Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ31299

A Survey of Deep Time Series Forecasting with Spectral Analysis

Hengxu Lai

School of Urban Design, Wuhan University, Wuhan, China
2021302093042@whu.edu.cn

Abstract. Time series forecasting is a core technology in critical domains such as energy dispatch and traffic management, yet its performance is hindered by challenges including long-term dependencies, multi-scale structures, and data non-stationarity. In recent years, integrating spectral analysis with deep learning has emerged as a significant trend for improving forecasting accuracy and efficiency. This paper systematically reviews progress in this field by introducing a challenge-oriented classification framework encompassing four dimensions: long-term dependency modeling, multi-scale feature extraction, lightweight design, and multi-task general-purpose modeling. Within this framework, we conduct a comparative analysis of representative methods, including Autoformer, FEDformer, and TimesNet, among others. These methods enhance modeling capacity for complex temporal patterns through mechanisms such as spectral sparsification, adaptive frequency filtering, and temporal multi-periodic transformations. We evaluate methods on eight mainstream benchmark datasets across multiple forecast horizons (96–720 steps). Results demonstrate that spectral sparsification and memory filtering mitigate error accumulation in long-term forecasting; multi-scale decomposition structures balance short-term fluctuations and long-term trends; and lightweight linear models achieve superior parameter efficiency on high-dimensional stationary data. By synthesizing technical pathways and scenario-based comparisons, this study offers practical guidance for model selection in engineering applications. Finally, we outline future research directions, including time-varying period detection and joint time-frequency representation, to enhance the robustness of forecasting models in non-stationary environments and facilitate their real-world deployment.

Keywords: time series forecasting, spectral analysis, deep learning, Fourier transform, frequency domain

1. Introduction

Time series forecasting is central to energy, finance, weather, and industrial monitoring. Modern data are often multivariate, long-horizon, and non-stationary; when periodicity coexists with disturbances (e.g., electricity, traffic, exchange rates), models must capture cross-scale structure, suppress noise, and remain computationally efficient.

Classical statistical and signal-processing methods (e.g., ARIMA/SARIMA, state-space models, and wavelet-based decomposition) can be effective on short-to-medium univariate series, but they

DOI: 10.54254/2755-2721/2026.TJ31299

often rely on stationarity/linearity assumptions, require hand-crafted pipelines, and struggle with high-dimensional correlations and long-horizon error accumulation. In many pipelines, FFT/DWT are used only as preprocessing instead of learnable, end-to-end components.

Deep models (RNN/CNN/Transformer variants) improve representation learning but still face difficulty with very long contexts, multi-scale coupling, and robustness under non-stationarity [1]. Since 2021, Fourier/wavelet/time–frequency mechanisms have been increasingly embedded into neural architectures as learnable modules, enabling spectral sparsity, adaptive filtering, and efficient long-range modeling. The contributions of this work are:

• A novel, challenge-oriented four-dimensional taxonomy. Departing from traditional classification schemes based solely on model architecture, this work constructs an analytical framework organized around four pivotal challenges: long-term dependency modeling, multi-scale feature extraction, lightweight design, and multi-task generalization. This taxonomy provides a unified, structured lens for systematically reviewing the integration of spectral analysis and deep learning in time series forecasting.

• A standardized benchmarking system accompanied by scenario-specific guidelines. Through a unified and systematic evaluation across multiple benchmark datasets and forecast horizons, this study offers an in-depth comparison of the performance and intrinsic mechanisms of representative methods. Building on this analysis, it develops practical algorithm selection guidelines tailored to typical application scenarios such as electricity load and traffic flow forecasting, thereby addressing the issues of fragmented experimental results and the lack of actionable guidance in the field.

• A clarification of technical boundaries and a roadmap for future research. This work delineates the performance boundary of existing spectral fusion methods, revealing their stronger suitability for stationary or quasi-stationary signals and identifying their primary limitations in handling time-varying periodicities and cross-scale causal relationships. Based on this critical analysis, it systematically proposes key research directions for the future, including joint time-frequency representation and adaptive periodicity detection, thereby charting a clear technical pathway for developing next-generation robust forecasting models capable of operating in complex, non-stationary environments.

