HARNESSING OSCILLATOR DYNAMICS FOR HIGH-FREQUENCY TIME SERIES REPRESENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

High-quality representation of high-frequency components remains challenging in multivariate time series forecasting (TSF), due to their noisy and non-stationary patterns. Inspired by coupled oscillator dynamics, we introduce a Kuramoto oscillator-based high-frequency refiner (KOR), which maps high-frequency signals into a multivariate oscillator system and extracts structured latent responses from coupling-constrained phase evolution. Its restoring dynamics suppress rapid micro-fluctuations, while the observed phase-locking and slipping regimes admit a local Adler-type interpretation through effective frequency mismatch. To control the residual energy processed by KOR and complement spectral details not fully retained by its macroscopic readout, we further develop KOR-centered frequency modeling framework (KFM), consisting of soft-gated frequency decomposition and spectral fusion. Extensive experiments demonstrate that KFM improves the large majority of evaluated backbone–dataset settings and achieves competitive performance. Code will be open sourced upon acceptance.

1 INTRODUCTION

Time series forecasting (TSF) is a fundamental problem in temporal data analysis, with broad applications spanning power systems (Li et al., 2022), retail (Li et al., 2025a), hospitality (Zhang et al., 2025b), and meteorology (Tang et al., 2026). Conventional TSF approaches involve statistical models such as autoregressive moving average (MA) models (Box et al., 2015), as well as classical machine learning methods including support vector machines and gradient boosting (Masini et al., 2023). Models such as multilayer perceptrons (MLPs), convolutional neural networks (CNNs), and Transformers model multivariate and long-range temporal dependencies.

Existing neural TSF models are generally effective at capturing low-frequency structures with significant periodicity, yet high-quality representation for high-frequency components remains a major concern (Eldele et al., 2024). Several studies extract high-frequency components based on MA (Wu et al., 2021), wavelet transforms (Murad et al., 2025), and frequency threshold (Eldele et al., 2024). As shown in Figure 1(a), neural networks like 1D CNNs (Stitsyuk & Choi, 2025) and MLPs (Murad et al., 2025) are employed to encode decomposed high-frequency sequences, as they excel at capturing local transients and mixing cross-step information, respectively. Yi et al. (2023) and Zhang et al. (2022) utilize frequency-domain MLPs and attention mechanisms to facilitate global frequency interactions. However, network-based methods often fail to differentiate the importance of various frequency components and overfit noisy patterns (Wang et al., 2025).

We draw inspiration from the Kuramoto oscillator model (Kuramoto, 2005; Acebrón et al., 2005), as shown in Figure 1(b). The high-frequency sequences are mapped into external forcing for the oscillator system via frozen random projections, producing latent representations governed by phase evolution trajectories. The restoring dynamics induced by inter-channel and intra-channel couplings attenuate rapid perturbations and retain slowly varying structure in the projected excitation. The effective mismatch $\Delta\omega$ between oscillator natural frequencies and the global effective frequency further provides a local Adler-type interpretation of the resulting phase difference trajectories $\sin(\delta(t))$ in latent oscillator space.

We design a **K**uramoto **O**scillator system-based high-frequency **R**efiner (**KOR**) for structured and noise-resilient dynamic patterns, and further develop **K**OR-centered **F**requency **M**odeling (**KFM**),

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.

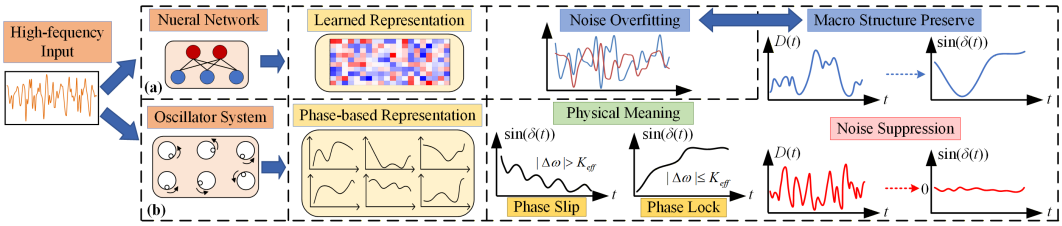


Figure 1: High-frequency sequence representation based on (a) neural networks and (b) the Kuramoto oscillator system, along with their respective characteristics.

a framework that integrates KOR with standard TSF backbones. Our main contributions are summarized as follows:

- We propose KOR, which adapts Kuramoto dynamics to multivariate TSF through intra- and inter-channel oscillator coupling, suppresses rapid micro-fluctuations via coupling-induced restoring dynamics, and provides a local Adler-type interpretation of latent phase responses through effective frequency mismatch.
- We develop KFM to integrate KOR with standard TSF backbones through soft-gated frequency decomposition, spectral fusion, and final representation fusion.
- Extensive experiments on real-world datasets demonstrate significant improvements over diverse TSF backbones, while ablation and module-replacement studies separate the contribution of KOR from complementary gains provided by remaining KFM components.

2 Related Work

2.1 Frequency Decomposition and Spectral Modeling in TSF

The complexity of real-world time series, characterized by coupled trends and high-frequency fluctuations, has motivated frequency-domain decomposition and modeling along three lines. First, component decomposition separates heterogeneous dynamics. Trend-seasonal decomposition based on MA (Wu et al., 2021; Chen et al., 2024; Behrouz et al., 2024; Qiu et al., 2025) or exponential MA (EMA) (Stitsyuk & Choi, 2025; Kong & Hong, 2026) acts as a low-pass filter, and has been extended to multi-scale subsequences (Wang et al., 2024), latent spaces (Tang & Zhang, 2025), and progressive filtering schemes (Huang et al., 2025) Low-pass and Butterworth zero-phase filters (Li et al., 2025b; Yin et al., 2025) and wavelet transforms (Yang et al., 2023) provide alternative decomposition bases. To achieve more adaptive decomposition, Xu et al. (2024) determine truncation via harmonic structure, Kou et al. (2025) use a median-frequency-based binary mask, while Wu et al. (2025b), Eldele et al. (2024), and An et al. (2026) learn parametric thresholds through MLPs, learnable scalars, and logarithmic frequency-band allocation, respectively.

Second, dual- and multi-branch architectures model the decomposed components with specific backbones. Zhang et al. (2025c) and Murad et al. (2025) assign dedicated predictors to wavelet coefficients at different resolutions (Zhang et al., 2025c; Murad et al., 2025). In trend-seasonal frameworks, stable low-frequency trends are captured by lightweight MLPs or polynomial models (Zhang et al., 2026; 2025a), whereas stochastic seasonal fluctuations are modeled with 1D convolutions or frequency-domain transformations (Stitsyuk & Choi, 2025; Zhang et al., 2022). Koopa further decouples time-varying and time-invariant dynamics into separate predictors (Liu et al., 2023).

Finally, some methods treat low-energy or high frequencies as noise and filter them out (Zhou et al., 2024; Wang et al., 2025; Zheng & Xie, 2025), with dynamic frequency selection via a spiking selector (Wu et al., 2025a) and adaptive filtering via Fourier derivative operators (Fan et al., 2024) Weak seasonal components are also assigned with low weight in representation fusion (Zhang et al., 2026). However, such deterministic filtering may discard informative high-frequency structures or ignore data heterogeneity, so Shibo et al. (2025) create shortcut to preserve high-frequency details.

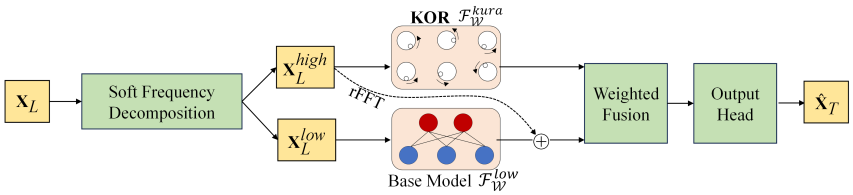


Figure 2: Overall pipeline of KFM, where $\oplus$ denotes the concatenation operation.

2.2 Application of Dynamical System in Deep Learning

Dynamical-system-based modeling has attracted attention in TSF. Neural differential equations learn continuous-time dynamics with external control paths for long-range dependency modeling (Oh et al., 2025), while Roth et al. (2025) impose Hamiltonian constraints to improve long-horizon stability. Koopman-based methods model non-stationary dynamics (Yang et al., 2024) and future trend evolution (Zhang et al., 2025b), including hierarchical formulations with global and local predictors (Liu et al., 2023). These dynamical TSF methods mainly target dependency modeling, trend extraction, or computational efficiency. In contrast, we adapt established Kuramoto dynamics to high-frequency TSF through task-specific signal-to-force mapping, phase readout, and inter-channel coupling.

Due to its phase locking and synchronization properties (Cooray, 2008), the Kuramoto model has been applied to model electroencephalogram signals characterized by multi-band oscillations (Ghosh, 2025), and interpret the oversmoothing phenomenon in graph neural networks. In diffusion models, Song et al. (2023) replace the standard forward noising process with Kuramoto-style phase synchronization. To the best of our knowledge, this is the first work to exploit Kuramoto dynamics for high-frequency representation in TSF.

3 Methodology

The objective of multivariate TSF is to predict the future target window $\mathbf{X}_T \in \mathbb{R}^{T\times C}$ with length T and C variates, given the historical look-back window $\mathbf{X}_L \in \mathbb{R}^{L\times C}$ with length L. Formally the task aims to learn a function $\mathcal{F}_{\mathcal{W}}$ which maps the historical observations to the future sequence $\hat{\mathbf{X}}_T$ through $\hat{\mathbf{X}}_T = \mathcal{F}_{\mathcal{W}}(\mathbf{X}_L)$. To mitigate distribution shifts observed in real-world time series, we apply instance-wise normalization (Kim et al., 2021) to the input sequence $\mathbf{X}_L$.

The normalized sequence, still denoted as $\mathbf{X}_L$ for brevity, is then fed into our KFM whose architecture is illustrated in Figure 2. After soft frequency decomposition, low-frequency $\mathbf{X}_L^{low}$ is fed into a conventional TSF base model $\mathcal{F}_{\mathcal{W}}^{low}$, while Kuramoto high-frequency module $\mathcal{F}_{\mathcal{W}}^{kura}$ models high-frequency $\mathbf{X}_L^{high}$. After low-frequency spectrum is mixed with representation from $\mathcal{F}_{\mathcal{W}}^{low}$, a learnable gated fusion of final dual-flow representations is performed.

3.1 Frequency Decomposition

To exploit conventional TSF models on smooth, periodic components, we decouple $\mathbf{X}_L^{low}$ and $\mathbf{X}_L^{high}$ from $\mathbf{X}_L$ to form a dual-flow architecture. $\mathbf{X}_L$ is transformed into the frequency domain via the real fast Fourier transform (rFFT) and denoted as $\mathbf{Z} \in \mathbb{C}^{L_f\times C}$, where $L_f = \lfloor L/2 \rfloor + 1$ is the number of frequency components. Then the Top-K lowest-frequency components are retained as the initial low-frequency representation $\mathbf{Z}_{top}^{low} \in \mathbb{C}^{K\times C}$.

Instead of discarding the remaining high-frequency spectrum $\mathbf{Z}_{res}^{high} \in \mathbb{R}^{(L_f-K)\times C}$, we introduce a channel-independent scoring network to generate a mask over the spectrum:

$$\mathbf{m}_c = \sigma(\text{MLP}(|\mathbf{Z}_{res}^{high,(c)}|)) \in (0,1)^{L_f-K}, \tag{1}$$

where $\sigma(\cdot)$ is the Sigmoid function. The selected high-frequency spectrum is obtained by applying the mask: $\tilde{\mathbf{Z}}_{sel}^{high} = \mathbf{Z}_{res}^{high} \odot \mathbf{M}$, where $\mathbf{M}$ is the stacked mask matrix for all channels. We reconstruct

the frequency-domain representations for the two decoupled branches:

$$\mathbf{Z}^{low} = \text{Concat}(\mathbf{Z}^{low}_{top}, \tilde{\mathbf{Z}}^{high}_{sel}), \mathbf{Z}^{high} = \mathbf{Z} - \mathbf{Z}^{low}. \tag{2}$$

The inverse rFFT (irFFT) operation transforms $\mathbf{Z}^{low}$ and $\mathbf{Z}^{high}$ back to the time domain, yielding the low-frequency base component $\mathbf{X}^{low}_L \in \mathbb{R}^{L\times C}$ and the high-frequency dynamic component $\mathbf{X}^{high}_L \in \mathbb{R}^{L\times C}$, respectively.

The soft mask conservatively routes informative residual high-frequency energy into the low-frequency flow. To prevent unreliable components from disrupting $\mathcal{F}^{low}_{\mathcal{W}}$ initially, we set the final MLP bias to a negative scalar $-\beta$. Consequently, the module begins as a strict Top-K hard filter and progressively learns frequency-wise selection that a fixed or shifted threshold cannot express.

