-stage adversarial framework. First, we train a generator to produce semantic-invariant perturbations that preserve the core semantic subspace while substantially disrupting the model's predictions. After the generator is trained, it is fixed and used to fine-tune the TSFM to maintain invariant predictions under such perturbations.

Stage 1: Generator Training. In this stage, the foundation model f is frozen. Our goal is to train a perturbation generator PG to produce effective adversarial perturbations that maximally disrupt the model's predictions while preserving temporal semantics.

The perturbation is generated as follows. Given an input X, the generator PG first produces a raw perturbation $\delta_{\text{raw}} = PG(X)$, where PG is a lightweight convolutional neural network (Ismail Fawaz et al., 2020; LeCun et al., 1998). To ensure semantic invariance, we project δ_{raw} into the subordinate semantic subspace using the projection matrix $\mathbf{P}_{\text{sub}} = \mathbf{U}_{\text{sub}}\mathbf{U}_{\text{sub}}^\top$, and the resulting semantic-invariant perturbation δ_{adv} is reconstructed as:

$$\delta_{\text{adv}} = \Psi^{-1}\left(\mathbf{P}_{\text{sub}} \cdot \Psi(\delta_{\text{raw}})\right) \tag{5}$$

where $\Psi(\cdot)$ denotes the trajectory operator that maps a time series into its trajectory matrix representation, and $\Psi^{-1}(\cdot)$ is the anti-diagonal averaging operator that reconstructs a time series from a trajectory matrix by averaging along its anti-diagonals. The final perturbed input is then $X_{\text{adv}} = X + \delta_{\text{adv}}$. To constrain leakage into the core semantic space, we define $\delta_{\text{core}} = \Psi^{-1}(\mathbf{U}_{\text{core}}\mathbf{U}_{\text{core}}^T\Psi(\delta_{\text{raw}}))$, which represents the projection of the raw perturbation into the core semantic subspace.

With X_{adv} and δ_{core} defined above, PG is optimized to minimize the following adversarial loss:

$$\min_{\theta_{PG}} \mathcal{L}_{PG} = -\underbrace{\|f(X_{\text{adv}}) - f(X)\|_2^2}_{\text{Adversarial Disruption}} + \underbrace{\|\delta_{\text{core}}\|_2^2}_{\text{Core Leakage Penalty}}. \tag{6}$$

Stage 2 Robust Fine-tuning. Once the generator is trained, the generator PG is fixed to serve as a dedicated module for augmentation, producing semantic-invariant perturbed inputs. During fine-tuning, these inputs are used to train the model, enforcing prediction consistency under such perturbations. To achieve this, we minimize the following consistency loss:

$$\mathcal{L}_{\text{consis}} = \|f(X_{\text{adv}}) - f(X)\|_2^2 \tag{7}$$

This encourages the model to focus on core semantic features rather than spurious patterns, thereby maintaining stable predictions, mitigating overfitting, and enhancing robustness.

3.2 Component-based Structural Fidelity Enhancement

Structural Component Mixup. To mitigate the mean-prediction failure mode where models produce overly smooth outputs, we first prioritize samples with high temporal volatility as they contain richer temporal dynamics beyond smooth trends. The volatility of a sample $X \in \mathbb{R}^{L\times C}$ is characterized by the Total Variation (TV) score, which measures the cumulative magnitude of temporal shifts:

$$TV(X) = \frac{1}{C}\sum_{c=1}^{C}\sum_{t=1}^{L-1}|x_{c,t+1} - x_{c,t}| \tag{8}$$

where $x_{c,t}$ denotes the observation value of the c-th channel at time step t. By ranking samples within each training batch according to their TV scores, we select the top-k most volatile segments. Formally, we define the indices of these selected segments as the high-volatility set I_{top}, which serves as the foundation samples for the structural mixup.

The mixed sample X_{mix} is synthesized by cross-pairing the constituent components of an anchor sample $X^{(a)}$ and an arbitrary sample $X^{(b)}$, both sampled from I_{top} ($a \neq b$). Using the Series Decomposition (SD) operation (Wu et al., 2021), we decouple them into their respective Trend (X_t) and Seasonal (X_s) components, denoted as $X_t, X_s = \text{SD}(X)$. The mixed sample X_{mix} is then synthesized by recombining these components:

$$X_{\text{mix}} = X_t^{(a)} + X_s^{(b)} \tag{9}$$

Component-wise Fidelity Reconstruction. To provide fine-grained structural supervision, we explicitly supervise that the model's prediction for the mixed input remains faithful to the corresponding structural components of its original sources. Given the prediction $\hat{Y}_{\text{mix}} = f(X_{\text{mix}})$, we first extract its predicted trend and seasonal components:

$$\hat{Y}_t, \hat{Y}_s = \text{SD}(\hat{Y}_{\text{mix}}) \tag{10}$$

The structural reinforcement loss is then defined by aligning these predicted components with their corresponding cross-sample ground truths:

$$\mathcal{L}_t = \left\|\hat{Y}_t - Y_t^{(a)}\right\|_1, \quad \mathcal{L}_s = \left\|\hat{Y}_s - Y_s^{(b)}\right\|_2^2 \tag{11}$$

We denote $Y_t^{(a)}$ and $Y_s^{(b)}$ as the ground-truth trend and seasonal components from source samples a and b, respectively. To balance the contributions of different structural components, we introduce an adaptive scaling factor $\alpha = \|X_s\|/\|X_t\|$, which normalizes the relative magnitudes of the seasonal and trend components. The integrated reconstruction loss is then formulated as:

$$\mathcal{L}_{\text{CFR}} = \mathcal{L}_s + \alpha \cdot \mathcal{L}_t \tag{12}$$

By providing explicit supervision for individual structural attributes, the model is discouraged from defaulting to a conservative but inaccurate mean prediction. Instead, the separate objective on $\mathcal{L}_t$ and $\mathcal{L}_s$ forces the model to preserve the sharp structural fidelity of both long-term trends and periodic oscillations, effectively preventing the predictive blurring even under complex structural variations.

3.3 Training and Inference

Training Phase. During training, we fine-tune the foundation model f by minimizing a joint objective. The total loss $\mathcal{L}_{\text{total}}$ is a weighted combination of standard forecasting supervision, the proposed semantic consistency loss and the reconstruction loss:

$$\mathcal{L}_{\text{total}} = \|f(X) - Y\|_2^2 + \lambda_{\text{consis}}\mathcal{L}_{\text{consis}} + \lambda_{\text{CFR}}\mathcal{L}_{\text{CFR}} \tag{13}$$

The first term denotes the task-level forecasting loss, instantiated as MSE in our experiments. The consistency loss $\mathcal{L}_{\text{consis}}$ encourages stable predictions under input perturbations, while the reconstruction loss $\mathcal{L}_{\text{CFR}}$ preserves the structural fidelity. Together, these regularizers prevent the foundation model from overfitting to spurious variations and mitigate the mean-prediction trap.

Inference Phase. Although the adversarial perturbation generator and the structural component mixup operations are crucial for guiding model training, they are not used during inference. During inference, the fine-tuned foundation model f generates predictions directly through its learned weights, ensuring that there is no additional computational overhead.

4 Experiments

4.1 Experimental Details

Datasets And Metrics. Our evaluation is conducted on TSFM-Bench (Li et al., 2025b), which comprises ten representative forecasting benchmarks, including four ETT subsets, Exchange, ZafNoo, Weather, Electricity (ECL), Solar, and AQShunyi, covering a wide range of real-world scenarios. We quantify predictive performance using Mean Squared Error (MSE) and Mean Absolute Error (MAE). Extensive details of the benchmark datasets are provided in the Appendix A.1.

Pre-trained Models And Baselines. To evaluate the effectiveness and versatility of our framework, we choose four representative TSFMs as the backbone models: Moirai (Woo et al., 2024), UniTS (Gao et al., 2024), Timer (Liu et al., 2024b), and Sundial (Liu et al., 2025). We compare our proposed method with several widely-used fine-tuning and adaptation baselines, including Linear (Kumar et al., 2022), LoRA (Hu et al., 2022), AdaLoRA (Zhang et al., 2023), and MSFT (Qiao et al., 2025).

Implementation Details. We follow the experimental protocol of TSFM-Bench (Li et al., 2025b), using prediction horizons $\{96, 192, 336, 720\}$ and selecting the optimal lookback length from $\{96, 336, 512\}$ based on validation performance. All methods are trained with Adam (Kingma & Ba, 2015) for at most 20 epochs, using an initial learning rate of 1×10^{-4}, an annealing schedule, and early stopping with a patience of 3. For fair comparison, we follow their official implementations and tune the hyperparameters on the validation set to obtain the best configuration. All reported results are averaged over five independent runs with different random seeds.

