FRACTALLOSS: FRACTAL STRUCTURE OF THE OVERLOOKED NON-DOMINANT COMPONENTS IN TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Despite rapid progress in time series forecasting (TSF), model training still largely relies on mean squared error (MSE), which often produces over-smoothed predictions thereby misses fine-grained temporal dynamics. We investigate this limitation from a spectral perspective and show that MSE is energy-preferential: it emphasizes high-energy frequency components while overlooking low-energy ones. Consequently, dominant spectral peaks are well fitted, whereas numerous low-energy background components remain underfitted. More importantly, extensive empirical analysis reveals that these individually weak yet collectively strong background components are not random residuals, but instead exhibit a clear fractal structure characterized by strong linearity in the log-frequency log-energy spectrum. Motivated by these observations, we propose FractalLoss, a model-agnostic regularizer that encourages the recovery of fractal structure for more faithful forecasting. We further theoretically show that FractalLoss provides energy-independent guidance, helping mitigate underfitting across background components irrespective of their spectral energy. Experiments on real-world benchmarks demonstrate forecasting performance gains and substantially improved fractal structure preservation.

1 INTRODUCTION

Time series forecasting (TSF) plays a crucial role in numerous applications, including weather (Ravuri et al., 2021; Wu et al., 2023; Fu et al., 2026), traffic (Kong et al., 2024; Fang et al., 2025; Ma et al., 2026b), and finance (Huang et al., 2024; Xia et al., 2024), where accurate forecasting facilitates informed decision making (Wang et al., 2024a; Williams et al., 2025). In recent years, substantial progress has been made in designing forecasting backbones, with particular attention to Transformer-based and MLP-based architectures. The former leverage self-attention to capture temporal and variate dependencies (Zhou et al., 2021; Zhang & Yan, 2023; Nie et al., 2022; Liu et al., 2024b; Piao et al., 2024; Chen et al., 2025; Niu et al., 2026), whereas the latter rely on lightweight projections, often achieving competitive performance with high efficiency (Zeng et al., 2023; Das et al., 2023; Wang et al., 2024b; Ni et al., 2026). Another line of research explores time series modeling in the Fourier domain (Zhou et al., 2022b;a; Xu et al., 2024; Yi et al., 2023b; Cao et al., 2026). Despite these advances, TSF models are still commonly trained with mean squared error (MSE) Loss.

However, MSE-based training often produces over-smoothed predictions, as also observed in prior studies (Le Guen & Thome, 2019; 2022; Kudrat et al., 2025). Fig. 1(a) provides empirical evidence: the resulting forecast fails to capture many local fluctuations and sharp variations. Such loss of fine-grained dynamics can be problematic in real-world applications. For example, in weather forecasting, failing to anticipate abrupt temperature extremes may undermine downstream planning and risk management (Bi et al., 2023; Zhang et al., 2023). To uncover the underlying mechanism of this behavior, we analyze MSE from a frequency-domain perspective via Parseval's theorem (Parseval des Chenes, 1806; Oppenheim & Schafer, 2010) and show that *MSE is energy-preferential*: it is more sensitive to relative errors in high-energy dominant components, leaving low-energy ones weakly supervised. As illustrated in Fig. 1(b), only a few dominant components are well learned, while numerous non-dominant components remain severely underfitted under MSE loss. For clarity, we refer to dominant and non-dominant components as peaks and backgrounds, respectively, with formal definitions in Section 4.1. Fig. 1(c) further quantifies this imbalance: backgrounds are individually

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.

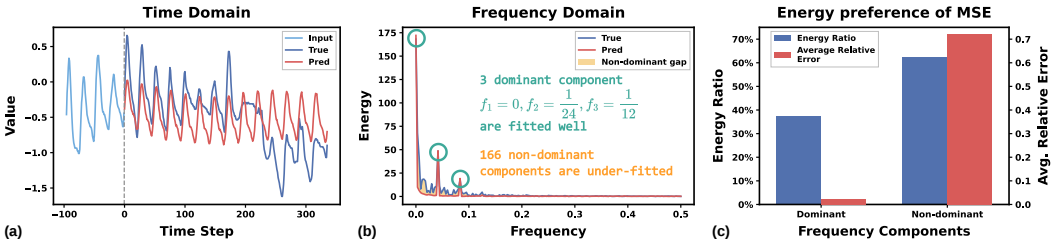

Figure 1: We train Fredformer (Piao et al., 2024) as the backbone with MSE loss on ETTh2, using a look-back window of 96 and a forecasting horizon of 336. (a) MSE loss produces noticeably over-smoothed predictions that fail to recover fine-grained temporal variations. (b) MSE loss fits 3 dominant high-energy components well, while a substantial fitting gap remains across 166 non-dominant components. (c) The non-dominant components are individually weak but collectively substantial, accounting for 62.5% of the total spectral energy. However, their average relative error reaches 0.722, far exceeding that of the dominant components at 0.022.

weak but collectively account for substantial energy, yet exhibit much larger relative errors than peaks. These backgrounds overlooked by MSE are related to the over-smoothing prediction.

We then extensively examine the backgrounds in Section 4 and uncover that they are not random noise but exhibit a clear *fractal structure, characterized by a strong linear relationship in log-frequency log-energy (log-log) spectrum.* We observed that backgrounds from real-world datasets can be well fit by a straight line in the log-log coordinates, which we refer to as the *Fractal Line*, the empirical support could be found in Fig. 2. This offers a new perspective on understanding time series: they contain not only a few peak components, but also numerous backgrounds that exhibit strong fractal structure. We further illustrate the generality of this structure in Fig. 3(a).

Motivated by this observation, we propose FractalLoss, a model-agnostic regularizer that complements MSE by recovering the fractal structure of backgrounds. We argue that naively compensating backgrounds through direct spectral matching is insufficient, because it still inherits the energy preference of MSE and thus mainly benefits relatively strong background components, as shown in Fig. 3(b-c). FractalLoss instead supervises predictions using the fractal line fitted from the ground truth. Our theoretical analysis shows that *FractalLoss provides energy-independent guidance*, distributing corrective pressure across backgrounds to mitigate underfitting without allowing high-energy components to dominate. We summarize our main contributions as follows:

- We analyze MSE from a spectral perspective and uncover that it inherently prioritizes a few high-energy components while underfitting numerous low-energy ones, thereby providing explanation for the over-smoothed predictions induced by MSE-based training.
- We reveal that the overlooked background components exhibit a clear fractal structure and extensively validate its generality across real-world time series. Building on this finding, we propose a model-agnostic FractalLoss and theoretically show that its energy-independent guidance mitigates backgrounds underfitting without favoring high-energy components.
- We evaluate FractalLoss on 7 real-world forecasting benchmarks with 5 representative backbones. Across 35 dataset-backbone scenarios, it achieves consistent gains in forecasting performance while markedly improving fractal structure fidelity.

2 Related Work

2.1 Backbones in Time Series Forecasting

Forecasting backbones have evolved to learn informative representations from historical sequences (Qiu et al., 2025; Wu et al., 2026; Zhao et al., 2026), including state-space models (Gu et al., 2021), CNNs (Luo & Wang, 2024), and graph neural networks (Yi et al., 2023a). Recent work has extensively explored Transformer and MLP architectures. Attention-based backbones, including Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), PatchTST (Nie et al., 2022),

iTransformer (Liu et al., 2024b), TQNet (Lin et al., 2025), TimeBridge (Liu et al., 2024a), and PhaseFormer (Niu et al., 2026), model temporal, channel, or cross-phase interactions. Meanwhile, MLP-based backbones such as DLinear (Zeng et al., 2023), TiDE (Das et al., 2023), TimeMixer (Wang et al., 2024b), CycleNet (Lin et al., 2024), OLinear (Yue et al., 2026), and MixLinear (Ma et al., 2026a) demonstrate that lightweight architectures can also achieve strong forecasting performance. Beyond these architectures, frequency-domain methods exploit spectral representations for sequence modeling (Yi et al., 2025), including FEDformer (Zhou et al., 2022b), FITS (Xu et al., 2024), FreTS (Yi et al., 2023b), Fredformer (Piao et al., 2024), CFPT (Kou et al., 2025), FreqMoE (Liu, 2025), Numerion (Cao et al., 2026), and FreIE (Sun et al., 2025a). Despite these advances, backbone training still largely relies on MSE, which often produces over-smoothed predictions. We theoretically analyze this behavior to reveal MSE's energy preference and are the first to identify fractal structure underlying the low-energy background components it overlooks.

2.2 Loss Functions in Time Series Forecasting

Compared with the extensive exploration of forecasting backbones, learning objectives for time series forecasting have received relatively limited attention Recent studies have introduced alternative loss functions for time series forecasting, broadly organized into types: shape-based and distribution-based. Shape-based objectives preserve temporal patterns beyond point-wise errors. Soft-DTW (Cuturi & Blondel, 2017), DILATE (Le Guen & Thome, 2019), and TILDE-Q (Lee et al., 2022) encourage temporal alignment and shape similarity, while PS Loss (Kudrat et al., 2025) captures patch-wise shape discrepancies. DBLoss (Qiu et al., 2026) and PMLF (Chen et al., 2026) further supervise seasonal and trend components. For the distribution-based group, FreDF (Wang et al., 2025) and Time-o1 (Wang et al., 2026a) account for label dependencies in transformed domains. DistDF (Wang et al., 2026b) and InfoGeo Loss (Fang et al., 2026) align distributions through Wasserstein distance and information geometry, respectively. However, these objectives still primarily minimize the discrepancy between predictions and labels from different perspectives, without accounting for the low-energy background components overlooked by MSE.

3 Analysis

In this section, we analyze MSE loss from the frequency-domain perspective and show that it implicitly favors dominant frequency components, leaving numerous non-dominant components under-fitted. This imbalance drives the model toward over-smoothed forecasts.

3.1 Preliminary

Time Series Forecasting. We use $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ to denote a univariate[1] time series and x_i to denote its ith time step. Given an observed historical window $\mathbf{x} \in \mathbb{R}^T$, time series forecasting (TSF) aims to predict the future values $\mathbf{y} \in \mathbb{R}^\tau$ over a target horizon τ. Let $f_\theta : \mathbb{R}^T \to \mathbb{R}^\tau$ be a forecasting model parameterized by θ, and its prediction is denoted $\hat{\mathbf{y}} = f_\theta(\mathbf{x})$.

Discrete Fourier Transform and Spectral Energy. We use $\mathcal{F}(\cdot)$ to denote the Discrete Fourier Transform (Oppenheim & Schafer, 2010), which transforms a real-value signal $\mathbf{y} = [y_0, y_1, \ldots, y_{\tau-1}]^\top \in \mathbb{R}^\tau$ to its complex-value representation in the frequency domain. The kth frequency component is:

$$Y_k = \mathcal{F}(\mathbf{y})_k = \sum_{n=0}^{\tau-1} y_n \, e^{-j2\pi kn/\tau} \in \mathbb{C}, \qquad k \in \{0, 1, \ldots, \tau-1\} \tag{1}$$

where $j = \sqrt{-1}$ is the imaginary unit and $f_k = k/\tau$ is the associated physical frequency with Y_k. We denote the frequency component of the ground truth and the prediction as $Y_k = \mathcal{F}(\mathbf{y})_k$ and $\hat{Y}_k = \mathcal{F}(\hat{\mathbf{y}})_k$, respectively. The spectral energy of the k-th component is defined as:

$$E_k = |Y_k|^2 \in \mathbb{R} \tag{2}$$

[1]We focus on the univariate case for simplicity. For multivariate data, the analysis is applied per channel.