2. Related work

2.1. Fundamentals of time series forecasting

Given a history window $\mathbf{X} \in \mathbb{R}^{L \times D}$, forecasting predicts the next H steps $\mathbf{Y} \in \mathbb{R}^{H \times D}$. Core challenges are long-range dependencies, multi-scale dynamics (trend vs. periodicity vs. fluctuations), non-stationarity, and efficiency constraints (latency/memory) in deployment.

2.2. Fundamentals of spectral analysis

2.2.1. Fourier transform

The discrete Fourier transform (DFT) maps a length- L time-domain sequence $x[n]$ to frequency coefficients

$$X\left[k\right] = \sum_{n=0}^{L-1} x\left[n\right] e^{-j2\pi kn/L}, \quad k = 0, \ldots, L-1, \tag{1}$$

DOI: 10.54254/2755-2721/2026.TJ31299

with inverse

$$x\left[n\right] = \frac{1}{L}\sum_{k=0}^{L-1} X\left[k\right] e^{j2\pi kn/L}. \tag{2}$$

FFT computes DFT in $O(L\log L)$. The magnitude $|X[k]|$ indicates spectral energy and the phase $\angle X[k]$ indicates phase; in forecasting, DFT/FFT is mainly used to identify dominant periods and perform denoising/compression via spectral sparsity.

2.2.2. Wavelet transform

The discrete wavelet transform (DWT) performs multi-scale decomposition with time–frequency localization. In practice, the Mallat filter bank applies a low-pass filter (approximation) and a high-pass filter (detail) followed by downsampling, producing approximation coefficients A_j and detail coefficients D_j at level j (with $A_0[n] = x[n]$ and typically $\left|A_j\right| \approx \left\lceil L/2^j \right\rceil$). After J levels, DWT yields $x\left[n\right] \xrightarrow{\text{DWT}} \left\{A_J, D_J, \ldots, D_1\right\}$, where A_J captures long-term trends and D_j captures localized fluctuations. The inverse transform (IDWT) reconstructs the time-domain signal using corresponding synthesis (dual) filters; under orthogonal/biorthogonal conditions, perfect reconstruction is possible. Key properties include: (1) time–frequency localization (long windows for low frequencies and short windows for high frequencies); (2) multi-resolution analysis (J levels yield $J+1$ sub-bands with total complexity $O(JL)$); and (3) basis flexibility (e.g., Haar, Daubechies, Symlets).

2.3. Fundamentals of deep learning architectures

2.3.1. Multilayer perceptron (MLP) and linear layers

A simple channel-wise linear predictor applies a learnable affine transformation to each time step independently:

$$Y = WX + b \tag{3}$$

where $\mathbf{X} \in \mathbb{R}^{L\times D}$ is the input sequence and $\mathbf{Y} \in \mathbb{R}^{H\times D}$ is the output. The cost and parameters scale as $O(HLD)$. Channel independence—processing each variable separately—allows the model to capture univariate temporal patterns without explicit cross-variable interaction, but limits the ability to exploit correlations across channels.

2.3.2. Convolutional Neural Networks (CNNs)

One-dimensional convolution extracts local temporal patterns with a sliding kernel of size K . For input $\mathbf{x} \in \mathbb{R}^L$ and kernel $\mathbf{w} \in \mathbb{R}^K$:

DOI: 10.54254/2755-2721/2026.TJ31299

$$y\left[i\right] = \sum_{k=0}^{K-1} w\left[k\right] \cdot x\left[i+k\right] + b \tag{4}$$

The receptive field is K; stacking N layers enlarges it roughly linearly, while dilated kernels expand it without increasing parameters. The per-layer cost is $O(LKD)$ for D channels. Two-dimensional convolution applies when sequences are reshaped into $H \times W$ patches (e.g., period-wise reshaping), with cost $O(HWK_{\mathrm{h}}K_{\mathrm{w}}D)$. CNNs provide parameter sharing and translation invariance, but locality limits long-range dependencies unless depth or dilation is used.