4.2 Experimental Results and Analysis

Main Results. Table 1 compares our method with various baselines across 10 datasets, 4 foundation models, and 4 prediction horizons of $\{96, 192, 336, 720\}$. SIFT achieves the best fine-tuning performance in the vast majority of evaluated backbone-dataset settings, outperforming full fine-tuning and existing adaptation baselines across a broad range of scenarios. Notably, it reduces the average MSE of Timer by 11.9% compared with full fine-tuning, while lower average MSE is also achieved on UniTS, Moirai, and Sundial. These results demonstrate that SIFT generalizes well across diverse TSFM architectures. Across datasets, SIFT also demonstrates consistent improvements over a broad range of forecasting benchmarks. For instance, compared with full fine-tuning, SIFT achieves 17.0% lower MSE on Weather, 9.3% on ETTm2, and 8.0% on Solar with Timer. On Exchange, it reduces MSE by 10.4% and 14.8% for UniTS and Moirai, respectively.

Ablation Study. We evaluate the contributions of SIAA and CSFE by progressively replacing their key components with simpler alternatives. As shown in Table 2, for SIAA, we compare random augmentation, unconstrained adversarial perturbation, and adversarial perturbation with semantic subspace projection. Random augmentation yields only marginal gains, while unconstrained adversarial perturbation is unstable across datasets. Adding the semantic projection consistently improves performance, demonstrating SIAA's effectiveness in improving semantic-invariant robustness. For CSFE, we compare naive mixup with component-aware mixup and reconstruction. The latter achieves better performance, showing the benefit of explicitly preserving temporal structure. Finally, combining SIAA and CSFE yields the best overall results, demonstrating their complementary effects.

Effectiveness of the adversarial perturbation generator. We conduct a comparison on a fully fine-tuned UniTS checkpoint to evaluate the perturbations produced by our adversarial generator PG against stochastic noise. As illustrated in Figure 3, with both perturbations constrained to the same

Table 1: Comparative results for long-term multivariate forecasting. Metrics represent the average MSE across horizons $\{96, 192, 336, 720\}$. **Bold** indicates the best fine-tuning performance for each TSFM, and the overall best across all backbones is in **red**. Our method is highlighted in green. See Appendix A.8 for full results.

Method	ETTh1	ETTh2	ETTm1	ETTm2	Exchange	ZafNoo	Weather	Solar	ECL	AQShunyi	Avg
PatchTST	0.469	0.387	0.387	0.269	0.450	0.512	0.265	0.207	0.216	0.808	0.397
iTransformer	0.439	0.370	0.361	0.269	0.360	0.522	0.258	0.207	0.182	0.706	0.367
TimeMixer	0.436	0.364	0.356	0.406	0.383	0.518	0.226	0.193	0.185	0.715	0.378
TimesNet	0.468	0.390	0.407	0.292	0.406	0.536	0.255	0.211	0.190	0.727	0.388
FedFormer	0.433	0.406	0.567	0.335	0.501	0.575	0.312	0.641	0.218	0.763	0.475
Timer	0.465	0.394	0.431	0.289	0.370	0.645	0.271	0.225	0.168	0.788	0.405
+ *LoRA*	0.452	0.382	0.792	0.318	0.367	0.612	0.537	0.345	0.421	0.989	0.522
+ *AdaLoRA*	0.457	0.380	0.764	0.313	0.369	0.585	0.391	0.305	0.265	0.774	0.460
+ *Linear*	0.456	0.377	0.808	0.325	0.358	0.649	0.380	0.280	0.222	0.767	0.462
+ *MSFT*	0.421	0.391	0.364	0.271	0.446	**0.515**	0.244	0.209	0.181	0.750	0.379
+ ***Ours***	**0.390**	**0.376**	**0.342**	**0.262**	**0.351**	**0.515**	**0.225**	**0.207**	**0.164**	**0.733**	**0.357**
UniTS	0.450	0.408	0.422	0.313	0.396	0.523	0.245	0.188	0.166	0.802	0.391
+ *LoRA*	0.458	0.403	0.432	0.327	0.389	0.631	0.250	1.866	0.188	0.897	0.584
+ *AdaLoRA*	0.436	0.384	0.392	0.309	0.366	0.566	0.256	0.313	0.180	0.820	0.402
+ *Linear*	0.438	0.399	0.392	0.330	0.367	0.614	0.273	0.215	0.200	0.820	0.405
+ *MSFT*	0.434	0.380	0.390	0.286	0.360	0.557	0.241	0.237	0.184	0.807	0.388
+ ***Ours***	**0.424**	**0.361**	**0.380**	**0.285**	**0.355**	**0.508**	**0.240**	**0.177**	**0.164**	**0.785**	**0.368**
MoiRai	0.522	0.382	0.424	0.340	0.452	0.598	0.246	0.205	0.259	0.725	0.415
+ *LoRA*	0.574	0.431	0.443	0.323	0.460	0.665	0.350	1.868	0.173	0.948	0.624
+ *AdaLoRA*	0.540	0.372	0.432	0.292	0.567	0.576	0.243	0.206	0.173	0.653	0.405
+ *Linear*	0.520	0.385	0.445	0.293	0.485	0.590	0.251	0.284	0.182	**0.599**	0.404
+ *MSFT*	**0.475**	0.372	0.417	0.295	0.390	0.558	0.241	0.207	0.173	0.686	0.381
+ ***Ours***	0.489	**0.370**	**0.407**	**0.284**	**0.385**	**0.524**	**0.238**	**0.196**	**0.171**	0.713	**0.378**
Sundial	0.455	0.592	0.402	0.477	0.472	0.515	0.259	0.211	0.188	0.821	0.439
+ *LoRA*	0.468	**0.517**	0.435	0.459	0.473	0.647	0.308	1.921	0.288	0.985	0.650
+ *AdaLoRA*	0.512	0.776	0.642	0.753	0.510	0.594	0.456	0.308	0.195	**0.712**	0.546
+ *Linear*	0.560	0.870	0.627	0.521	0.504	0.565	0.278	0.334	0.199	0.715	0.517
+ *MSFT*	0.598	1.066	0.425	0.424	0.557	0.629	0.259	0.259	0.191	0.718	0.513
+ ***Ours***	**0.450**	0.550	**0.396**	**0.387**	**0.363**	**0.486**	**0.253**	**0.204**	**0.187**	**0.712**	**0.399**

Table 2: Ablation study of the proposed Semantic-Invariant Adversarial Augmentation (SIAA) and Component-based Structural Fidelity Enhancement (CSFE).

Method	SIAA Variant	CSFE Variant	ETTh1	ETTm1	Exchange	ZafNoo
Full-FT	–	–	0.450	0.422	0.396	0.523
+ Random Aug.	Random	–	0.431	0.389	0.386	0.521
+ Adv. w/o Proj.	Adv.	–	0.455	0.385	0.370	0.554
+ SIAA	Adv. + Proj.	–	0.427	0.382	0.360	0.517
+ Naive Mixup	–	Mixup	0.450	0.390	0.399	0.542
+ CSFE	–	Mixup + Decomp.	0.428	0.386	0.390	0.514
Ours Full	**Adv. + Proj.**	**Mixup + Decomp.**	**0.424**	**0.380**	**0.355**	**0.508**

magnitude, the learned perturbations induce substantially larger deviations from the original forecasts than stochastic noise. This suggests that P_G identifies model-sensitive temporal variations rather than introducing arbitrary disturbances.

Sensitivity Analysis. To analyze the influence of the loss coefficients, we conduct a sensitivity analysis by varying λ_{consis} and λ_{CFR} independently. We evaluate $\lambda_{\text{consis}} \in \{0, 5, 10, 15\}$ and $\lambda_{\text{CFR}} \in$

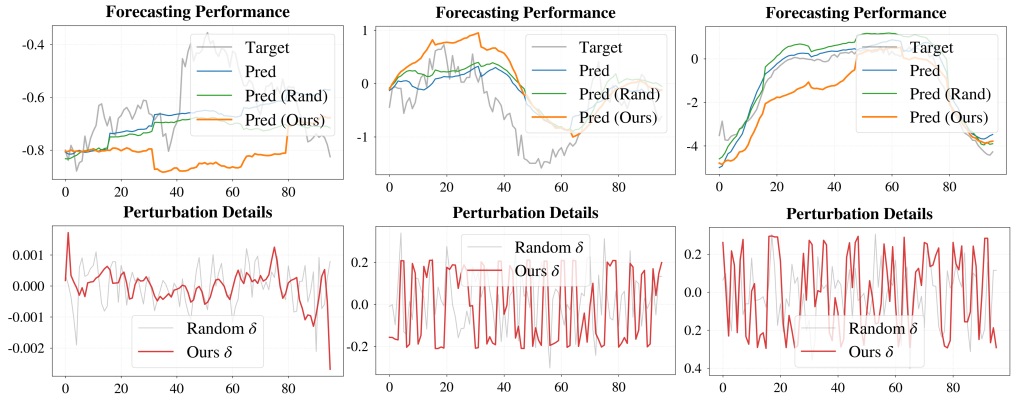
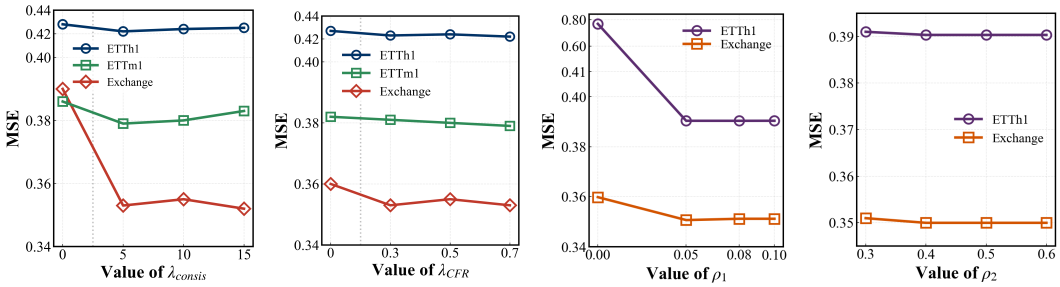


Figure 3: Effectiveness analysis of perturbation generator.