3.2 High-Frequency Representation via Kuramoto Dynamics

The low-frequency sequence $\mathbf{X}^{low}_L$ is fed into $\mathcal{F}^{low}_{\mathcal{W}}$ to extract a hidden representation $\mathbf{H}^{low} \in \mathbb{R}^{C\times d}$. For the high-frequency branch, we use the Kuramoto oscillator system (Acebrón et al., 2005).

The Kuramoto model is a non-linear dynamical system describing phase synchronization in networks of coupled oscillators. For a system of N oscillators, the phase θ_i of the i-th oscillator evolves in continuous time t as:

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i), \tag{3}$$

where ω_i is the natural frequency governing the free evolution rate, and K is the coupling strength. In complex physical networks, sparse interactions between oscillators are governed by specific topological structures, naturally modeled by an adjacency matrix $\mathbf{A}$. Furthermore, if the open dynamical systems constantly interact with environmental interfere, an external continuous stimulus $\mathbf{D}(t)$ with a control coefficient γ is added to the system:

$$\frac{d\theta_i}{dt} = \underbrace{\omega_i}_{\text{Natural Frequency}} + \underbrace{\frac{K}{N}\sum_{j=1}^{N}\mathbf{A}_{ij}\sin(\theta_j - \theta_i)}_{\text{Coupling Force}} + \underbrace{\mathbf{D}_i(t)}_{\text{External Drive}}. \tag{4}$$

Motivated by this, as illustrated in Figure 3(a), we first map the multivariate time series into a high-dimensional virtual variable space. Within each virtual channel, we then segment the sequence and project it onto the integration step $s \in \{1, \ldots, S\}$ to act as the external forcing $\mathbf{D}^{(s)}$ of the **O**scillator **S**ystem (**OS**). To isolate the net effect of the external drive from the intrinsic dynamics, we then record the phase difference of the n-th oscillator within the v-th virtual variable channel between its free evolution and forced dynamics:

$$\delta^{(s)}_{v,n} = \theta^{driven,(s)}_{v,n} - \theta^{free,(s)}_{v,n}. \tag{5}$$

We transform these phase difference trajectories into the following macroscopic representation:

$$h^{macro}_{v,n} = \frac{1}{S}\sum_{s=1}^{S}\sin(\delta^{(s)}_{v,n}). \tag{6}$$

We assemble the macroscopic phase differences of the oscillators across all channels into $\mathbf{H}^{macro} \in \mathbb{R}^{V,N}$. To align the dimensions with the low-frequency representation, $\mathbf{H}^{macro}$ is mapped to $\mathbf{Z}^{high}_{kura}$ by frozen linear projections.

This representation admits a locally tractable dynamical interpretation. Suppose that the external excitation contains a dominant frequency component Ω_D, whose phase-relevant action interacts with the coupling force to form a global effective force with strength K_{eff} and phase Φ_{eff}. Under the local effective-harmonic approximation detailed in Appendix C, defining ϕ as the deviation of the oscillator phase θ_n from Φ_{eff} and $\Delta\omega$ as the effective mismatch between the oscillator's natural frequency and the local rotational velocity Ω_{eff} of Φ_{eff} yields an Adler-type equation (Pikovsky et al., 1985):

$$\frac{d\phi}{dt} = \Delta\omega - K_{eff}\sin(\phi). \tag{7}$$

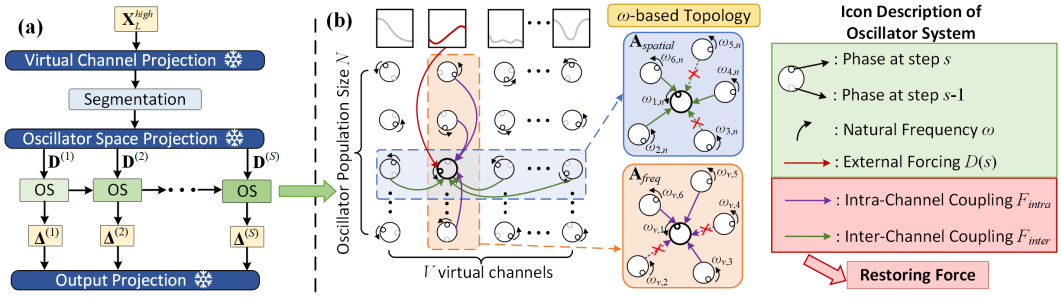


Figure 3: Architecture of (a) KOR and its internal (b) OS structure.

For locally quasi-stationary K_{eff} and Ω_{eff}, $|\Delta\omega| \leq K_{eff}$ admits a steady-state solution, signifying local phase locking. In this regime, $\sin(\delta_n(t))$ can vary slowly and approach a constant when the driven and free phase velocities align. Conversely, $|\Delta\omega| > K_{eff}$ produces a local phase-slipping regime. We attribute the continuous but highly non-uniform oscillation of $\sin(\delta_n(t))$ in this regime to the dynamical bottleneck effect (Strogatz, 2001). Within this local interpretation, the effective frequency mismatch $\Delta\omega$ explains $\sin(\delta_n(t))$ as a latent response in the projected oscillator space rather than a physical phase directly associated with an original time-series variable.

3.2.1 Formulation and Evolution of OS

As illustrated in Figure 3(b), OS is initialized with V virtual channels and N oscillators per channel. The phases of all oscillators within OS constitute the matrix $\mathbf{\Theta} \in \mathbb{R}^{V\times N}$, and the elements in the initial phase matrix $\mathbf{\Theta}^{(0)}$ are randomly generated within $[0, 2\pi)$. Each oscillator is assigned an fixed natural frequency, forming a frequency matrix $\mathbf{\Omega} \in \mathbb{R}^{V\times N}$ uniformly sampled from $[\omega_{min}, \omega_{max}]$.

For each integration step, $\mathbf{X}_L^{high}$ acts as an external drive $\mathbf{D}^{(s)} \in \mathbb{R}^{V\times N}$ on OS through projection and segmentation detailed in Subsubsection 3.2.2. The discrete-time phase evolution of the n-th oscillator in the v-th virtual channel is updated via the following form modified from Equation 4:

$$\theta_{v,n}^{(s)} = \theta_{v,n}^{(s-1)} + \Delta t\Big(\omega_{v,n} + F_{intra} + F_{inter} + \gamma \mathbf{D}_{v,n}^{(s)}\Big), \tag{8}$$

where Δt is the integration step size, and γ is a predefined scaling factor for the external drive.

The oscillator coupling term in Equation 8 is divided into intra- and inter-channel components, formulated as:

$$F_{intra} = \frac{K_{intra}}{N}\sum_{j=1}^{N} \mathbf{A}_{freq}^{j,n} \sin(\theta_{v,j}^{(s-1)} - \theta_{v,n}^{(s-1)}), \tag{9}$$

$$F_{inter} = \frac{K_{inter}}{V}\sum_{u=1}^{V} \mathbf{A}_{spatial}^{u,v} \sin(\theta_{u,n}^{(s-1)} - \theta_{v,n}^{(s-1)}), \tag{10}$$

where K_{intra} and K_{inter} are non-negative hyperparameters controlling the local and global synchronization coupling strengths, respectively.

The topological interactions are governed by the predefined matrices $\mathbf{A}_{freq} \in \{0,1\}^{N\times N}$ and $\mathbf{A}_{spatial} \in \{0,1\}^{V\times V}$, where $\mathbf{A}_{freq}$ is shared across virtual channels and $\mathbf{A}_{spatial}$ is shared across oscillator indices. To capture the proximity resonance, the existence of an edge is modeled via a Bernoulli distribution $\mathcal{B}$ parameterized by the exponential negative distance of their natural frequencies measured in per-step phase increments (scaled by Δt):

$$\mathbf{A}_{freq}^{i,j} \sim \mathcal{B}\left(\exp\left(-\frac{\Delta t}{V}\sum_{q=1}^{V}|\omega_{q,i} - \omega_{q,j}|\right)\right), \mathbf{A}_{spatial}^{u,v} \sim \mathcal{B}\left(\exp\left(-\frac{\Delta t}{N}\sum_{n=1}^{N}|\omega_{u,n} - \omega_{v,n}|\right)\right), \tag{11}$$

where each sampled edge is mirrored, so both adjacency matrices are symmetric.

The first-order expansion of Equations 9 and 10 yields a cosine-weighted Laplacian $\mathbf{L}(t)$ and the restoring term $-\mathbf{L}(t)\boldsymbol{\Delta}(t)$. On locally cohesive intervals where the magnitudes of the wrapped free-state phase differences on active edges are smaller than $\pi/2$, the symmetric effective edge weights are non-negative, making $\mathbf{L}(t)$ positive semidefinite. Its zero mode represents a common phase shift, whereas its positive modes contract oscillator disagreement transverse to this common-phase direction. Under low-frequency forcing, this transverse restoring effect confines the disagreement response to a bounded oscillatory regime, which is inefficient for abstracting the macroscopic periodic structure. Under high-frequency forcing, it attenuates the transverse contribution of rapid micro-fluctuations $N_{fast}(t)$ as $|\Omega| \to \infty$, while preserving more slowly varying shared responses. The derivation processes are detailed in Appendices B.2 to B.4.

Equivalently, around a locally cohesive free-state manifold, the linearized internal forcing field is negative semidefinite and strictly contractive only on its stable transverse modes (Appendix D.2). The resulting leaky integration attenuates instantaneous disagreement noise while retaining structured shared responses, and a larger transverse contraction rate lowers the variance contributed by these stable modes (Appendix D.3).

3.2.2 Signal-to-Force Mapping

During the input stage of KOR, we linearly project $\mathbf{X}_L^{high} \in \mathbb{R}^{L\times C}$ from the original C real-world channels into V virtual latent channels, yielding $\mathbf{X}_V^{high} \in \mathbb{R}^{L\times V}$. This projection serves a dual purpose. When $V > C$, expanding to a higher-dimensional latent space provides additional capacity for representing high-order entangled phase interactions. Conversely, when $V \ll C$, compressing the original channels reduces computational complexity and mitigates spurious correlations within the oscillator space.

To alleviate the computational overhead of executing L iterative steps, we avoid using $\mathbf{X}_V^{high}$ directly as an L-step external stimulus. Instead, we partition it into S non-overlapping segments of length P and jointly project each segment $\mathbf{z}^{(s)} \in \mathbb{R}^P$ into the N-dimensional oscillator space. This segment-level mapping is not point-wise subsampling, because every value within a segment contributes to the projected excitation. However, random projection does not always guarantee preservation of individual Fourier components, phase structure, or immunity to aliasing. Appendix B.1 instead analyzes the more limited property that Gaussian projection approximately preserves geometric differences between adjacent segments (Johnson et al., 1984; Dasgupta & Gupta, 2003), which provides indirect intuition for retaining rapid temporal variation.

The input-mapping and output-projection linear layers of KOR are randomly initialized from a scaled-variance Gaussian and kept frozen during training, rendering KOR a learning-free module that only aligns dimensions with the virtual-channel and oscillator space rather than learning features. This preserves observation invariance by avoiding learned biases (Rahimi & Recht, 2007) and mitigates collapse to trivial solutions. The initialization and forward pass are provided in Algorithm 1 of Appendix A.1.

3.3 Spectral Mixing and Gated Dual-Branch Fusion

To preserve information in the residual spectral energy, we explicitly extract both the real and imaginary parts of the frequency-domain high-frequency residual, $\mathbf{Z}^{high}$ directly concatenating them with low-frequency representation $\mathbf{H}^{low}$ through a shortcut connection, since base models generally lack the capacity to encode high-frequency spectra and tend to overfit. This hybrid feature is subsequently sent to a learnable linear layer, aligning the output dimension to $2L$:

$$\mathbf{Z}_{fused}^{low} = \text{Linear}\Big(\text{Concat}\big(\text{Re}(\mathbf{Z}^{high})^\top, \text{Im}(\mathbf{Z}^{high})^\top, \mathbf{H}^{low}\big)\Big) \in \mathbb{R}^{C\times 2L}. \tag{12}$$

This shortcut provides raw frequency-domain information and compensates for KOR's representation from time-domain inspiration.

A weighted fusion for dual branch representations is performed through a learnable weight w.

$$\mathbf{Z}_{final} = \text{LayerNorm}\big(\sigma(w)\mathbf{Z}_{kura}^{high} + (1-\sigma(w))\mathbf{Z}_{fused}^{low}\big). \tag{13}$$

Table 1: KFM performance against baselines at $L = 96$ and $T = 336$. All hyperparameters are extensively tuned except for a universally fixed batch size. The best results are highlighted in bold, and the second best are underlined.

