SMLDR: Spectral Memory Learner with Dual-Retrieval for Time Series Forecasting

Zhenxin Li, **Longquan Liao**, **Wenchang Zhang**, **Jiaying Zhang**, **Kaiwen Wei**, **Jiang Zhong** and **Linjiang Zheng**
Chongqing University
li.zhenxin@outlook.com, jefflongquan@foxmail.com, a592006923@gmail.com, zjy@cqu.edu.cn, weikaiwen@cqu.edu.cn, zhongjiang@cqu.edu.cn, zlj_cqu@cqu.edu.cn

Abstract

Time series forecasting aims to predict future values using historical observations, which is crucial for many practical applications with complex temporal dynamics. Recent frequency-domain forecasting methods have utilized spectral representations to model periodicity, but they usually rely on an implicit assumption: periodic components and aperiodic residuals are orthogonal in the frequency domain. However, in practice, this assumption often does not hold due to issues such as spectral leakage, superposition of multiple periods, and the gradual periodic changes. This paper re-examines frequency-domain modeling from the perspective of spectral non-identifiability and proposes a Spectral Memory Learner with Dual Retrieval (SMLDR) for time series forecasting. Instead of modeling periodicity in the frequency dimension, SMLDR represents periodic components as reusable spectral prototypes stored in learnable memory. The dual retrieval mechanism extracts historical periodic components and predicts them in the frequency domain, while a lightweight time-domain branch is used to model aperiodic dynamics to predict future residuals. Extensive experiments on multiple time series forecasting benchmarks show that SMLDR consistently achieves excellent performance, and the learnable memory can capture complex periodicity.

1 Introduction

Time series forecasting plays an important role in many practical applications, such as energy management [Zhang *et al.*, 2024; Yang *et al.*, 2025], weather monitoring [Price *et al.*, 2025; Conti, 2024], traffic flow control [Qiu *et al.*, 2025; Cao *et al.*, 2025] and economic planning [Wu *et al.*, 2025]. The core challenge of prediction is to model the periodic components intertwined with non-cyclical fluctuations [Zhou *et al.*, 2022b; Yi *et al.*, 2025].

Early statistical models, such as ARIMA [Ho and Xie, 1998] and exponential smoothness [Gardner Jr, 1985], assume that data has stable or linear dynamic characteristics, so it is usually difficult to generalize to complex, nonlinear and multi-cycle situations. In recent years, the deep learning model has alleviated these limitations by learning flexible time representation. In view of the importance of periodicity, a series of studies are committed to separating periodic components from dynamic characteristics [Wu *et al.*, 2021; Zhou *et al.*, 2022b; Liu *et al.*, 2024; Lin *et al.*, 2024a]. These methods are usually decomposed by time-domain smoothing operations, such as moving average [Ansley *et al.*, 1977], STL local weighted regression [Cleveland *et al.*, 1990] or exponential smoothing. This smoothing operation is equivalent to the low-pass filtering in the frequency domain [Oppenheim, 1999; Zhang *et al.*, 2025], which inhibits the high-frequency component. At the same time, frequency domain prediction methods often learn periodicity by selecting, filtering or re-weighting frequency components, such as random sampling, predefined filters or Top-K amplitude selection [Zhou *et al.*, 2022b; Xu *et al.*, 2023; Zhou *et al.*, 2022a; Piao *et al.*, 2024; Yue *et al.*, 2025; Fei *et al.*, 2025]. Both methods rely on the operation of the frequency dimension to model periodicity.

These strategies assume that meaningful periodic information can be separated by selecting frequencies. However, in real scenarios, due to spectral leakage [Oppenheim, 1999; Glorot and Bengio, 2010], multi-period harmonic overlap And periodic gradient, cyclic energy tends to spread to multiple frequency bands and overlap with non-periodic components. As shown in figure 1, this makes the frequency selection ambiguous, that is to say, a single frequency component will contain both periods and residuals. We call this phenomenon spectral non-identifiability. Therefore, strategies based on frequency selection may discard valuable periodic signals or inadvertently retain noise, resulting in performance decline in multi-cycle and long-term prediction scenarios.

We propose SMLDR (spectral memory learner with dual retrieval), which is a new frequency domain prediction framework aimed at overcoming spectral non-identifiability. A key observation is that although the time-domain waveforms with similar periodicity may be completely different, their normalized spectral prototypes are highly similar. Instead of trying to design more complex frequency selection rules, SMLDR uses a learnable spectral memory library to explicitly store these recurring spectral prototypes. Given an input sequence, SMLDR uses its spectral representation as a query to retrieve the most similar cycle patterns, which alleviates the ambiguity introduced by spectral leakage. Components that cannot be explained by the retrieved prototype are regarded as non-periodic residues, and a lightweight time domain branch is used for modeling. Through this double retrieval design, SMLDR explicitly decouples the periodic component and the residual component in the spectral space.

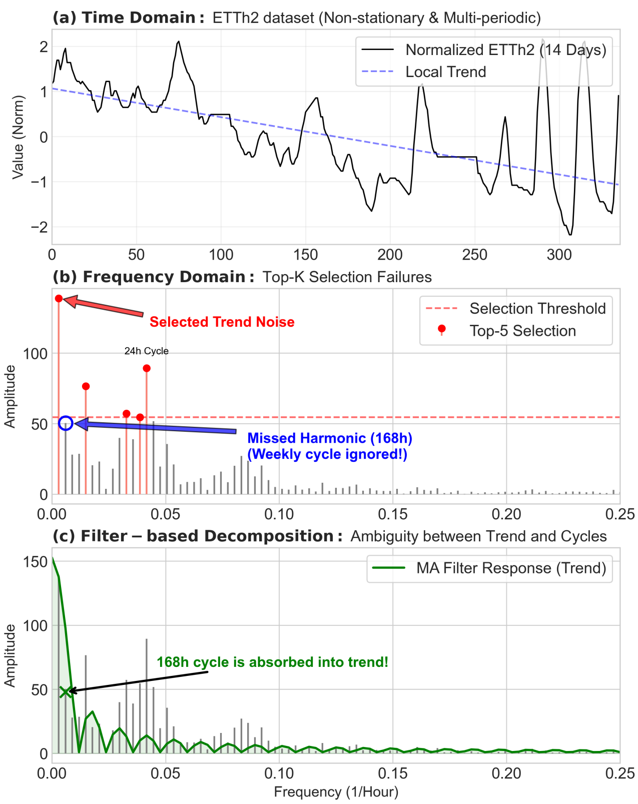

Figure 1: Illustration of spectral non-identifiability in time series forecasting. (a) A representative segment of the ETT time series exhibiting complex dynamics. (b) The corresponding frequency spectral. The Top-K strategy (red dots above the dashed threshold) fails to disentangle periodical components from noise: it erroneously selects high-energy structured noise (red arrow) while discarding low-energy but informative periodic components (green arrow) due to energy diffusion. (c) The filter used to extract trends mistakenly treats important weekly cycles as trends.