(a) Sensitivity of λ_{consis} (b) Sensitivity of λ_{CFR} (c) Sensitivity of ρ_1 (d) Sensitivity of ρ_2

Figure 4: Parameter sensitivity analysis across different datasets.

$\{0, 0.3, 0.5, 0.7\}$. As shown in Figure 4a and 4b, the sensitivity analysis reveals that the forecasting results remain stable across a wide range of non-zero values of λ_{consis} and λ_{CFR}. The model achieves its best performance at $\lambda_{\text{consis}} = 5$ and $\lambda_{\text{CFR}} = 0.7$. These results indicate that both losses contribute meaningfully to performance, while the model is not sensitive to the exact choice of loss weights within a reasonable range.

Impact of Rank Selection Ratios. We evaluate the sensitivity of rank ratios by partitioning the singular spectrum into a core semantic space ($< r_1$), a subordinate semantic perturbation zone ($[r_1, r_2]$), and a residual noise space ($> r_2$). Following the definition in 3.1, the partition boundaries are governed by the ratios ρ_1 and ρ_2 such that $r_1 = \lfloor d \cdot \rho_1 \rfloor$ and $r_2 = \lfloor d \cdot \rho_2 \rfloor$, where $d = \min(W, K)$ denotes the maximum possible rank. We vary $\rho_1 \in \{0, 0.05, 0.08, 0.1\}$ and $\rho_2 \in \{0.3, 0.4, 0.5, 0.6\}$. As shown in Figure 4c and 4d, the model suffers a huge performance drop at $\rho_1 = 0$, underscoring the necessity of preserving the core semantic backbone. Once $\rho_1 > 0$, the performance remains broadly stable within a wide range of ρ_1 and ρ_2 configurations, demonstrating that our method is robust to the selection of rank thresholds.

5 Conclusion

We introduce SIFT, a novel fine-tuning method designed to address the challenge of overfitting and the mean-prediction trap in TSFMs, which are primarily caused by the low-dimensional nature and diverse structural patterns of time series data. Our method employs semantic-invariant adversarial augmentation to enhance the model's robustness against semantic-preserving perturbations, mitigating overfitting; simultaneously, it adopts a component-based structural fidelity enhancement to improve the model's ability to retain fine-grained dynamics, alleviating the mean-prediction trap. Our experiments demonstrate that SIFT can significantly boost the fine-tuning performance of TSFMs.

AI USE STATEMENT

Generative AI tools were used only as writing assistance during the preparation of this manuscript, mainly for language polishing, grammatical refinement, and improving readability. The research problem, technical motivation, methodological design, experimental setup, and scientific conclusions were developed by us independently. All AI-assisted text was manually inspected, revised, and verified to ensure technical accuracy and consistency with the authors' intended claims. We are fully responsible for the content of the final manuscript, including all statements and materials that were prepared with the assistance of generative AI.

ETHICS STATEMENT

This study uses only publicly available time-series benchmark datasets and does not involve human participants, private records, or personally identifiable information. The proposed method is developed for general-purpose time-series forecasting and does not rely on sensitive attributes or user-specific data. We are not aware of any substantial ethical risks arising directly from the methodology or experimental setting considered in this work.

REPRODUCIBILITY STATEMENT

We describe the experimental protocol, model configurations, training settings, and evaluation procedures in the main paper and supplementary material. The datasets evaluated in this work are publicly accessible, and the experiments follow the benchmark settings specified in the paper. To facilitate independent reproduction, we include the source code and instructions for reproducing the experiments in the supplementary material.

REFERENCES

Xinyang Chen, Sinan Wang, Bo Fu, Mingsheng Long, and Jianmin Wang. Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning. In *NeurIPS*, pp. 1906–1916, 2019.

Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *ICML*, pp. 10148–10167, 2024.

Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

Vijay Ekambaram, Arindam Jati, Pankaj Dayama, Sumanta Mukherjee, Nam 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 *NeurIPS*, 2024.

Shanghua Gao, Teddy Koker, Owen Queen, Thomas Hartvigsen, Theodoros Tsiligkaridis, and Marinka Zitnik. Units: Building a unified time series model. *CoRR*, abs/2403.00131, 2024. URL https://doi.org/10.48550/arXiv.2403.00131.

Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *ICML*, pp. 16115–16152, 2024.

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 *ICLR*, 2022.

Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. Timebase: The power of minimalism in efficient long-term time series forecasting. In *ICML*, 2025.

Hassan Ismail Fawaz, Benjamin Lucas, Germain Forestier, Charlotte Pelletier, Daniel F Schmidt, Jonathan Weber, Geoffrey I Webb, Lhassane Idoumghar, Pierre-Alain Muller, and François Petitjean. Inceptiontime: Finding alexnet for time series classification. *Data mining and knowledge discovery*, 34(6):1936–1962, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1412.6980.

Ananya Kumar, Aditi Raghunathan, Robbie Matthew Jones, Tengyu Ma, and Percy Liang. Fine-tuning can distort pretrained features and underperform out-of-distribution. In *ICLR*, 2022.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *SIGIR*, pp. 95–104. ACM, 2018.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791. URL https://doi.org/10.1109/5.726791.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *EMNLP*, pp. 3045–3059. Association for Computational Linguistics, 2021.

Beibu Li, Qichao Shentu, Yang Shu, Hui Zhang, Ming Li, Ning Jin, Bin Yang, and Chenjuan Guo. Crossad: Time series anomaly detection with cross-scale associations and cross-window modeling. *CoRR*, abs/2510.12489, 2025a. URL https://doi.org/10.48550/arXiv.2510.12489.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. *CoRR*, abs/2101.00190, 2021. URL https://arxiv.org/abs/2101.00190.

Zhe Li, Xiangfei Qiu, Peng Chen, Yihang Wang, Hanyin Cheng, Yang Shu, Jilin Hu, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, and Bin Yang. Tsfm-bench: A comprehensive and unified benchmark of foundation models for time series forecasting. In *SIGKDD*, pp. 5595–5606. ACM, 2025b.

Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X. Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *ICLR*, 2022.

Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *WWW*, pp. 4095–4106. ACM, 2024a.

Xvyuan Liu, Xiangfei Qiu, Xingjian Wu, Zhengyu Li, Chenjuan Guo, Jilin Hu, and Bin Yang. Rethinking irregular time series forecasting: A simple yet effective baseline. In *AAAI*, pp. 23873–23881, 2026.

Yong Liu, Haoran Zhang, Chenyu Li, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Timer: Generative pre-trained transformers are large time series models. In *ICML*, pp. 32369–32399, 2024b.

Yong Liu, Guo Qin, Zhiyuan Shi, Zhi Chen, Caiyin Yang, Xiangdong Huang, Jianmin Wang, and Mingsheng Long. Sundial: A family of highly capable time series foundation models. In *ICML*, 2025.

Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967.

Jishnu Mukhoti, Yarin Gal, Philip Torr, and Puneet K. Dokania. Fine-tuning can cripple your foundation model; preserving features may be the solution. *Trans. Mach. Learn. Res.*, 2024, 2024.