Models	**KFM (Ours)**		FilterTS (2025)		CFPT (2025)		xPatch (2025)		Affirm (2025)		FITS (2024)		VCformer (2024)		Koopa (2023)	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.463**	**0.441**	0.467	0.445	0.471	0.451	0.465	0.442	0.477	0.459	0.468	0.452	0.479	0.455	0.480	0.448
ETTh2	**0.406**	**0.417**	0.411	0.423	0.413	0.429	0.409	0.418	0.429	0.433	0.409	0.423	0.419	0.429	0.411	0.424
ETTm1	0.400	**0.395**	0.402	0.409	**0.393**	0.407	0.404	0.408	0.411	0.402	0.409	0.401	0.405	0.409	0.404	0.404
ETTm2	0.296	**0.333**	0.300	0.391	**0.294**	0.338	0.299	0.342	0.310	0.344	0.304	0.365	0.309	0.349	0.301	0.394
Traffic	**0.432**	**0.268**	0.468	0.309	0.505	0.321	0.439	0.281	0.438	0.278	0.455	0.291	0.481	0.319	0.477	0.311
Electricity	**0.171**	**0.264**	0.179	0.277	0.173	0.271	0.172	0.268	0.173	0.275	0.180	0.274	0.181	0.272	0.183	0.273
Weather	0.266	**0.288**	0.270	0.297	**0.264**	0.292	0.269	0.293	0.267	0.289	0.277	0.296	0.280	0.299	0.265	0.296
Exchange	**0.334**	**0.412**	0.344	0.420	0.351	0.428	0.365	0.436	0.338	0.423	0.363	0.428	0.338	0.417	0.335	0.417

Table 2: Performance before and after replacing the seasonal branch (or the high-frequency detail branch) with KOR under $L = 336$ and $T = 336$.

Models	interPDN		+KOR		xPatch		+KOR		WPMixer		+KOR	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.419	0.422	**0.411**	**0.416**	0.437	0.430	**0.428**	**0.423**	0.402	0.418	**0.400**	**0.415**
ETTh2	0.329	0.378	**0.321**	**0.375**	0.340	0.384	**0.335**	**0.381**	**0.335**	**0.393**	0.337	0.394
ETTm1	0.352	0.375	**0.348**	**0.369**	0.365	0.381	**0.364**	**0.376**	0.367	0.391	**0.363**	**0.389**
ETTm2	0.271	0.317	**0.265**	**0.313**	0.275	0.322	**0.270**	**0.320**	0.275	0.330	**0.272**	**0.326**
Weather	0.224	0.261	**0.224**	**0.260**	0.226	0.264	**0.222**	**0.262**	0.241	0.276	**0.237**	**0.271**
Traffic	**0.392**	0.243	0.393	**0.240**	0.396	0.248	**0.393**	**0.246**	**0.391**	0.273	0.392	**0.269**
Electricity	0.154	0.245	**0.150**	**0.241**	0.162	0.251	**0.155**	**0.244**	0.167	0.262	**0.163**	**0.259**

This joint encoding can be utilized for the linear probing of $\hat{\mathbf{X}}_T$ after which it is denormalized back to the original data space. We employ the L1 distance as KFM's objective function.

4 Results

4.1 Main Results

To verify KFM, we conduct experiments on nine public benchmarks: ETT (h1, h2, m1, m2), Traffic, Electricity, Weather, Exchange, and Illness (Wu et al., 2021). Dataset and device details are provided in Appendices A.2 and A.3. We utilize the iTransformer as base model $\mathcal{F}_{\mathcal{W}}^{low}$ within KFM architecture to conduct a comprehensive comparison against baseline methods, involving frequency decomposition from Subsection 2.1 such as FilterTS and CFPT, alongside VCformer and Koopa from Subsection 2.2 which incorporate Koopman dynamical models. In Table 1, the MAEs of KFM achieve state of the art (SOTA) performance across all evaluated datasets. Although KFM fails to reach SOTA performance in MSE metric on ETTm1, ETTm2, and Weather datasets, the discrepancy between its MSEs and the best is strictly constrained to approximately 1%. Appendix Table 11 further extends this direct comparison to four prediction horizons on ETTh1 and ETTm1, where KFM is best or tied best in 8 of the 10 tasks.

To further validate the effectiveness of KOR module, we modify several multi-branch TSF models that utilize frequency decomposition. We replace the modules for high-frequency or seasonal components representation with KOR. In Table 2, we replace the CNN-based seasonal branches in interPDN (Kong & Hong, 2026) and xPatch (Stitsyuk & Choi, 2025), along with the MLP-based module for the finest detail coefficients in WPMixer (Murad et al., 2025), with KOR. All other hyperparameters and objective functions remain identical to the original models. After replacement, the MSE values for interPDN, xPatch, and WPMixer decrease by 1.20%, 1.59%, and 0.91%, respectively. These results indicate that even when detached from KFM framework, KOR can serve as an effective and readily integrable high-frequency module.

Six representative TSF models (PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024), CATS (Kim et al., 2024), PDF (Dai et al., 2024), TimesNet (Wu et al., 2023), and TSMixer (Chen et al.,

Table 3: Average MSE of baselines and their KFM variants. Results are averaged over four horizons: $T \in 24, 36, 48, 60$ for Illness and $T \in 96, 192, 336, 720$ for others. The look-back window is $L = 104$ for Illness. For other datasets, $L = 336$ for PatchTST, CATS, and PDF, and $L = 96$ for iTransformer, TSMixer, and TimesNet. These are pre-specified paired settings rather than best-performing look-back selections for KFM. Avg. Δ (%) is the arithmetic mean of relative error reductions for each dataset. Better results in each group are bolded.

Dataset	PatchTST		iTransformer		CATS		PDF		TimesNet		TSMixer		Avg.
	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Δ (%)
ETTh1	0.417	**0.396**	0.456	**0.435**	0.412	**0.401**	0.414	**0.395**	0.460	**0.432**	0.463	**0.435**	4.84
ETTh2	0.334	**0.329**	0.385	**0.372**	0.342	**0.331**	0.343	**0.329**	0.407	**0.378**	0.407	**0.377**	4.44
ETTm1	0.357	**0.348**	0.409	**0.389**	0.359	**0.354**	0.346	**0.343**	0.417	**0.384**	0.395	**0.384**	3.39
ETTm2	0.260	**0.248**	0.293	**0.278**	0.268	**0.261**	0.258	**0.251**	0.296	**0.284**	0.288	**0.280**	3.65
Traffic	0.410	**0.404**	**0.426**	0.438	0.414	**0.410**	**0.393**	0.394	0.625	**0.603**	0.526	**0.501**	1.27
Electricity	0.168	**0.161**	0.179	**0.171**	0.171	**0.164**	0.166	**0.159**	0.206	**0.186**	0.209	**0.186**	6.28
Weather	0.238	**0.227**	0.261	**0.250**	0.231	**0.227**	0.231	**0.225**	0.259	**0.248**	0.248	**0.240**	3.44
Exchange	0.412	**0.375**	0.374	**0.369**	0.372	**0.320**	0.432	**0.395**	0.407	**0.396**	**0.367**	0.368	5.88
Illness	1.748	**1.626**	2.318	**1.663**	4.461	**2.796**	2.265	**1.753**	2.140	**1.945**	2.921	**1.624**	24.78

Table 4: Average MAE of baselines and their KFM variants. Other experimental settings and the definition of Avg. Δ are identical to Table 3.

Dataset	PatchTST		iTransformer		CATS		PDF		TimesNet		TSMixer		Avg.
	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Δ (%)
ETTh1	0.429	**0.417**	0.447	**0.432**	0.425	**0.419**	0.424	**0.417**	0.455	**0.428**	0.453	**0.425**	3.56
ETTh2	0.378	**0.374**	0.405	**0.394**	0.384	**0.375**	0.386	**0.375**	0.421	**0.399**	0.420	**0.398**	3.24
ETTm1	0.376	**0.368**	0.411	**0.386**	0.380	**0.374**	0.374	**0.366**	0.420	**0.388**	0.399	**0.385**	3.84
ETTm2	0.312	**0.305**	0.333	**0.319**	0.323	**0.315**	0.313	**0.306**	0.333	**0.322**	0.332	**0.320**	3.01
Traffic	0.263	**0.249**	0.281	**0.271**	0.266	**0.262**	0.256	**0.243**	0.332	**0.324**	0.337	**0.301**	4.76
Electricity	0.252	**0.247**	0.270	**0.260**	0.255	**0.251**	0.252	**0.247**	0.302	**0.284**	0.298	**0.276**	3.76
Weather	0.265	**0.253**	0.280	**0.268**	0.264	**0.259**	0.260	**0.253**	0.286	**0.272**	0.278	**0.265**	3.83
Exchange	0.437	**0.412**	0.413	**0.408**	0.407	**0.386**	0.441	**0.421**	0.438	**0.420**	0.408	**0.406**	3.54
Illness	0.866	**0.821**	1.039	**0.812**	1.497	**1.107**	1.051	**0.849**	0.986	**0.880**	1.169	**0.804**	19.05

2023)) are selected as baselines, whose architectures span Transformer, CNN, and MLP (detailed descriptions in Appendix A.5). By integrating these models into KFM as the low-frequency module $\mathcal{F}_{\mathcal{W}}^{low}$, we report their performance improvements in Tables 3 and 4. For every (Base, +KFM) pair, we use the same predefined backbone-specific look-back window rather than selecting L separately for KFM. Appendix E further evaluates robustness across multiple look-back lengths. All baseline entries in Tables 1, 2, 3, and 4 are our reproductions from official repositories. Paired variants use identical backbone parameter settings, and Tables 3–4 report mean error of five runs. Detailed protocols, full results, and standard deviations are provided in Appendices A.4, G, and H.

Across the 108 dataset–backbone–metric comparisons in Tables 3 and 4, KFM improves 105 cases, including 51 of 54 MSE and all 54 MAE comparisons. The three dataset-averaged regressions are confined to MSE and account for 2.78% of the comparisons, indicating broad but not universal gains. For example, on TimesNet, MSE and MAE across the four ETT datasets decrease by 6.27% and 5.44% on average, with nearly 10% improvement on Electricity. For CATS, MAE is reduced by over 1.4% on ETT, Traffic, and Weather. Dataset–backbone-specific reduction ratios are reported in Appendix Table 12. Visualizations of some predicted sequences and their comparisons with the ground truth are illustrated in Appendix J to intuitively demonstrate the improvements of baseline performance achieved by KFM.

4.2 Ablation Study and Hyperparameter Analysis

Table 5 separates the contributions of the three frequency-related components in KFM. Removing KOR causes the largest overall degradation, with several error metrics increasing by more than 5%, indicating that oscillator-based high-frequency representation is the principal source of improvement within this configuration. Spectral fusion provides additional gains exceeding 2.5% on ETTm1 and Electricity by supplying residual amplitude and phase information through a complementary

Table 5: Ablation of the frequency-related modules and the internal modules in OS, with the best results highlighted in bold and the second best underlined.

Dataset	ETTh1		ETTm1		Weather		Electricity		Traffic		Illness	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Full Model	**0.432**	**0.428**	0.384	**0.388**	**0.248**	**0.272**	**0.186**	**0.284**	0.603	**0.324**	**1.945**	**0.880**
w/o Eq.1 Soft Gating Mask	0.437	0.435	**0.383**	0.389	0.252	0.276	0.190	0.289	0.607	0.326	1.997	0.948
w/o Eq.12 Spectral Fusion	0.433	0.432	0.394	0.395	0.249	0.274	0.194	0.291	0.609	0.328	1.965	0.914
w/o KOR	0.447	0.445	0.408	0.410	0.253	0.281	0.196	0.294	0.618	0.328	2.001	0.932
w/o $\mathbf{A}_{freq}$&$\mathbf{A}_{spatial}$	0.435	0.432	0.388	0.392	0.250	0.272	0.189	0.290	0.610	0.326	1.955	0.909
w/o F_{inter}	0.439	0.438	0.386	0.389	0.252	0.273	0.192	0.287	**0.602**	0.324	1.959	0.901
TimesNet (baseline)	0.460	0.455	0.417	0.420	0.259	0.286	0.206	0.302	0.625	0.332	2.140	0.986

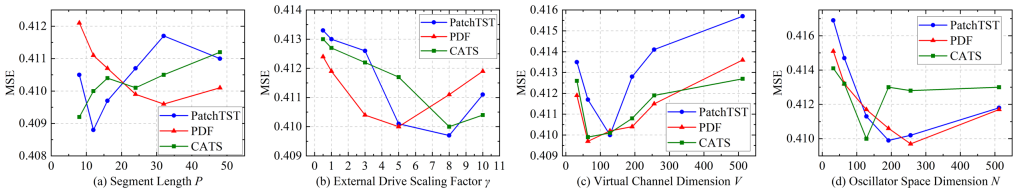


Figure 4: Sensitivity analysis of hyperparameters in KOR on ETTh1 when $L = 336$ and $T = 336$.

bypass. Soft gating yields a smaller average MSE reduction of 1.31%, consistent with its role as a conservative input-routing mechanism initialized from the Top-K hard threshold. These results suggest that KOR provides the largest individual contribution, while the complete performance gain arises from its interaction with decomposition and spectral fusion.