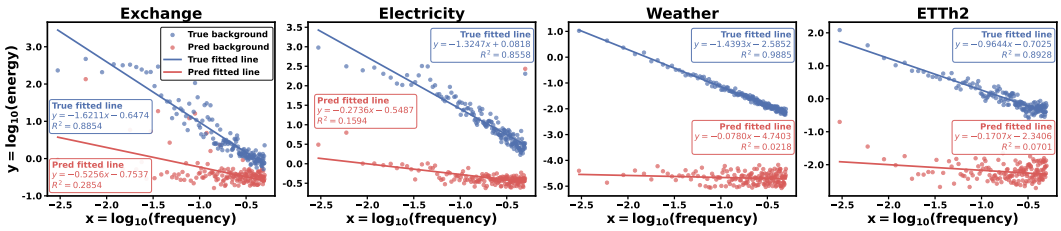

Figure 2: Energy spectra of background components from four real-world datasets with a window size of 336, shown in the log-frequency log-energy scale. Blue and red points denote the ground-truth and MSE-predicted background components, respectively, while the corresponding lines are their linear fits. Two observations stand out: (1) ground-truth spectra exhibit strong linearity with an average R^2 of 0.906, whereas prediction spectra are poorly fitted with an average R^2 of 0.134; (2) the fitted lines differ markedly in slope and intercept. These results illustrate that the fractal structure present in the ground-truth backgrounds is poorly preserved by MSE-trained predictions. More visualizations across different forecasting horizons are provided in Appendix C.3.

Theorem 1. Parseval's Theorem. *Under the standard DFT convention, the energy of a signal in the time domain is equal to its energy in the frequency domain up to a normalization factor:*

$$\|\mathbf{y}\|_2^2 = \frac{1}{\tau}\sum_{k=0}^{\tau-1} E_k \tag{3}$$

Fractal Structure in Time Series. Fractals are ubiquitous in nature: a coastline appears similarly intricate whether viewed from a satellite or a few meters away (Mandelbrot, 1967), while tree branches (Eloy et al., 2017) and snowflakes (Demange et al., 2017) exhibit similar patterns across scales. This property, known as self-similarity, is the signature characteristic of fractals. In time series, this self-similarity extends from spatial to temporal scale. Concretely, given a time series $\mathbf{y} \in \mathbb{R}^\tau$ with energy spectrum $E_k = |\mathcal{F}(\mathbf{y})_k|^2$ at frequencies $f_k = k/\tau$, self-similarity across temporal scales is reflected in a power-law spectrum, with the derivation provided in Appendix A.4:

$$E_k \propto f_k^{-\alpha} \quad \Longrightarrow \quad \log E_k = -\alpha \log f_k + \beta \tag{4}$$

which defines the linear *fractal line* in the log-frequency log-energy (abbreviated as log-log below) coordinate. Here, $-\alpha$ governs frequency-wise energy decay and β sets the energy level.

3.2 MSE Implicitly Favors Dominant Frequency Components

Define the prediction error as $\mathbf{e} = \hat{\mathbf{y}} - \mathbf{y}$. By the linearity of the DFT, its k-th frequency component:

$$\mathcal{F}(\mathbf{e})_k = \mathcal{F}(\hat{\mathbf{y}} - \mathbf{y})_k = \mathcal{F}(\hat{\mathbf{y}})_k - \mathcal{F}(\mathbf{y})_k = \hat{Y}_k - Y_k \tag{5}$$

Applying Parseval's Theorem to the error signal $\mathbf{e}$, the MSE loss can be rewritten:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{\tau}\|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 = \frac{1}{\tau}\|\mathbf{e}\|_2^2 \overset{\text{Parseval's Theorem}}{=\!=\!=} \frac{1}{\tau^2}\sum_{k=0}^{\tau-1}|\mathcal{F}(\mathbf{e})_k|^2 = \frac{1}{\tau^2}\sum_{k=0}^{\tau-1}|\hat{Y}_k - Y_k|^2 \tag{6}$$

Proposition 1 (MSE provides energy-preferential supervision). *For each frequency component Y_k with $|Y_k| > 0$, let $\rho_k = |\hat{Y}_k - Y_k|/|Y_k|$ denote its relative error in the frequency domain. Then the sensitivity of $\mathcal{L}_{\text{MSE}}$ to each frequency component satisfies*

$$\frac{\partial \mathcal{L}_{\text{MSE}}}{\partial \rho_k} = \frac{2}{\tau^2} \cdot |Y_k|^2 \cdot \rho_k \tag{7}$$

where τ represents the target forecasting horizon.

Remark. The factor $2/\tau^2$ is constant across frequency components, while the energy weighting term $|Y_k|^2$ indicates that, at the same relative error, MSE is more sensitive to higher-energy components. This helps explain why MSE-based training preferentially fits dominant components while neglecting low-energy non-dominant ones. Fig. 1(b–c) further provides empirical support.

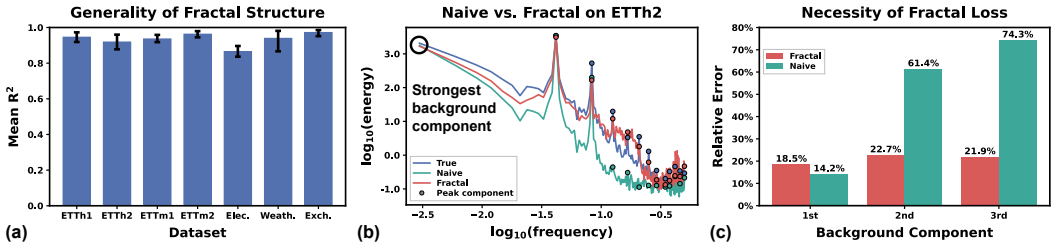

Figure 3: (a) Fractal Line fits across seven datasets with the window size of 336: all mean R^2 exceed 0.8with an average 0.92, error bars denote 95% bootstrap confidence intervals. (b) For the underfitted backgrounds set $\mathcal{B}$, Naive Loss mainly compensate the strongest ones in $\mathcal{B}$, whereas the energy-independent FractalLoss yields a more balanced recovery of the entire background spectrum. (c) Relative errors of the top-3 highest-energy backgrounds: under Naive Loss, errors increase sharply as component energy decreases, whereas those under FractalLoss remain comparatively stable.

4 Methodology

4.1 Fractal Structure in Time Series

The above analysis reveals that MSE neglects low-energy components. To further investigate these parts, we first define peak and background components. More importantly, We further show that the background components are not random residuals, but exhibit strong fractal structure, motivating targeted supervision. Fig. 1(b) provides an intuitive example of peak components: the salient f_1, f_2, f_3 in the prediction all appear as local maxima in the spectrum. Moreover, local energy maximality provides a more stable criterion across different datasets and prediction horizons than simple top-k energy selection. Therefore, we identify peak components as spectral frequencies with locally maximal energy and treat the rest as background ones.

Definition 1 (Peak Components). Given the frequency spectrum $\{Y_k\}_{k\in\Omega} = \mathcal{F}(\mathbf{y})$ of a time series signal $\mathbf{y}$ with length τ, here we use $\Omega = \{0, 1, \ldots, \tau - 1\}$ denote the frequency index set. A frequency component Y_k is called peak component if its energy is a local maximum in the spectrum. Formally, we define the set of peak components as follows:

$$\mathcal{P} = \{Y_k \mid E_k > E_j,\ k \in \Omega,\ \forall j \in \mathcal{N}_r(k)\} \tag{8}$$

where $\mathcal{N}_r(k) = \{j \in \Omega \mid 0 < |j - k| \leq r\}$ denotes the neighboring frequency indices of k within a radius r, sensitivity analysis of r can be found in Fig. 4(c-d).

Definition 2 (Background Components). The remaining are defined as the background components:

$$\mathcal{B} = \{Y_k \mid k \in \Omega,\ Y_k \notin \mathcal{P}\} \tag{9}$$

For notational convenience, we denote its index set as $\mathcal{I}_\mathcal{B} = \{k \in \Omega \mid Y_k \in \mathcal{B}\}$.

Fractal Structure of Background Components is Discarded by MSE Loss. We analyze the under-fitted background components $\mathcal{B}$ in log-log coordinates. Remarkably, Fig. 2 reveals strong linearity in the ground-truth spectra across 4 datasets from different applications, consistent with the fractal characteristics described in Preliminary 3.1. Here, we use the coefficient of determination R^2 (Draper & Smith, 1998) to quantify how well the background spectrum fits a fractal line, with larger values indicating a better fit. However, predictions trained with the energy-preferential MSE loss fail to preserve this fractal structure. Beyond the visualization, Fig. 3(a) reports mean R^2 values above 0.8 across all 7 displayed datasets, with an average of 0.92. The 95% bootstrap confidence intervals (Efron, 1992) are narrow, with widths ranging from 0.034 to 0.115, further supporting the generality of the fractal structure across diverse time series.

4.2 Fractal Loss

To recover the discarded fractal structure of background components $\mathcal{B}$, a straightforward approach is to align prediction with groundtruth on $\mathcal{B}$. Given the future label $\mathbf{y}$ and prediction $\hat{\mathbf{y}}$, let Y_k and $\hat{Y}_k$

denote their Fourier coefficients at frequency f_k, respectively, we formulate this Naive Loss as:

$$\mathcal{L}_{\text{naive}} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|} \sum_{k \in \mathcal{I}_{\mathcal{B}}} \left| \hat{Y}_k - Y_k \right|^2 \tag{10}$$

However, restricting supervision to $\mathcal{I}_{\mathcal{B}}$ does not eliminate energy preference. Components in $\mathcal{B}$ still differ in energy and $\mathcal{L}_{\text{naive}}$ primarily compensates for the strongest ones rather than recovering the overall spectral structure, as shown in Fig. 3(b-c). We therefore propose **FractalLoss**, which leverages the *fractal line* fitted to the ground-truth background spectrum to provide energy-independent supervision, enabling all background components to receive guidance irrespective of their spectral energy. We first map the background spectrum into log-log coordinates:

$$u_k = \log f_k, \qquad v_k = \log(|Y_k|^2 + \epsilon), \qquad \hat{v}_k = \log(|\hat{Y}_k|^2 + \epsilon), \tag{11}$$

where ϵ is a small constant for numerical stability. Next, we fit the ground-truth fractal line $L(u) = -\alpha u + \beta$ of Eq. 4 using ordinary least squares on $\{(u_k, v_k)\}_{k \in \mathcal{I}_{\mathcal{B}}}$:

$$\alpha = -\frac{\sum_{k \in \mathcal{I}_{\mathcal{B}}} (u_k - \bar{u})(v_k - \bar{v})}{\sum_{k \in \mathcal{I}_{\mathcal{B}}} (u_k - \bar{u})^2}, \qquad \beta = \bar{v} + \alpha\, \bar{u}, \tag{12}$$

where $\bar{u}$ and $\bar{v}$ are the means of $\{u_k\}$ and $\{v_k\}$ over $\mathcal{I}_{\mathcal{B}}$. Finally, FractalLoss is defined as the mean squared deviation between the predicted background log-spectrum and the fitted fractal line:

$$\mathcal{L}_{\text{fractal}} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|} \sum_{k \in \mathcal{I}_{\mathcal{B}}} \left[\hat{v}_k - L(u_k) \right]^2 = \frac{1}{|\mathcal{I}_{\mathcal{B}}|} \sum_{k \in \mathcal{I}_{\mathcal{B}}} \left[\hat{v}_k + \alpha u_k - \beta \right]^2. \tag{13}$$

FractalLoss improves recovery of the overall spectral structure of $\mathcal{B}$ by overcoming the energy preference, as shown in Fig. 3(b–c). The following theoretical analysis further explains its effectiveness.

Proposition 2 (Fractal loss provides energy-independent supervision). *For each background frequency f_k, let $\tilde{E}_k = \exp(-\alpha u_k + \beta)$ denote the energy prescribed by the fractal line at f_k, let $\hat{E}_k = |\hat{Y}_k|^2$ denote the predicted energy, and define the relative error at f_k as $\xi_k = (\hat{E}_k - \tilde{E}_k)/\tilde{E}_k$. Then the sensitivity of $\mathcal{L}_{\text{frac}}$ to each background frequency satisfies:*

$$\frac{\partial \mathcal{L}_{\text{fractal}}}{\partial \xi_k} = \frac{2}{|\mathcal{I}_{\mathcal{B}}|} \cdot \frac{\log(1 + \xi_k)}{1 + \xi_k} \tag{14}$$

where $|\mathcal{I}_{\mathcal{B}}|$ denotes the number of background components.