In summary, the contributions of this paper are as follows:

- We point out the core limitations of existing time series decomposition and frequency domain modeling methods, namely the ambiguity in frequency-level selection caused by spectral leakage and harmonic overlap.
- We propose SMLDR, which models periodicity through a learnable spectral memory bank and achieves the decoupling of periodic components and residuals in the frequency domain space.
- The dual-branch design explicitly separates the modeling of reusable periodic prototypes from lightweight time-domain residual modeling, making it highly suitable for prediction in multi-period scenarios.

2 Related Work

Time Series Forecasting. Research on Time Series Forecasting (TSF) is extensive, ranging from statistical models such as ARIMA [Ho and Xie, 1998] and exponential smoothing [Gardner Jr, 1985] to various deep learning methods nowadays. Recurrent and convolutional models form an important branch, including LSTNet [Lai *et al.*, 2018], DeepAR [Salinas *et al.*, 2020], SCINet [Liu *et al.*, 2022a], and TimesNet [Wu *et al.*, 2023]. Recently, Transformer-based methods have become a hot topic in time series forecasting, mainly improving prediction performance by enhancing temporal representation and attention mechanisms. Representative models include Informer [Zhou *et al.*, 2021], Autoformer [Wu *et al.*, 2021], FEDformer [Zhou *et al.*, 2022b], Pyraformer [Liu *et al.*, 2022b], PatchTST [Nie *et al.*, 2023], and iTransformer [Liu *et al.*, 2024]. Meanwhile, MLP-based models, such as TIDE [Das *et al.*, 2023], DLinear [Zeng *et al.*, 2023], SparseTSF [Lin *et al.*, 2024b], TTM [Ekambaram *et al.*, 2024] and XLinear [Chen *et al.*, 2026], demonstrate that competitive prediction accuracy can be achieved using simplified models.

Time-series decomposition. Explicit modeling of periodic components is an important research direction in time series forecasting (TSF). A large number of studies have adopted traditional Seasonal-Trend Decomposition (STD) methods [Yi *et al.*, 2025] to decompose time series into seasonal components and trend components, such as Transformer-based models Autoformer [Wu *et al.*, 2021] and FEDformer [Zhou *et al.*, 2022b]. Subsequently, DLinear [Zeng *et al.*, 2023] further adopted fixed decomposition operators combined with linear prediction, while Leddam [Yu *et al.*, 2024] replaced manually designed filters with learnable decomposition kernels. Other methods model periodicity more explicitly, including DEPTS [Fan *et al.*, 2022], which parameterizes periodic functions through time; and CycleNet [Lin *et al.*, 2024a], which learns stable periodic patterns through global periodic indices and uses lightweight Multi-Layer Perceptrons (MLPs) to predict residuals.

Frequency-domain modeling Frequency-domain modeling has become a popular paradigm for capturing global periodic structures in time series. Based on FFT-based spectral representations [Duhamel and Vetterli, 1990], , a series of recent works have explored frequency-domain forecasting from different perspectives. FEDformer [Zhou *et al.*, 2022b] implements a fast attention mechanism through low-rank approximate transformations in the frequency domain. FiLM[Zhou *et al.*, 2022a] applies Fourier analysis to preserve historical information while mitigating noise. FITS[Xu *et al.*, 2023] uses frequency domain interpolation for prediction, essentially acting as a low-pass filter. FreTS[Yi *et al.*, 2023] retains the simplicity and efficiency of the MLP architecture, while combining the global perspective and energy aggregation properties of the frequency domain. FAN[Ye *et al.*, 2024] mitigates the effects of nonstationary by filtering the top K dominant components for each input instance in the Fourier domain. Amplifier[Fei *et al.*, 2025] and Fredformer[Piao *et al.*, 2024] use different approaches to solve the problem that models tend to ignore low-energy components. FreEformer[Yue *et al.*, 2025] solves the low-rank problem of vanilla attention in frequency domain transformers.

Unlike those existing methods that mainly focus on frequency selection or filtering strategies, we re-examine periodic modeling from the perspective of spectral non-identifiability. In our opinion, the difficulty lies in distinguishing reusable and stable periodic components in the highly coupled spectral space of period and residual. Based on this insight, we have proposed a frequency domain prototype retrieval framework, which models the cycle pattern through the learned spectral prototypes and avoids relying on fixed period assumptions or unified frequency rules.

3 Preliminaries

3.1 Time Series Forecasting

Let $\mathbf{X} \in \mathbb{R}^{C\times L}$ denote a multivariate time series segment with C variates (channels) and a look-back length L. The forecasting task aims to predict the future horizon $\mathbf{Y} \in \mathbb{R}^{C\times T}$ of length T. Given a dataset $\mathcal{D} = \{(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)})\}_{i=1}^{N}$, the goal is to learn a mapping function

$$\mathcal{F} : \mathbb{R}^{C\times L} \to \mathbb{R}^{C\times T}. \tag{1}$$

that minimizes the expected forecasting error under a predefined loss function. Long-term forecasting becomes particularly challenging when the underlying dynamics exhibit multiple coexisting periodicities, gradual period variation, and non-stationary residual perturbations.

3.2 Fast Fourier Transform

Frequency-domain modeling is a common approach for capturing global periodic structures in time series. Given a discrete-time signal $x(t)$ of length L, its frequency-domain representation is obtained via the Fast Fourier Transform (FFT):

$$X(\omega_k) = \sum_{t=0}^{L-1} x(t)e^{-j2\pi kt/L}, \quad k = 0, \dots, L-1. \tag{2}$$

FFT provides an efficient transformation from the time domain to the spectral domain with a computational complexity of $O(L\log L)$, enabling scalable modeling of long sequences. Many frequency-based forecasting methods leverage this transformation to perform filtering, frequency selection, or spectral interpolation, under the assumption that meaningful periodic structures are compactly represented in the spectral space.