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

Rafael Poyatos, Víctor Granda, Víctor Flo, Mark A. Adams, Balázs Adorján, David Aguadé, Marcos P. M. Aidar, Scott Allen, M. Susana Alvarado-Barrientos, Kristina J. Anderson-Teixeira, Luiza Maria Aparecido, M. Altaf Arain, Ismael Aranda, Heidi Asbjornsen, Robert Baxter, Eric Beamesderfer, Z. Carter Berry, Daniel Berveiller, Bethany Blakely, Johnny Boggs, Gil Bohrer, Paul V. Bolstad, Damien Bonal, Rosvel Bracho, Patricia Brito, Jason Brodeur, Fernando Casanoves, Jérôme Chave, Hui Chen, Cesar Cisneros, Kenneth Clark, Edoardo Cremonese, Hongzhong Dang, Jorge S. David, Teresa S. David, Nicolas Delpierre, Ankur R. Desai, Frederic C. Do, Michal Dohnal, Jean-Christophe Domec, Sebinasi Dzikiti, Colin Edgar, Rebekka Eichstaedt, Tarek S. El-Madany, Jan Elbers, Cleiton B. Eller, Eugénie S. Euskirchen, Brent Ewers, Patrick Fonti, Alicia Forner, David I. Forrester, Helber C. Freitas, Marta Galvagno, Omar Garcia-Tejera, Chandra Prasad Ghimire, Teresa E. Gimeno, John Grace, André Granier, Anne Griebel, Yan Guangyu, Mark B. Gush, Paul J. Hanson, Niles J. Hasselquist, Ingo Heinrich, Virginia Hernandez-Santana, Valentine Herrmann, Teemu Hölttä, Friso Holwerda, James Irvine, Supat Isarangkool Na Ayutthaya, Paul G. Jarvis, Hubert Jochheim, Carlos A. Joly, Julia Kaplick, Hyun Seok Kim, Leif Klemedtsson, Heather Kropp, Fredrik Lagergren, Patrick Lane, Petra Lang, Andrei Lapenas, Víctor Lechuga, Minsu Lee, Christoph Leuschner, Jean-Marc Limousin, Juan Carlos Linares, Maj-Lena Linderson, Anders Lindroth, Pilar Llorens, Álvaro López-Bernal, Michael M. Loranty, Dietmar Lüttschwager, Cate Macinnis-Ng, Isabelle Maréchaux, Timothy A. Martin, Ashley Matheny, Nate McDowell, Sean McMahon, Patrick Meir, Ilona Mészáros, Mirco Migliavacca, Patrick Mitchell, Meelis Mölder, Leonardo Montagnani, Georgianne W. Moore, Ryogo Nakada, Furong Niu, Rachael H. Nolan, Richard Norby, Kimberly Novick, Walter Oberhuber, Nikolaus Obojes, A. Christopher Oishi, Rafael S. Oliveira, Ram Oren, Jean-Marc Ourcival, Teemu Paljakka, Oscar Perez-Priego, Pablo L. Peri, Richard L. Peters, Sebastian Pfautsch, William T. Pockman, Yakir Preisler, Katherine Rascher, George Robinson, Humberto Rocha, Alain Rocheteau, Alexander Röll, Bruno H. P. Rosado, Lucy Rowland, Alexey V. Rubtsov, Santiago Sabaté, Yann Salmon, Roberto L. Salomón, Elisenda Sánchez-Costa, Karina V. R. Schäfer, Bernhard Schuldt, Alexandr Shashkin, Clément Stahl, Marko Stojanović, Juan Carlos Suárez, Ge Sun, Justyna Szatniewska, Fyodor Tatarinov, Miroslav Tesař, Frank M. Thomas, Pantana Tor-ngern, Josef Urban, Fernando Valladares, Christiaan van der Tol, Ilja van Meerveld, Andrej Varlagin, Holm Voigt, Jeffrey Warren, Christiane Werner, Willy Werner, Gerhard Wieser, Lisa Wingate, Stan Wullschleger, Koong Yi, Roman Zweifel, Kathy Steppe, Maurizio Mencuccini, and Jordi Martínez-Vilalta. Global transpiration data from sap flow measurements: the sapfluxnet database. *Earth System Science Data*, 13(6):2607–2649, June 2021. ISSN 1866-3516.

Zhongzheng Qiao, Chenghao Liu, Yiming Zhang, Ming Jin, Quang Pham, Qingsong Wen, Ponnuthurai N. Suganthan, Xudong Jiang, and Savitha Ramasamy. Multi-scale finetuning for encoder-based time series foundation models. *CoRR*, abs/2506.14087, 2025. URL https://doi.org/10.48550/arXiv.2506.14087.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. DUET: dual clustering enhanced multivariate time series forecasting. In *SIGKDD*, pp. 1185–1196. ACM, 2025.

Andrew L. Rukhin. Analysis of time series structure SSA and related techniques. *Technometrics*, 44 (3):290, 2002.

Yanru Sun, Zongxia Xie, Dongyue Chen, Emadeldeen Eldele, and Qinghua Hu. Hierarchical classification auxiliary network for time series forecasting. In *AAAI*, pp. 20743–20751, 2025.

Artur Trindade. Electricityloaddiagrams20112014. https://doi.org/10.24432/C58C86, March 2015. URL https://doi.org/10.24432/C58C86.

Shu Wan, Miles Ma, Hank Zhu, Guangqi Liu, Stephen Wang, Qingsong Wen, and Huan Liu. Forecast collapse in time-series foundation models. *CoRR*, abs/2608.14106, 2026. doi: 10.48550/ARXIV.2608.14106. URL https://doi.org/10.48550/arXiv.2608.14106.

Huiqiang Wang, Jian Peng, Feihu Huang, Jince Wang, Junhui Chen, and Yifei Xiao. MICN: multi-scale local and global context modeling for long-term series forecasting. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=zt53IDUR1U.

Yihang Wang, Yuying Qiu, Peng Chen, Kai Zhao, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. ROSE: register assisted general time series forecasting with decomposed frequency learning. *CoRR*, abs/2405.17478, 2024. URL https://doi.org/10.48550/arXiv.2405.17478.

Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified training of universal time series forecasting transformers. In *ICML*, pp. 53140–53164, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *NeurIPS*, pp. 22419–22430, 2021.

Annan Yu, Danielle C. Maddix, Boran Han, Xiyuan Zhang, Abdul Fatir Ansari, Oleksandr Shchur, Christos Faloutsos, Andrew Gordon Wilson, Michael W. Mahoney, and Yuyang Wang. Understanding transformers for time series: Rank structure, flow-of-ranks, and compressibility. *CoRR*, abs/2510.03358, 2025. URL https://doi.org/10.48550/arXiv.2510.03358.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. *CoRR*, abs/2303.10512, 2023. URL https://doi.org/10.48550/arXiv.2303.10512.

Shuyi Zhang, Bin Guo, Anlan Dong, Jing He, Ziping Xu, and Song Xi Chen. Cautionary tales on air-quality improvement in beijing. *Proceedings of the Royal Society. A, Mathematical, physical, and engineering sciences*, 473(2205):20170457–20170457, 2017. ISSN 1364-5021.

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 *AAAI*, pp. 11106–11115, 2021.

A APPENDIX

A.1 DATASETS

To conduct a rigorous and unbiased assessment of our models, we evaluate their performance across ten well-established forecasting benchmarks. These datasets encompass a diverse range of domains: (1) the ETT suite (Zhou et al., 2021), which captures seven variates from two electric transformers between 2016 and 2018 across four subsets of varying granularity (hourly ETTh1/ETTh2 and 15-minute ETTm1/ETTm2); (2) the Electricity dataset (Trindade, 2015), providing hourly consumption logs for 321 customers from 2016 to 2019; (3) the Exchange dataset (Lai et al., 2018), which tracks daily exchange rates for eight countries from 1990 to 2016; (4) the Solar dataset (Lai et al., 2018), featuring 10-minute interval power generation from 137 PV plants in 2006; (5) the Weather dataset (Wu et al., 2021), documenting 21 meteorological indicators for Germany in 2020 at 10-minute intervals; (6) the AQShunyi dataset(Zhang et al., 2017) which offers comprehensive multi-year air quality measurements; and (7) the ZafNoo dataset(Poyatos et al., 2021), which provides sap flow data along with associated environmental variables. Comprehensive details regarding these benchmarks are provided in the table 3.

Table 3: Statistics of the benchmark datasets used in our experiments.

Dataset	Domain	Frequency	Samples	Dim	Train/Val/Test
ETTh1	Electricity	1h	14,400	7	6:2:2
ETTh2	Electricity	1h	14,400	7	6:2:2
ETTm1	Electricity	15m	57,600	7	6:2:2
ETTm2	Electricity	15m	57,600	7	6:2:2
Electricity	Electricity	1h	26,304	321	7:1:2
Solar	Energy	10m	52,560	137	6:2:2
Weather	Environment	10m	52,696	21	7:1:2
AQShunyi	Environment	1h	35,064	11	6:2:2
ZafNoo	Nature	30m	19,225	11	7:1:2
Exchange	Economic	1d	7,588	8	7:1:2

A.2 TSFMS

Timer (Liu et al., 2024b): Timer is a GPT-style autoregressive model for time series analysis that treats time series forecasting as a next-token prediction task. By training on large-scale time series datasets, it demonstrates strong sequence modeling capabilities, effectively supporting various downstream tasks such as forecasting, imputation, and anomaly detection across diverse domains.