Remark. The constant term $2/|\mathcal{I}_{\mathcal{B}}|$ is shared by all components, and the remaining term depends only on the relative error ξ_k. Therefore, fractal loss assigns energy-independent supervision to every background components for preserving the fractal structure.

Total Training Objective. We complement MSE loss with the proposed FractalLoss:

$$\mathcal{L}_{\text{total}} = \lambda\, \mathcal{L}_{\text{fractal}} + (1 - \lambda) \mathcal{L}_{\text{MSE}}, \tag{15}$$

where λ controls the trade-off. $\mathcal{L}_{\text{MSE}}$ provides temporal supervision capturing both energy and phase information, while $\mathcal{L}_{\text{fractal}}$ leverages the intrinsic fractal structure of the time series to provide dedicated guidance for background components.

5 Experiments

To comprehensively evaluate FractalLoss, we conduct experiments from the following perspectives:

1. **Overall Performance:** In Section 5.2, we first evaluate the performance of the same backbone models trained with MSE alone and with additional FractalLoss on 7 datasets spanning diverse domains. We then compare FractalLoss with other learning objectives. All experiments report Mean Squared Error(MSE), Mean Absolute Error(MAE), and fractal structure related metrics.
2. **Full Ablation Study:** In Section 5.3, we compare FractalLoss with two variants: (1) directly applying a frequency-domain loss to the background components, and (2) fitting the fractal line without removing the peak component, to examine the individual design choices of FractalLoss.
3. **Sensitivity Analysis and Case Study:** In Section 5.4, we first assess the sensitivity of FractalLoss to the loss weight λ and the peak detection radius r. We then conduct a case study on the highly non-stationary Exchange dataset to investigate why FractalLoss is effective in this setting. We further provide visual examples to illustrate its effectiveness.

Table 1: Long-term multivariate forecasting comparison across 7 datasets and 5 representative backbones. We evaluate forecasting performance using MSE and MAE, and fractal-structure fidelity using FSE and FFE. All experiments use a fixed look-back window 96, and results are averaged over forecasting horizons $T \in \{96, 192, 336, 720\}$, the full results is provided in Appendix C.2.

Dataset	Loss	iTransformer				PatchTST				DLinear				Fredformer				PhaseFormer			
		MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE
ETTh1	MSE	0.459	0.450	0.694	0.305	**0.440**	0.444	0.367	0.176	0.452	0.448	0.178	0.101	0.463	0.445	0.303	0.237	0.434	0.428	0.257	0.495
	+Fractal	**0.444**	**0.439**	**0.038**	**0.077**	0.441	**0.438**	**0.061**	**0.087**	**0.442**	**0.434**	**0.047**	**0.022**	**0.439**	**0.430**	**0.016**	**0.059**	**0.427**	**0.425**	**0.044**	**0.079**
ETTh2	MSE	0.383	0.408	0.657	0.472	0.378	0.404	0.313	0.329	0.438	0.460	0.650	0.771	0.378	0.402	0.203	0.300	0.388	0.408	0.253	0.518
	+Fractal	**0.372**	**0.399**	**0.099**	**0.116**	**0.368**	**0.392**	**0.128**	**0.152**	**0.424**	**0.428**	**0.059**	**0.060**	**0.362**	**0.390**	**0.060**	**0.049**	**0.378**	**0.401**	**0.086**	**0.030**
ETTm1	MSE	0.419	0.419	0.274	0.178	0.396	0.406	0.113	0.142	0.403	0.406	0.152	0.237	0.385	0.398	0.362	0.191	0.407	0.405	0.592	0.290
	+Fractal	**0.403**	**0.406**	**0.124**	**0.060**	**0.383**	**0.390**	**0.095**	**0.069**	**0.395**	**0.389**	**0.070**	**0.041**	**0.375**	**0.390**	**0.051**	**0.030**	**0.402**	**0.401**	**0.019**	**0.070**
ETTm2	MSE	0.293	0.336	0.647	0.261	0.282	0.329	0.612	0.150	0.343	0.394	0.580	0.356	0.287	0.328	0.495	0.273	0.292	0.338	0.256	0.176
	+Fractal	**0.285**	**0.326**	**0.039**	**0.037**	**0.278**	**0.318**	**0.046**	**0.018**	**0.335**	**0.371**	**0.175**	**0.020**	**0.276**	**0.319**	**0.160**	**0.127**	**0.287**	**0.327**	**0.085**	**0.086**
Weather	MSE	0.262	0.283	0.288	0.186	**0.254**	0.277	0.212	0.259	0.265	0.294	0.575	0.395	0.249	0.275	0.242	0.279	0.272	0.291	0.198	0.325
	+Fractal	**0.260**	**0.279**	**0.079**	**0.065**	0.255	**0.270**	**0.085**	**0.094**	**0.263**	**0.266**	**0.088**	**0.069**	**0.245**	**0.273**	**0.051**	**0.096**	**0.268**	**0.287**	**0.051**	**0.109**
Electricity	MSE	0.178	0.269	0.398	0.185	0.213	0.315	0.174	0.273	0.209	0.296	0.536	0.211	0.178	**0.274**	0.280	0.210	0.204	0.281	0.146	0.451
	+Fractal	**0.172**	**0.261**	**0.085**	**0.091**	**0.202**	**0.304**	**0.092**	**0.077**	**0.191**	**0.266**	**0.048**	**0.056**	**0.176**	**0.274**	**0.100**	**0.079**	**0.195**	**0.267**	**0.020**	**0.061**
Exchange	MSE	0.363	0.408	0.605	0.536	0.387	0.416	0.578	0.376	0.347	0.412	0.586	0.711	0.358	0.404	0.173	0.181	0.377	0.417	0.450	0.329
	+Fractal	**0.356**	**0.399**	**0.079**	**0.045**	**0.377**	**0.408**	**0.137**	**0.009**	**0.335**	**0.402**	**0.115**	**0.316**	**0.343**	**0.395**	**0.087**	**0.028**	**0.360**	**0.409**	**0.083**	**0.035**

Note: For each dataset-backbone scenario, the better result is **boldfaced** with a blue background for MSE/MAE and with a red background for FSE/FFE. Lower is better for all metrics.

5.1 Experimental Setup

Datasets. We evaluate FractalLoss on 7 standard public long-term forecasting benchmarks following (Kudrat et al., 2025), including ETT series (ETTh1, ETTh2, ETTm1, ETTm2), Weather, ECL, and Exchange. For data preprocessing, we adhere to the standard procedure adopted in prior studies. A more detailed description of the datasets can be found in the Appendix B.1.

Backbones and Other Learning Objectives. We first evaluate FractalLoss on 5 representative forecasting backbones: iTransformer (Liu et al., 2024b), Fredformer (Piao et al., 2024), PhaseFormer (Niu et al., 2026), DLinear (Zeng et al., 2023), and PatchTST (Nie et al., 2022). Using the same backbone Fredformer, we then compare FractalLoss against alternative learning objectives: FreDF (Wang et al., 2025), PS Loss (Kudrat et al., 2025), DBLoss (Qiu et al., 2026), and DistDF (Wang et al., 2026b).

Metrics. We evaluate forecasting accuracy using MSE and MAE, and fractal-structure recovery using Fractal Slope Error (FSE) and Fractal Fit Error (FFE), detailed calculation is provided in Appendix B.4:

$$\mathrm{FSE} = |\hat{\alpha} - \alpha|, \qquad \mathrm{FFE} = |\hat{R}^2 - R^2|, \tag{16}$$

where (α, R^2) and $(\hat{\alpha}, \hat{R}^2)$ denote the slope and coefficient of determination of the fractal lines fitted to the ground-truth $\mathbf{y}$ and the prediction $\hat{\mathbf{y}}$, respectively.

Implementation Details. We use the official implementations of all forecasting backbones and existing learning objectives. For fair comparison, we follow their original experimental settings when integrating FractalLoss, while only tuning the loss weight λ and peak-detection radius r. All models are trained with the Adam optimizer (Kingma & Ba, 2014). Experiments are conducted on an Intel Xeon Silver 4210R CPU and an NVIDIA RTX A5000 GPU with 24 GB of memory. Further implementation details are provided in the Appendix B.3.

5.2 Main Results

Table 1 reports forecasting performance and fractal-structure fidelity for 5 backbones across 7 datasets. (1) FractalLoss achieves substantial improvements on DLinear for Electricity, with reductions of approximately 8.61% in MSE and 10.14% in MAE. Similar gains are observed across other dataset–backbone scenarios, with FractalLoss consistently enhancing forecasting performance and outperforming MSE-only training in 33/35 cases for MSE and 34/35 cases for MAE. Such consistent gains highlight the generality of FractalLoss across diverse backbones and datasets. (2) Regarding structural fidelity, FractalLoss consistently outperforms MSE-only training in both FSE and FFE across all 35 dataset–backbone combinations, reducing the average FSE and FFE by approximately

Table 2: Comparison with other learning objectives using Fredformer as the backbone. Values are averaged over forecasting horizons $T \in \{96, 192, 336, 720\}$ with a fixed look-back window of 96.

Loss	ETTh2				ETTm2				Weather				Electricity				Exchange			
	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE
MSE	0.378	0.402	0.203	0.300	0.287	0.328	0.495	0.273	0.249	0.275	0.242	0.279	0.178	**0.274**	0.280	0.210	0.358	0.404	0.173	0.181
FreDF	0.369	0.401	0.291	0.449	0.282	0.324	0.306	0.199	0.255	0.278	0.321	0.245	0.178	0.275	0.277	0.241	0.374	0.407	0.167	0.154
PSLoss	0.371	0.394	0.252	0.465	0.279	0.320	0.422	0.224	0.256	0.276	0.271	0.159	0.177	0.280	0.357	0.206	0.352	0.406	0.164	0.173
DBLoss	0.371	0.392	0.375	0.479	0.277	0.320	0.417	0.218	0.254	**0.273**	0.307	0.336	0.178	0.277	0.395	0.201	0.369	0.409	0.299	0.147
DistDF	0.372	0.397	0.115	0.294	0.279	0.323	0.503	0.212	0.260	0.282	0.225	0.378	0.181	0.276	0.323	0.207	0.375	0.410	0.229	0.185
FractalLoss	**0.362**	**0.390**	**0.060**	**0.049**	**0.276**	**0.319**	**0.160**	**0.127**	**0.245**	**0.273**	**0.051**	**0.096**	**0.176**	**0.274**	**0.100**	**0.079**	**0.343**	**0.395**	**0.087**	**0.028**

Note: Best/second-best are **bold**/underlined, with blue (MSE/MAE) or red (FSE/FFE) backgrounds, lighter shades mark second-best values.

Table 3: Comparison of two variants using Fredformer and DLinear as backbones. Values are averaged over forecasting horizons $T \in \{96, 192, 336, 720\}$ with a fixed look-back window of 96.