3.3 Spectral Non-identifiability

While frequency-domain models typically assume that periodic signals $s(t)$ and residuals $r(t)$ are spectrally orthogonal, this ideal separation is rarely achieved in practice. We formalize this challenge as *Spectral Non-identifiability*, which stems from two primary factors:

1) Spectral Leakage: Due to the finite look-back window L, the observed spectral $\mathbf{Z}$ is a convolution of the true spectrum and the Fourier transform of the window function $W_L(\omega)$. This spreads energy across neighboring bins, blurring the boundaries of periodic components [Yang and Zhou, 2022].

2) Component Entanglement: Real-world residuals often possess structured spectral content that overlaps with periodicities. The observed power spectrum satisfies:

$$|\mathbf{Z}(\omega)|^2 = |S(\omega)|^2 + |R(\omega)|^2 + \underbrace{2\mathrm{Re}\{S(\omega)R^*(\omega)\}}_{\text{Cross-term}}, \tag{3}$$

The non-zero cross-term indicates that individual frequency coefficients are coupled. Consequently, simple energy-based filtering (e.g., Top-k selection) cannot uniquely identify periodic components, leading to accumulated decomposition errors. This motivates SMLDR to bypass frequency selection or filtering and instead employ a memory-based dual-retrieval mechanism to distill identifiable spectral prototypes from a global perspective.

4 Methodology

4.1 Overall Architecture

The SMLDR proposed in this paper adopts a decomposition-based prediction paradigm, which decomposes the input time series into periodic components and residual components in the frequency domain. The design motivation of SMLDR stems from spectral unidentifiability. Traditional time series analysis regards observations as the superposition of periodic patterns and residual changes. However, due to spectral leakage and multi-period superposition, the periods and residuals in the frequency domain are highly coupled, resulting in unstable frequency selection strategies.

Algorithm 1 Forward Pass of SMLDR

Require: Historical series $\mathbf{X} \in \mathbb{R}^{C\times L}$.
Ensure: Predicted future values $\hat{\mathbf{Y}} \in \mathbb{R}^{C\times T}$.

Step 1: Normalize and Frequency Transform
$\mathbf{X}_{\text{norm}} \leftarrow \mathrm{RevIN}(\mathbf{X})$
$\mathbf{X}_{\text{fft}} \leftarrow \mathrm{rFFT}(\mathbf{X}_{\text{norm}})$
Step 2: Spectral Representation
$\mathbf{S}_{\text{raw}} \leftarrow \mathrm{Concat}(\mathrm{Re}(\mathbf{X}_{\text{fft}}), \mathrm{Im}(\mathbf{X}_{\text{fft}}))$
$\mathbf{S} \leftarrow \mathrm{Normalize}(\mathbf{S}_{\text{raw}})$
$\mathbf{Q} \leftarrow \mathbf{S}\mathbf{W}_Q$
Step 3: Spectral Memory Dual-Retrieval
$\mathbf{Z} \leftarrow \mathrm{MemoryRetrieve}(\mathbf{Q}, \mathcal{M}_K)$ {Eq. (15)}
$\mathbf{E}^{\text{hist}} \leftarrow \mathbf{Z}\mathcal{M}_V^{\text{hist}}$ {Eq. (16)}
$\mathbf{E}^{\text{pred}} \leftarrow \mathbf{Z}\mathcal{M}_V^{\text{pred}}$ {Eq. (17)}
Step 4: Seasonal Reconstruction
$\mathbf{X}_{\text{sea}} \leftarrow \mathbf{W}_{\text{hist}}\mathbf{E}^{\text{hist}}$
$\hat{\mathbf{Y}}_{\text{sea}} \leftarrow \mathbf{W}_{\text{pred}}\mathbf{E}^{\text{pred}}$
Step 5: Aperiodic Residual Modeling
$\mathbf{X}_{\text{res}} \leftarrow \mathbf{X}_{\text{norm}} - \mathbf{X}_{\text{sea}}$
$\hat{\mathbf{Y}}_{\text{res}} \leftarrow \mathrm{MLP}(\mathbf{X}_{\text{res}})$
Step 6: Aggregation and Inverse Normalization
$\hat{\mathbf{Y}}_{\text{norm}} \leftarrow \hat{\mathbf{Y}}_{\text{sea}} + \hat{\mathbf{Y}}_{\text{res}}$
$\hat{\mathbf{Y}} \leftarrow \mathrm{Denorm}(\hat{\mathbf{Y}}_{\text{norm}})$
return $\hat{\mathbf{Y}}$

To solve this problem, SMLDR explicitly separates the periodic components across the entire spectral dimension, represents the periodic components with reusable spectral prototypes, and uses the input context as query to retrieve them, thereby activating multiple periodic patterns simultaneously.

As shown in the figure 2, SMLDR consists of two branches. The spectral memory branch captures globally reusable periodic components and predicted it through dual retrieval. Meanwhile, the aperiodic residual branch models the remaining aperiodic residuals in the time domain. This approach decouples the periodic components from residual changes in the frequency domain and alleviates the entanglement between periodicity and residuals.

To alleviate distribution shift across different time windows, Reversible Instance Normalization (RevIN) is applied at both input and output stages. The final prediction is obtained as:

$$\hat{\mathbf{Y}} = \mathrm{Denorm}\left(\hat{\mathbf{Y}}_{\text{sea}} + \hat{\mathbf{Y}}_{\text{res}}\right), \tag{4}$$

where $\hat{\mathbf{Y}}_{\text{sea}} \in \mathbb{R}^{C\times T}$ and $\hat{\mathbf{Y}}_{\text{res}} \in \mathbb{R}^{C\times T}$ denote the outputs of the spectral and residual branches, respectively.