2.3.3. Transformers and self-attention

Standard self-attention measures pairwise similarity inside a sequence through projected queries, keys, and values. For input $\mathbf{X} \in \mathbb{R}^{L\times D}$:

$$extAttention\left(Q,K,V\right) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right)V \tag{5}$$

The scaling by d_k stabilizes dot-product magnitudes. Complexity is $O\left(L^2D\right)$ due to the dense $L \times L$ score matrix. Multi-head attention runs h heads in parallel, concatenates their outputs, and applies an output projection to mix heads; this enables modeling multiple interaction patterns. Stacking encoder layers provides global receptive fields, but the quadratic cost becomes the main bottleneck on very long sequences; see [2].

3. A problem-oriented taxonomy of models

A problem-oriented taxonomy of spectral methods: four core objectives and representative approaches.

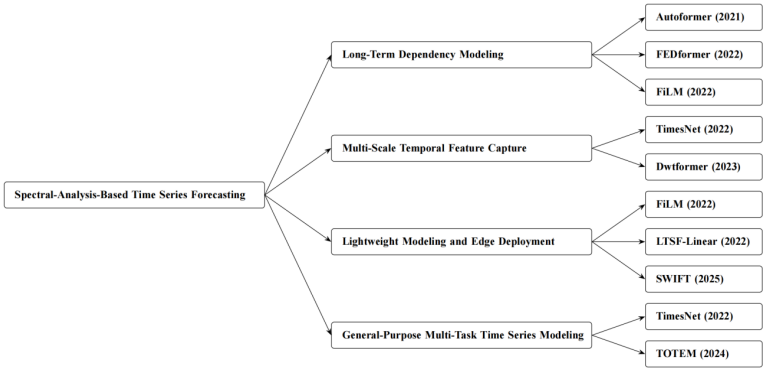

Figure 1. A problem-oriented taxonomy of spectral methods: four core objectives and representative approaches

3.1. Long-term dependency modeling

Long-horizon forecasting mainly requires (i) capturing global dependencies over long contexts (e.g., periodicity across hundreds of steps) and (ii) compressing history without excessive information

DOI: 10.54254/2755-2721/2026.TJ31299

loss. While self-attention offers global modeling, its $O(L^2)$ cost limits very long sequences; recurrent models may also forget long-range information due to recursive updates. To reduce complexity while preserving long-range representations, we summarize three routes: autocorrelation-driven delay aggregation, frequency-domain sparse modeling, and low-dimensional memory with spectral filtering.

Autoformer [3] introduces an autocorrelation-driven delay aggregation mechanism for long-horizon forecasting. Leveraging a periodicity prior, it builds an autocorrelation operator, selects the Top- k most informative delays in the time domain, and aggregates them to avoid explicitly constructing an $L \times L$ attention matrix. Together with trend/seasonality decomposition and denoising, Autoformer achieves more stable performance on long sequences with strong periodicity. Figure 2 illustrates the overall workflow.