Backbone	Variant	ETTh2				ETTm2				Weather				Electricity				Exchange			
		MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE
Fredformer	MSE	0.378	0.402	0.203	0.300	0.287	0.328	0.495	0.273	0.249	0.275	0.242	0.279	0.178	**0.274**	0.280	0.210	0.358	0.404	0.173	0.181
	FractalLoss†	0.368	0.398	0.160	0.291	0.280	0.323	0.395	0.265	0.247	0.274	0.185	0.271	0.177	0.276	0.146	0.114	0.349	0.399	0.147	0.176
	FractalLoss‡	0.374	0.399	0.171	0.122	0.281	0.326	0.311	0.168	0.249	0.276	0.188	0.149	0.179	0.277	0.154	0.117	0.350	0.402	0.153	0.082
	FractalLoss	**0.362**	**0.390**	**0.060**	**0.049**	**0.276**	**0.319**	**0.160**	**0.127**	**0.245**	**0.273**	**0.051**	**0.096**	**0.176**	**0.274**	**0.100**	**0.079**	**0.343**	**0.395**	**0.087**	**0.028**
DLinear	MSE	0.438	0.460	0.650	0.771	0.343	0.394	0.580	0.356	0.265	0.294	0.575	0.395	0.209	0.296	0.536	0.211	0.347	0.412	0.586	0.711
	FractalLoss†	0.430	0.441	0.403	0.748	0.338	0.379	0.459	0.345	0.264	0.288	0.329	0.383	0.205	0.289	0.390	0.205	0.337	0.405	0.278	0.390
	FractalLoss‡	0.435	0.453	0.336	0.266	0.339	0.380	0.297	0.118	**0.263**	0.277	0.234	0.163	0.198	0.278	0.194	0.101	0.344	0.410	0.356	0.428
	FractalLoss	**0.424**	**0.428**	**0.059**	**0.060**	**0.335**	**0.371**	**0.175**	**0.020**	**0.263**	**0.266**	**0.088**	**0.069**	**0.191**	**0.266**	**0.048**	**0.056**	**0.335**	**0.402**	**0.115**	**0.316**

Note: Best/second-best are **bold**/underlined, with blue (MSE/MAE) or red (FSE/FFE) backgrounds; lighter shades mark second-best values.

79.83% and 75.99%, respectively. These results demonstrate that FractalLoss effectively preserves the fractal structure of background components while improving forecasting performance.

Table 2 further evaluates FractalLoss with advanced loss functions under Fredformer backbone. FreDF directly supervises in the frequency domain, PSLoss and DBLoss incorporate structural or seasonality-trend regularization, and DistDF minimizes the Wasserstein discrepancy between the joint distributions of forecast and target. FractalLoss achieves the best MSE and MAE across all 5 datasets, including ties, while consistently attaining the lowest FSE and FFE. Compared with conventional frequency-domain objectives like FreDF, FractalLoss provides dedicated structured guidance for the overlooked background components, which follows the characteristic of real-world time series.

5.3 Ablation Study

Table 3 comprare FractalLoss with its 2 variants to examine the contributions of background fractal-line fitting and peak-background separation under same backbones. The main findings are as follows:

- FractalLoss† augments MSE loss by directly aligning predicted background components with their ground-truth counterparts in the frequency domain rather than with the fitted fractal line. It generally reduces MSE/MAE and FSE/FFE over MSE-only training but underperforms FractalLoss on all metrics. Consistent with Fig. 3(b-c), this variant retains the energy preference, suggesting that point-wise spectral supervision is insufficient to preserve the fractal structure.
- FractalLoss‡ fits the fractal line without removing peaks. It reduces MSE/MAE in 8/10 dataset–backbone scenarios and FSE/FFE in all 10 relative to MSE-only training. However, retaining peaks can distort the fitted line: its average FSE is 153.87% higher than that of FractalLoss, highlighting fractal structure is exhibited by background components rather than peaks.
- FractalLoss combines peak removal with background fractal-line fitting. It achieves the best MSE/MAE across all dataset–backbone scenarios, including ties, and the lowest FSE/FFE throughout, supporting the effectiveness of integrating both design choices.

5.4 Sensitivity and Case Study

Hyperparameter Sensitivity. We analyze the sensitivity of FractalLoss to the loss weight λ in Eq. 15 and peak-detection radius r in Eq. 8. As shown in Fig. 4(a-b), increasing λ from 0 generally

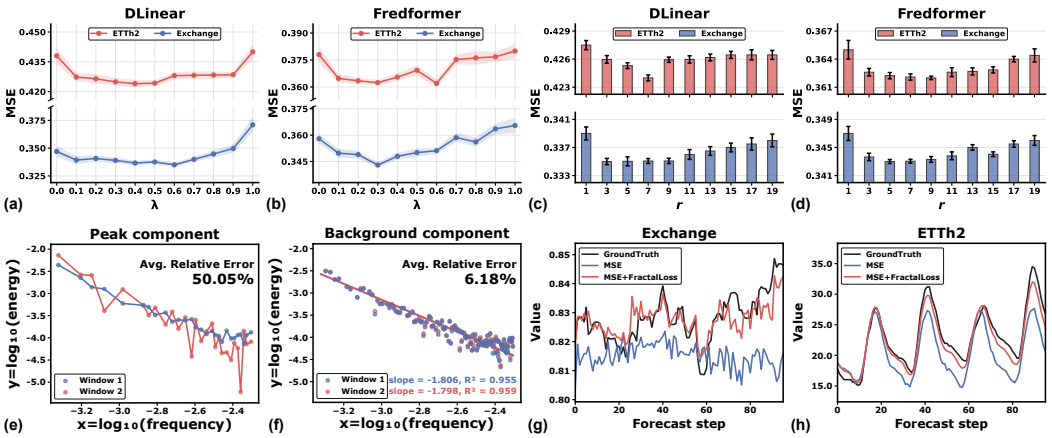

Figure 4: Hyperparameter sensitivity and case study. (a,b) Sensitivity to the loss weight λ under DLinear and Fredformer backbones, with the best performance falling within a moderate range $\lambda \in [0.3, 0.6]$. (c,d) Sensitivity to the peak-detection radius r for the same backbones, with $r \in [3, 9]$ yielding the lowest MSE, while $r = 1$ produces comparatively high MSE. (e,f) Log-log energy spectra of two adjacent 336-step windows from Exchange: peak components differ substantially (the average relative error reaches 50.05%), whereas background components remain relatively stable with only 6.18% error. (g,h) Visualizations of forecasting examples on Exchange and ETTh2, respectively, using window size of 96 to highlight local details.

reduces MSE, with the best results obtained within $\lambda \in [0.3, 0.6]$. However, excessively large weights degrade performance, particularly at $\lambda = 1$, where the MSE term is removed. This highlights the complementary roles of point-wise MSE supervision and fractal-structure guidance. Fig. 4(c-d) shows that peak detection is overly permissive at $r = 1$ and misclassifies many overlooked background components as peaks. This weakens fractal-structure guidance and results in relatively high MSE. For $r > 1$, performance becomes relatively stable, with the best results achieved within $r \in [3, 9]$.

Case Study. Table 1 shows that FractalLoss remains effective on the highly non-stationary Exchange dataset (Ye et al., 2024; Liu et al., 2025; Sun et al., 2025b). To investigate the underlying reason of the success, we compare two adjacent windows with window size of 336 and examine their shared peak and background components with peak-detection radius $r = 3$. We quantify differences using the per-frequency average relative error of the second window with respect to the first. Fig. 4(e-f) shows that peak components vary substantially while background components remain stable with nearly overlapping fractal lines (FSE=0.008, FFE=0.004), suggesting the fractal structure of backgrounds keeps stable even on non-stationarity data. Finally, Fig. 4(g–h) shows that MSE-only training misses the abrupt peak on ETTh2 and even predicts an opposite trend after time step 62 on Exchange, whereas FractalLoss better follows the ground truth in both examples.

6 Conclusion

In this paper, we analyze MSE from a spectral perspective to explain the mechanism underlying over-smoothed forecasts. Our analysis reveals that MSE favors dominant high-energy components while leaving low-energy background components underfitted. We further identify a clear fractal structure underlying these overlooked components in real-world time series. Building on this finding, we propose FractalLoss, a model-agnostic regularizer that guides background spectral recovery toward the ground-truth fractal line, mitigating underfitting irrespective of spectral energy and compensating for MSE's energy preference.

Limitations. While this work demonstrates the effectiveness of FractalLoss in time series forecasting, its applicability to other tasks remains unexplored. The background fractal structure we identify is a property of time series data rather than a task-specific characteristic, suggesting its potential relevance to broader applications, such as imputation and reasoning. Extending this structural prior to these tasks is a promising direction for future work.

AI USE STATEMENT

We declare that LLMs were used solely for language polishing, including grammar correction, readability improvements, and terminology standardization. All ideas, theoretical analyses, the design of FractalLoss, and empirical demonstrations were conceived and conducted by the authors.

ETHICS STATEMENT

This work is conducted in accordance with the ICLR Code of Ethics. Our study aims to improve time series forecasting reliability by examining the limitations of MSE training and introducing FractalLoss to recover the fractal structure underlying data. This research supports more faithful predictions in real-world applications. Our experiments use publicly available benchmark datasets and do not involve human-subject research or the collection of personal data. We transparently report our methodology, experimental settings, and findings to support reproducibility and responsible use.

REPRODUCIBILITY STATEMENT

We take several measures to ensure the reproducibility of our work. All datasets used in our experiments are publicly available and described in the main text and Appendix B.1. Implementation details of FractalLoss, including forecasting backbones, training schedules, hyperparameter settings, and evaluation procedures, are well-documented. We also provide the core code at the review stage.

REFERENCES

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

Hanzhong Cao, Wenbo Yan, and Ying Tan. Numerion: A multi-hypercomplex model for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2026.

Xinghong Chen, Weilin Wu, Kunping Yang, and Guannan Chen. Pmlf: A physics-guided multiscale loss framework for structurally heterogeneous time series. *Advances in Neural Information Processing Systems*, 2026.

Yu Chen, Nathalia Céspedes, and Payam Barnaghi. A closer look at transformers for time series forecasting: Understanding why they work and where they struggle. In *Forty-second international conference on machine learning*, 2025.

Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *International Conference on Machine Learning (ICML)*, 2017.

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.

Gilles Demange, Helena Zapolsky, Renaud Patte, and Marc Brunel. A phase field model for snow crystal growth in three dimensions. *NPJ Computational Materials*, 3(1):15, 2017.

Norman R Draper and Harry Smith. *Applied regression analysis*. John Wiley & Sons, 1998.

Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pp. 569–593. Springer, 1992.

Christophe Eloy, Meriem Fournier, André Lacointe, and Bruno Moulia. Wind loads and competition for light sculpt trees into self-similar structures. *Nature communications*, 8(1):1014, 2017.

Jiayu Fang, Xuande Liu, Sangsha Fang, Ernie Tian, Hongwei Ma, Zhiqi Shao, and Junbin Gao. Information geometry loss for time series forecasting. In *International Conference on Learning Representations*, 2026.

Yuchen Fang, Yuxuan Liang, Bo Hui, Zezhi Shao, Liwei Deng, Xu Liu, Xinke Jiang, and Kai Zheng. Efficient large-scale traffic forecasting with transformers: A spatial data management perspective. In *Proceedings of the 31st ACM SIGKDD conference on knowledge discovery and data mining v. 1*, pp. 307–317, 2025.

Yisong Fu, Fei Wang, Zezhi Shao, Boyu Diao, Lin Wu, Zhulin An, Chengqing Yu, Yujie Li, and Yongjun Xu. On the integration of spatial-temporal knowledge: A lightweight approach to atmospheric time series forecasting. *Advances in Neural Information Processing Systems*, 38: 82509–82538, 2026.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021.

Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In *International Conference on Learning Representations*, volume 2024, pp. 56374–56394, 2024.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Weiyang Kong, Ziyu Guo, and Yubao Liu. Spatio-temporal pivotal graph neural networks for traffic flow forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, 2024.

Feifei Kou, Jiahao Wang, Lei Shi, Yuhan Yao, Yawen Li, Suguo Zhu, Zhongbao Zhang, and Junping Du. Cfpt: Empowering time series forecasting through cross-frequency interaction and periodic-aware timestamp modeling. In *ICML*, 2025.

Dilfira Kudrat, Zongxia Xie, Yanru Sun, Tianyu Jia, and Qinghua Hu. Patch-wise structural loss for time series forecasting. *arXiv preprint arXiv:2503.00877*, 2025.

Vincent Le Guen and Nicolas Thome. Shape and time distortion loss for training deep time series forecasting models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Vincent Le Guen and Nicolas Thome. Deep time series forecasting with shape and temporal criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):342–355, 2022.