Removing either the inter-channel coupling term F_{inter} or the oscillator adjacency matrices $\mathbf{A}_{freq}, \mathbf{A}_{spatial}$ increases the error by 1%. The sparse structure of $\mathbf{A}$ promotes frequency-local synchronization while preventing global phase homogenization, whereas F_{inter} captures dynamic cross-variable dependencies and mitigates phase divergence, playing a complementary role. Additional analysis, including complete combination-wise ablation, module substitution, and visualizations of soft-gated high-frequency extraction, is provided in Appendices I.1, I.2, and I.3, respectively.

Figure 4 shows the MSE sensitivity to segment length P, driving scale γ, virtual channel number V, and oscillator number N per channel under three $\mathcal{F}^{low}\mathcal{W}$ on ETTh1. A large P induces numerical instability and trajectory divergence, while an excessively small P, despite improving performance (e.g., with CATS), incurs high computational cost. Appendix F shows that P provides an explicit accuracy–efficiency trade-off, and that KFM attains both better accuracy and lower training, inference, and memory costs than recent heavier TSF models. With $K_{inter} = K_{intra} = 1.5$, γ performs best in the range $[5, 8]$. Too small values lead to coupling-dominated dynamics with $\mathbf{H}^{macro} \approx 0$, whereas overly large values reduce the system to a trivial signal transmission regime. The virtual channel size should satisfy $V < 128$, while the oscillator dimension should be $N > 128$, ensuring sufficient spectral coverage and avoiding sparsity in the natural frequency distribution of oscillators.

5 Conclusion

This paper introduces KOR, a high-frequency time series representation module that adapts Kuramoto oscillator dynamics to multivariate TSF through intra- and inter-channel coupling. Its coupling-induced restoring dynamics suppress rapid micro-fluctuations, while the observed phase-locking and slipping regimes admit a local Adler-type interpretation through effective frequency mismatch. We further develop KFM to integrate KOR with standard forecasting backbones through soft-gated frequency decomposition, spectral fusion, and final representation fusion. Experiments on multiple benchmark datasets demonstrate that KFM improves the large majority of evaluated forecasting settings and achieves competitive and often state-of-the-art performance. Future work may adapt KOR to more advanced Fourier- or wavelet-based decomposition methods and reduce the iterative simulation cost for forecasting with ultra-long contexts.

AI USE STATEMENT

In this work, we did not use generative AI tools for any task requiring disclosure: we did not use them to generate synthetic data sets, to help develop theoretical models or conceptual frameworks, to formulate mathematical claims or write proofs, to propose or refine hypotheses, to design or provide feedback on the research methodology or experiments, to implement methods, to assist with translation, to clean or reformat the data sets, to support qualitative or thematic data analysis, or to interpret results. Additionally, we used generative AI tools to draft parts of the manuscript and to edit the text to improve clarity and readability. All method design, experiments, and the analysis and organization of results were performed by the authors. We have reviewed all AI-assisted work, and every AI-generated suggestion was verified for correctness by the authors. We take full responsibility for the final content of this work, including all text, claims, and artifacts produced with or without the aid of generative AI.

REFERENCES

Juan A Acebrón, Luis L Bonilla, Conrad J Pérez Vicente, Félix Ritort, and Renato Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of modern physics*, 77(1):137–185, 2005.

Zhongde An et al. FreDN: Spectral disentanglement for time series forecasting via learnable frequency decomposition. In *Proc. AAAI Conf. Artif. Intell.*, 2026.

Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics reports*, 469(3):93–153, 2008.

Ali Behrouz, Michele Santacatterina, and Ramin Zabih. Chimera: Effectively modeling multivariate time series with 2-dimensional state space models. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 119886–119918, 2024.

George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

Peng Chen et al. Pathformer: Multi-scale Transformers with adaptive pathways for time series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2024.

Si-An Chen, Chun-Liang Li, Nate Yoder, Sercan O. Arik, and Tomas Pfister. TSMixer: An all-MLP architecture for time series forecasting. *Trans. Mach. Learn. Res.*, 2023.

G. Cooray. The kuramoto model. Technical Report 2008:23, Department of Mathematics, Uppsala University, Uppsala, Sweden, 2008.

Tao Dai et al. Periodicity decoupling framework for long-term series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2024.

Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. TSLANet: Rethinking Transformers for time series representation learning. In *Proc. Int. Conf. Mach. Learn.*, 2024.

Wei Fan, Kun Yi, Hangting Ye, Zhiyuan Ning, Qi Zhang, and Ning An. Deep frequency derivative learning for non-stationary time series forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 3944–3952, 2024.

Sayan Ghosh. Deep coupled kuramoto oscillatory neural network (dckonn): A biologically inspired deep neural model for eeg signal analysis. *bioRxiv*, pp. 2025–09, 2025.

Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. TimeKAN: KAN-based frequency decomposition learning architecture for long-term time series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2025.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with Gumbel-Softmax. In *Proc. Int. Conf. Learn. Representations*, 2017.

William B Johnson, Joram Lindenstrauss, et al. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

Dongbin Kim, Jinseong Park, Jaewook Lee, and Hoki Kim. Are self-attentions effective for time series forecasting? In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 114180–114209, 2024.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *Proc. Int. Conf. Learn. Representations*, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. Int. Conf. Learn. Representations*, 2015.

Linghao Kong and Xiaopeng Hong. Time series forecasting via direct per-step probability distribution modeling. In *Proc. AAAI Conf. Artif. Intell.*, 2026.

Feifei Kou et al. CFPT: Empowering time series forecasting through cross-frequency interaction and periodic-aware timestamp modeling. In *Proc. Int. Conf. Mach. Learn.*, 2025.

Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International symposium on mathematical problems in theoretical physics: January 23–29, 1975, kyoto university, kyoto/Japan*, pp. 420–422. Springer, 2005.

Fan Li, Guoxuan Wang, Huiyu Chu, Dawei Cheng, and Xiaoyang Wang. Pcan: A pandemic-compatible attentive neural network for retail sales forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 9232–9240, 2025a.

Ruikun Li, Dai Shi, Ye Xiao, and Junbin Gao. UFGTime: Mining intertwined dependencies in multivariate time series via an efficient pure graph approach. *Proc. VLDB Endow.*, 18(9):3175–3188, 2025b.

Wendi Li, Xiao Yang, Weiqing Liu, Yingce Xia, and Jiang Bian. DDG-DA: Data distribution generation for predictable concept drift adaptation. In *Proc. AAAI Conf. Artif. Intell.*, pp. 4092–4100, 2022.

Xinyu Li, Yuchen Luo, Hao Wang, Haoxuan Li, Liuhua Peng, Feng Liu, Yandong Guo, Kun Zhang, and Mingming Gong. Towards accurate time series forecasting via implicit decoding. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 38, pp. 10003–10031, 2026.

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with Koopman predictors. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 12271–12290, 2023.

Yong Liu et al. iTransformer: Inverted Transformers are effective for time series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2024.

Ricardo P. Masini, Marcelo C. Medeiros, and Eduardo F. Mendes. Machine learning advances for time series forecasting. *J. Econ. Surveys*, 37(1):76–111, 2023.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. WPMixer: Efficient multi-resolution mixing for long-term time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 19581–19588, 2025.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with Transformers. In *Proc. Int. Conf. Learn. Representations*, 2023.

YongKyung Oh, Seungsu Kam, Jonghun Lee, Dong-Young Lim, Sungil Kim, and Alex Bui. Comprehensive review of neural differential equations for time series analysis. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 10631–10621, 2025.

Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. *Proc. Adv. Neural Inf. Process. Syst.*, 32, 2019.

Arkady Pikovsky, Michael Rosenblum, Jürgen Kurths, and A Synchronization. *A universal concept in nonlinear sciences*. Cambridge University Press, Princeton, 1985.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. DUET: Dual clustering enhanced multivariate time series forecasting. In *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 1185–1196, 2025.

Ali Rahimi and Benjamin Recht. Random features for large-scale kernel machines. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 20, 2007.

Fabian J Roth, Dominik K Klein, Maximilian Kannapinn, Jan Peters, and Oliver Weeger. Stable port-hamiltonian neural networks. *arXiv preprint arXiv:2502.02480*, 2025.

Feng Shibo, Wanjin Feng, Xingyu Gao, Peilin Zhao, and Zhiqi Shen. TS-LIF: A temporal segment spiking neuron network for time series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2025.

Yue Song et al. Kuramoto orientation diffusion models. In *Proc. Adv. Neural Inf. Process. Syst.*, 2023.

Artyom Stitsyuk and Jaesik Choi. xPatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition. In *Proc. AAAI Conf. Artif. Intell.*, pp. 20601–20609, 2025.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering (studies in nonlinearity)*, volume 1. Westview press, 2001.

Dahao Tang, Nan Yang, Yanli Li, Zhiyu Zhu, Zhibo Jin, and Dong Yuan. Optimal look-back horizon for time series forecasting in federated learning. In *Proc. AAAI Conf. Artif. Intell.*, pp. 25823–25830, 2026.

Peiwang Tang and Weitai Zhang. Unlocking the power of patch: Patch-based MLP for long-term time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 12640–12648, 2025.

Suxin Tong and Jingling Yuan. Efficiently enhancing long-term series forecasting via ultra-long lookback windows. In *Proc. AAAI Conf. Artif. Intell.*, volume 39, pp. 20912–20920, 2025.

Shiyu Wang et al. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *Proc. Int. Conf. Learn. Representations*, 2024.

Yulong Wang, Yushuo Liu, Xiaoyi Duan, and Kai Wang. FilterTs: Comprehensive frequency filtering for multivariate time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 21375–21383, 2025.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *Proc. Int. Conf. Learn. Representations*, 2023.

Wenjie Wu, Dexuan Huo, and Hong Chen. SpikF: Spiking Fourier network for efficient long-term prediction. In *Proc. Int. Conf. Mach. Learn.*, 2025a.

Yuhan Wu, Xiyu Meng, Huajin Hu, Junru Zhang, Yabo Dong, and Dongming Lu. Affirm: Interactive mamba with adaptive fourier filters for long-term time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 21599–21607, 2025b.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *Proc. Int. Conf. Learn. Representations*, 2024.

Fuhao Yang, Xin Li, Min Wang, Hongyu Zang, Wei Pang, and Mingzhong Wang. WaveForM: Graph enhanced wavelet learning for long sequence forecasting of multivariate time series. In *Proc. AAAI Conf. Artif. Intell.*, pp. 10754–10761, 2023.

Yingnan Yang, Qingling Zhu, and Jianyong Chen. VCformer: Variable correlation Transformer with inherent lagged correlation for multivariate time series forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 5335–5343, 2024.

Kun Yi et al. Frequency-domain MLPs are more effective learners in time series forecasting. In *Proc. Adv. Neural Inf. Process. Syst.*, volume 36, pp. 76656–76679, 2023.

Tianyi Yin et al. Apollo-Forecast: Overcoming aliasing and inference speed challenges in language models for time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 22173–22181, 2025.

Hu Zhang, Zhien Dai, Zhaohui Tang, and Yongfang Xie. MDMLP-EIA: Multi-domain Dynamic MLPs with energy invariant attention for time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, 2026.

Jintao Zhang, Mingyue Cheng, Xiaoyu Tao, Zhiding Liu, and Daoyu Wang. Conditional denoising meets polynomial modeling: A flexible decoupled framework for time series forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 6993–7001, 2025a.

Xiyuan Zhang et al. First de-trend then attend: Rethinking attention for time-series forecasting. In *Proc. Adv. Neural Inf. Process. Syst.*, 2022.

Yingwei Zhang et al. Craft: Time series forecasting with cross-future behavior awareness. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 7056–7064, 2025b.

Zichi Zhang et al. WaveletMixer: A multi-resolution wavelets based MLP-mixer for multivariate long-term time series forecasting. In *Proc. AAAI Conf. Artif. Intell.*, pp. 22741–22749, 2025c.

Jiankai Zheng and Liang Xie. A dynamic stiefel graph neural network for efficient spatio-temporal time series forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 7155–7163, 2025.

Ziyu Zhou, Gengyu Lyu, Yiming Huang, Zihao Wang, Ziyu Jia, and Zhen Yang. SDformer: Transformer with spectral filter and dynamic attention for multivariate time series long-term forecasting. In *Proc. Int. Joint Conf. Artif. Intell.*, pp. 5689–5697, 2024.

A Implementation Details

A.1 Implementation for KOR

As a supplement to Subsection 3.2, we present the forward architecture of proposed KOR in Algorithm 1, where $\odot$ represents element-wise multiplication. During the offline initialization phase, various random parameters including natural frequencies, initial phases, and projection layers must be computed according to the hyperparameters. ω_{min} denotes the minimum frequency required for an oscillator to accumulate a phase shift of at least 1 radian over the entire evolution period, while ω_{max} represents the frequency that induces a phase shift of $\pi/2$ within a single iteration step.