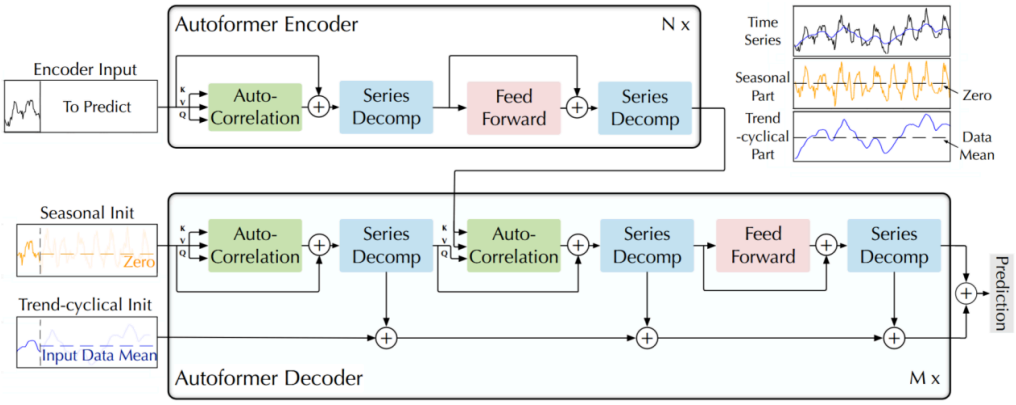

Figure 2. Overall architecture of Autoformer

FEDformer [4] is a frequency-enhanced decomposed Transformer that can be summarized by three components. (1) FEB (Frequency-Enhanced Block): the input is linearly projected and transformed by FFT; a small set of key frequencies is retained via random sampling; a learnable complex-valued kernel modulates both magnitude and phase, followed by IFFT back to the time domain. (2) FEA (Frequency-Enhanced Attention): FFT and sampling are applied to Q/K/V; attention is computed directly in the frequency domain and combined with V without constructing an $L \times L$ matrix, then transformed back via IFFT. (3) MOEDecomp: a mixture-of-experts decomposition dynamically separates and recombines trend/seasonal components to better handle non-stationarity. Overall, frequency subsampling and complex-domain operations yield near-linear complexity while preserving both magnitude and phase information.

FiLM [5] follows the principle of compress first, then filter in the frequency domain. Specifically, a length- L history is projected onto a low-dimensional memory state $s \in \mathbb{R}^N$ using Legendre polynomial bases ($N \ll L$). FiLM then applies FFT to s and uses a learnable frequency mask to select informative bands and suppress noise, followed by IFFT to reconstruct a denoised memory representation. In addition, low-rank parameterization (e.g., factorizing a frequency mask or related linear mappings as $UV^\top$) reduces parameters and computation. By avoiding spectral operations on the original high-dimensional sequence and focusing computation on compact memory states, FiLM preserves long-range dependencies while substantially reducing memory and compute cost, which is particularly suitable for noise-sensitive domains such as finance and meteorology.

DOI: 10.54254/2755-2721/2026.TJ31299

3.2. Multi-scale temporal feature capture

Time series are inherently multi-scale: low frequencies capture long-term trends, whereas high frequencies reflect short-term fluctuations. Classical models often trade off global context and local detail (e.g., fixed-window convolutions vs. global attention). We highlight two spectral routes to mitigate this tension: wavelet transforms that decompose signals into multiple scales for separate modeling, and implicit frequency-domain modeling that discovers dominant periods and reshapes sequences into 2D forms to jointly capture intra-period and inter-period variations.

TimesNet [6] introduces an FFT-based mechanism to automatically discover dominant periods. Given an input sequence, TimesNet computes its spectrum $X[k]$ via FFT and selects several dominant periods $p_i=L/k_i$ according to the magnitudes $|X[k]|$. The sequence is then reshaped into multiple 2D patches based on these periods, and 2D convolutions are applied to jointly model intra-period and inter-period variations. Finally, multi-scale blocks are aggregated for unified prediction. Figure 3 summarizes the three-step pipeline: dominant frequency/period discovery by FFT, period-based 2D reshaping, and 2D convolutional modeling over the intra-period $\times$ inter-period axes.