Hyunwook Lee, Chunggi Lee, Hongkyu Lim, and Sungahn Ko. Tilde-q: a transformation invariant loss function for time-series forecasting. *arXiv preprint arXiv:2210.15050*, 2022.

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.

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.

Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-tao Xia. Timebridge: Non-stationarity matters for long-term time series forecasting. *arXiv preprint arXiv:2410.04442*, 2024a.

Tianyu Liu, Kai Sun, Fuchun Sun, Yu Luo, and Yuanlong Zhang. Slimming the fat-tail: Morphing-flow for adaptive time series modeling. In *International Conference on Machine Learning (ICML)*, 2025.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024b.

Ziqi Liu. Freqmoe: Enhancing time series forecasting through frequency decomposition mixture of experts. *arXiv preprint arXiv:2501.15125*, 2025.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *International conference on learning representations*, volume 2024, pp. 31728–31770, 2024.

Aitian Ma, Dongsheng Luo, and Mo Sha. Mixlinear: Extreme low resource multivariate time series forecasting with $0.1k$ parameters. In *International Conference on Learning Representations*, volume 2026, pp. 77764–77789, 2026a.

Jiaming Ma, Binwu Wang, Guanjun Wang, Kuo Yang, Zhengyang Zhou, Pengkun Wang, Xu Wang, and Yang Wang. Less but more: Linear adaptive graph learning empowering spatiotemporal forecasting. *Advances in Neural Information Processing Systems*, 38:58382–58418, 2026b.

Benoit Mandelbrot. How long is the coast of britain? statistical self-similarity and fractional dimension. *science*, 1967.

Juntong Ni, Zewen Liu, Shiyu Wang, Ming Jin, and Wei Jin. Timedistill: Efficient long-term time series forecasting with mlp via cross-architecture distillation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1113–1124, 2026.

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.

Yiming Niu, Jinliang Deng, and Yongxin Tong. Phaseformer: From patches to phases for efficient and effective time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2026.

Alan V Oppenheim and Ronald W Schafer. *Discrete-time signal processing*. Pearson, 2010.

MA Parseval des Chenes. Mémoires présentés à l'institut des sciences, lettres et arts, par divers savans, et lus dans ses assemblées. *Sciences, mathématiques et physiques (Savans étrangers)*, 1806.

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD conference on knowledge discovery and data mining v. 1*, pp. 1185–1196, 2025.

Xiangfei Qiu, Xingjian Wu, Hanyin Cheng, Xvyuan Liu, Chenjuan Guo, Jilin Hu, and Bin Yang. Dbloss: Decomposition-based loss function for time series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2026.

Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597(7878):672–677, 2021.

Jialong Sun, Xinpeng Ling, Jiaxuan Zou, Jiawen Kang, and Kejia Zhang. Freie: Low-frequency spectral bias in neural networks for time-series tasks. In *2025 IEEE International Conference on Data Mining (ICDM)*, pp. 1545–1554. IEEE, 2025a.

Yanru Sun, Zongxia Xie, Emadeldeen Eldele, Dongyue Chen, Qinghua Hu, and Min Wu. Learning pattern-specific experts for time series forecasting under patch-level distribution shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025b.

Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, and Jianxin Liao. Rethinking the power of timestamps for robust time series forecasting: A global-local fusion perspective. *Advances in neural information processing systems*, 37:22206–22232, 2024a.

Hao Wang, Lichen Pan, Yuan Shen, Zhichao Chen, Degui Yang, Yifei Yang, Sen Zhang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. Fredf: Learning to forecast in the frequency domain. In *International Conference on Learning Representations (ICLR)*, 2025.

Hao Wang, Licheng Pan, Zhichao Chen, Xu Chen, Qingyang Dai, Lei Wang, Haoxuan Li, and Zhouchen Lin. Time-o1: Time-series forecasting needs transformed label alignment. *Advances in Neural Information Processing Systems*, 38:8660–8690, 2026a.

Hao Wang, Licheng Pan, Yuan Lu, Zhixuan Chu, Xiaoxi Li, Shuting He, Zi Chan, Qingsong Wen, Haoxuan Li, and Zhouchen Lin. Distdf: Time-series forecasting needs joint-distribution wasserstein alignment. In *International Conference on Learning Representations (ICLR)*, 2026b.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024b.

Andrew Robert Williams, Arjun Ashok, Étienne Marcotte, Valentina Zantedeschi, Jithendaraa Subramanian, Roland Riachi, James Requeima, Alexandre Lacoste, Irina Rish, Nicolas Chapados, and Alexandre Drouin. Context is key: A benchmark for forecasting with essential textual information. In *Proceedings of the 42nd International Conference on Machine Learning(ICML)*, 2025.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Haixu Wu, Hang Zhou, Mingsheng Long, and Jianmin Wang. Interpretable weather forecasting for worldwide stations with a unified deep model. *Nature Machine Intelligence*, 5(6):602–611, 2023.

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. *Advances in Neural Information Processing Systems*, 38:23328–23354, 2026.

Hongjie Xia, Huijie Ao, Long Li, Yu Liu, Sen Liu, Guangnan Ye, and Hongfeng Chai. Ci-sthpan: Pre-trained attention network for stock selection with channel-independent spatio-temporal hypergraph. In *Proceedings of the AAAI conference on artificial intelligence*, 2024.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations (ICLR)*, 2024.

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in neural information processing systems*, 36:69638–69660, 2023a.

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*, 2023b.

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*, pp. 6206–6215, 2025.

Wenzhen Yue, Yong Liu, Hao Wang, Haoxuan Li, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. Olinear: A linear model for time series forecasting in orthogonally transformed domain. *Advances in Neural Information Processing Systems*, 38:60573–60632, 2026.

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 (AAAI)*, 2023.

Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I Jordan, and Jianmin Wang. Skilful nowcasting of extreme precipitation with nowcastnet. *Nature*, 619 (7970):526–532, 2023.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2023.

Zhiyuan Zhao, Juntong Ni, Shangqing Xu, Haoxin Liu, Wei Jin, and B Aditya Prakash. Timerecipe: A time-series forecasting recipe via benchmarking module level effectiveness. In *International Conference on Learning Representations*, volume 2026, pp. 150745–150792, 2026.

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 (AAAI)*, 2021.

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. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022a.

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 (ICML)*, 2022b.

APPENDIX: TABLE OF CONTENTS

A Theoretical Justification 16

A.1 Notation Table 16

A.2 Proof of Proposition 1 16

A.3 Proof of Proposition 2 17

A.4 Derivation of the Power-Law Energy Spectrum 18

B Reproduction Details 18

B.1 Dataset Descriptions 18

B.2 Forecasting Backbones 19

B.3 Implementation Details and Hyperparameter Selection 19

B.4 Evaluation Metrics 19

C Additional Experiment Results 21

C.1 RANDOM SEED SENSITIVITY 21

C.2 Full horizon results of Long-term multivariate forecasting comparison 22

C.3 More Visualization of Fractal Structure on real-world time series 23

A THEORETICAL JUSTIFICATION

A.1 NOTATION TABLE

The table below summarizes the main notation, organized by conceptual role, together with its meaning and first appearance in the paper.

Notations	Descriptions	First appearance
$\mathbf{x},\ T$	Historical input window and its length	Section 3.1
$\mathbf{y},\ \hat{\mathbf{y}},\ \tau$	Ground-truth future, prediction, and forecasting horizon	Section 3.1
$f_\theta,\ \theta$	Forecasting model and its parameters	Section 3.1
$\mathcal{F}(\cdot)$	Discrete Fourier transform	Section 3.1, Eq. 1
Ω	Frequency index set $\{0, 1, \ldots, \tau - 1\}$	Section 4.1, Eq. 8
$f_k = k/\tau$	Frequency associated with the k th Fourier component	Section 3.1
$Y_k,\ \hat{Y}_k$	Fourier coefficients of the ground truth and prediction at index k	Section 3.1
$E_k = \lvert Y_k\rvert^2$	Ground-truth spectral energy at index k	Section 3.1, Eq. 2
$\mathcal{N}_r(k),\ r$	Neighboring frequency indices of k and the peak-detection radius	Section 4.1, Eq. 8
$\mathcal{P},\ \mathcal{B}$	Sets of peak components and background components, respectively	Section 4.1, Eqs. 8–9
$\mathcal{I}_\mathcal{B},\ \lvert\mathcal{I}_\mathcal{B}\rvert$	Index set of background components and its cardinality	Section 4.1
ϵ	Positive numerical stability constant inside the energy logarithm	Section 4.2, Eq. 11
$u_k,\ v_k,\ \hat{v}_k$	Log-frequency $\log f_k$, ground-truth log-energy $\log(\lvert Y_k\rvert^2 + \epsilon)$, and predicted log-energy $\log(\lvert\hat{Y}_k\rvert^2 + \epsilon)$	Section 4.2, Eq. 11
$\bar{u},\ \bar{v}$	Means of u_k and v_k over $k \in \mathcal{I}_\mathcal{B}$	Section 4.2, Eq. 12
$\alpha,\ \beta$	Power-law exponent and log-energy intercept; the fractal-line slope is $-\alpha$	Section 3.1, Eq. 4
$L(u) = -\alpha u + \beta$	Fractal line fitted to the ground-truth background log-spectrum	Section 4.2
$\tilde{E}_k,\ \hat{E}_k$	Fractal-line energy $\exp(-\alpha u_k + \beta)$ and predicted energy $\lvert\hat{Y}_k\rvert^2$	Section 4.2, Proposition 2
$\mathbf{e} = \hat{\mathbf{y}} - \mathbf{y}$	Prediction error in the time domain	Section 3.2
ρ_k	Relative Fourier error: $\lvert\hat{Y}_k - Y_k\rvert/\lvert Y_k\rvert$, for $\lvert Y_k\rvert > 0$	Section 3.2, Proposition 1
ξ_k	Relative energy error with respect to the fractal line: $(\hat{E}_k - \tilde{E}_k)/\tilde{E}_k$	Section 4.2, Proposition 2
$\mathcal{L}_{\text{MSE}}$	Mean squared error in the time domain	Section 3.2, Eq. 6
$\mathcal{L}_{\text{naive}}$	Mean squared Fourier error over background components	Section 4.2, Eq. 10
$\mathcal{L}_{\text{fractal}}$	Mean squared deviation of predicted background log-energy from the fitted line	Section 4.2, Eq. 13
$\mathcal{L}_{\text{total}},\ \lambda$	Combined training objective and the weight assigned to FractalLoss	Section 4.2, Eq. 15
R^2	Coefficient of determination for the ground-truth background line fit	Section 4.1
$\hat{\alpha},\ \hat{R}^2$	Power-law exponent and coefficient of determination fitted to the predicted background spectrum	Section 5.1, Eq. 16
FSE, FFE	Fractal Slope Error $\lvert\hat{\alpha} - \alpha\rvert$ and Fractal Fit Error $\lvert\hat{R}^2 - R^2\rvert$	Section 5.1, Eq. 16

A.2 PROOF OF PROPOSITION 1

Proposition 1 (MSE provides energy-preferential supervision). *For each frequency component Y_k with $\lvert Y_k\rvert > 0$, let $\rho_k = \lvert\hat{Y}_k - Y_k\rvert/\lvert Y_k\rvert$ denote its relative error in the frequency domain. Then the*

sensitivity of $\mathcal{L}_{\mathrm{MSE}}$ to each frequency component satisfies

$$\frac{\partial \mathcal{L}_{\mathrm{MSE}}}{\partial \rho_k} = \frac{2}{\tau^2} \cdot |Y_k|^2 \cdot \rho_k \tag{17}$$

where τ represents the target forecasting horizon.