4.2 Spectral Memory Bank

Spectral Representation

Given the normalized input time series $\mathbf{X}_{\text{norm}} \in \mathbb{R}^{C\times L}$, we transform it into the frequency domain to obtain a compact representation of its global periodic structure. Specifically, we apply the real-valued Fast Fourier Transform (FFT) [Duhamel and Vetterli, 1990] along the temporal dimension:

$$\mathbf{X}_{\text{fft}} = \mathrm{rFFT}(\mathbf{X}) \in \mathbb{C}^{C\times F}, \tag{5}$$

where $F = \lfloor L/2 \rfloor + 1$ due to conjugate symmetry. To facilitate learning with real-valued networks, we construct spectral features by concatenating the real and imaginary parts:

$$\mathbf{S}_{\text{raw}} = [\mathrm{Re}(\mathbf{X}_{\text{fft}}), \mathrm{Im}(\mathbf{X}_{\text{fft}})] \in \mathbb{R}^{C\times D_{\text{freq}}}, \tag{6}$$

where $D_{\text{freq}} = 2F$. Frequency-domain magnitude is highly sensitive to amplitude variations and non-stationary scaling [Ye *et al.*, 2024], which tend to mask consistent spectral structures across different instance. To decouple spectral shape from instance-specific energy, we perform channel-wise normalization:

$$\begin{aligned}\mu_c &= \frac{1}{D_{\text{freq}}}\sum_{j=1}^{D_{\text{freq}}} \mathbf{S}_{\text{raw}}^{(c,j)},\\ \sigma_c &= \sqrt{\frac{1}{D_{\text{freq}}}\sum_{j=1}^{D_{\text{freq}}}\left(\mathbf{S}_{\text{raw}}^{(c,j)} - \mu_c\right)^2 + \epsilon}.\end{aligned} \tag{7}$$

The normalized spectral representation is then given by:

$$\mathbf{S}^{(c)} = \frac{\mathbf{S}_{\text{raw}}^{(c)} - \mu_c}{\sigma_c}, \tag{8}$$

yielding $\mathbf{S} \in \mathbb{R}^{C\times D_{\text{freq}}}$. The normalization removes amplitude-induced variance while preserving spectral shape, which is crucial for stable prototype matching under non-stationary conditions. The scale factor σ_c is retained for adaptive energy restoration.

Finally, the normalized spectral signal is projected into a latent embedding space to serve as a query:

$$\mathbf{Q} = \mathbf{S}\mathbf{W}_Q \in \mathbb{R}^{C\times d_{\text{model}}}, \tag{9}$$

where $\mathbf{W}_Q$ is a learnable linear projection. For multi-head retrieval, $\mathbf{Q}$ is reshaped into:

$$\mathbf{Q} \to \mathbf{Q}_h \in \mathbb{R}^{C\times H\times d_{\text{head}}}, \tag{10}$$

with $d_{\text{head}} = d_{\text{model}}/H$.

Prototype Retrieval in the Frequency Domain

We introduce a learnable Spectral Memory Bank to store reusable spectral prototypes that summarize recurring periodic patterns across the dataset. The memory consists of N prototypes with keys and values defined as:

$$\mathcal{M}_K \in \mathbb{R}^{N\times C\times d_{\text{model}}}, \tag{11}$$

$$\mathcal{M}_V^{\text{hist}}, \mathcal{M}_V^{\text{pred}} \in \mathbb{R}^{N\times C\times d_{\text{model}}}. \tag{12}$$

For multi-head retrieval, all memory tensors are reshaped into:

$$\mathcal{M}_{K,h}, \mathcal{M}_{V,h} \in \mathbb{R}^{N\times C\times H\times d_{\text{head}}}. \tag{13}$$

Given the query embedding, similarity scores between the input spectrum and all spectral prototypes are computed independently for each channel and head:

$$Z_{c,h,n} = \frac{\mathbf{Q}_{c,h}\cdot \mathcal{M}_{K,c,h,n}}{\sqrt{d_{\text{head}}}}, \quad n = 1,\dots,N. \tag{14}$$

Unlike the traditional attention mechanism, we have not applied Softmax normalization to the prototype dimension. This design choice is inspired by the spectral non-identifiability discussed in Section 2: due to spectral leakage and harmonic wave overlap, the real-world spectral is usually composed of multiple coupled periodic components. Forcing competitive normalization will artificially inhibit the coexistence model. On the contrary, the original similarity score $\mathbf{Z} \in \mathbb{R}^{C\times H\times N}$ acts as a retrieval weight with bias guidance, promoting multiple prototypes to work at the same time. The potential embedding retrieved for historical reconstruction and future forecasting is then calculated as follows:

$$\mathbf{E}_{c,h}^{\text{hist}} = \sum_{n=1}^{N} Z_{c,h,n}\cdot \mathcal{M}_{V,c,h,n}^{\text{hist}}, \tag{15}$$

$$\mathbf{E}_{c,h}^{\text{pred}} = \sum_{n=1}^{N} Z_{c,h,n}\cdot \mathcal{M}_{V,c,h,n}^{\text{pred}}. \tag{16}$$

The outputs from all heads are concatenated to form:

$$\mathbf{E}_c^{\text{hist}}, \mathbf{E}_c^{\text{pred}} \in \mathbb{R}^{d_{\text{model}}}. \tag{17}$$

Time-domain Reconstruction with Energy Restoration

The retrieved prototype embeddings are projected back to the time domain to reconstruct periodic components:

$$\mathbf{X}_{\text{sea}}^{(c)} = \mathbf{W}_{\text{hist}}\mathbf{E}_c^{\text{hist}} \in \mathbb{R}^{L}, \tag{18}$$

$$\hat{\mathbf{Y}}_{\text{sea}}^{(c)} = \mathbf{W}_{\text{pred}}\mathbf{E}_c^{\text{pred}} \in \mathbb{R}^{T}, \tag{19}$$

where $\mathbf{W}_{\text{hist}}$ and $\mathbf{W}_{\text{pred}}$ are learnable linear projections. Finally, the instance-specific energy is restored using the preserved scale factor:

$$\mathbf{X}_{\text{sea}}^{(c)} \leftarrow \mathbf{X}_{\text{sea}}^{(c)}\cdot\sigma_c, \quad \hat{\mathbf{Y}}_{\text{sea}}^{(c)} \leftarrow \hat{\mathbf{Y}}_{\text{sea}}^{(c)}\cdot\sigma_c. \tag{20}$$

By decoupling spectral shape selection from amplitude modeling, the Spectral Memory Bank enables stable retrieval of periodic prototypes under non-stationary conditions. The reconstructed seasonal component captures reusable periodic structures, while unexplained dynamics are naturally deferred to the residual modeling branch, forming a principled decomposition tailored for long-horizon forecasting.