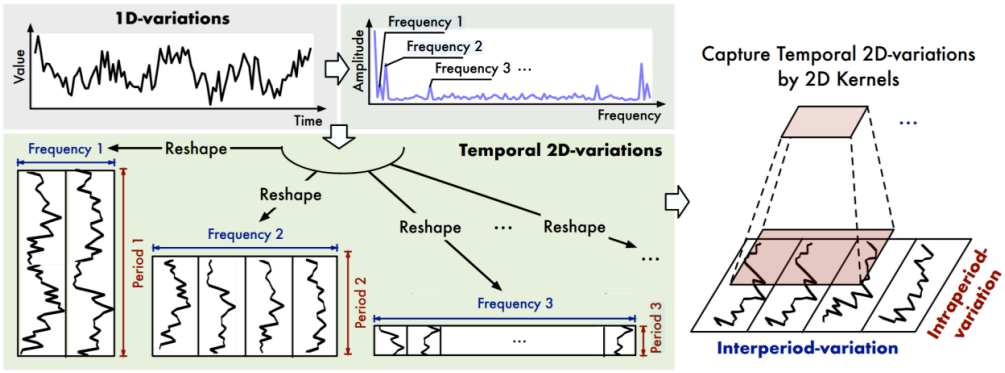

Figure 3. Overall architecture of TimesNet

Dwtformer [7] applies multi-level discrete wavelet transform (DWT) to decompose the input into $\{A_J,D_J,\ldots,D_1\}$, where the low-frequency component A_J captures long-term trends and high-frequency components D_j capture local details. At each scale, FFT is used to extract dominant periods and reshape representations into 2D tensors, which are then processed by 2D convolutions to jointly model intra-period and inter-period structures. Finally, inverse DWT reconstructs the output back to the time domain. By combining DWT downsampling + spectral period discovery + 2D convolution, Dwtformer aims to capture both long-period patterns and short-term fluctuations in multi-scale scenarios such as electricity and traffic forecasting.

3.3. Lightweight modeling and edge deployment

Edge/online deployment favors low latency and small memory. In contrast, complex spectral designs (e.g., deep wavelet hierarchies or high-dimensional spectral attention) can be costly. We therefore summarize three lightweight routes: low-dimensional memory with spectral filtering, implicit frequency-domain linear modeling, and minimal wavelet structures.

FiLM [5] is representative of the compress-then-filter paradigm: it projects a length- L history into a low-dimensional memory state $s\in\mathbb{R}^N$ ($N\ll L$), applies FFT in the memory space, and uses a

DOI: 10.54254/2755-2721/2026.TJ31299

learnable frequency mask to select informative bands and suppress noise, followed by IFFT reconstruction. Low-rank parameterization further reduces parameters and computation. By concentrating spectral operations on compact memory representations, FiLM achieves favorable efficiency for deployment, especially in noise-sensitive settings such as finance and meteorology.

LTSF-Linear [8] models the input–output relationship using channel-independent linear mappings, $\hat{Y}=WX+b$. Each row of the weight matrix can be viewed as a time-domain FIR filter; by the convolution theorem, this is equivalent to constructing low-pass or band-pass filters in the frequency domain, which tends to preserve low/mid-frequency trends and dominant periodicities while suppressing high-frequency noise. This minimalist design scales linearly with sequence length and often matches or even surpasses more complex attention-based models on multiple benchmarks, serving as a lightweight baseline from an implicit frequency-domain perspective.

SWIFT-MLP [9] adopts a minimal wavelet structure to compress sequence length and align scales. It first performs a single-level discrete wavelet decomposition to obtain approximation and detail sub-sequences. Each branch is then processed by lightweight mappings and upsampled to align with the original resolution. Finally, outputs from the original/low-frequency/high-frequency branches are fused by weighted aggregation. With the minimum number of decomposition levels, SWIFT-MLP reduces computation and memory usage while retaining multi-scale representations, making it suitable for edge devices and online inference.

3.4. General-purpose multi-task time series modeling

Time-series tasks span forecasting, classification, imputation, and anomaly detection. Training separate models for each task increases maintenance cost and underuses shared temporal regularities. A practical goal is thus to learn task-agnostic representations (e.g., periodicity, trend, and local fluctuations), for which frequency-domain or implicit spectral views are often natural. We introduce two representative frameworks: TimesNet (shared 2D backbone via implicit period reshaping) and TOTEM (tokenization + general-purpose backbone for reusable embeddings).