UniTS (Gao et al., 2024): UniTS is a unified time series model that supports general task specifications, accommodating classification, forecasting, imputation, and anomaly detection. This is achieved through a novel unified backbone that integrates both sequential and variate attention with dynamic linear operators, all trained as a single unified model.

Moirai (Woo et al., 2024): Moirai is a universal time series forecasting model based on a patch-based (Nie et al., 2022) Transformer architecture. It addresses three unique challenges of time series foundation models: cross-frequency learning, accommodating an arbitrary number of variates, and handling varying distributional properties in large-scale data.

Sundial (Liu et al., 2025): Sundial is a family of time series foundation models that introduces TimeFlow Loss, a flow-matching based training objective that enables native pre-training of Transformers on continuous-valued time series without discrete tokenization. Unlike prior methods that rely on parametric densities, Sundial learns arbitrary distributions flexibly and can generate multiple probable predictions. Sundial achieves state-of-the-art zero-shot performance on both point and probabilistic forecasting benchmarks, with efficient inference speed.

Table 4: Cross-domain affinity between the decomposed spectral subspaces. Higher affinity indicates stronger similarity between the corresponding subspaces.

Dataset comparison	Core	Subordinate	Noise
ETTh1–ETTh2	0.993	0.850	0.940
ETTm1–ETTm2	0.999	0.978	0.994
Avg. of related ETT pairs	0.995	0.915	0.967
Avg. of ETT–external pairs	0.812	0.855	0.946
Related–external gap	**0.183**	0.060	0.021

A.3 Fine-tuning baselines

Full Fine-tune and Linear Probe. Full fine-tuning updates all parameters of the pretrained model. We use a shared learning rate for both full fine-tuning and linear probing. Linear probing (Kumar et al., 2022) only updates the output head while keeping the backbone frozen.

LoRA and AdaLoRA. LoRA (Hu et al., 2022) introduces trainable low-rank decomposition matrices into attention layers, enabling parameter-efficient fine-tuning by injecting updates into a low-rank subspace. AdaLoRA (Zhang et al., 2023) further improves this process by dynamically allocating the rank during training based on parameter importance. In our implementation, we apply LoRA and AdaLoRA to the query, key, and value projection layers. For LoRA and AdaLoRA, we follow the official configurations as the starting point and tune the key hyperparameters on the validation set.

MSFT. For MSFT (Qiao et al., 2025), we follow the official implementation and tune its key hyperparameters on the validation set. The selected configuration uses rank $r = 16$ and scaling number $scale_num = 3$.

A.4 Visualizations of Semantic Spectrum Decomposition

To provide an intuitive interpretation of the proposed spectral decomposition, we visualize the rank-wise reconstructed components of an ETTh1 sample in Figure 5. Each component is reconstructed from an individual singular direction and is annotated with its relative spectral energy. The leading components account for the majority of the signal energy and exhibit smooth, large-scale temporal variations, capturing the dominant trend and long-period seasonal patterns of the original sequence.

As the rank increases, the reconstructed components gradually transition toward higher-frequency and more localized oscillations. In particular, the intermediate-rank components retain structured and non-negligible temporal variations while contributing substantially less energy than the dominant components, corresponding to the subordinate semantic space used for adversarial perturbation.

In contrast, the high-rank components have substantially smaller energy contributions and predominantly exhibit irregular, high-frequency fluctuations, which are assigned to the noise space.

Overall, the visualization reveals a progressive transition from dominant trend and periodic structure to structured intermediate variations and finally to weak and noisy high-frequency fluctuations across the singular spectrum, providing an intuitive interpretation of the three spectral subspaces adopted in SIFT. In practice, the partition boundaries are selected on the validation set and then fixed for downstream fine-tuning and evaluation.

A.5 Semantic Interpretation of Spectral Subspaces

While the rank-wise visualization above reveals distinct temporal characteristics across the spectral spectrum, it remains important to examine whether the decomposed subspaces capture domain-related temporal information rather than merely reflecting differences in frequency or energy. We therefore measure the affinity between corresponding subspaces across related ETT datasets and external datasets.

As shown in Table 4, the core subspace exhibits very high affinity across related ETT datasets, while its affinity decreases substantially when comparing ETT with external datasets. This yields



Figure 5: Rank-wise visualization of the proposed spectral decomposition on an ETTh1 sample. The top panel shows the original time series, while the remaining panels visualize components reconstructed from individual singular directions together with their relative spectral energy. Low-rank components capture dominant large-scale temporal structure, intermediate components preserve structured but non-dominant variations, and high-rank components mainly contain weak high-frequency fluctuations. Colors denote the core semantic space ($\mathcal{S}_{\text{core}}$), subordinate semantic space ($\mathcal{S}_{\text{sub}}$), and noise space ($\mathcal{S}_{\text{noise}}$), respectively.

a related–external affinity gap of 0.183, markedly larger than those of the subordinate (0.060) and noise (0.021) subspaces. These results suggest that the dominant spectral directions capture more domain-shared temporal structure, whereas the subordinate subspace retains more dataset-specific

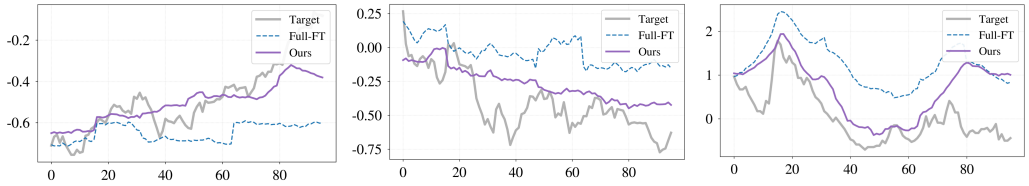


Figure 6: Forecasting visualization comparing our method and standard fine-tuning. The purple line denotes our method, while the blue line represents Full-FT.

but still structured variations. The noise subspace, in contrast, shows little discrimination between related and external datasets.

Together with the rank-wise reconstruction in Figure 5, this analysis indicates that the proposed decomposition is not merely an energy-based partition of the spectrum, but separates temporal components with distinct structural and domain-related characteristics. This provides empirical support for preserving the dominant core structure while applying adversarial perturbations within the subordinate subspace.

A.6 Forecasting Visualization.

Figure 6 illustrates the visualization results of representative samples from the ETTm1 and Exchange datasets under both standard fine-tuning and our proposed method. We can observe that our approach aligns more closely with the ground truth. This is primarily because our method enables the model to anchor the primary semantics while maintaining structural fidelity.

A.7 Limitation of Series Decomposition Assumption

Although series decomposition techniques are widely adopted in time series modeling, they provide only an approximate separation of trend and seasonal components rather than a fully identifiable or exact factorization. In practice, the decomposition quality may vary across datasets, particularly for highly non-stationary or multi-scale signals where temporal patterns are strongly entangled. As a result, the effectiveness of the proposed structural fidelity enhancement may be influenced by the accuracy of the underlying decomposition.

A.8 Full Result

Table 5: Full MSE results on Timer and UniTS.