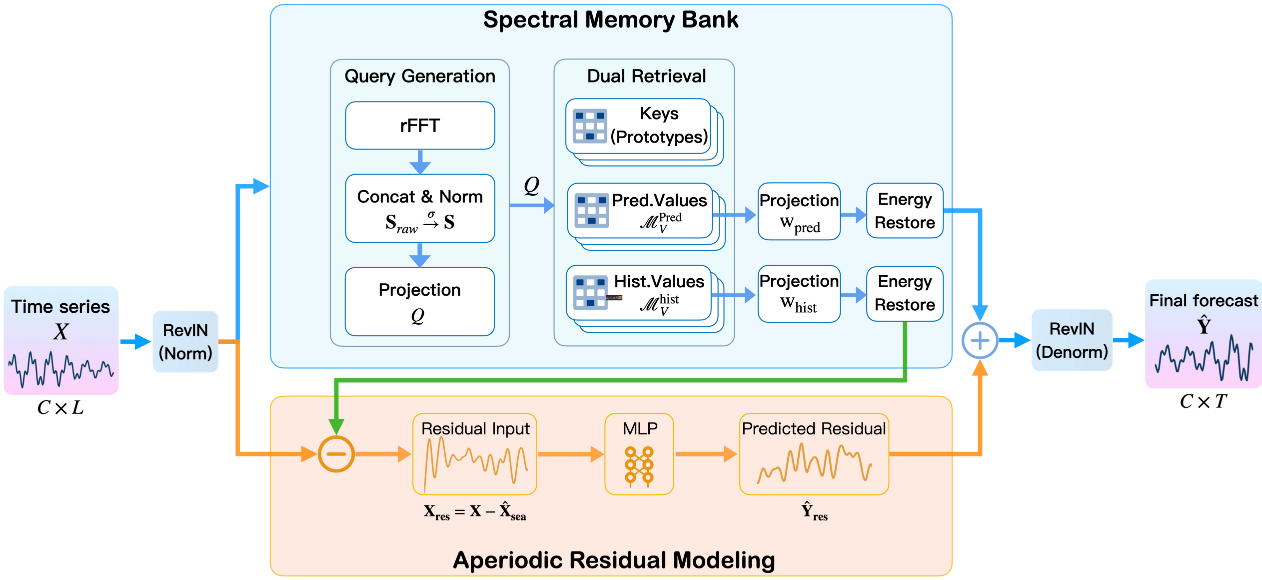

Figure 2: The architecture of the proposed SMLDR model.

4.3 Residual Modeling Branch

Once reusable spectral structures are extracted, the remaining signal primarily reflects aperiodic fluctuations, which are more effectively modeled in the time domain. The residual component of the input is computed as:

$$\mathbf{X}_{\text{res}} = \mathbf{X}_{\text{norm}} - \mathbf{X}_{\text{sea}} \in \mathbb{R}^{C \times L}. \tag{21}$$

We employ a lightweight temporal MLP to model residual dynamics:

$$\hat{\mathbf{Y}}_{\text{res}} = \text{MLP}(\mathbf{X}_{\text{res}}) \in \mathbb{R}^{C \times T}. \tag{22}$$

By explicitly separating reusable spectral prototypes from non-periodic residuals, SMLDR achieves efficient long-term forecasting under multi-period superposition and non-stationary conditions. The final forecast is obtained by additively combining the seasonal and residual predictions:

$$\hat{\mathbf{Y}} = \text{Denorm}\left(\hat{\mathbf{Y}}_{\text{sea}} + \hat{\mathbf{Y}}_{\text{res}}\right). \tag{23}$$

This formulation explicitly reflects the assumption that observed time series can be decomposed into reusable periodic prototypes and a complementary aperiodic residual component, both learned jointly in an end-to-end manner.

4.4 Computational Efficiency Analysis

We analyzed the efficiency of the proposed method and focused on the double retrieval mechanism.

Dual Retrieval. Our dual retrieval decouples attention calculation from the sequence length L. The input sequence is projected into a compact query vector, which interacts with N fixed patterns. The complexity of this retrieval process is $\mathcal{O}(N \cdot C \cdot d_{\text{model}})$. Since N is a small constant (for example, $N = 16$), and usually $N \ll L$, the retrieval cost is actually $\mathcal{O}(1)$ relative to the time dimension L.

Overall Scalability. The rest of the operations (including FFT and linear projection) are logarithmically linear or linearly extended. Specifically, FFT requires $\mathcal{O}(L \log L)$, and projection requires $\mathcal{O}(L \cdot d_{\text{model}})$. Therefore, the total time complexity is $\mathcal{O}(L \log L + L \cdot C \cdot d_{\text{model}})$, which is better than the quadratic extension of the standard Transformer.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate our model on ten widely used real-world benchmarks: **ETT** [Zhou *et al.*, 2023](ETTh1, ETTh2, ETTm1, ETTm2), **Weather** [Wu *et al.*, 2021], **ILI**, **Exchange** [Lai *et al.*, 2018], **Electricity**, **Covid-19**, **NN5**. These datasets cover diverse domains and exhibit varying levels of periodicity and noise. Following prior work [Nie *et al.*, 2023; Liu *et al.*, 2024], all datasets are preprocessed and normalized using standard normalization. We split each dataset into training, validation, and test sets by the ratio of 6:2:2 (ETT 7:1:2). The look-back window is fixed to L = 96, and prediction horizons are selected from $T \in \{96, 192, 336, 720\}$

Baselines. We compare SMLDR against a comprehensive set of SOTA models: (1) *Transformer-based*: iTransformer [Liu *et al.*, 2024], FEDformer [Zhou *et al.*, 2022b]; (2) *Frequency-based*: Amplifier [Yue *et al.*, 2025], Fredformer [Piao *et al.*, 2024], TimesNet [Wu *et al.*, 2023]; (3) *Linear/MLP-based*: DLinear [Zeng *et al.*, 2023], CycleNet [Lin *et al.*, 2024a], TQNet [Lin *et al.*, 2025]

Implementation Details. All experiments in this study were carried out using PyTorch [Paszke *et al.*, 2019] on one single NVIDIA RTX 3090 GPU with 24GB. Optimization is performed using Adam [Adam and others, 2014] and early stopping patience of 5 epochs. We use Mean Squared Error (MSE) as the loss function and report the results using both MSE and Mean Absolute Error (MAE) as evaluation metrics.