Proof. Fix the ground truth $\mathbf{y}$ and let $\mathbf{e} = \hat{\mathbf{y}} - \mathbf{y}$. By Parseval's theorem and the linearity of the DFT,

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{\tau}\|\mathbf{e}\|_2^2 = \frac{1}{\tau^2}\sum_{\ell=0}^{\tau-1} |\mathcal{F}(\mathbf{e})_\ell|^2 = \frac{1}{\tau^2}\sum_{\ell=0}^{\tau-1} |\hat{Y}_\ell - Y_\ell|^2 \tag{18}$$

For $|Y_k| > 0$, the definition of ρ_k gives $|\hat{Y}_k - Y_k|^2 = |Y_k|^2\rho_k^2$. Separating the k th term yields

$$\mathcal{L}_{\mathrm{MSE}} = \frac{|Y_k|^2}{\tau^2}\rho_k^2 + \frac{1}{\tau^2}\sum_{\substack{\ell=0 \\ \ell\neq k}}^{\tau-1} |\hat{Y}_\ell - Y_\ell|^2 \tag{19}$$

Only the first term depends on ρ_k. Thus,

$$\frac{\partial \mathcal{L}_{\mathrm{MSE}}}{\partial \rho_k} = \frac{|Y_k|^2}{\tau^2}\frac{\partial \rho_k^2}{\partial \rho_k} = \frac{2}{\tau^2}|Y_k|^2\rho_k \tag{20}$$

At the same positive relative error, the sensitivity is proportional to $|Y_k|^2$. Hence, MSE favors higher-energy components in relative-error coordinates, completing the proof. □

A.3 Proof of Proposition 2

Proposition 2 (Fractal loss provides energy-independent supervision). *For each background frequency f_k, let $\tilde{E}_k = \exp(-\alpha u_k + \beta)$ denote the energy prescribed by the fractal line at f_k, let $\hat{E}_k = |\hat{Y}_k|^2$ denote the predicted energy, and define the relative error at f_k as $\xi_k = (\hat{E}_k - \tilde{E}_k)/\tilde{E}_k$. Then the sensitivity of $\mathcal{L}_{\mathrm{fractal}}$ to each background frequency satisfies:*

$$\frac{\partial \mathcal{L}_{\mathrm{fractal}}}{\partial \xi_k} = \frac{2}{|\mathcal{I}_{\mathcal{B}}|} \cdot \frac{\log(1+\xi_k)}{1+\xi_k} \tag{21}$$

where $|\mathcal{I}_{\mathcal{B}}|$ denotes the number of background components.

Proof. For theoretical analysis, we omit the numerical stabilizer ϵ and assume strictly positive spectral energies, so that all logarithms are well defined. Fix the nonempty background index set $\mathcal{I}_{\mathcal{B}}$ and the fitted parameters α and β, which depend only on the ground truth. Starting from Eq. 13,

$$\mathcal{L}_{\mathrm{fractal}} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\sum_{\ell\in\mathcal{I}_{\mathcal{B}}} [\hat{v}_\ell - L(u_\ell)]^2 = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\sum_{\ell\in\mathcal{I}_{\mathcal{B}}} [\hat{v}_\ell + \alpha u_\ell - \beta]^2 \tag{22}$$

Using $\hat{v}_k = \log \hat{E}_k$ and $L(u_k) = -\alpha u_k + \beta = \log \tilde{E}_k$, we obtain

$$\mathcal{L}_{\mathrm{fractal}} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\sum_{\ell\in\mathcal{I}_{\mathcal{B}}} \left[\log \hat{E}_\ell - \log \tilde{E}_\ell\right]^2 = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\sum_{\ell\in\mathcal{I}_{\mathcal{B}}} \left[\log \frac{\hat{E}_\ell}{\tilde{E}_\ell}\right]^2 \tag{23}$$

The definition of ξ_k gives $\hat{E}_k/\tilde{E}_k = 1 + \xi_k > 0$. Substituting into the loss yields

$$\mathcal{L}_{\mathrm{fractal}} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\sum_{\ell\in\mathcal{I}_{\mathcal{B}}} [\log(1+\xi_\ell)]^2 \tag{24}$$

Only the k th term depends on ξ_k, we then apply the chain rule,

$$\frac{\partial \mathcal{L}_{\mathrm{fractal}}}{\partial \xi_k} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|}\frac{\partial}{\partial \xi_k}[\log(1+\xi_k)]^2 = \frac{2}{|\mathcal{I}_{\mathcal{B}}|}\frac{\log(1+\xi_k)}{1+\xi_k} \tag{25}$$

The constant $2/|\mathcal{I}_{\mathcal{B}}|$ is shared by all background components, and the remaining term depends only on ξ_k. Thus, components with the same relative error have identical sensitivities regardless of $\tilde{E}_k$, establishing energy-independent supervision in relative-error coordinates. □

A.4 Derivation of the Power-Law Energy Spectrum

We provide a scaling argument for the power-law spectrum in Eq. 4. Let $E(f) > 0$ denote a spectral envelope at positive frequency f. By the Fourier scaling property, stretching time by a factor $\lambda > 0$ changes the frequency scale inversely:

$$t \to \lambda t, \qquad f \to f/\lambda. \tag{26}$$

Assumption (Spectral scale invariance). Frequency rescaling preserves the spectral shape, multiplying energy at every frequency by the same factor λ^α, where α is a fixed scaling exponent:

$$E(f/\lambda) = \lambda^\alpha E(f), \qquad f > 0, \quad \lambda > 0. \tag{27}$$

The factor λ^α is independent of f, so all frequencies follow the same scaling rule. Choose a reference frequency $f_0 > 0$ and set $\lambda = f/f_0$. Substituting into Eq. 27 gives

$$E(f_0) = \left(\frac{f}{f_0}\right)^\alpha E(f) \quad \Longrightarrow \quad E(f) = E(f_0)\left(\frac{f}{f_0}\right)^{-\alpha}. \tag{28}$$

With $E(f_0)$ and f_0 fixed, evaluating at positive DFT frequencies $f_k = k/\tau$ gives

$$E_k \propto f_k^{-\alpha}. \tag{29}$$

Taking logarithms and defining $\beta = \log E(f_0) + \alpha \log f_0$ yields

$$\log E_k = -\alpha \log f_k + \beta. \tag{30}$$

Spectral implication. Under this assumption, energy follows a power law in frequency, giving a straight line in log-frequency log-energy coordinates with slope $-\alpha$ and intercept β. For finite time series, this relation approximates the background spectral envelope; individual components may deviate from the line. We exclude $f_k = 0$, where $\log f_k$ is undefined. □

B Reproduction Details

This appendix reports the datasets, forecasting backbones, implementation details, server configuration, evaluation metrics, and hyperparameter settings used in our experiments.

B.1 Dataset Descriptions

We evaluate FractalLoss on 7 widely used real-world multivariate time series forecasting benchmarks, including ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, and Exchange. These datasets cover diverse forecasting scenarios, including electricity transformer temperature, meteorology, electricity consumption, and exchange rates. Following standard long-term forecasting protocols (Piao et al., 2024), we fix the lookback window length to $T = 96$ and evaluate forecasting horizons $\tau \in \{96, 192, 336, 720\}$ across all datasets. The table below summarizes their main characteristics.

Dataset	*D*	Forecast length	Train / validation / test	Frequency	Domain
ETTh1	7	96, 192, 336, 720	8545/2881/2881	Hourly	Health
ETTh2	7	96, 192, 336, 720	8545/2881/2881	Hourly	Health
ETTm1	7	96, 192, 336, 720	34465/11521/11521	15 min	Health
ETTm2	7	96, 192, 336, 720	34465/11521/11521	15 min	Health
Weather	21	96, 192, 336, 720	36792/5271/10540	10 min	Weather
Electricity	321	96, 192, 336, 720	18317/2633/5261	Hourly	Electricity
Exchange	8	96, 192, 336, 720	5120/665/1422	Daily	Economy

Note: D denotes the number of variates. *Frequency* denotes the sampling interval. *Train, validation, and test* indicate the number of samples in each split.

ETT Datasets. The Electricity Transformer Temperature (ETT) datasets contain two-year monitoring data collected from electricity transformers. They include four subsets: ETTh1, ETTh2, ETTm1, and

ETTm2. The "h" subsets are recorded hourly, while the "m" subsets are recorded every 15 minutes. Each subset contains seven variables, including power load related features and oil temperature.

Weather Dataset. The Weather dataset contains meteorological indicators collected at a 10-minute interval. It includes multiple weather-related variables, such as temperature, pressure, humidity, and wind information. This dataset is used to evaluate whether FractalLoss can recover fine-grained temporal dynamics in real-world meteorological forecasting.

Electricity Dataset. The Electricity Consuming Load dataset records hourly electricity consumption from 321 clients. It is a high-dimensional multivariate forecasting benchmark and is commonly used to evaluate long-term forecasting models under complex electricity usage patterns.

Exchange Dataset. The Exchange dataset contains daily exchange rates from eight countries. Compared with other datasets, it has fewer variables but contains complex non-stationary variations, making it suitable for evaluating the robustness of forecasting objectives.

B.2 Forecasting Backbones

To verify that FractalLoss is model-agnostic, we evaluate it on representative forecasting backbones of different types. For each backbone, we compare the standard MSE training objective with the same backbone trained by additionally incorporating FractalLoss. The table below summarizes the forecasting backbones used in our experiments.

Backbone	Type	Code
iTransformer Liu et al. (2024b)	Transformer based	https://github.com/thuml/iTransformer
PatchTST Nie et al. (2022)	Patch based	https://github.com/yuqinie98/PatchTST
DLinear Zeng et al. (2023)	Linear based	https://github.com/cure-lab/LTSF-Linear
Fredformer Piao et al. (2024)	Frequency domain	https://github.com/chenzRG/Fredformer
PhaseFormer Niu et al. (2026)	Phase domain	https://github.com/neumyor/PhaseFormer_TSL

B.3 Implementation Details and Hyperparameter Selection

All experiments are implemented in PyTorch and conducted on a server equipped with an Intel® Xeon® Silver 4210R CPU @ 2.40 GHz and an NVIDIA RTX A5000 GPU with 24 GB memory. For each dataset and backbone, models are trained on the training split, selected based on validation performance, and evaluated on the test split. We fix the lookback window length to $T = 96$ and evaluate forecasting horizons $\tau \in \{96, 192, 336, 720\}$. All reported results are averaged over these four horizons unless otherwise specified. All models are trained using the Adam optimizer. Unless otherwise specified, the batch size is 32, and the initial learning rate is selected from $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}\}$ based on validation performance. The learning rate decay coefficient is fixed at $\eta = 0.5$, with the learning rate halved every two epochs. For fair comparison, applying FractalLoss preserves the backbone architecture and training protocol, modifying only the training objective through the proposed fractal regularization term. Table 4 summarizes the hyperparameters.

B.4 Evaluation Metrics

We evaluate forecasting accuracy using mean squared error (MSE) and mean absolute error (MAE), and fractal structure preservation using Fractal Slope Error (FSE) and Fractal Fit Error (FFE). We present the single-channel definitions below. Superscript (n) indexes test samples, while subscripts t and k index time steps and frequencies, respectively. Let $\mathbf{y}^{(n)}, \hat{\mathbf{y}}^{(n)} \in \mathbb{R}^{\tau}$ denote the ground-truth and predicted sequences, where N is the number of test samples and τ is the forecasting horizon.

Mean Squared Error (MSE). MSE measures squared prediction errors, averaged over test samples and future time steps:

$$\mathrm{MSE} = \frac{1}{N\tau} \sum_{n=1}^{N} \sum_{t=1}^{\tau} \left(\hat{y}_t^{(n)} - y_t^{(n)}\right)^2. \tag{31}$$

Table 4: Summary of the main hyperparameter settings and sensitivity analysis results.