Based on these parameters, the free phase evolution trajectory without external driving is subsequently calculated. During the forward pass, the phase at each integration step is iteratively computed before aggregating them to form the macroscopic phase representation.

Algorithm 1 Forward Pass of KOR

Given: Input: $\mathbf{X}_L^{high} \in \mathbb{R}^{L\times C}$
Hyperparameters: segment length P, virtual channel dimension V,
oscillator dimension N, scaling drive factor γ, coupling coefficient K_{intra}, K_{inter}

Initialization (offline):

$S \leftarrow L/P, \quad dt \leftarrow 1/P$
$\omega_{min} \leftarrow \frac{1}{S\cdot dt}, \quad \omega_{max} \leftarrow \frac{\pi}{2\cdot dt}$
$\mathbf{W}_{vir} \sim \mathcal{N}(0, \frac{1}{C}),\ \mathbf{W}_{osc} \sim \mathcal{N}(0, \frac{1}{P}),\ \mathbf{W}_{dim} \sim \mathcal{N}(0, \frac{1}{N}),\ \mathbf{W}_{chan} \sim \mathcal{N}(0, \frac{1}{V})$
$\mathbf{\Omega} \sim \mathcal{U}(\omega_{min}, \omega_{max})^{V\times N}$
$\mathbf{\Theta}^{(0)} \sim \mathcal{U}(0, 2\pi)^{V\times N}$
$\mathbf{A}_{freq}, \mathbf{A}_{spatial} \leftarrow$ Eq. 11
$\mathbf{\Theta}^{free,(0)} \leftarrow \mathbf{\Theta}^{(0)}$ ▷ Compute free phase evolution $\mathbf{\Theta}^{free}$
for $s = 1$ **to** S **do**
  $\mathbf{S}^{(s-1)} \leftarrow \sin(\mathbf{\Theta}^{free,(s-1)}), \quad \mathbf{C}^{(s-1)} \leftarrow \cos(\mathbf{\Theta}^{free,(s-1)})$
  $\mathbf{F}_{intra} \leftarrow \mathbf{C}^{(s-1)} \odot (\mathbf{S}^{(s-1)}\mathbf{A}_{freq}) - \mathbf{S}^{(s-1)} \odot (\mathbf{C}^{(s-1)}\mathbf{A}_{freq})$
  $\mathbf{F}_{inter} \leftarrow \mathbf{C}^{(s-1)} \odot (\mathbf{A}_{spatial}\mathbf{S}^{(s-1)}) - \mathbf{S}^{(s-1)} \odot (\mathbf{A}_{spatial}\mathbf{C}^{(s-1)})$
  $d\mathbf{\Theta}^{free} \leftarrow \mathbf{\Omega} + (K_{intra}/N)\mathbf{F}_{intra} + (K_{inter}/V)\mathbf{F}_{inter}$
  $\mathbf{\Theta}^{free,(s)} \leftarrow \mathbf{\Theta}^{free,(s-1)} + d\mathbf{\Theta}^{free} \cdot dt$
end for

Forward Pass Phase 1: Driven Evolution & Phase Deviation

$\mathbf{X}_V^{high} \leftarrow \mathbf{X}_L^{high}\mathbf{W}_{vir}^{\top}$ ▷ Map original C channels to V virtual channels
Split $\mathbf{X}_V^{high}$ into segments $\{\mathbf{z}^{(1)}, \mathbf{z}^{(2)} \ldots, \mathbf{z}^{(S)}\}$, where $\mathbf{z}^{(s)} \in \mathbb{R}^{V\times P}$
$\mathbf{\Theta}^{driven,(0)} \leftarrow \mathbf{\Theta}^{(0)}$ ▷ Compute driven phase evolution $\mathbf{\Theta}^{free}$
for $s = 1$ **to** S **do**
  $\mathbf{S}^{(s-1)} \leftarrow \sin(\mathbf{\Theta}^{driven,(s-1)}), \quad \mathbf{C}^{(s-1)} \leftarrow \cos(\mathbf{\Theta}^{driven,(s-1)})$
  $\mathbf{F}_{intra} \leftarrow \mathbf{C}^{(s-1)} \odot (\mathbf{S}^{(s-1)}\mathbf{A}_{freq}) - \mathbf{S}^{(s-1)} \odot (\mathbf{C}^{(s-1)}\mathbf{A}_{freq})$
  $\mathbf{F}_{inter} \leftarrow \mathbf{C}^{(s-1)} \odot (\mathbf{A}_{spatial}\mathbf{S}^{(s-1)}) - \mathbf{S}^{(s-1)} \odot (\mathbf{A}_{spatial}\mathbf{C}^{(s-1)})$
  External drive: $\mathbf{D}^{(s)} \leftarrow \gamma\left(\mathbf{z}^{(s)}\mathbf{W}_{osc}^{\top}\right)$ ▷ Map segment P to oscillator space N
  $d\mathbf{\Theta}^{driven} \leftarrow \mathbf{\Omega} + (K_{intra}/N)\mathbf{F}_{intra} + (K_{inter}/V)\mathbf{F}_{inter} + \mathbf{D}^{(s)}$
  Update phase: $\mathbf{\Theta}^{driven,(s)} \leftarrow \mathbf{\Theta}^{driven,(s-1)} + d\mathbf{\Theta}^{driven} \cdot dt$
  Calculate phase deviation: $\mathbf{\Delta}^{(s)} \leftarrow \mathbf{\Theta}^{driven,(s)} - \mathbf{\Theta}^{free,(s)}$
end for

Forward Pass Phase 2: Macroscopic Aggregation & Readout

$\mathbf{H}^{macro} \leftarrow \frac{1}{S}\sum_{s=1}^{S}\sin(\mathbf{\Delta}^{(s)})$ ▷ Group phase representation $\mathbf{H}^{macro} \in \mathbb{R}^{V\times N}$
$\mathbf{Z}_{kura}^{H} \leftarrow \mathbf{W}_{chan}\mathbf{H}^{macro}\mathbf{W}_{dim}^{\top}$ ▷ Map back to $\mathbb{R}^{C\times 2L}$
return Final representation $\mathbf{Z}_{kura}^{H}$

Table 6: Detailed dataset descriptions. *Dim* denotes the variate number of each dataset. *Dataset Size* is organized in (Train, Validation, Test). *Frequency* denotes the sampling interval of time points.

Dataset	Dim	Prediction Length	Dataset Size	Frequency	Domain
ETTh1, ETTh2	7	{96, 192, 336, 720}	(8545, 2881, 2881)	1 hour	Electricity
ETTm1, ETTm2	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15 min	Electricity
Weather	21	{96, 192, 336, 720}	(36792, 5271, 10540)	10 min	Weather
Traffic	862	{96, 192, 336, 720}	(12185, 1757, 3509)	1 hour	Transportation
Electricity	321	{96, 192, 336, 720}	(18317, 2633, 5261)	1 hour	Electricity
Exchange	8	{96, 192, 336, 720}	(5120, 665, 1422)	1 day	Finance
Illness	7	{24, 36, 48, 60}	(617, 74, 170)	1 week	Healthcare

A.2 Dataset Descriptions

The detailed statistics of the nine benchmark datasets, including the number of variables and dataset sizes, are summarized in Table 6.

- **ETT (h1, h2, m1, m2):** Four subsets consisting of load and oil temperature data collected from two power transformer stations.
- **Traffic:** A collection of road occupancy rates recorded on highways across the San Francisco Bay Area.
- **Electricity:** Hourly electricity consumption records of customers in Portugal.
- **Weather:** Meteorological observations collected by the weather station at the Max Planck Institute for Biogeochemistry in Germany, including temperature, humidity, air pressure, and wind speed information.
- **Exchange:** Daily fluctuations in exchange rates between multiple countries (e.g., USD relative to other currencies).
- **Illness:** The weekly ratio of influenza-like illness (ILI) patients to the total number of patients, as reported by the U.S. Centers for Disease Control and Prevention from 2002 to 2021.

A.3 Computing Device Configuration

All experiments were conducted on a Linux server running Ubuntu 20.04. The hardware configuration consists of an Intel Xeon Silver 4309Y CPU (@ 2.80GHz) and four NVIDIA GeForce RTX 3090 GPUs (24GB VRAM per GPU). The software environment was built with Python 3.11.5 and PyTorch 2.4.0 (Paszke et al., 2019), accelerated by CUDA 12.2. During the training phase, the network parameters were updated using the Adam optimizer (Kingma & Ba, 2015).

A.4 Baseline Reproduction Description

All baseline entries in Tables 3 and 4 are independently reproduced in our environment. None of them are copied from the original papers. For each baseline–KFM pair, we keep the backbone architecture, look-back window, data split, optimizer, and backbone-specific hyperparameters identical, and modify only the components introduced by KFM. Backbone-specific architectural hyperparameters, such as the numbers of attention layers and embedding dimensions, are taken from the parameter-configuration scripts in the official repositories of PatchTST, iTransformer, CATS, and PDF. TimesNet and TSMixer use the configurations provided in the official Time-Series-Library repository.

We optimize all paired models using the same L_1 objective. The purpose of this unified objective is to isolate the effect of introducing KFM, rather than confounding it with a change in the optimization target. Controlled backbone-extension evaluations in prior TSF studies similarly keep the training objective and other configurations fixed across the original and augmented models (Li et al., 2026; Tong & Yuan, 2025). Differences from numbers reported in the original papers may therefore arise

from the unified objective, standardized look-back windows, hyperparameter tuning, and execution environment.

To assess whether replacing the original objectives affects the comparison, we additionally train PDF and CATS using both their original objectives and L_1 on Weather and ETTm1, while keeping all other settings fixed. Across the resulting MSE and MAE comparisons, the changes caused by the objective replacement are consistently below 0.6%. This indicates that the unified objective has limited influence on the reported paired improvements.

To quantify the deviation from originally reported results, Table 7 compares our reproduced iTransformer results with those in the original paper. Most differences are below 1%, indicating that our controlled protocol largely preserves baseline performance.

Table 7: Comparison between iTransformer results reported in the original paper and our reproductions. Differences are computed as $(\text{reproduced} - \text{reported})/\text{reported} \times 100\%$.

Dataset	T	Reported MSE	Reported MAE	MSE difference (%)	MAE difference (%)
ETTh1	96	0.386	0.405	1.30	−1.23
ETTh1	192	0.441	0.436	0.23	−0.23
ETTh1	336	0.487	0.458	0.62	0.22
ETTh1	720	0.503	0.491	−0.40	0.20
ETTh1	Avg	0.454	0.448	0.39	−0.22
ETTh2	96	0.297	0.349	2.36	−0.57
ETTh2	192	0.380	0.400	0.26	−1.50
ETTh2	336	0.428	0.432	−0.93	0.00
ETTh2	720	0.427	0.445	0.70	0.45
ETTh2	Avg	0.383	0.407	0.46	−0.37

A.5 Baseline Model Descriptions

We briefly introduce the TSF baseline methods used in Table 3 and Table 4:

- **PatchTST** (Nie et al., 2023): Based on Transformer, it innovatively introduces a patching mechanism to replace the traditional step-wise token organization, reducing memory consumption and enhancing the ability to capture long-range dependencies. It also proposes a channel-independent design to avoid learning spurious correlations among channels.
- **iTransformer** (Liu et al., 2024): An inverted Transformer architecture that embeds the complete look-back sequence of each independent variable as a token and interacts based on self-attention. It applies the feed-forward network (FFN) independently to each variable token, breaking the practice of traditional models that fuse multiple variables per time step.
- **CATS** (Kim et al., 2024): A Transformer architecture entirely based on cross-attention. It argues that the self-attention model possesses permutation invariance and disrupts the correlation between the context and the target, thus it virtualizes the target horizon as independent queries and introduces an adaptive mask, performing cross-attention computation using historical data as keys and values.
- **PDF** (Dai et al., 2024): A parallel architecture based on multi-periodicity decoupling. It performs patch segmentation based on periods and cross-period points with the same phase, realizing the organization of input units containing short-term and long-term information, which are respectively encoded by 1D CNN and Transformer.
- **TimesNet** (Wu et al., 2023): It discovers the prominent periods of time series based on FTT, based on which the 1D signal is reconstructed into a 2D tensor. Then a 2D CNN is utilized to capture intra-period and inter-period variations.
- **TSMixer** (Chen et al., 2023): A lightweight and efficient TSF architecture based on MLP, which alternately performs linear feature mixing in the time dimension and the channel dimension.