		Timer						UniTS					
		Full	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*	*Full*	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*
ETTh1	96	0.396	0.382	0.382	0.381	0.399	**0.366**	0.395	0.407	0.372	0.372	0.376	**0.371**
	192	0.487	0.441	0.451	0.438	0.413	**0.384**	0.446	0.443	0.421	0.419	0.428	**0.414**
	336	0.483	0.487	0.479	0.486	0.426	**0.398**	0.479	0.478	0.465	0.468	0.463	**0.451**
	720	0.497	0.497	0.517	0.519	0.446	**0.414**	0.478	0.503	0.488	0.491	0.469	**0.459**
	Avg	0.465	0.452	0.457	0.456	0.421	**0.390**	0.450	0.458	0.436	0.438	0.434	**0.424**
ETTh2	96	0.305	0.295	0.293	**0.291**	0.300	**0.291**	0.311	0.339	0.308	0.329	**0.302**	0.303
	192	0.385	0.377	0.376	0.374	0.386	**0.373**	0.470	0.416	0.387	0.414	0.375	**0.347**
	336	0.437	0.418	0.423	**0.415**	0.436	0.416	0.429	0.429	0.416	0.414	0.416	**0.387**
	720	0.448	0.438	0.429	0.427	0.441	**0.424**	0.424	0.429	0.426	0.438	0.425	**0.406**
	Avg	0.394	0.382	0.380	0.377	0.391	**0.376**	0.408	0.403	0.384	0.399	0.380	**0.361**
ETTm1	96	0.335	0.327	0.327	0.327	0.308	**0.289**	0.332	0.415	0.326	0.324	0.336	**0.315**
	192	0.424	0.371	0.378	0.372	0.343	**0.317**	0.383	0.413	0.368	0.371	0.366	**0.360**
	336	0.450	1.215	1.079	1.215	0.375	**0.351**	0.467	0.431	0.408	0.420	0.396	**0.395**
	720	0.514	1.255	1.271	1.319	0.432	**0.410**	0.507	0.467	0.466	0.453	0.461	**0.450**
	Avg	0.431	0.792	0.764	0.808	0.364	**0.342**	0.422	0.432	0.392	0.392	0.390	**0.380**
ETTm2	96	0.185	0.186	0.177	0.187	0.189	**0.165**	0.198	0.242	0.190	0.193	0.179	**0.174**
	192	0.257	0.260	0.246	0.262	0.237	**0.227**	0.252	0.282	0.258	0.280	0.248	**0.246**
	336	0.313	0.353	0.354	0.359	0.287	**0.284**	0.334	0.341	0.335	0.352	0.313	**0.310**
	720	0.402	0.474	0.476	0.492	**0.372**	**0.372**	0.468	0.442	0.451	0.497	**0.405**	0.410
	Avg	0.289	0.318	0.313	0.325	0.271	**0.262**	0.313	0.327	0.309	0.330	0.286	**0.285**
Exchange	96	0.098	0.088	0.087	0.084	0.094	**0.082**	0.106	0.143	0.090	0.093	0.090	**0.085**
	192	0.181	0.180	0.183	0.175	0.256	**0.172**	0.222	0.218	0.187	0.199	0.185	**0.181**
	336	0.332	0.337	0.336	0.320	0.421	**0.314**	0.365	0.357	0.345	0.335	0.344	**0.326**
	720	0.870	0.863	0.871	0.853	1.012	**0.834**	0.891	0.840	0.842	0.839	**0.820**	0.827
	Avg	0.370	0.367	0.369	0.358	0.446	**0.351**	0.396	0.389	0.366	0.367	0.360	**0.355**
ZafNoo	96	0.503	0.481	0.479	0.479	0.441	**0.431**	0.448	0.550	0.460	0.583	0.465	**0.431**
	192	0.615	0.569	**0.481**	0.563	0.490	0.484	0.492	0.614	0.538	0.606	0.524	**0.489**
	336	0.682	0.636	0.630	0.629	**0.534**	**0.534**	0.545	0.642	0.585	0.585	0.565	**0.537**
	720	0.781	0.762	0.750	0.925	**0.595**	0.612	0.608	0.716	0.682	0.683	0.674	**0.575**
	Avg	0.645	0.612	0.585	0.649	**0.515**	**0.515**	0.523	0.631	0.566	0.614	0.557	**0.508**
Weather	96	0.172	0.498	0.174	0.176	0.178	**0.148**	0.146	0.163	0.173	0.164	0.158	**0.145**
	192	0.235	0.536	0.442	0.337	0.210	**0.193**	**0.200**	0.214	0.221	0.223	0.204	0.203
	336	0.296	0.564	0.449	0.447	0.259	**0.244**	0.272	0.274	0.278	0.301	**0.261**	**0.261**
	720	0.380	0.551	0.499	0.561	0.328	**0.316**	0.362	0.349	0.353	0.403	**0.342**	0.350
	Avg	0.271	0.537	0.391	0.380	0.244	**0.225**	0.245	0.250	0.256	0.273	0.241	**0.240**
Electricity	96	**0.134**	0.442	0.135	0.182	0.152	0.139	**0.133**	0.163	0.152	0.171	0.154	**0.133**
	192	0.155	0.369	0.269	0.201	0.167	**0.148**	**0.153**	0.170	0.168	0.184	0.169	**0.153**
	336	0.172	0.423	0.319	0.224	0.183	**0.165**	0.175	0.192	0.183	0.202	0.186	**0.174**
	720	0.211	0.448	0.336	0.280	0.223	**0.206**	0.204	0.228	0.218	0.242	0.227	**0.197**
	Avg	0.168	0.421	0.265	0.222	0.181	**0.164**	0.166	0.188	0.180	0.200	0.184	**0.164**
Solar	96	0.183	0.369	0.207	0.222	0.191	**0.182**	0.162	1.812	0.248	0.197	0.214	**0.161**
	192	0.209	0.337	0.323	0.268	0.202	**0.198**	0.195	1.871	0.255	0.198	0.237	**0.171**
	336	**0.210**	0.338	0.339	0.300	0.216	0.211	0.194	1.929	0.274	0.217	0.246	**0.181**
	720	0.297	0.338	0.350	0.329	**0.226**	0.236	0.201	1.853	0.477	0.248	0.252	**0.196**
	Avg	0.225	0.345	0.305	0.280	0.209	**0.207**	0.188	1.866	0.313	0.215	0.237	**0.177**
AQshunyi	96	0.699	0.846	0.846	0.895	0.814	**0.662**	0.739	0.890	0.755	0.762	0.745	**0.722**
	192	0.787	1.040	0.718	**0.687**	0.693	0.765	0.784	0.941	0.816	0.817	0.805	**0.776**
	336	0.809	1.034	0.738	**0.712**	0.717	0.752	0.829	0.862	0.831	0.830	0.838	**0.801**
	720	0.855	1.034	0.794	0.774	0.777	**0.753**	0.857	0.896	0.876	0.871	**0.842**	0.843
	Avg	0.788	0.989	0.774	0.767	0.750	**0.733**	0.802	0.897	0.820	0.820	0.807	**0.785**

Table 6: Full MSE results on Moirai and Sundial.

		Moirai						Sundial					
		Full	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*	*Full*	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*
ETTh1	96	0.495	0.575	**0.379**	0.422	0.447	**0.379**	0.377	0.376	0.374	0.470	0.508	**0.370**
	192	0.540	0.519	0.587	0.531	0.497	**0.418**	0.422	0.493	0.443	0.537	0.571	**0.419**
	336	0.491	0.500	0.463	0.543	**0.459**	0.544	0.477	0.492	0.495	0.588	0.630	**0.469**
	720	0.561	0.703	0.732	0.586	**0.497**	0.614	0.543	**0.512**	0.736	0.643	0.685	0.542
	Avg	0.522	0.574	0.540	0.520	**0.475**	0.489	0.455	0.468	0.512	0.560	0.598	**0.450**
ETTh2	96	0.319	0.435	0.314	0.308	0.290	**0.285**	0.393	0.412	0.363	0.526	0.633	**0.343**
	192	0.385	0.420	0.376	0.388	0.348	**0.345**	0.548	**0.459**	0.503	0.746	0.896	0.474
	336	0.391	0.395	0.408	0.379	0.381	**0.361**	0.519	**0.482**	0.556	0.910	1.139	0.495
	720	0.432	0.474	**0.390**	0.466	0.468	0.489	0.906	**0.714**	1.684	1.297	1.597	0.886
	Avg	0.382	0.431	0.372	0.385	0.372	**0.370**	0.592	**0.517**	0.776	0.870	1.066	0.550
ETTm1	96	**0.312**	0.410	0.358	0.373	0.368	0.336	0.334	0.407	0.806	0.575	0.363	**0.326**
	192	0.422	0.417	0.393	0.412	**0.379**	0.393	0.379	0.410	0.381	0.613	0.407	**0.369**
	336	0.449	0.450	0.441	0.444	**0.434**	0.443	0.427	0.427	0.426	0.636	0.438	**0.415**
	720	0.515	0.493	0.536	0.552	0.488	**0.456**	**0.469**	0.495	0.953	0.685	0.492	0.476
	Avg	0.424	0.443	0.432	0.445	0.417	**0.407**	0.402	0.435	0.642	0.627	0.425	**0.396**
ETTm2	96	0.224	0.242	0.195	0.193	0.207	**0.183**	0.246	0.249	0.569	0.385	0.252	**0.210**
	192	0.308	0.282	0.267	0.279	0.268	**0.239**	0.354	**0.281**	0.308	0.385	0.346	0.322
	336	0.369	0.341	0.310	**0.300**	0.310	0.330	0.456	**0.371**	0.571	0.539	0.457	0.452
	720	0.460	0.425	0.397	0.402	0.397	**0.382**	0.852	0.936	1.565	0.775	0.639	**0.563**
	Avg	0.340	0.323	0.292	0.293	0.295	**0.284**	0.477	0.459	0.753	0.521	0.424	**0.387**
Exchange	96	0.096	0.162	0.119	0.121	0.082	**0.081**	0.120	0.160	0.221	0.155	0.204	**0.109**
	192	0.200	0.229	0.249	0.292	0.182	**0.177**	0.267	0.214	0.255	0.300	0.373	**0.190**
	336	0.381	0.421	0.475	0.490	**0.336**	0.433	0.511	0.457	0.485	0.507	0.605	**0.365**
	720	1.130	1.025	1.427	1.036	0.959	**0.849**	0.988	1.061	1.076	1.054	1.046	**0.788**
	Avg	0.452	0.460	0.567	0.485	0.390	**0.385**	0.472	0.473	0.510	0.504	0.557	**0.363**
ZafNoo	96	0.541	0.717	0.539	0.534	0.612	**0.424**	0.478	0.761	0.570	0.484	0.539	**0.419**
	192	0.604	0.646	0.582	0.542	**0.486**	0.499	0.516	0.648	0.520	0.548	0.612	**0.479**
	336	0.573	0.665	0.612	0.625	**0.567**	0.588	0.522	0.599	0.529	0.587	0.656	**0.505**
	720	0.675	0.632	0.571	0.659	**0.565**	0.586	0.545	0.581	0.758	0.639	0.710	**0.542**
	Avg	0.598	0.665	0.576	0.590	0.558	**0.524**	0.515	0.647	0.594	0.565	0.629	**0.486**
Weather	96	0.161	0.333	0.160	0.160	**0.148**	0.149	0.171	0.276	0.329	0.198	0.181	**0.167**
	192	**0.191**	0.323	0.223	0.220	0.205	0.207	0.222	0.265	0.422	0.244	0.228	**0.229**
	336	0.278	0.350	0.251	0.264	0.270	**0.247**	0.281	0.296	0.478	0.296	0.278	**0.269**
	720	0.353	0.395	**0.340**	0.361	**0.340**	0.348	0.363	0.395	0.594	0.375	**0.349**	**0.349**
	Avg	0.246	0.350	0.243	0.251	0.241	**0.238**	0.259	0.308	0.456	0.278	0.259	**0.253**
Electricity	96	0.170	0.142	0.142	0.155	0.149	**0.136**	**0.159**	0.341	0.171	0.176	0.168	**0.159**
	192	0.191	0.160	0.163	0.169	0.164	**0.158**	0.172	0.278	0.179	0.184	0.176	**0.172**
	336	0.264	0.178	0.178	0.184	0.176	**0.173**	0.190	0.266	0.198	0.201	0.192	**0.189**
	720	0.412	0.210	0.207	0.221	**0.203**	0.218	0.232	0.268	0.233	0.236	**0.227**	0.231
	Avg	0.259	0.173	0.173	0.182	0.173	**0.171**	0.188	0.288	0.195	0.199	0.191	**0.187**
Solar	96	0.207	1.921	0.184	**0.172**	0.184	0.189	**0.184**	1.944	0.199	0.281	0.227	0.189
	192	0.196	1.838	**0.195**	0.197	**0.195**	0.196	0.213	1.890	0.409	0.334	0.263	**0.200**
	336	0.193	1.892	0.209	0.206	0.209	**0.193**	0.221	1.877	0.401	0.351	0.270	**0.214**
	720	0.222	1.820	0.237	0.562	0.240	**0.207**	0.228	1.974	0.224	0.371	0.277	**0.211**
	Avg	0.205	1.868	0.206	0.284	0.207	**0.196**	0.211	1.921	0.308	0.334	0.259	**0.204**
AQshunyi	96	0.704	1.234	0.580	**0.542**	0.635	0.647	0.765	1.239	0.668	**0.663**	0.683	0.669
	192	0.691	0.883	0.637	**0.580**	0.668	0.702	0.799	1.002	0.703	0.703	0.707	**0.702**
	336	0.723	0.862	0.682	**0.612**	0.704	0.727	0.837	0.888	**0.721**	0.726	0.727	0.723
	720	0.782	0.811	0.712	**0.663**	0.735	0.776	0.881	0.810	0.758	0.769	0.757	**0.754**
	Avg	0.725	0.948	0.653	**0.599**	0.686	0.713	0.821	0.985	**0.712**	0.715	0.718	**0.712**