Parameter	Description	Setting and sensitivity result
Optimizer	Optimization algorithm.	Adam optimizer (Kingma & Ba, 2014) is used for all models.
B	Batch size.	Fixed at 32 in all experiments.
lr	Initial learning rate.	Search range: $1\times10^{-4}, 3\times10^{-4}, 5\times10^{-4}, 1\times10^{-3}, 3\times10^{-3}, 5\times10^{-3}\}$; selected based on validation performance.
η	Learning rate decay coefficient.	$\eta = 0.5$; the learning rate is halved every 2 epochs.
p_{drop}	Dropout rate.	Fixed at 0.2 to prevent overfitting.
Seeds	Random seeds for independent runs.	Results are averaged over 5 random seeds: $\{2026, 2027, 2028, 2029, 2030\}$.
λ	Weight balancing MSE and FractalLoss.	Search range: $[0, 1]$. Sensitivity analysis shows that the best results are obtained within $\lambda \in [0.3, 0.6]$. Excessively large weights degrade performance, particularly at $\lambda = 1$, where the MSE term is removed.
r	Neighborhood radius for spectral peak detection.	Search range: $[1, 19]$. Sensitivity analysis shows that the best results are obtained within $r \in [3, 9]$. At $r = 1$, overly permissive peak detection results in relatively high MSE.
ϵ	Numerical stability constant	Fixed at 10^{-8}.

Mean Absolute Error (MAE). MAE measures absolute prediction errors over the same samples and time steps:

$$\text{MAE} = \frac{1}{N\tau}\sum_{n=1}^{N}\sum_{t=1}^{\tau}\big|\hat{y}_t^{(n)} - y_t^{(n)}\big|. \tag{32}$$

To compute FSE and FFE, we first transform each sequence into the frequency domain using Eq. 1:

$$Y_k^{(n)} = \mathcal{F}(\mathbf{y}^{(n)})_k, \qquad \hat{Y}_k^{(n)} = \mathcal{F}(\hat{\mathbf{y}}^{(n)})_k. \tag{33}$$

We identify the background index set $\mathcal{I}_{\mathcal{B}}^{(n)}$ from the ground truth using Eqs. 8–9, excluding the zero-frequency component. The same indices are used for both spectra. Following Eq. 11, we obtain

$$u_k = \log f_k, \qquad v_k^{(n)} = \log\big(|Y_k^{(n)}|^2 + \epsilon\big), \qquad \hat{v}_k^{(n)} = \log\big(|\hat{Y}_k^{(n)}|^2 + \epsilon\big). \tag{34}$$

We then fit separate fractal lines over $\mathcal{I}_{\mathcal{B}}^{(n)}$ using Eq. 12:

$$L^{(n)}(u) = -\alpha^{(n)}u + \beta^{(n)}, \qquad \hat{L}^{(n)}(u) = -\hat{\alpha}^{(n)}u + \hat{\beta}^{(n)}. \tag{35}$$

For the predicted fit, $v_k^{(n)}$ is replaced by $\hat{v}_k^{(n)}$ in the fitting formula.

Fractal Slope Error (FSE). FSE measures the discrepancy between the fitted power-law exponents:

$$\text{FSE} = \frac{1}{N}\sum_{n=1}^{N}\big|\hat{\alpha}^{(n)} - \alpha^{(n)}\big|. \tag{36}$$

Because the fractal-line slope is $-\alpha$, this also equals the mean absolute difference between the fitted slopes.

Fractal Fit Error (FFE). FFE measures the discrepancy in how closely the ground-truth and predicted background spectra follow their respective fractal lines. For clarity, let $Q^{(n)}$ and $\hat{Q}^{(n)}$ denote their R^2 values:

$$Q^{(n)} = 1 - \frac{\sum_{k\in\mathcal{I}_{\mathcal{B}}^{(n)}}[v_k^{(n)} - L^{(n)}(u_k)]^2}{\sum_{k\in\mathcal{I}_{\mathcal{B}}^{(n)}}[v_k^{(n)} - \bar{v}^{(n)}]^2}, \qquad \hat{Q}^{(n)} = 1 - \frac{\sum_{k\in\mathcal{I}_{\mathcal{B}}^{(n)}}[\hat{v}_k^{(n)} - \hat{L}^{(n)}(u_k)]^2}{\sum_{k\in\mathcal{I}_{\mathcal{B}}^{(n)}}[\hat{v}_k^{(n)} - \bar{\hat{v}}^{(n)}]^2}, \tag{37}$$

where $\bar{v}^{(n)}$ and $\bar{\hat{v}}^{(n)}$ are the respective mean log-energies over $\mathcal{I}_{\mathcal{B}}^{(n)}$. These fits require at least two distinct positive frequencies, and the R^2 expressions assume nonzero log-energy variance. FFE is defined as

$$\mathrm{FFE} = \frac{1}{N}\sum_{n=1}^{N}\left|\hat{Q}^{(n)} - Q^{(n)}\right|. \tag{38}$$

Lower values are better for all four metrics. For multivariate time series, each metric is computed separately for each channel and then averaged across channels.

C Additional Experiment Results

C.1 Random Seed Sensitivity

Table 5: Random-seed (2026,2027,2028,2029,2030) results (mean±std) using Fredformer backbone on ETTh2 and Exchange datasets.

τ	Loss	ETTh2				Exchange			
		MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE
96	MSE	$0.299_{\pm 0.001}$	$0.345_{\pm 0.000}$	$0.142_{\pm 0.004}$	$0.457_{\pm 0.007}$	$0.082_{\pm 0.003}$	$0.200_{\pm 0.001}$	$0.395_{\pm 0.004}$	$0.392_{\pm 0.005}$
	FractalLoss	$\mathbf{0.281}_{\pm 0.002}$	$\mathbf{0.330}_{\pm 0.002}$	$\mathbf{0.043}_{\pm 0.001}$	$\mathbf{0.029}_{\pm 0.001}$	$\mathbf{0.080}_{\pm 0.001}$	$\mathbf{0.199}_{\pm 0.002}$	$\mathbf{0.241}_{\pm 0.000}$	$\mathbf{0.070}_{\pm 0.002}$
192	MSE	$0.371_{\pm 0.003}$	$0.392_{\pm 0.004}$	$0.432_{\pm 0.007}$	$0.391_{\pm 0.011}$	$0.180_{\pm 0.001}$	$0.303_{\pm 0.001}$	$0.107_{\pm 0.005}$	$0.239_{\pm 0.006}$
	FractalLoss	$\mathbf{0.357}_{\pm 0.002}$	$\mathbf{0.381}_{\pm 0.001}$	$\mathbf{0.107}_{\pm 0.001}$	$\mathbf{0.055}_{\pm 0.002}$	$\mathbf{0.168}_{\pm 0.000}$	$\mathbf{0.296}_{\pm 0.002}$	$\mathbf{0.051}_{\pm 0.002}$	$\mathbf{0.003}_{\pm 0.004}$
336	MSE	$0.416_{\pm 0.001}$	$0.427_{\pm 0.000}$	$0.038_{\pm 0.005}$	$0.204_{\pm 0.006}$	$0.332_{\pm 0.003}$	$0.418_{\pm 0.002}$	$0.048_{\pm 0.011}$	$0.073_{\pm 0.006}$
	FractalLoss	$\mathbf{0.404}_{\pm 0.002}$	$\mathbf{0.419}_{\pm 0.001}$	$\mathbf{0.031}_{\pm 0.004}$	$\mathbf{0.077}_{\pm 0.001}$	$\mathbf{0.320}_{\pm 0.002}$	$\mathbf{0.409}_{\pm 0.001}$	$\mathbf{0.045}_{\pm 0.001}$	$\mathbf{0.024}_{\pm 0.001}$
720	MSE	$0.426_{\pm 0.001}$	$0.444_{\pm 0.001}$	$0.200_{\pm 0.009}$	$0.148_{\pm 0.004}$	$0.838_{\pm 0.003}$	$0.695_{\pm 0.000}$	$0.142_{\pm 0.003}$	$0.020_{\pm 0.007}$
	FractalLoss	$\mathbf{0.406}_{\pm 0.001}$	$\mathbf{0.430}_{\pm 0.001}$	$\mathbf{0.059}_{\pm 0.003}$	$\mathbf{0.035}_{\pm 0.000}$	$\mathbf{0.804}_{\pm 0.005}$	$\mathbf{0.676}_{\pm 0.001}$	$\mathbf{0.011}_{\pm 0.001}$	$\mathbf{0.015}_{\pm 0.002}$

Note: The lookback window is 96. FractalLoss denotes MSE combined with the proposed regularizer. For each dataset, horizon, and metric, the lower mean is bolded, with blue backgrounds for MSE/MAE and red backgrounds for FSE/FFE. Lower is better.

C.2 FULL HORIZON RESULTS OF LONG-TERM MULTIVARIATE FORECASTING COMPARISON

Table 6: Long-term multivariate forecasting comparison.