5.2 Long-term Forecasting Performance

Table 1 reports the average prediction results of all methods on different datasets. SMLDR achieves lower prediction errors

Dataset	SMLDR ours		TQNet 2025		Amplifier 2025		iTranformer 2024		FredFormer 2024		CycleNet 2024		DLinear 2023		TimesNet 2023		FEDformer 2022	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.421**	**0.421**	0.441	0.434	0.444	0.435	0.459	0.451	0.436	0.426	0.457	0.441	0.456	0.450	0.458	0.450	0.440	0.460
ETTh2	**0.362**	**0.390**	0.378	0.402	0.384	0.408	0.383	0.405	0.365	0.394	0.388	0.409	0.530	0.499	0.414	0.427	0.437	0.449
ETTm1	**0.371**	**0.392**	0.377	0.393	0.379	0.396	0.407	0.410	0.380	0.387	0.379	0.396	0.402	0.405	0.400	0.406	0.448	0.452
ETTm2	0.271	0.317	0.277	0.322	0.272	0.318	0.288	0.332	0.279	0.324	**0.266**	**0.314**	0.340	0.391	0.291	0.333	0.305	0.349
Weather	**0.241**	**0.267**	0.242	0.269	0.243	0.271	0.258	0.278	0.246	0.273	0.243	0.271	0.265	0.317	0.259	0.287	0.309	0.360
Exchange	**0.351**	**0.397**	0.367	0.405	0.361	0.402	0.360	0.403	0.374	0.408	0.366	0.404	0.354	0.414	0.416	0.443	0.519	0.500
ILI	**1.785**	**0.841**	1.949	0.855	1.820	0.904	2.276	0.988	2.269	1.201	2.431	0.984	2.616	1.090	2.139	0.931	2.847	1.144
Electricity	0.173	0.262	**0.165**	**0.259**	0.171	0.265	0.178	0.270	0.175	0.269	0.168	0.259	0.212	0.300	0.193	0.295	0.213	0.327
Covid-19	**1.404**	**0.048**	1.523	0.052	2.148	0.075	1.785	0.213	1.572	0.052	1.681	0.163	136.948	0.805	2.930	0.093	2.565	0.187
NN5	**0.677**	**0.568**	0.800	0.641	1.005	0.740	0.734	0.601	0.740	0.609	0.945	0.729	1.399	0.915	1.675	1.028	0.870	0.673
1st Count	**8**	**8**	**1**	**1**	0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0

Table 1: Time series forecasting performance comparison. Results are averaged over horizons $\{24, 36, 48, 60\}$ for ILI, Covid-19, and NN5, and $\{96, 192, 336, 720\}$ otherwise, with input length $L = 36$ and $L = 96$, respectively. Best results are shown in **bold**, and the second best are underlined. The 1st Count row indicates the number of best-performing cases. Detailed results for individual horizons are provided in Table 6 of Appendix B.

on the vast majority of datasets, indicating its strong effectiveness in modeling complex temporal dynamics. The ETT dataset has significant multi-scale periodicity, while Exchange and NN5 contain complex trends and noise with relatively hidden periodicity. Compared with Transformer-based models, SMLDR explicitly models periodic components in the frequency domain, can capture stable periodic patterns on datasets with different characteristics, and has higher computational efficiency than Transformer-like models. Compared with existing frequency-domain baseline methods, SMLDR does not rely on frequency selection-based strategies. Instead, it retrieves spectral prototypes from a learnable memory bank, which alleviates the ambiguity caused by the coupling between periodicity and residuals. In conclusion, SMLDR can flexibly model multiple coexisting periodic components and maintain stable performance on datasets with different periodicities.

5.3 Ablation Study

To evaluate the effectiveness of the key components in SMLDR, we conducted ablation studies on the ETTh1 and Weather datasets. And we designed three variants: (1) *w/o SMB*, which removes the spectral memory bank; (2) *TimeMatch*, which performs prototype retrieval in the time domain without using spectral prototypes; (3) *w/o Residual*, which removes the residual modeling branch.

Datasets	Metrics	TimeMatch		w/o MemoryBank		w/o Residual MLP		SMLDR (Original)	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	0.278	0.329	0.296	0.345	0.284	0.333	**0.276**	**0.328**
	192	0.361	0.382	0.384	0.397	0.376	0.393	**0.355**	**0.380**
	336	0.415	0.426	0.442	0.446	0.434	0.433	**0.400**	**0.417**
	720	0.436	0.448	0.452	0.461	0.450	0.456	**0.416**	**0.436**
	Avg	0.373	0.396	0.393	0.412	0.386	0.404	**0.362**	**0.390**
Weather	96	0.162	0.204	0.192	0.237	0.162	0.210	**0.155**	**0.199**
	192	0.209	0.252	0.233	0.267	0.210	0.252	**0.205**	**0.245**
	336	0.266	0.293	0.290	0.310	0.266	0.294	**0.261**	**0.286**
	720	0.350	0.347	0.360	0.352	0.347	0.346	**0.343**	**0.340**
	Avg	0.247	0.274	0.269	0.292	0.246	0.276	**0.241**	**0.267**

Table 2: Ablation Study of SMLDR on ETTh2 and Weather Datasets

As shown in Table 2, removing the Spectral Memory Bank leads to consistent performance degradation, indicating that explicit spectral prototype retrieval is essential for capturing stable periodic structures. Time-domain matching further degrades performance, especially for long-horizon forecasting, demonstrating the advantage of frequency-domain representations in handling phase variation and spectral coupling. Removing the residual branch also harms performance, confirming the necessity of explicitly modeling aperiodic components to prevent interference with periodic prototype learning.

5.4 Spectral Memory Activation Analysis

The activation distribution of spectral memory patterns in Figure 3 illustrates how SMLDR performs content-adaptive periodic modeling in the frequency domain. Each memory entry corresponds to a learned spectral prototype encoding a reusable periodic structure, while activation scores reflect its relevance to the current input. The activation dynamics show that different prototypes are selectively engaged across time, rather than relying on a single dominant frequency, which is consistent with the presence of multi-period structures and their gradual evolution in real-world time series. Moreover, the smoothly varying activation intensities indicate that SMLDR adapts prototype retrieval continuously.

This behavior due to the design of SMLDR: stable periodic patterns are captured by reusable spectral prototypes, while components not explained by these prototypes are handled by the residual branch at the same frequencies. Such a flexible separation mitigates spectral identifiability issues inherent to frequency selection or filtering rules.