Next, we briefly introduce the multi-branch architectures involving high-frequency component decomposition in Table 2:

- **InterPDN** (Kong & Hong, 2026): Models the time series regression problem as a probability distribution prediction problem on a predefined support set. It employs a multi-branch architecture for dual-scale self-supervision and self-supervision on interleaved support sets. Within each branch, it extracts the smooth trend term and the high-frequency, noisy residual term based on MA method, encoding them with MLP and CNN, respectively.
- **XPatch** (Stitsyuk & Choi, 2025): It performs trend-seasonal separation based on EMA. It then encodes them based on MLP and 1D convolution with residual connections, respectively, to focus on the long-term evolution of the trend component and the short-term fluctuation of the high-frequency component. A decayed L1 objective function is adopted to make the model focus on time steps closer to the context.
- **WPMixer** (Murad et al., 2025): High- and low-frequency components are decomposed based on DWT. For the obtained coefficients at each level, it adopts independent branches for encoding to avoid information aliasing across different frequency bands. Within each branch, it performs a two-stage MLP mixing on the patch dimension and the embedding dimension.
- **FilterTS** (Wang et al., 2025): A TSF architecture based on frequency-domain filtering. It designs static band-pass filters for dominant stable frequencies and leverages inter-variable representations as cross-filters, realizing the selective extraction of both persistent and shared variable frequency components.
- **CFPT** (Kou et al., 2025): A dual-branch architecture that employs cross-frequency interaction and timestamp periodic pattern recognition. It decomposes signals into high- and low-frequency components via discrete Fourier transform, employing a coupling network to model the interactive dynamics between long-term trends and short-term fluctuations before frequency recombination.
- **FITS** (Xu et al., 2024): A lightweight architecture that forecast in the complex frequency domain. It employs a low-pass filter to discard high-frequency components and noise, isolating the essential base frequencies. This compact, filtered representation is then processed by a single complex-valued linear layer to learn amplitude scaling and phase shifting.
- **Affirm** (Wu et al., 2025b): A model combining adaptive frequency filtering with state space models. It applies learnable high- and low-pass filters in the frequency domain to adaptively separate essential spectral patterns from noise, followed by an interactive dual Mamba block to efficiently capture multi-scale temporal dependencies.
- **Koopa** (Liu et al., 2023): An architecture grounded in modern Koopman theory to model non-stationary time series dynamics. It utilizes a Fourier Filter to disentangle time-invariant and time-variant components, which are hierarchically advanced by modular Koopman predictors using globally learned and locally context-aware linear operators.
- **VCformer** (Yang et al., 2024): A dual-module architecture capturing lagged cross-variable dependencies and temporal non-stationarity. It projects localized time series segments as dynamic snapshots into a Koopman space, approximating infinite-dimensional linear operators for forward state prediction.

B Theoretical Justification I: Different Dynamical Responses to Projected Frequencies

Appendices B–D provide local interpretations of KOR rather than a global solution of its nonlinear dynamics. We use first-order response around a partially coherent free trajectory, locally freeze the coupling operator when needed, and introduce stylized oscillatory or stochastic inputs only to isolate the corresponding mechanisms. None of these approximations is imposed during training.

B.1 Inter-segment Variation under Random Projection

After the spatial projection $z_v(t) = \sum_{c=1}^{C} W_{vir}^{v,c} x_c(t)$, let $\mathbf{z}_v^{(s)} \in \mathbb{R}^P$ be the s-th segment. At Gaussian initialization, a row $\mathbf{w}_n \sim \mathcal{N}(\mathbf{0}, \frac{1}{P}\mathbf{I})$ of $\mathbf{W}^{in}$ produces $D_{v,n}^{(s)} = \mathbf{w}_n^\top \mathbf{z}_v^{(s)}$, and isotropy

gives

$$\mathbb{E}_{\mathbf{w}_n}\left[\left(D_{v,n}^{(s+1)} - D_{v,n}^{(s)}\right)^2\right] = \frac{1}{P}\left\|\mathbf{z}_v^{(s+1)} - \mathbf{z}_v^{(s)}\right\|_2^2. \tag{14}$$

Thus, gradual segment changes induce gradual excitation changes on average, whereas rapidly varying segments tend to produce larger changes. This is a statement about inter-segment geometry, not exact preservation of Fourier components, phase, or immunity to aliasing.

A controlled diagnostic with equal-amplitude sinusoids and $P = 16$ supports this tendency: normalized excitation variation has Spearman correlation $\rho = 0.79$ with input frequency, and its mean in the highest-frequency third is $2.2\times$ that in the lowest-frequency third.

B.2 Restoring Force and ODE Formulation

To analyze the phase evolution, we stack the driven–free deviations from Equation 5 into $\mathbf{\Delta}(t) \in \mathbb{R}^M$, where $M = VN$. For sufficiently small deviations, subtracting the free dynamics from the driven dynamics and applying a first-order expansion gives

$$\frac{d\mathbf{\Delta}(t)}{dt} \approx -\mathbf{L}(t)\mathbf{\Delta}(t) + \gamma\mathbf{D}(t). \tag{15}$$

On a locally cohesive interval, the effective weights satisfy $w_{ij}(t) \propto A_{ij}\cos(\theta_j^{free}(t) - \theta_i^{free}(t)) \geq 0$. Together with the symmetric adjacency matrices, this makes the cosine-weighted Laplacian $\mathbf{L}(t)$ symmetric positive semidefinite. Let $\mathbf{P} = \mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^\top$ and $\mathbf{\Delta}_\perp = \mathbf{P}\mathbf{\Delta}$. The zero eigenmode of $\mathbf{L}(t)$ is the neutral common-phase direction, while $-\mathbf{L}(t)\mathbf{\Delta}_\perp(t)$ restores disagreement transverse to it. When the locally cohesive graph is connected, the spectral gap $\lambda_2(t) > 0$ provides strict transverse contraction, as further discussed in Appendix D.2.

B.3 Low-Frequency Tracking Response

For a stable transverse eigenmode of a locally frozen Laplacian, Equation 15 reduces to $\dot{\delta}_k = -\lambda_k\delta_k + \gamma D_k$. If $D_k(t) = C_k\cos(\omega_{low}t)$, its steady response is

$$\delta_k(t) = \frac{\gamma C_k}{\sqrt{\lambda_k^2 + \omega_{low}^2}}\cos\left(\omega_{low}t - \tan^{-1}\frac{\omega_{low}}{\lambda_k}\right). \tag{16}$$

At low frequency, the response therefore remains a bounded signal that closely tracks the drive instead of strongly compressing its slow oscillation. This behavior motivates assigning low-frequency structure to the forecasting backbone rather than relying on KOR alone. The oscillatory readouts in Figures 5–7 are consistent with this interpretation.

B.4 High-Frequency Attenuation

For a connected local graph, freeze the symmetric Laplacian over a short interval and write $\bar{\mathbf{L}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ with $0 = \lambda_1 < \lambda_2 \leq \cdots$. The Fourier response of each stable transverse mode is

$$\hat{\delta}_k(\Omega) = \frac{\gamma}{\lambda_k + i\Omega}\hat{D}_k(\Omega), \qquad k \geq 2. \tag{17}$$

Hence its gain decays as $\mathcal{O}(|\Omega|^{-1})$, so rapid disagreement components contribute only a small residual when their input amplitude is bounded. Equivalently, if the transverse drive contains a slow envelope and fast fluctuations, the stable modes retain a smoother response while attenuating the fast component. This mechanism applies to transverse disagreement, not to the neutral common-phase mode, and explains the envelope-dominated readouts observed in Figures 5–7.

C Theoretical Justification II: Adler Equation and Explainable Phase Dynamics

To interpret phase locking and phase slip in OS, we consider the continuous-time analogue of its discrete update:

$$\frac{d\theta_{v,n}}{dt} = \omega_{v,n} + F_{intra}(t) + F_{inter}(t) + \gamma\mathbf{D}_{v,n}(t), \tag{18}$$

where F_{intra} and F_{inter} are the two topological coupling terms.

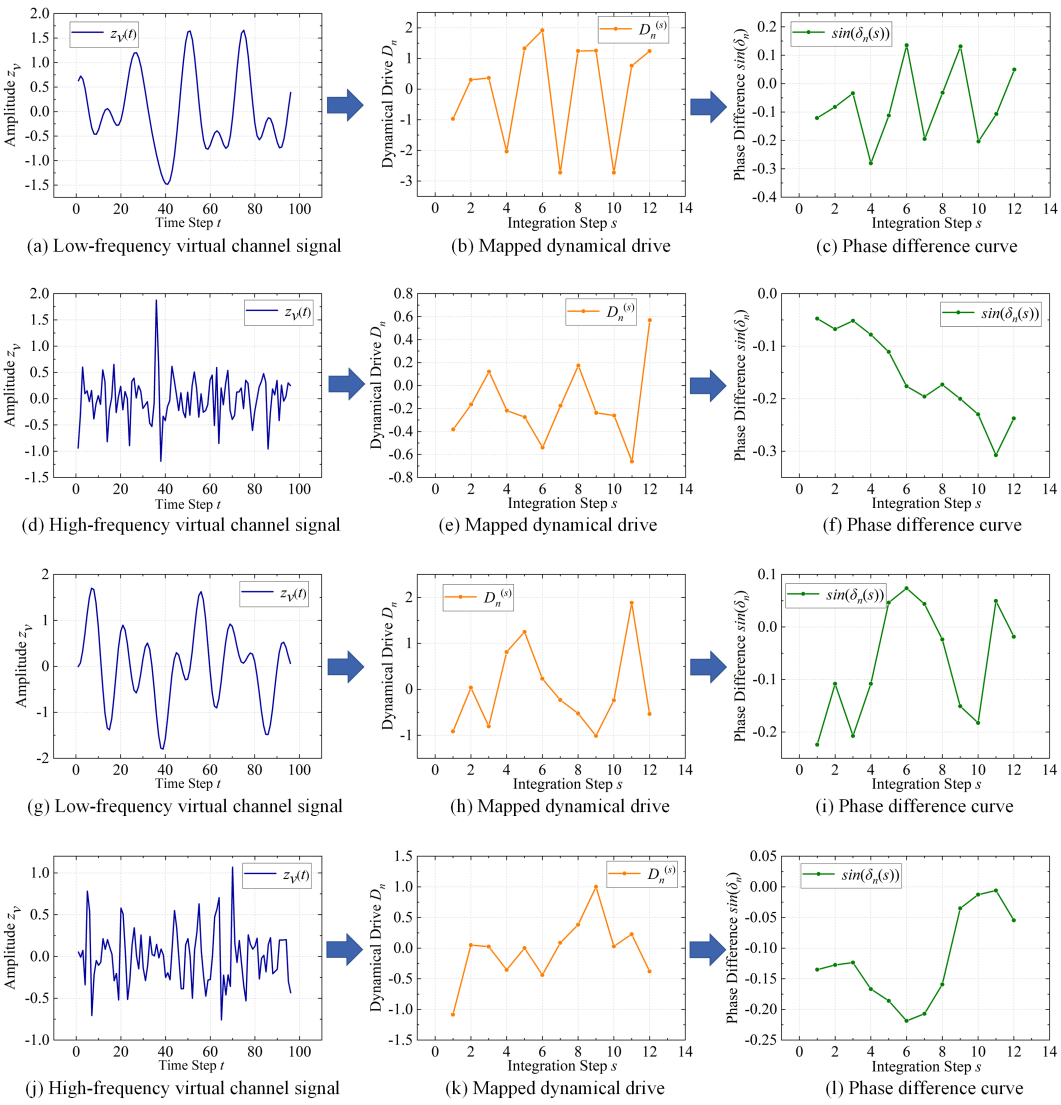


Figure 5: (a)–(f) and (g)–(l) respectively illustrate the mapping process for two samples from the ETTh1 dataset: from the time series (blue) to OS external drive (orange), and subsequently to the derived phase differences $\sin(\delta_n)$ (green). Rows 1 and 3 present the decomposed low-frequency sequences, whereas rows 2 and 4 show the decomposed high-frequency sequences.