Dataset	T	Loss	iTransformer				PatchTST				DLinear				Fredformer				PhaseFormer			
			MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE	MSE	MAE	FSE	FFE
ETTh1	96	MSE	0.392	0.410	0.116	0.267	0.389	0.404	0.116	0.227	0.383	0.396	0.292	0.100	0.383	0.397	0.203	0.303	0.380	0.394	0.351	0.614
		+Fractal	**0.381**	**0.401**	**0.037**	**0.145**	**0.384**	**0.398**	**0.061**	**0.122**	**0.378**	**0.387**	**0.043**	**0.022**	**0.381**	**0.393**	**0.005**	**0.096**	**0.376**	**0.392**	**0.033**	**0.027**
	192	MSE	0.446	0.436	0.674	0.446	0.433	0.432	0.296	0.245	0.439	0.435	0.188	0.115	0.439	0.429	0.380	0.322	0.426	0.421	0.229	0.639
		+Fractal	**0.431**	**0.426**	**0.056**	**0.122**	**0.430**	**0.425**	**0.036**	**0.162**	**0.429**	**0.419**	**0.047**	**0.020**	**0.423**	**0.422**	**0.012**	**0.034**	**0.424**	**0.420**	**0.004**	**0.154**
	336	MSE	0.488	0.459	0.949	0.252	**0.466**	0.456	0.537	0.137	0.481	0.458	0.102	0.102	0.495	0.459	0.328	0.142	0.462	0.438	0.218	0.413
		+Fractal	**0.472**	**0.445**	**0.012**	**0.039**	**0.466**	**0.448**	**0.056**	**0.049**	**0.473**	**0.446**	**0.039**	**0.026**	**0.478**	**0.442**	**0.004**	**0.051**	**0.460**	**0.436**	**0.060**	**0.040**
	720	MSE	0.510	0.495	1.037	0.255	**0.472**	0.484	0.519	0.095	0.505	0.503	0.130	0.087	0.535	0.495	0.301	0.181	0.468	0.459	0.230	0.314
		+Fractal	**0.492**	**0.484**	**0.047**	**0.002**	0.484	**0.481**	**0.091**	**0.015**	**0.488**	**0.484**	**0.059**	**0.020**	**0.474**	**0.463**	**0.043**	**0.055**	**0.448**	**0.452**	**0.079**	**0.095**
ETTh2	96	MSE	0.301	0.352	0.488	0.514	0.290	0.341	0.221	0.401	0.337	0.392	0.365	0.625	0.293	0.342	0.150	0.282	0.305	0.352	0.147	0.631
		+Fractal	**0.288**	**0.338**	**0.080**	**0.194**	**0.283**	**0.331**	**0.107**	**0.248**	**0.319**	**0.347**	**0.058**	**0.062**	**0.281**	**0.330**	**0.046**	**0.029**	**0.293**	**0.341**	**0.063**	**0.012**
	192	MSE	0.380	0.400	0.583	0.521	0.372	0.393	0.272	0.397	0.449	0.463	0.718	0.915	0.375	0.393	0.172	0.360	0.382	0.398	0.170	0.574
		+Fractal	**0.369**	**0.391**	**0.103**	**0.170**	**0.362**	**0.382**	**0.090**	**0.216**	**0.421**	**0.405**	**0.060**	**0.059**	**0.357**	**0.380**	**0.029**	**0.055**	**0.372**	**0.390**	**0.099**	**0.018**
	336	MSE	0.423	0.434	0.744	0.411	0.420	0.433	0.232	0.261	0.486	0.490	0.736	0.807	0.415	0.428	0.274	0.344	0.425	**0.430**	0.226	0.534
		+Fractal	**0.410**	**0.425**	**0.116**	**0.011**	**0.408**	**0.420**	**0.146**	**0.094**	**0.473**	**0.460**	**0.059**	**0.061**	**0.405**	**0.420**	**0.021**	**0.077**	**0.422**	**0.430**	**0.145**	**0.014**
	720	MSE	0.428	0.446	0.813	0.442	0.430	0.449	0.527	0.257	**0.480**	**0.495**	0.781	0.737	0.429	0.445	0.216	0.214	0.440	0.452	0.469	0.333
		+Fractal	**0.421**	**0.442**	**0.097**	**0.089**	**0.419**	**0.435**	**0.169**	**0.050**	0.483	0.500	**0.059**	**0.058**	**0.405**	**0.430**	**0.144**	**0.035**	**0.425**	**0.443**	**0.037**	**0.076**
ETTm1	96	MSE	0.342	0.376	0.145	0.113	0.336	0.374	0.116	0.094	0.344	0.371	0.216	0.198	0.319	0.360	0.469	0.108	0.348	0.374	0.586	0.198
		+Fractal	**0.332**	**0.364**	**0.025**	**0.027**	**0.314**	**0.347**	**0.101**	**0.039**	**0.331**	**0.351**	**0.069**	**0.037**	**0.316**	**0.350**	**0.094**	**0.013**	**0.338**	**0.367**	**0.026**	**0.018**
	192	MSE	0.395	0.403	0.273	0.241	0.372	0.392	0.105	0.183	0.381	0.390	0.177	0.350	0.365	0.384	0.510	0.314	0.384	0.390	0.373	0.380
		+Fractal	**0.377**	**0.388**	**0.189**	**0.053**	**0.366**	**0.378**	**0.092**	**0.087**	**0.374**	**0.374**	**0.086**	**0.046**	**0.356**	**0.377**	**0.073**	**0.037**	**0.379**	**0.387**	**0.006**	**0.052**
	336	MSE	0.444	0.433	0.243	0.149	0.410	0.414	0.102	0.109	0.414	0.414	0.104	0.198	0.398	0.407	0.243	0.152	0.421	0.411	0.618	0.238
		+Fractal	**0.416**	**0.414**	**0.150**	**0.066**	**0.397**	**0.400**	**0.088**	**0.064**	**0.407**	**0.396**	**0.071**	**0.043**	**0.384**	**0.399**	**0.029**	**0.034**	**0.419**	**0.409**	**0.041**	**0.039**
	720	MSE	0.495	0.464	0.435	0.209	0.466	0.444	0.129	0.182	0.473	0.449	0.111	0.202	0.458	0.441	0.226	0.190	0.475	0.445	0.791	0.344
		+Fractal	**0.487**	**0.458**	**0.132**	**0.094**	**0.455**	**0.435**	**0.099**	**0.086**	**0.468**	**0.435**	**0.054**	**0.038**	**0.444**	**0.434**	**0.008**	**0.036**	**0.472**	**0.441**	**0.003**	**0.171**
ETTm2	96	MSE	0.183	0.268	0.521	0.263	0.177	0.261	0.504	0.185	0.192	0.291	0.321	0.461	0.177	0.259	0.502	0.276	0.184	0.272	0.215	0.155
		+Fractal	**0.180**	**0.263**	**0.021**	**0.003**	**0.173**	**0.250**	**0.040**	**0.025**	**0.190**	**0.279**	**0.162**	**0.016**	**0.175**	**0.254**	**0.125**	**0.099**	**0.180**	**0.263**	**0.008**	**0.064**
	192	MSE	0.252	0.312	1.016	0.425	0.243	0.305	0.740	0.174	0.281	0.358	0.549	0.303	0.250	0.307	0.734	0.380	0.256	0.319	0.289	0.194
		+Fractal	**0.245**	**0.303**	**0.072**	**0.034**	**0.239**	**0.294**	**0.044**	**0.016**	**0.276**	**0.338**	**0.209**	**0.022**	**0.238**	**0.298**	**0.168**	**0.145**	**0.250**	**0.306**	**0.189**	**0.096**
	336	MSE	0.317	0.353	0.439	0.147	0.304	0.346	0.502	0.118	0.370	0.418	0.735	0.339	0.306	0.342	0.342	0.200	0.315	0.355	0.240	0.132
		+Fractal	**0.305**	**0.340**	**0.033**	**0.032**	**0.300**	**0.334**	**0.026**	**0.015**	**0.361**	**0.392**	**0.173**	**0.021**	**0.299**	**0.335**	**0.191**	**0.124**	**0.308**	**0.341**	**0.011**	**0.065**
	720	MSE	0.420	0.411	0.612	0.209	0.404	0.404	0.702	0.123	0.529	0.509	0.715	0.321	0.415	0.404	0.402	0.236	0.413	0.406	0.280	0.223
		+Fractal	**0.410**	**0.398**	**0.030**	**0.079**	**0.400**	**0.394**	**0.074**	**0.016**	**0.513**	**0.475**	**0.156**	**0.021**	**0.392**	**0.389**	**0.156**	**0.140**	**0.410**	**0.398**	**0.132**	**0.119**
Weather	96	MSE	0.178	0.219	0.306	0.137	**0.172**	0.214	0.269	0.241	**0.196**	0.236	0.377	0.272	0.165	0.208	0.267	0.228	0.193	0.232	0.271	0.236
		+Fractal	**0.174**	**0.214**	**0.004**	**0.056**	0.174	**0.207**	**0.112**	**0.110**	0.198	**0.214**	**0.109**	**0.072**	**0.163**	**0.204**	**0.022**	**0.062**	**0.185**	**0.226**	**0.069**	**0.023**
	192	MSE	0.226	0.259	0.186	0.164	**0.217**	0.255	0.289	0.245	0.236	0.272	0.400	0.266	0.211	0.250	0.230	0.252	0.239	0.269	0.127	0.219
		+Fractal	**0.223**	**0.255**	**0.101**	**0.064**	0.220	**0.248**	**0.181**	**0.120**	**0.235**	**0.246**	**0.076**	**0.067**	**0.208**	**0.249**	**0.045**	**0.050**	**0.235**	**0.267**	**0.058**	**0.173**
	336	MSE	0.285	0.302	0.219	0.176	**0.275**	0.296	0.187	0.241	0.282	0.312	0.774	0.545	0.271	0.295	0.249	0.281	0.291	0.309	0.161	0.285
		+Fractal	**0.282**	**0.298**	**0.095**	**0.060**	**0.275**	**0.288**	**0.026**	**0.073**	**0.279**	**0.280**	**0.106**	**0.070**	**0.266**	**0.293**	**0.029**	**0.114**	**0.289**	**0.303**	**0.065**	**0.121**
	720	MSE	**0.359**	0.352	0.441	0.267	0.352	0.343	0.103	0.309	0.346	0.356	0.749	0.497	0.349	0.347	0.222	0.355	0.365	0.354	0.233	0.560
		+Fractal	0.361	**0.349**	**0.116**	**0.080**	**0.351**	**0.337**	**0.021**	**0.073**	**0.340**	**0.324**	**0.061**	**0.067**	**0.343**	**0.346**	**0.108**	**0.158**	**0.363**	**0.352**	**0.012**	**0.119**
Electricity	96	MSE	0.151	0.243	0.302	0.240	0.188	0.293	0.133	0.294	0.195	0.277	0.196	0.263	0.140	0.237	0.151	0.247	0.187	0.264	0.150	0.410
		+Fractal	**0.146**	**0.236**	**0.017**	**0.059**	**0.178**	**0.283**	**0.083**	**0.085**	**0.178**	**0.249**	**0.027**	**0.032**	**0.139**	**0.233**	**0.101**	**0.075**	**0.179**	**0.251**	**0.013**	**0.053**
	192	MSE	0.167	0.259	0.467	0.206	0.193	0.298	0.161	0.306	0.194	0.281	0.768	0.161	0.158	**0.254**	0.171	0.273	0.178	0.260	0.197	0.448
		+Fractal	**0.161**	**0.251**	**0.095**	**0.071**	**0.183**	**0.287**	**0.091**	**0.057**	**0.178**	**0.252**	**0.029**	**0.058**	**0.157**	0.255	**0.037**	**0.120**	**0.170**	**0.247**	**0.003**	**0.097**
	336	MSE	0.182	0.274	0.355	0.130	0.213	0.317	0.173	0.239	0.206	0.296	0.574	0.241	0.188	0.286	0.354	0.176	0.209	0.286	0.130	0.615
		+Fractal	**0.176**	**0.266**	**0.123**	**0.111**	**0.202**	**0.306**	**0.098**	**0.089**	**0.188**	**0.266**	**0.043**	**0.099**	**0.185**	**0.284**	**0.137**	**0.074**	**0.200**	**0.272**	**0.047**	**0.023**
	720	MSE	0.212	0.300	0.468	0.164	0.258	0.352	0.229	0.253	0.241	0.330	0.606	0.179	0.226	**0.319**	0.444	0.144	0.242	0.314	0.107	0.331
		+Fractal	**0.205**	**0.291**	**0.105**	**0.123**	**0.245**	**0.340**	**0.096**	**0.077**	**0.220**	**0.297**	**0.093**	**0.035**	**0.223**	0.324	**0.125**	**0.047**	**0.231**	**0.298**	**0.017**	**0.071**
Exchange	96	MSE	0.086	0.207	0.325	0.431	0.088	0.206	0.255	0.315	0.082	0.208	0.292	0.400	0.087	0.206	0.184	0.169	0.093	0.217	0.600	0.341
		+Fractal	**0.083**	**0.203**	**0.067**	**0.043**	**0.083**	**0.202**	**0.190**	**0.012**	**0.076**	**0.203**	**0.097**	**0.340**	**0.080**	**0.201**	**0.052**	**0.032**	**0.084**	**0.213**	**0.109**	**0.014**
	192	MSE	0.187	0.312	0.349	0.421	0.186	0.307	0.801	0.447	0.210	0.345	0.241	0.772	0.180	0.303	0.217	0.211	0.188	0.312	0.218	0.425
		+Fractal	**0.182**	**0.305**	**0.055**	**0.049**	**0.178**	**0.300**	**0.200**	**0.008**	**0.200**	**0.337**	**0.070**	**0.346**	**0.168**	**0.295**	**0.063**	**0.020**	**0.174**	**0.305**	**0.097**	**0.020**
	336	MSE	0.330	0.417	0.521	0.526	0.340	0.422	0.791	0.409	0.324	0.426	0.624	0.833	0.336	0.419	0.124	0.168	0.342	0.426	0.675	0.314
		+Fractal	**0.322**	**0.407**	**0.052**	**0.031**	**0.329**	**0.413**	**0.077**	**0.011**	**0.311**	**0.415**	**0.114**	**0.271**	**0.320**	**0.410**	**0.091**	**0.024**	**0.324**	**0.418**	**0.082**	**0.051**
	720	MSE	0.849	0.696	1.225	0.766	0.934	0.729	0.465	0.333	0.772	0.669	1.187	0.839	0.829	0.688	0.167	0.176	0.885	0.713	0.307	0.236
		+Fractal	**0.837**	**0.681**	**0.142**	**0.057**	**0.918**	**0.717**	**0.081**	**0.005**	**0.753**	**0.653**	**0.179**	**0.307**	**0.804**	**0.674**	**0.142**	**0.036**	**0.858**	**0.700**	**0.044**	**0.055**

Note: Values are reported separately for prediction lengths $T \in \{96, 192, 336, 720\}$ with input length 96. MSE and +Fractal denote the baseline and augmented results, respectively. For each dataset, backbone, prediction length, and metric, the better result is shown in **bold**, with a blue background for MSE/MAE and a red background for FSE/FFE; ties at the displayed precision are both highlighted. Lower is better for all metrics.

C.3 More Visualization of Fractal Structure on Real-world Time Series