5.5 Impact of hyperparameter N

We conducted a sensitivity analysis of the number of memory patterns N to test its impact on predictive performance. The experiment was conducted on three benchmarks including ETTh1, Exchange and Weather. For all datasets, the length of the backwindow is fixed at 96, and the prediction horizons considers $H \in \{96, 192, 336, 720\}$. The results of the report are obtained by taking an average of all predicted lengths.

As shown in Figure 4, there is no monotonous relationship between predictive performance and the number of memory patterns N. When $N = 1$, the performance of the model on each data set is generally poor, indicating that a single periodic

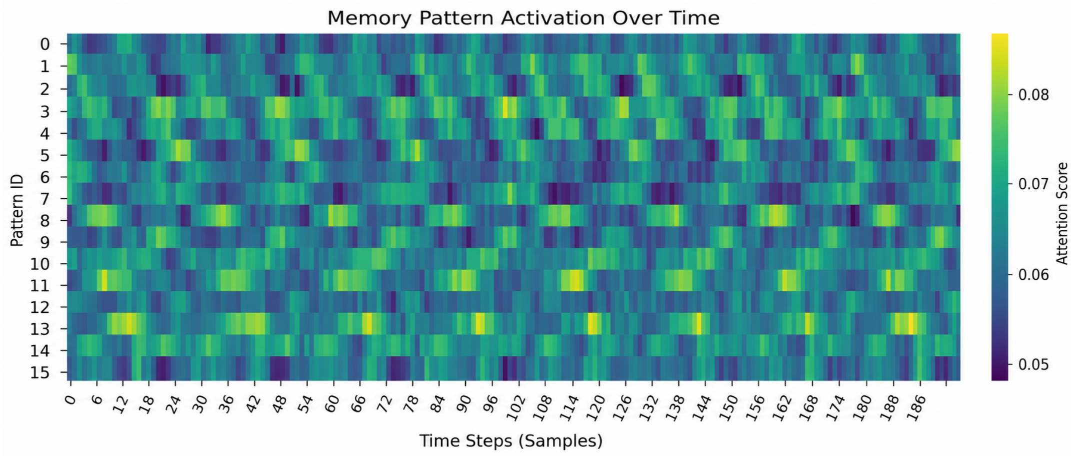

Figure 3: Memory patterns activation on weather dataset. The horizontal axis represents the time steps, and the vertical axis represents the different periodic patterns.

prototype is not enough to capture the multi-periodic structure in the real-world time series. When N is in the range of 8 to 16, the model achieves good results on different data sets and has low sensitivity to the specific selection of N, which shows that a relatively small set of spectral prototypes is enough to represent the main periodic pattern.

Further increase of N does not bring additional benefits, and may even lead to a slight decline in performance, but the overall result is still close to the best. This phenomenon shows that there is redundancy in the periodic structure in the spectral space, and too many prototypes may introduce unnecessary characterizations and bring noise.

In general, the proposed spectral memory mechanism can effectively model the multi-period structure through a compact prototype set, and the selection of N is robust.

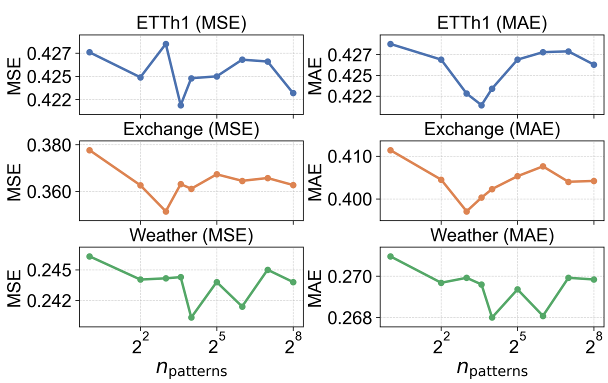

Figure 4: Model effects vary with the number of memory patterns N on different datasets

5.6 Performance with varied look-back length

We conduct experiments on the ETTh1 and Weather datasets, in which the look-back length is set to 48 to 720, and the prediction horizon was fixed at 96. Longer input windows often bring redundant or noisy historical information, increasing the difficulty of modeling.

As shown in Figure 5, the performance of many comparison methods fluctuates significantly with the increase of the look-back length, which indicates that they are insufficient in continuously capturing long-range historical dependencies. In contrast, SMLDR achieves the lowest mean square error (MSE) in all look-back lengths. With increasing the look-back length, the performance of SMLDR remains stable and even improves. These phenomena show that SMLDR can effectively extract reusable and valuable cycle patterns from long input sequences while reducing the negative impact of noise.

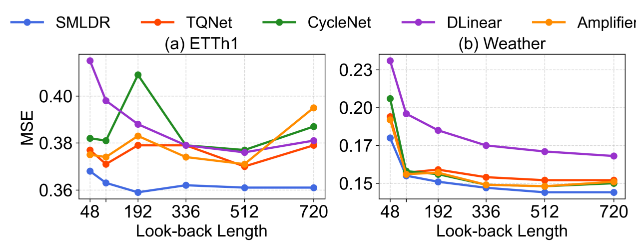

Figure 5: Performance evaluation of SMLDR and competing models with varying look-back lengths, with a forecast horizon of 96.

6 Conclusion

We propose SMLDR, a time-frequency prediction framework that models periodic components through reusable spectral prototypes instead of selecting multiple frequency components. By introducing a spectral memory bank and performing dual retrieval, SMLDR avoids frequency selection strategies and mitigates the strong coupling between periodicity and residuals caused by spectral leakage. Furthermore, SMLDR decouples the modeling of periodic prototypes from a lightweight time-domain residual branch, maintaining high computational efficiency. Extensive experiments on various benchmark datasets demonstrate that this prototype-centered spectral modeling paradigm can capture complex periodicities and exhibit excellent performance across different context lengths.

References

[Adam and others, 2014] Kingma DP Ba J Adam et al. A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 1412(6), 2014.

[Ansley *et al.*, 1977] Craig F Ansley, W Allen Spivey, and William J Wrobleski. On the structure of moving average processes. *Journal of Econometrics*, 6(1):121–134, 1977.

[Cao *et al.*, 2025] Lingxiao Cao, Bin Wang, Guiyuan Jiang, Yanwei Yu, and Junyu Dong. Spatiotemporal-aware trend-seasonality decomposition network for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11463–11471, 2025.