Table 7: Full MAE results on UniTS and Timer.

		Timer						UniTS					
		Full	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*	*Full*	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*
ETTh1	96	0.424	**0.395**	0.397	**0.395**	0.414	0.402	0.404	0.406	0.395	0.395	**0.392**	0.397
	192	0.434	0.429	0.440	0.428	0.438	**0.414**	0.435	0.427	0.423	0.424	**0.421**	0.424
	336	0.458	0.456	0.450	0.449	0.447	**0.425**	0.458	**0.443**	0.447	0.451	0.446	0.445
	720	0.480	0.484	0.486	0.483	0.470	**0.445**	0.486	0.480	0.474	0.482	**0.461**	0.469
	Avg	0.449	0.441	0.443	0.439	0.442	**0.422**	0.446	0.439	0.435	0.438	**0.430**	0.434
ETTh2	96	0.366	0.346	0.344	0.344	0.349	**0.342**	0.365	0.367	0.348	0.363	0.353	**0.343**
	192	0.410	0.396	0.395	0.397	0.398	**0.394**	0.453	0.412	0.402	0.427	0.395	**0.391**
	336	0.438	0.435	0.440	0.432	0.437	**0.428**	0.437	0.433	0.428	0.427	0.427	**0.420**
	720	0.456	0.454	0.447	0.446	0.451	**0.441**	0.458	0.444	0.444	0.461	0.444	**0.433**
	Avg	0.417	0.408	0.407	0.405	0.409	**0.401**	0.428	0.414	0.406	0.419	0.405	**0.397**
ETTm1	96	0.359	0.361	0.365	0.361	0.357	**0.347**	0.361	0.401	0.366	0.361	0.372	**0.357**
	192	0.406	0.388	0.392	0.388	0.377	**0.362**	0.398	0.399	0.388	0.387	0.388	**0.384**
	336	0.428	0.703	0.673	0.704	0.397	**0.384**	0.436	0.421	0.414	0.415	0.408	**0.406**
	720	0.465	0.748	0.729	0.745	0.428	**0.418**	0.467	0.446	0.449	0.447	0.445	**0.440**
	Avg	0.415	0.550	0.540	0.550	0.390	**0.378**	0.415	0.417	0.404	0.402	0.403	**0.396**
ETTm2	96	0.264	0.266	0.260	0.266	0.276	**0.254**	0.278	0.296	0.275	0.276	0.267	**0.258**
	192	0.311	0.316	0.305	0.317	0.306	**0.294**	0.322	0.325	0.320	0.337	0.311	**0.307**
	336	0.351	0.413	0.391	0.413	0.339	**0.335**	0.375	0.365	0.366	0.376	0.349	**0.346**
	720	0.408	0.469	0.460	0.468	**0.393**	0.396	0.450	0.425	0.432	0.451	**0.403**	0.408
	Avg	0.334	0.366	0.354	0.366	0.328	**0.320**	0.356	0.353	0.348	0.360	0.333	**0.330**
Exchange	96	0.233	0.208	0.207	0.202	0.215	**0.199**	0.234	0.276	0.211	0.214	0.212	**0.204**
	192	0.349	0.302	0.304	0.298	0.379	**0.294**	0.325	0.345	0.309	0.320	0.307	**0.304**
	336	0.419	0.422	0.422	0.411	0.488	**0.406**	0.437	0.444	0.426	0.421	0.427	**0.412**
	720	0.819	0.703	0.706	0.699	0.770	**0.689**	0.712	0.695	0.689	0.688	**0.680**	0.687
	Avg	0.455	0.409	0.410	0.402	0.463	**0.397**	0.427	0.440	0.409	0.411	0.407	**0.402**
ZafNoo	96	0.428	0.420	0.419	0.419	0.418	**0.408**	0.414	0.452	**0.403**	**0.403**	0.412	0.405
	192	0.479	0.470	0.467	0.463	0.449	**0.434**	0.441	0.483	0.447	0.478	0.445	**0.439**
	336	0.510	0.509	0.507	0.501	0.472	**0.462**	0.467	0.498	0.474	0.474	**0.465**	0.468
	720	0.556	0.566	0.568	0.555	0.503	**0.502**	0.501	0.533	0.520	0.520	0.519	**0.488**
	Avg	0.493	0.491	0.490	0.485	0.460	**0.452**	0.456	0.491	0.461	0.469	0.460	**0.450**
Weather	96	0.218	0.402	0.217	0.218	0.231	**0.196**	0.218	0.211	0.211	0.207	0.208	**0.196**
	192	0.261	0.442	0.419	0.360	0.257	**0.238**	0.261	0.256	0.254	0.264	0.249	**0.249**
	336	0.305	0.471	0.423	0.422	0.294	**0.279**	0.305	0.299	0.296	0.321	**0.290**	0.292
	720	0.356	0.472	0.452	0.476	0.343	**0.330**	0.356	0.348	0.347	0.384	**0.342**	0.351
	Avg	0.285	0.447	0.378	0.369	0.281	**0.261**	0.285	0.279	0.277	0.294	0.272	**0.272**
Electricity	96	**0.224**	0.410	0.252	0.276	0.253	0.239	**0.231**	0.259	0.243	0.264	0.252	0.239
	192	0.244	0.411	0.266	0.295	0.266	**0.243**	**0.249**	0.266	0.259	0.280	0.266	0.260
	336	0.262	0.410	0.366	0.318	0.281	**0.261**	**0.273**	0.285	0.275	0.294	0.282	0.276
	720	**0.297**	0.417	0.401	0.364	0.313	**0.297**	**0.295**	0.316	0.307	0.326	0.315	0.297
	Avg	**0.257**	0.412	0.321	0.313	0.278	0.260	**0.262**	0.282	0.271	0.291	0.279	0.268
Solar	96	0.245	0.272	0.310	0.310	0.262	**0.225**	0.219	0.968	0.334	0.227	0.241	**0.213**
	192	0.252	0.327	0.405	0.341	0.270	**0.237**	0.230	0.991	0.332	0.232	0.249	**0.225**
	336	**0.244**	0.327	0.413	0.366	0.279	0.248	0.238	1.018	0.331	0.242	0.260	**0.235**
	720	0.282	0.328	0.428	0.397	0.288	**0.264**	**0.240**	0.994	0.478	0.248	0.273	0.250
	Avg	0.256	0.314	0.389	0.354	0.275	**0.244**	0.232	0.993	0.369	0.237	0.256	**0.231**
AQshunyi	96	0.497	0.590	0.583	0.566	0.572	**0.484**	0.501	0.568	0.510	0.513	0.505	**0.495**
	192	0.538	0.647	0.527	**0.505**	0.511	0.539	0.517	0.578	0.532	0.532	0.526	**0.516**
	336	0.550	0.645	0.537	**0.520**	0.524	0.531	0.537	0.554	0.542	0.541	0.540	**0.529**
	720	0.575	0.647	0.562	0.549	0.551	**0.509**	0.553	0.569	0.563	0.560	0.550	**0.549**
	Avg	0.540	0.632	0.552	0.535	0.539	**0.516**	0.527	0.567	0.537	0.537	0.530	**0.522**