TimesNet [6] can serve as a general-purpose multi-task backbone by reusing its FFT-based period discovery and period-wise 2D reshaping (see Section 3.2). Different tasks are then handled by attaching lightweight, task-specific heads on top of the same backbone with minimal additional parameters. This view highlights Fourier-based period discovery as a task-agnostic period coordinate system that naturally supports a shared spectral backbone + multi-task heads paradigm.

TOTEM [10] discretizes time series into token sequences with a fixed vocabulary via a self-supervised VQ-VAE tokenizer, and then uses a general-purpose Transformer backbone plus lightweight heads to support multiple tasks (forecasting, imputation, anomaly detection, etc.). Concretely, the input is segmented along the time axis and vector quantization is used to learn a time-series vocabulary, mapping continuous signals to discrete tokens. Tokens with positional encodings are fed into a shared encoder to obtain embeddings that can be reused across tasks, and lightweight task heads are attached for adaptation. Although this tokenization $\rightarrow$ general backbone $\rightarrow$ task heads design does not explicitly introduce Fourier or wavelet modules, it can be viewed as implicitly absorbing periodic and local structures in token space and leveraging cross-domain self-supervised pretraining to obtain a unified representation with good generalization.

4. Experimental results and discussion

The benchmark datasets used in this paper are summarized below:(1) ETT: Transformer Temperature Series (ETTh/ETTm) with clear daily/weekly periodicity; widely used to evaluate

DOI: 10.54254/2755-2721/2026.TJ31299

long-range forecasting and period modeling [11]. (2) Electricity: hourly load of 321 clients; high-dimensional and often trend-dominated, where simple linear baselines can be strong [11,12]. (3) Exchange-Rate: daily exchange rates with strong non-stationarity and noise; a difficult benchmark for robustness [13]. (4) Traffic: hourly flow from 862 sensors; high-dimensional with local correlations and sparsity, often favoring lightweight models [13]. (5) Weather: 21 variables sampled every 10 minutes; relatively regular patterns, useful for assessing stability across horizons [11].

Based on Table 1, frequency-sparse and memory-based models (e.g., FEDformer/FiLM) are generally more stable at long horizons on strongly periodic datasets (ETT/Weather). On high-dimensional datasets (Electricity/Traffic), simple linear or lightweight MLP baselines can be competitive. TimesNet and Dwtformer provide a reasonable trade-off when both intra-period and inter-period variations matter. Overall, method suitability is largely driven by periodicity strength, dimensionality, and non-stationarity.(Results in Table 1 are taken from the original papers when available; missing entries are marked as "–/–". For TOTEM [10] we keep the reported long-term aggregate; for other methods we report the mean over horizons 96/192/336/720.)

The ratings in Table 2 combine empirical error trends across horizons in Table 1 and mechanism matching to data characteristics (periodicity, dimensionality, noise). They are intended as coarse deployment-oriented guidance rather than a definitive ranking.Grades: A=good, B=fairly good, C=average, D=poor, E=very poor.

Table 1. Multi-dataset performance comparison (MSE/MAE) of representative spectrum-related methods under different forecasting horizons