C.1 Mean-Field Approximation for Network Coupling

Following standard network mean-field notation (Arenas et al., 2008), define the local order parameters

$$R_{intra}e^{i\Psi_{intra}} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{A}_{freq}^{j,n}e^{i\theta_{v,j}}, \tag{19}$$

$$R_{inter}e^{i\Psi_{inter}} = \frac{1}{V}\sum_{u=1}^{V}\mathbf{A}_{spatial}^{u,v}e^{i\theta_{u,n}}. \tag{20}$$

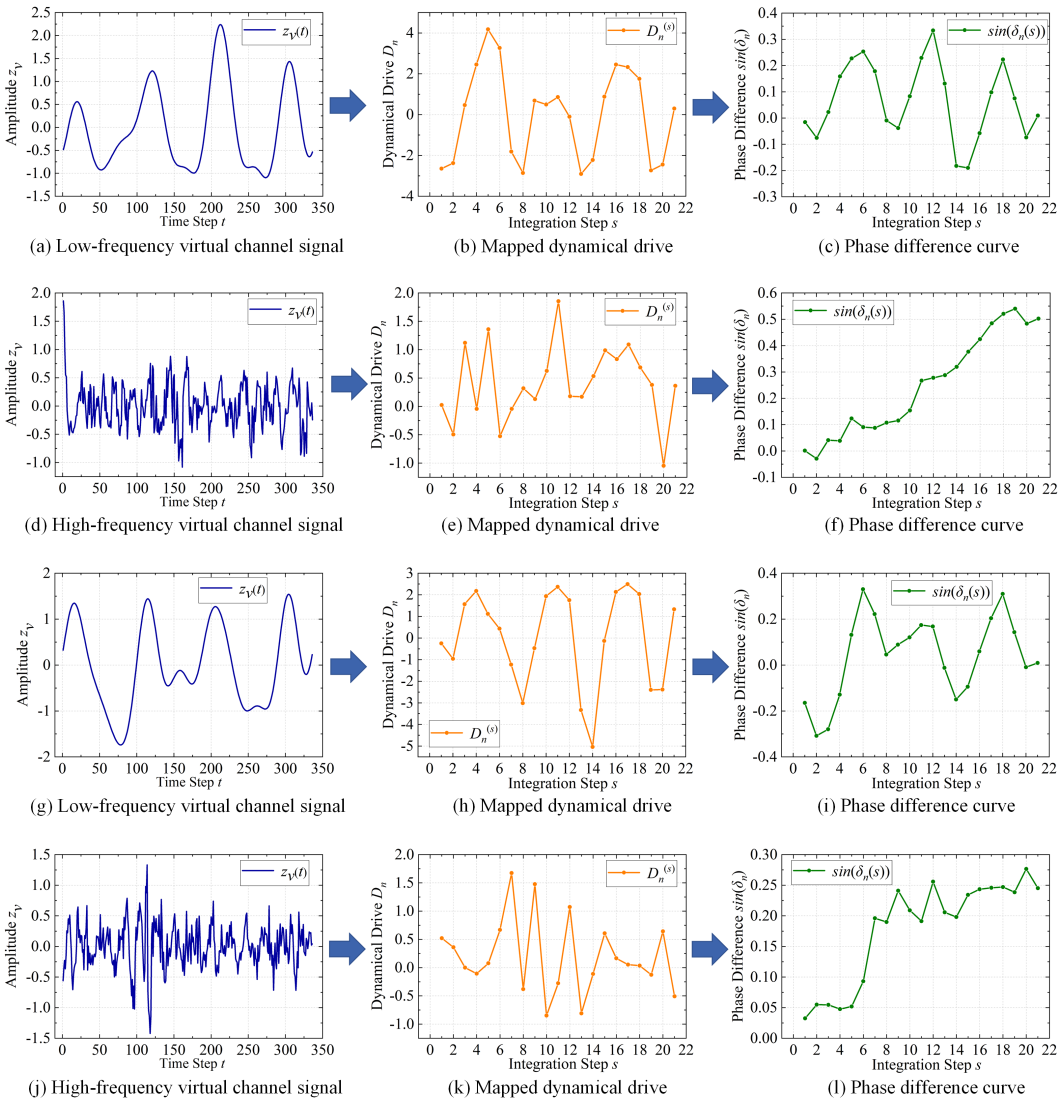


Figure 6: (a)–(f) and (g)–(l) respectively illustrate the mapping process for two samples from the ETTm1 dataset: from the time series (blue) to OS external drive (orange), and subsequently to the derived phase differences $\sin(\delta_n)$ (green). Rows 1 and 3 present the decomposed low-frequency sequences, whereas rows 2 and 4 show the decomposed high-frequency sequences.

Taking the corresponding imaginary parts rewrites Equations 9–10 as

$$F_{intra} = K_{intra} R_{intra} \sin(\Psi_{intra} - \theta_{v,n}), \tag{21}$$

$$F_{inter} = K_{inter} R_{inter} \sin(\Psi_{inter} - \theta_{v,n}). \tag{22}$$

C.2 Harmonic Addition and the Effective Forcing Field

Over a local response interval, represent the phase-relevant action of a dominant projected component with amplitude D and phase $\Theta_D(t)$ by its leading effective harmonic

$$F_D^{\text{eff}}(\theta_{v,n}, t) = \gamma D \sin(\Theta_D - \theta_{v,n}). \tag{23}$$

This approximation targets local locking and slipping behavior, not the complete driven trajectory. The two mean fields and this effective harmonic combine by phasor addition:

$$K_{eff} e^{i\Phi_{eff}} = K_{intra} R_{intra} e^{i\Psi_{intra}} + K_{inter} R_{inter} e^{i\Psi_{inter}} + \gamma D e^{i\Theta_D}. \tag{24}$$

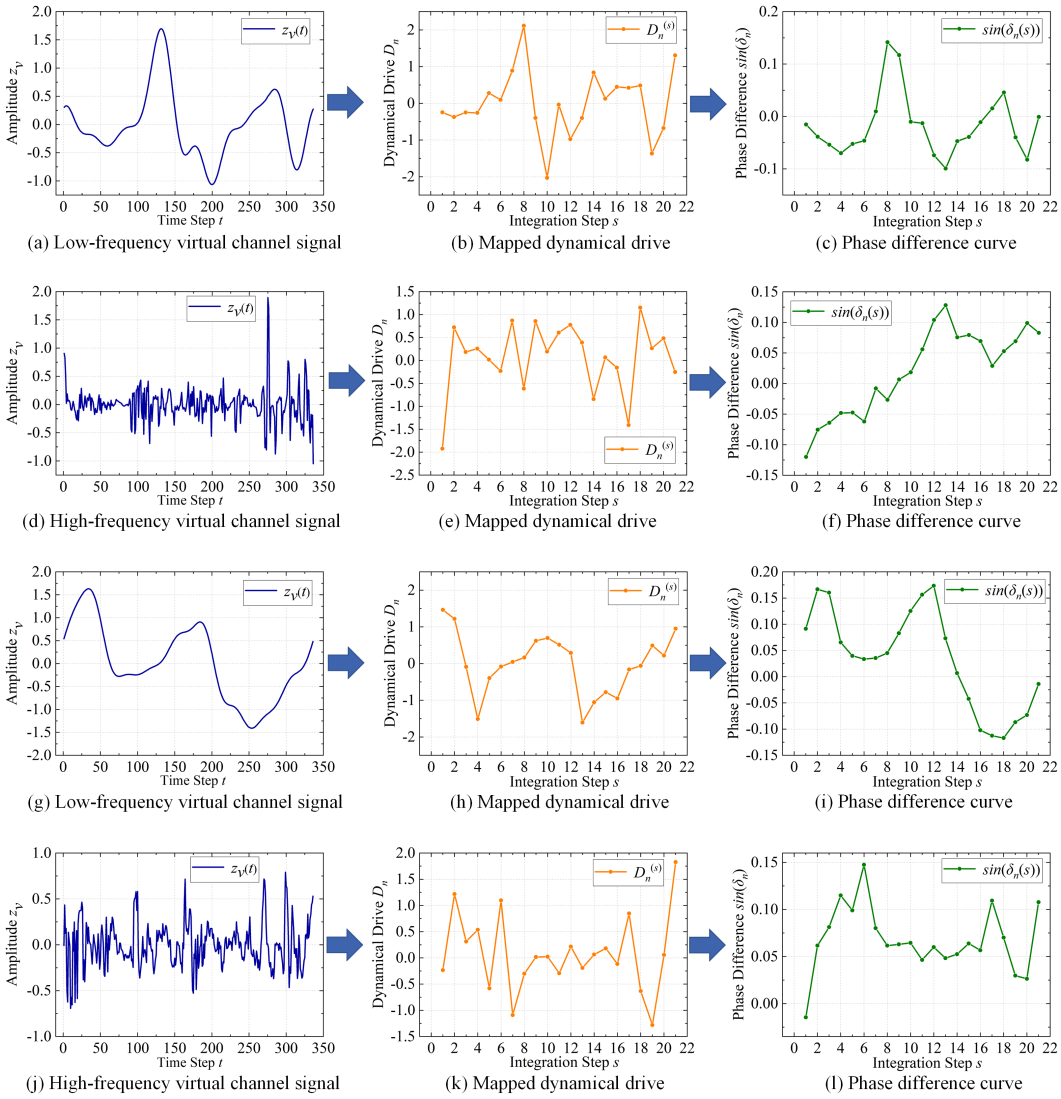


Figure 7: (a)–(f) and (g)–(l) respectively illustrate the mapping process for two samples from the Weather dataset: from the time series (blue) to OS external drive (orange), and subsequently to the derived phase differences $\sin(\delta_n)$ (green). Rows 1 and 3 present the decomposed low-frequency sequences, whereas rows 2 and 4 show the decomposed high-frequency sequences.

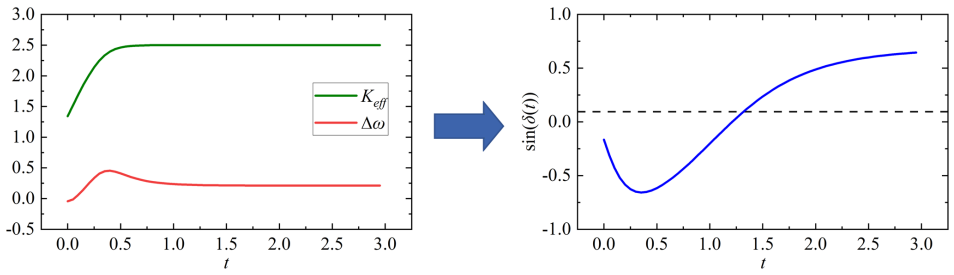
The local phase equation is therefore

$$\frac{d\theta_{v,n}}{dt} = \omega_{v,n} + K_{eff}\sin(\Phi_{eff} - \theta_{v,n}). \tag{25}$$

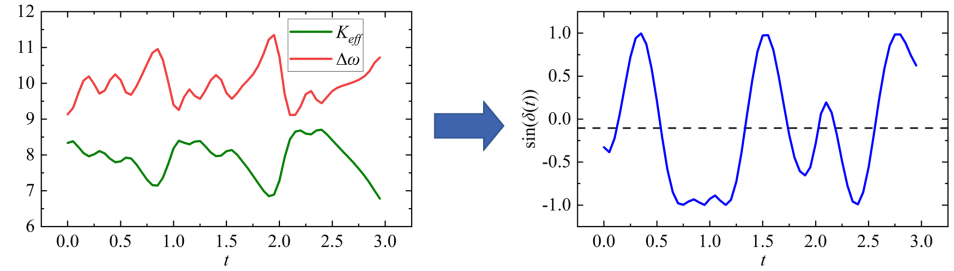
C.3 Reduction to Adler Equation

Let $\phi = \Phi_{eff} - \theta_{v,n}$. When K_{eff} and $\dot{\Phi}_{eff} \approx \Omega_{eff}$ vary slowly over the response interval, Equation 25 reduces to the Adler equation (Pikovsky et al., 1985):

$$\frac{d\phi}{dt} = \Delta\omega - K_{eff}\sin\phi, \qquad \Delta\omega = \Omega_{eff} - \omega_{v,n}. \tag{26}$$



(a) Phase Lock with Ω_D=3.0



(b) Phase Slip with Ω_D=11.0

Figure 8: The evolution of the effective coupling strength K_{eff} and frequency mismatch $\Delta\omega$ over integration steps for a small-scale oscillator system, alongside the corresponding $\sin(\delta(t))$ curve. The dashed line represents its mean value.

C.4 Dynamical Analysis and Asymptotic Mean Representation

For $|\Delta\omega| < K_{eff}$, the stable branch is phase-locked at $\phi^* = \arcsin(\Delta\omega/K_{eff})$. For $|\Delta\omega| > K_{eff}$, ϕ slips continuously, with equality marking the transition. Although ϕ differs from the KOR readout phase $\delta_n = \theta_n^{driven} - \theta_n^{free}$, a locally co-rotating free reference links the two: locking aligns their velocities and makes $\sin(\delta_n)$ slowly varying or nearly constant, as illustrated in the locking panels of Figures 8–10.

In the slipping regime, $dt = d\phi/(\Delta\omega - K_{eff}\sin\phi)$ creates a bottleneck (Strogatz, 2001). For locally frozen parameters and $\Delta\omega > K_{eff} > 0$, this gives

$$\begin{aligned}\langle \sin\phi \rangle_{T_{slip}} &= \frac{1}{T_{slip}} \int_0^{2\pi} \frac{\sin\phi}{\Delta\omega - K_{eff}\sin\phi} d\phi \\ &= \frac{\Delta\omega - \sqrt{\Delta\omega^2 - K_{eff}^2}}{K_{eff}} > 0.\end{aligned} \tag{27}$$

Relative to a smoothly evolving free reference, this asymmetric dwell time can induce a finite-window bias in the readout,

$$\mu_n = \frac{1}{S}\int_0^S \sin(\delta_n(s))ds, \tag{28}$$

without implying that every slipping trajectory has non-zero mean. The slip panels in Figures 8–10 show this finite-window behavior in both toy and real-data trajectories.