[Chen *et al.*, 2026] Xinyang Chen, Huidong Jin, Yu Huang, and Zaiwen Feng. Xlinear: A lightweight and accurate mlp-based model for long-term time series forecasting with exogenous inputs. *arXiv preprint arXiv:2601.09237*, 2026.

[Cleveland *et al.*, 1990] Robert B Cleveland, William S Cleveland, Jean E McRae, Irma Terpenning, et al. Stl: A seasonal-trend decomposition. *J. off. Stat*, 6(1):3–73, 1990.

[Conti, 2024] Silvia Conti. Artificial intelligence for weather forecasting. *Nature Reviews Electrical Engineering*, 1(1):8–8, 2024.

[Das *et al.*, 2023] Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*, 2023.

[Duhamel and Vetterli, 1990] Pierre Duhamel and Martin Vetterli. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.

[Ekambaram *et al.*, 2024] 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. *Advances in Neural Information Processing Systems*, 37:74147–74181, 2024.

[Fan *et al.*, 2022] Wei Fan, Shun Zheng, Xiaohan Yi, Wei Cao, Yanjie Fu, Jiang Bian, and Tie Yan Liu. Depts: Deep expansion learning for periodic time series forecasting. In *10th International Conference on Learning Representations, ICLR 2022*, 2022.

[Fei *et al.*, 2025] Jingru Fei, Kun Yi, Wei Fan, Qi Zhang, and Zhendong Niu. Amplifier: Bringing attention to neglected low-energy components in time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11645–11653, 2025.

[Gardner Jr, 1985] Everette S Gardner Jr. Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1):1–28, 1985.

[Glorot and Bengio, 2010] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[Ho and Xie, 1998] Siu Lau Ho and Min Xie. The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216, 1998.

[Lai *et al.*, 2018] 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*, pages 95–104, 2018.

[Lin *et al.*, 2024a] Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37:106315–106345, 2024.

[Lin *et al.*, 2024b] Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: Modeling long-term time series forecasting with 1k parameters. *arXiv preprint arXiv:2405.00946*, 2024.

[Lin *et al.*, 2025] Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. *arXiv preprint arXiv:2505.12917*, 2025.

[Liu *et al.*, 2022a] Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022.

[Liu *et al.*, 2022b] 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 *# PLACEHOLDER_PARENT_METADATA_VALUE#*, 2022.

[Liu *et al.*, 2024] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. 2024.

[Nie *et al.*, 2023] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *International Conference on Learning Representations*, 2023.

[Oppenheim, 1999] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Piao *et al.*, 2024] Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2400–2410, 2024.

[Price *et al.*, 2025] Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, et al. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025.

[Qiu *et al.*, 2025] Xiangfei Qiu, Liu Yang, Hanyin Cheng, Xingjian Wu, Rongjia Wu, Zhigang Zhang, Ding Tu, Chenjuan Guo, Bin Yang, Christian S Jensen, et al. Multi-scale spatial-temporal hypergraph network with lead-lag structures for stock time series forecasting. *arXiv preprint arXiv:2509.23668*, 2025.

[Salinas *et al.*, 2020] 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.

[Wu *et al.*, 2021] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. volume 34, pages 22419–22430, 2021.

[Wu *et al.*, 2023] Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

[Wu *et al.*, 2025] Xingjian Wu, Xiangfei Qiu, Hanyin Cheng, Zhengyu Li, Jilin Hu, Chenjuan Guo, and Bin Yang. Enhancing time series forecasting through selective representation spaces: A patch perspective. *arXiv preprint arXiv:2510.14510*, 2025.

[Xu *et al.*, 2023] Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. *arXiv preprint arXiv:2307.03756*, 2023.

[Yang and Zhou, 2022] Jun Yang and Zhou Zhou. Spectral inference under complex temporal dynamics. *Journal of the American Statistical Association*, 117(537):133–155, 2022.

[Yang *et al.*, 2025] Mao Yang, Yutong Huang, Chuanyu Xu, Chenyu Liu, and Bozhi Dai. Review of several key processes in wind power forecasting: Mathematical formulations, scientific problems, and logical relations. *Applied Energy*, 377:124631, 2025.

[Ye *et al.*, 2024] Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for nonstationary time series forecasting. *Advances in Neural Information Processing Systems*, 37:31350–31379, 2024.

[Yi *et al.*, 2023] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

[Yi *et al.*, 2025] Kun Yi, Qi Zhang, Wei Fan, Longbing Cao, Shoujin Wang, Hui He, Guodong Long, Liang Hu, Qingsong Wen, and Hui Xiong. A survey on deep learning based time series analysis with frequency transformation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 6206–6215, 2025.

[Yu *et al.*, 2024] Guoqi Yu, Jing Zou, Xiaowei Hu, Angelica I Aviles-Rivero, Jing Qin, and Shujun Wang. Revitalizing multivariate time series forecasting: learnable decomposition with inter-series dependencies and intra-series variations modeling. In *Proceedings of the 41st International Conference on Machine Learning*, pages 57818–57841, 2024.

[Yue *et al.*, 2025] Wenzhen Yue, Yong Liu, Xianghua Ying, Bowei Xing, Ruohao Guo, and Ji Shi. Freeformer: Frequency enhanced transformer for multivariate time series forecasting. *CoRR*, 2025.

[Zeng *et al.*, 2023] Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

[Zhang *et al.*, 2024] Yagang Zhang, Xue Kong, Jingchao Wang, Hui Wang, and Xiaodan Cheng. Wind power forecasting system with data enhancement and algorithm improvement. *Renewable and Sustainable Energy Reviews*, 196:114349, 2024.

[Zhang *et al.*, 2025] Q Zhang, YUTING SUN, HONGGANG WEN, PENG YANG, XINZHU LI, MING LI, KWOK-YAN LAM, PIETRO LIÒ, SIU-MING YIU, and HONGZHI YIN. Time series analysis in frequency domain: A survey of open challenges opportunities and benchmarks. *arXiv: 2504.07099*, 2025.

[Zhou *et al.*, 2021] 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 *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

[Zhou *et al.*, 2022a] Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in neural information processing systems*, 35:12677–12690, 2022.

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

[Zhou *et al.*, 2023] H Zhou, S Zhang, J Peng, S Zhang, J Li, H Xiong, and W Zhang Informer. Beyond efficient transformer for long sequence time-series forecasting., 2021. *DOI: https://doi. org/10.1609/aaai. v35i12*, 17325, 2023.