Method	Horizon	ETTh1		ETTh2		ETTm1		ETTm2		Electricity		Exchange		Traffic		Weather	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Autoformer [3]	96	–/–	–/–	–/–	–/–	–/–	–/–	0.255	0.339	0.201	0.317	0.197	0.323	0.613	0.388	0.206	0.336
	192	–/–	–/–	–/–	–/–	–/–	–/–	0.281	0.340	0.222	0.334	0.300	0.369	0.616	0.382	0.367	0.367
	336	–/–	–/–	–/–	–/–	–/–	–/–	0.339	0.423	0.251	0.368	0.509	0.524	0.622	0.377	0.419	0.395
	720	–/–	–/–	–/–	–/–	–/–	–/–	0.422	0.419	0.234	0.331	1.497	0.941	0.660	0.408	0.359	0.428
	Avg	–/–	–/–	–/–	–/–	–/–	–/–	0.324	0.380	0.227	0.338	0.626	0.539	0.628	0.389	0.338	0.382
FEDformer-f [4]	96	–/–	–/–	–/–	–/–	–/–	–/–	0.203	0.287	0.193	0.183	0.148	0.179	0.179	0.362	0.217	0.224
	192	–/–	–/–	–/–	–/–	–/–	–/–	0.262	0.313	0.202	0.195	0.271	0.250	0.604	0.373	0.272	0.256
	336	–/–	–/–	–/–	–/–	–/–	–/–	0.325	0.356	0.214	0.212	0.460	0.426	0.621	0.380	0.339	0.338
	720	–/–	–/–	–/–	–/–	–/–	–/–	0.421	0.435	0.236	0.231	1.195	0.891	0.626	0.392	0.403	0.424
	Avg	–/–	–/–	–/–	–/–	–/–	–/–	0.303	0.348	0.211	0.205	0.496	0.437	0.610	0.377	0.308	0.311
FiLM [5]	96	–/–	–/–	–/–	–/–	–/–	–/–	0.165	0.256	0.154	0.267	0.086	0.204	0.416	0.284	0.199	0.262
	192	–/–	–/–	–/–	–/–	–/–	–/–	0.222	0.296	0.164	0.258	0.188	0.292	0.408	0.298	0.228	0.288
	336	–/–	–/–	–/–	–/–	–/–	–/–	0.271	0.339	0.186	0.283	0.356	0.433	0.425	0.338	0.269	0.323
	720	–/–	–/–	–/–	–/–	–/–	–/–	0.377	0.385	0.238	0.322	0.377	0.469	0.520	0.285	0.317	0.361
	Avg	–/–	–/–	–/–	–/–	–/–	–/–	0.259	0.319	0.186	0.283	0.252	0.350	0.442	0.301	0.253	0.309

DOI: 10.54254/2755-2721/2026.TJ31299

Table 1. (continued)

	96	0.458	0.450	0.414	0.427	0.400	0.406	0.291	0.333	0.192	0.295	0.416	0.443	0.620	0.336	0.259	0.287
	192	0.502	0.485	0.456	0.472	0.445	0.452	0.328	0.371	0.221	0.328	0.468	0.492	0.675	0.372	0.301	0.325
TimesNet [6]	336	0.546	0.523	0.498	0.510	0.488	0.495	0.365	0.408	0.249	0.362	0.520	0.540	0.720	0.405	0.342	0.362
	720	0.598	0.567	0.542	0.553	0.532	0.538	0.365	0.451	0.278	0.401	0.572	0.590	0.765	0.438	0.385	0.401
	Avg	0.526	0.506	0.478	0.491	0.466	0.473	0.346	0.391	0.235	0.346	0.494	0.516	0.695	0.380	0.322	0.344
	96	–/–	–/–	–/–	–/–	0.420	0.440	0.432	0.389	0.190	0.306	0.558	0.346	0.154	0.284	–/–	–/–
	192	–/–	–/–	–/–	–/–	0.439	0.450	0.423	0.434	0.215	0.327	0.597	0.376	0.269	0.351	–/–	–/–
Dwtformer [7]	336	–/–	–/–	–/–	–/–	0.477	0.481	0.458	0.465	0.225	0.335	0.611	0.376	0.368	0.395	–/–	–/–
	720	–/–	–/–	–/–	–/–	0.450	0.496	0.470	0.484	0.241	0.350	0.633	0.390	0.417	0.428	–/–	–/–
	Avg	–/–	–/–	–/–	–/–	0.447	0.467	0.446	0.443	0.218	0.330	0.600	0.372	0.302	0.365	–/–	–/–
	96	0.378	0.416	0.370	0.419	0.352	0.366	0.179	0.213	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
	192	0.416	0.447	0.419	0.469	0.366	0.414	0.150	0.214	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
SWIFT-MLP [8]	336	0.447	0.477	0.469	0.476	0.414	0.440	0.179	0.268	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
	720	0.477	0.490	0.476	0.492	0.440	0.455	0.353	0.355	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
	Avg	0.430	0.458	0.434	0.464	0.393	0.419	0.215	0.263	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–
	96	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	0.143	0.237	0.082	0.207	0.040	0.280	0.178	0.236
	192	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	0.150	0.250	0.138	0.302	0.423	0.287	0.216	0.276
LTSF-Linear* [9]	336	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	0.169	0.268	0.331	0.415	0.436	0.307	0.271	0.301
	720	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	0.203	0.301	0.640	0.681	0.436	0.296	0.323	0.319
	Avg	–/–	–/–	–/–	–/–	–/–	–/–	–/–	–/–	0.166	0.264	0.298	0.401	0.334	0.293	0.247	0.283
TOTEM-Generalist [10]	Long-term	0.010	0.020	0.014	0.025	0.044	0.132	0.040	0.125	0.054	0.164	–/–	–/–	–/–	–/–	0.028	0.046