Table 8: Full MAE results on Moirai and Sundial.

		Moirai						**Sundial**					
		Full	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*	*Full*	*LoRA*	*AdaLoRA*	*Linear*	*MSFT*	*Ours*
ETTh1	96	0.436	0.466	**0.407**	0.423	0.431	**0.407**	0.399	**0.397**	0.398	0.459	0.475	**0.397**
	192	0.472	0.458	0.492	0.468	0.457	**0.436**	0.434	0.451	0.442	0.500	0.514	**0.432**
	336	0.476	0.462	0.463	0.488	**0.458**	0.520	0.467	0.465	0.478	0.530	0.548	**0.463**
	720	0.519	0.563	0.618	0.532	**0.489**	0.540	0.525	**0.504**	0.625	0.581	0.596	0.528
	Avg	0.476	0.487	0.495	0.478	**0.459**	0.476	0.456	**0.454**	0.486	0.517	0.533	0.455
ETTh2	96	0.371	0.418	0.357	0.356	0.350	**0.341**	0.406	0.415	0.384	0.465	0.501	**0.375**
	192	0.409	0.419	0.400	0.406	0.382	**0.376**	0.489	**0.439**	0.469	0.568	0.611	0.458
	336	0.418	0.426	0.434	0.412	0.420	**0.402**	0.486	**0.460**	0.501	0.644	0.709	0.472
	720	0.461	0.476	**0.424**	0.469	0.463	0.486	0.634	**0.564**	0.892	0.787	0.861	0.627
	Avg	0.415	0.434	0.404	0.411	0.404	**0.401**	0.504	**0.470**	0.562	0.616	0.671	0.483
ETTm1	96	**0.353**	0.379	0.371	0.374	0.364	0.358	0.371	0.381	0.560	0.493	0.375	**0.365**
	192	0.403	0.393	0.391	0.401	**0.371**	0.395	0.395	0.401	0.403	0.516	0.402	**0.395**
	336	0.422	0.419	0.416	0.419	0.452	**0.413**	0.440	**0.420**	0.437	0.532	0.425	0.432
	720	0.457	0.449	0.469	0.468	0.452	**0.437**	0.462	0.472	0.670	0.565	**0.460**	0.469
	Avg	0.409	0.410	0.412	0.416	0.410	**0.400**	0.417	0.419	0.517	0.527	**0.415**	**0.415**
ETTm2	96	0.283	0.296	0.270	0.270	0.282	**0.268**	0.307	0.301	0.478	0.346	0.316	**0.293**
	192	0.335	0.325	0.322	0.321	0.317	**0.304**	0.382	**0.330**	0.356	0.417	0.379	0.359
	336	0.374	0.365	0.352	**0.350**	0.352	0.371	0.426	**0.386**	0.459	0.501	0.443	0.428
	720	0.430	0.442	0.404	0.407	0.404	**0.402**	0.592	0.594	0.871	0.614	**0.504**	0.530
	Avg	0.356	0.357	0.337	0.337	0.339	**0.336**	0.427	0.403	0.541	0.469	0.410	**0.402**
Exchange	96	0.213	0.300	0.246	0.245	**0.198**	0.204	0.252	0.298	0.292	0.257	0.281	**0.239**
	192	0.314	0.353	0.353	0.406	**0.301**	0.302	0.369	0.341	0.369	0.377	0.410	**0.330**
	336	0.439	0.478	0.508	0.510	**0.419**	0.487	0.531	0.510	0.522	0.502	0.538	**0.448**
	720	0.778	0.767	0.903	0.780	0.726	**0.689**	0.769	0.804	0.757	0.746	0.747	**0.680**
	Avg	0.436	0.474	0.502	0.485	**0.411**	0.421	0.480	0.488	0.485	0.470	0.494	**0.424**
ZafNoo	96	0.427	0.514	0.431	0.433	0.395	**0.392**	0.416	0.511	0.448	0.421	0.434	**0.391**
	192	0.465	0.493	0.463	0.455	**0.418**	0.436	0.443	0.489	0.448	0.460	0.476	**0.428**
	336	0.470	0.497	0.477	0.476	**0.467**	0.469	0.452	0.480	0.456	0.483	0.501	**0.444**
	720	0.510	0.527	0.470	0.509	**0.458**	0.466	**0.467**	0.479	0.554	0.513	0.531	0.468
	Avg	0.468	0.508	0.460	0.468	**0.435**	0.441	0.445	0.490	0.476	0.469	0.485	**0.433**
Weather	96	0.204	0.331	0.204	0.202	**0.195**	0.200	0.219	0.308	0.314	0.257	0.230	**0.213**
	192	**0.241**	0.339	0.255	0.252	0.255	0.242	0.269	0.311	0.382	0.302	0.278	**0.255**
	336	0.295	0.367	0.287	0.293	0.300	**0.279**	0.314	0.340	0.422	0.341	0.316	**0.297**
	720	0.342	0.390	**0.339**	0.354	**0.339**	0.360	0.367	0.388	0.494	0.398	0.370	**0.358**
	Avg	0.271	0.357	0.271	0.275	0.272	**0.270**	0.292	0.337	0.403	0.324	0.298	**0.281**
Electricity	96	0.258	0.234	0.235	0.246	0.248	**0.234**	0.247	0.350	0.258	0.265	0.256	**0.246**
	192	0.278	**0.252**	0.256	0.258	0.262	0.254	0.260	0.325	0.268	0.274	0.266	**0.260**
	336	0.334	0.269	**0.267**	0.272	0.278	0.272	0.282	0.326	0.285	0.291	0.284	**0.280**
	720	0.337	0.296	0.296	0.301	**0.293**	0.309	**0.314**	0.339	0.317	0.325	0.316	**0.314**
	Avg	0.302	**0.263**	0.264	0.269	0.270	0.267	0.276	0.335	0.282	0.289	0.280	**0.275**
Solar	96	0.222	0.990	0.243	**0.208**	0.216	0.216	0.239	1.015	0.253	0.337	0.277	**0.230**
	192	0.219	0.980	0.247	0.209	**0.205**	**0.205**	0.263	1.032	0.375	0.375	0.305	**0.252**
	336	0.209	1.004	0.241	0.226	0.219	**0.219**	0.272	1.081	0.371	0.385	0.310	**0.254**
	720	0.234	0.987	0.237	0.577	0.253	**0.229**	0.275	1.092	0.274	0.340	0.321	**0.254**
	Avg	0.221	0.990	0.242	0.305	0.223	**0.217**	0.262	1.055	0.318	0.359	0.303	**0.247**
AQshunyi	96	0.495	0.655	**0.431**	0.437	0.451	0.483	0.527	0.661	0.491	0.501	0.501	**0.489**
	192	0.502	0.565	**0.460**	**0.460**	0.468	0.501	0.543	0.602	0.510	0.512	0.510	**0.507**
	336	0.514	0.561	**0.482**	0.491	0.492	0.518	0.557	0.574	0.524	0.534	**0.519**	0.523
	720	0.539	0.545	0.492	**0.491**	0.504	0.536	0.575	0.551	0.542	0.541	0.542	**0.534**
	Avg	0.512	0.581	**0.466**	0.470	0.479	0.510	0.551	0.597	0.517	0.522	0.518	**0.513**