Table 2. Algorithm recommendations for typical application scenarios

Method	Electric load(hourly)	Highway traffic(multi-sensor)	Weather(10-min)	Exchange rate(highly non-stationary)	Industrialtemperature (ETT)
Autoformer [3]	C	C	C	D	B
FEDformer [4]	C	C	B	C	B
FiLM [5]	C	C	B	C	B
TimesNet [6]	C	C	C	C	C
Dwtformer [7]	C	C	D	D	D
SWIFT-MLP [8]	B	B	C	D	B
LTSF-Linear [9]	A	B	C	C	C
TOTEM-Generalist [10]	B	B	A	—	B

5. Conclusions and future directions

This paper reviews spectral-analysis-based deep forecasting methods and proposes a taxonomy centered on long-term dependency, multi-scale modeling, lightweight deployment, and multi-task generalization. We summarize core spectral principles and mechanisms, and provide benchmark-based comparisons and scenario-oriented recommendations for practical model selection.

Future directions include: (1) time-varying period detection and adaptive spectral partitioning; (2) cross-scale transfer and structured modeling of interactions across scales; (3) multivariate spectral denoising and interference localization in multi-sensor systems; (4) frequency-domain anomaly detection under extreme non-stationarity; and (5) causal reasoning to avoid spurious spectral

DOI: 10.54254/2755-2721/2026.TJ31299

correlations. Spectral methods are most effective when periodic or multi-scale structure is present; highly irregular signals may require time-domain or time–frequency modeling.

References

[1] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388-427.
[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
[3] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems, 34, 22419-22430.
[4] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022, June). Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International conference on machine learning (pp. 27268-27286). PMLR.
[5] Zhou, T., Ma, Z., Wen, Q., Sun, L., Yao, T., Yin, W., & Jin, R. (2022). Film: Frequency improved legendre memory model for long-term time series forecasting. Advances in neural information processing systems, 35, 12677-12690.
[6] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv: 2210.02186.
[7] Cao, Y., & Zhao, X. (2023, February). Dwtformer: Wavelet decomposition Transformer with 2D Variation for Long-Term Series Forecasting. In 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (Vol. 6, pp. 1548-1558). IEEE.
[8] Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023, June). Are transformers effective for time series forecasting?. In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 9, pp. 11121-11128).
[9] Xie, W., & Cao, F. (2025). SWIFT: Mapping Sub-series with Wavelet Decomposition Improves Time Series Forecasting. arXiv preprint arXiv: 2501.16178.
[10] Talukder, S., Yue, Y., & Gkioxari, G. (2024). Totem: Tokenized time series embeddings for general time series analysis. arXiv preprint arXiv: 2402.16412.
[11] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 12, pp. 11106-11115).
[12] Asuncion, A., & Newman, D. (2007, November). UCI machine learning repository.
[13] Lai, G., Chang, W. C., Yang, Y., & Liu, H. (2018, June). Modeling long-and short-term temporal patterns with deep neural networks. In The 41st international ACM SIGIR conference on research & development in information retrieval (pp. 95-104).