D Theoretical Justification III: Local Synchronization and Noise Attenuation

We finally connect the local restoring dynamics to stochastic perturbations in the stable transverse modes.

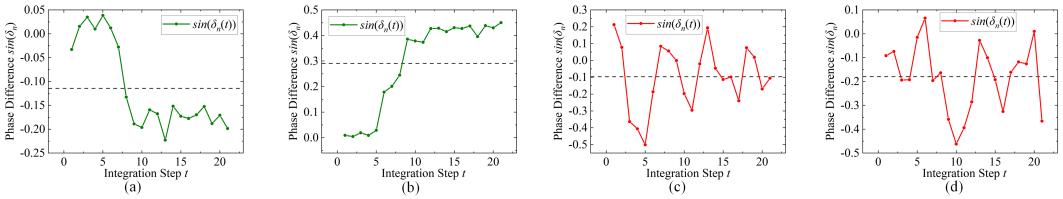


Figure 9: (a), (b): Phase-locking and (c), (d): phase slip phenomena exhibited by different oscillators within a specific virtual channel for a single sample from the ETTm1 dataset. The dashed line represents the mean of $\sin(\delta_n)$ across all integration steps.

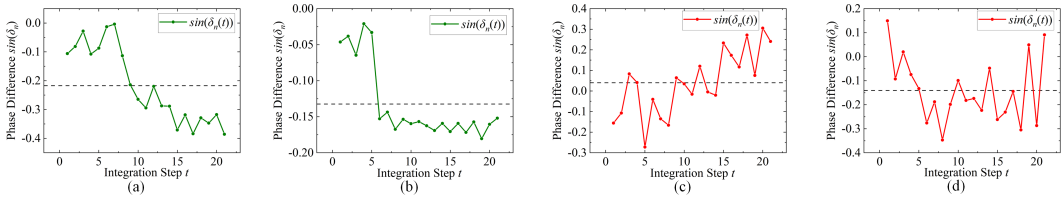


Figure 10: (a), (b): Phase-locking and (c), (d): phase slip phenomena exhibited by different oscillators within a specific virtual channel for a single sample from the Electricity dataset. The dashed line represents the mean of $\sin(\delta_n)$ across all integration steps.

D.1 Local Free-State Coherence

The free network only needs to provide a locally coherent reference, not global phase locking. For the Kuramoto mean field in Appendix C.1, classical results give a partially synchronized regime when coupling is sufficiently strong relative to frequency dispersion (Acebrón et al., 2005). Locally locked oscillators satisfy $|\omega - \dot{\Psi}| < K_{local}R$ on the stable branch. This motivates the locally cohesive intervals used by the Laplacian analysis in Appendix B.2. Figure 11 shows representative free-evolution phase trajectories.

On ETTh1, the mean pairwise circular distance during free evolution decreases by 26.2%. Under external drive, 16.8% of oscillators retain a stable relative phase, falling to 8.7% among the strongest 10% of drives. These observations are consistent with coexisting locally locked and slipping oscillators, rather than global synchronization.

D.2 Stable-Mode Reduction

Equation 15 already gives the linearization around the free trajectory: $\mathbf{L}$ is positive semidefinite locally, so $-\mathbf{L}$ is strictly contractive only on its stable transverse modes. For a connected local graph, projecting it onto a transverse eigenvector $\mathbf{q}_k$ gives

$$\frac{d\delta_k}{dt} = -\kappa_k \delta_k(t) + \gamma D_k(t), \qquad k \geq 2. \tag{29}$$

Here, $\delta_k = \mathbf{q}_k^\top \mathbf{\Delta}$, $D_k = \mathbf{q}_k^\top \mathbf{D}$, and $\kappa_k > 0$ is the mode-specific contraction rate. The neutral common-phase mode is not included in this reduction.

D.3 Modewise Noise Attenuation

Let $D_k(t) = S_k(t) + \xi_k(t)$, with $\xi_k(t)dt = \sigma_k dW_k(t)$. Equation 29 is then an Ornstein–Uhlenbeck-type response whose mean follows the leaky integration of S_k. For a deterministic initial state, Itô's isometry gives

$$\mathrm{Var}[\delta_k(t)] = \frac{\gamma^2\sigma_k^2}{2\kappa_k}\left(1 - e^{-2\kappa_k t}\right) \leq \frac{\gamma^2\sigma_k^2}{2\kappa_k}, \qquad k \geq 2. \tag{30}$$

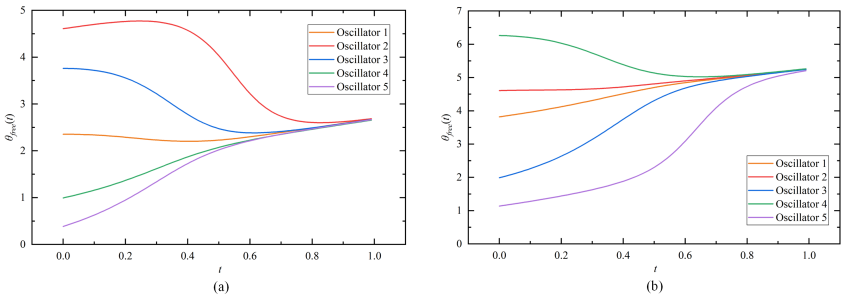


Figure 11: Phase trajectories $\theta_{free}(t)$ of a subset of oscillators under free evolution of different random seeds.

Table 8: MSE of PatchTST and its KFM variant across look-back lengths for $T = 336$. Values in parentheses denote the relative error reduction of KFM. Each comparison uses the same L for PatchTST and KFM.

	$L = 96$		$L = 336$		$L = 512$		$L = 720$	
Dataset	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM
ETTh1	0.495	**0.471** (4.85%)	0.428	**0.410** (4.21%)	0.424	**0.409** (3.54%)	0.416	**0.399** (4.09%)
ETTm1	0.403	**0.396** (1.74%)	0.372	**0.361** (2.96%)	0.368	**0.360** (2.17%)	0.365	**0.358** (1.92%)
Weather	0.276	**0.267** (3.26%)	0.259	**0.247** (4.63%)	0.245	**0.238** (2.86%)	0.240	**0.236** (1.67%)
Electricity	0.206	**0.180** (12.62%)	0.173	**0.164** (5.20%)	0.164	**0.159** (3.05%)	0.162	**0.151** (6.79%)

Thus white-noise variance is bounded in each stable disagreement mode and decreases with stronger contraction. This does not imply noise-free realizations or attenuation of the neutral common-phase component.

E Robustness to Look-back Length

The paired experiments in Tables 3 and 4 use the same pre-specified look-back length for each baseline and its KFM variant, rather than selecting the best L separately for KFM. To test whether the gains depend on a single look-back setting, we additionally use PatchTST as $\mathcal{F}_{\mathcal{W}}^{low}$ and evaluate the $T = 336$ task under $L \in \{96, 336, 512, 720\}$. As shown in Table 8, KFM improves all 16 dataset–L settings, while the reduction magnitude varies from 1.67% to 12.62%. Thus, the benefit is not restricted to the look-back length used in the main paired experiments, although its magnitude remains sensitive to L.

F Computational Efficiency

When employing PatchTST as $\mathcal{F}_{\mathcal{W}}^{low}$, Table 9 provides a vertical comparison on the ETTm1 dataset for the $T = 96$ task. Increasing the segment length P reduces the number of iterative integration steps and thus the training time. From $P = 16$ to $P = 32$, MSE and MAE increase by only 0.39% and 0.12%, while the per-epoch and total training time decrease by 30.2% and 19.5%, respectively. We therefore use $P = 32$ in the main experiments as an accuracy–efficiency balance.

Compared with the original PatchTST, KFM requires more time per epoch and during inference, but reaches its optimum in substantially fewer epochs. At P values of 32, 48, and 64, the total training time of PatchTST exceeds KFM by 2.97%, 12.36%, and 18.55%, respectively.

In the scenario where $P = 64$, the increased memory footprint of our model primarily originates from KOR and the linear layers utilized for the shortcut spectral fusion in Equation 12. These components account for 21.15 MB and 14.59 MB of memory respectively. Their corresponding trainable parameter counts are 0 and 3,827,040.

Table 10 further compares KFM with PatchTST as $\mathcal{F}_{\mathcal{W}}^{low}$ against recent TSF models on ETTh1 for the $T = 96$ task under the same environment. KFM achieves the lowest training time per epoch,

Table 9: Average training and inference times along with errors for the original model and our framework across five random seeds under different segment lengths P. In the Kuramoto model, $N = 256$ and $V = 64$.

Model		Mean Epochs	Training Time for Each Epoch (s)	Total Training Time (s)	Inference Time (s)	MSE	MAE
PatchTST +KFM	P=16	10.40	99.46	1034.38	5.187	0.2857	0.3281
	P=32	12.00	69.42	833.04	5.106	0.2868	0.3285
	P=48	11.60	65.81	763.40	5.087	0.2874	0.3284
	P=64	11.40	63.47	723.56	4.842	0.2889	0.3295
PatchTST		22.00	38.99	857.78	1.788	0.2976	0.3402

Table 10: Horizontal efficiency comparison with recent TSF models on ETTh1 for the $T = 96$ task under the same environment. The best results are highlighted in bold.

Model	Epochs	Time/Epoch (s)	Total Train (s)	Inference (s)	Params	Memory (MB)	MSE	MAE
KFM (+PatchTST)	**12.8**	**10.47**	**134.02**	**1.08**	**4.47M**	**17.43**	**0.360**	**0.387**
interPDN	22.4	13.17	295.01	1.23	5.84M	22.26	0.368	0.391
CFPT	13.6	19.85	269.96	1.43	5.36M	20.44	0.366	0.394
VCformer	17.2	509.72	8767.18	91.16	6.68M	25.48	0.371	0.395

total training time, inference time, parameter count, and memory footprint among the compared methods, while also obtaining the best forecasting accuracy. This comparison indicates that the additional cost relative to a lightweight backbone does not preclude competitive practical efficiency against recent models.

G Full Results

As a supplement to Tables 3 and 4, we provide the test results across four prediction lengths for all datasets in Tables 13 and 14. Our forecasting framework provides improvements at both short and long prediction lengths, although the gains are not universal across all settings. Taking PatchTST as an example, the average reduction rates for MSE and MAE at the shortest prediction length are 6.36% and 4.21% respectively. At the longest prediction length, the average MSE and MAE reduction rates are 4.83% and 3.48%, respectively. Across a total of 270 forecasting tasks on 9 datasets, only 19 MSE metrics and 8 MAE metrics exceed those of the original models. These account for a mere 7.03% and 2.96% respectively. These results indicate broad, though not universal, generalization across various TSF models.

To evaluate whether the direct-comparison results in Table 1 generalize beyond $T = 336$, Table 11 reports MSE at four standard prediction horizons on ETTh1 and ETTm1. This experiment follows the protocol of Table 1, where the low-frequency encoder and its hyperparameters are independently tuned. It is therefore distinct from the paired plug-in protocol in Tables 13 and 14, where each KFM variant retains the configuration of its reproduced backbone.

Table 12 complements the absolute errors in Tables 3 and 4 with dataset–backbone-specific relative reductions.

H Robustness Analysis

Since we utilize frozen random linear layers in KOR, model robustness becomes a primary concern. We therefore use PatchTST and iTransformer as $\mathcal{F}_{\mathcal{W}}^{low}$ and report the corresponding evaluation metrics in Table 15. This table presents the average errors and standard deviations from five repeated experiments under different random seeds for both the baseline reproductions and proposed embedded KFM framework. Metrics exhibiting a standard deviation greater than 0.004 are relatively scarce and are primarily concentrated on the Traffic, Electricity, and ETTh2 datasets. Only 6 forecasting metrics reach a standard deviation of 0.006. Furthermore, the variance of our TSF model demonstrates high consistency with the original models. This indicates that the overall model robustness

Table 11: MSE comparison with recent forecasting methods across four prediction horizons on ETTh1 and ETTm1. Following the direct-comparison protocol of Table 1, the look-back length is $L = 96$, and the low-frequency encoder and its hyperparameters are independently tuned. The best results are highlighted in bold.

Dataset	T	KFM	FilterTS	CFPT	xPatch	Affirm	FITS	VCformer	Koopa
ETTh1	96	**0.376**	0.381	0.377	0.379	0.382	0.385	0.379	0.384
	192	**0.425**	0.430	0.429	0.431	0.426	0.431	0.428	0.429
	336	**0.463**	0.467	0.471	0.465	0.477	0.468	0.479	0.480
	720	**0.459**	0.471	0.481	0.475	0.479	0.472	0.480	0.484
	Avg.	**0.431**	0.437	0.440	0.438	0.441	0.439	0.442	0.444
ETTm1	96	0.327	0.333	**0.325**	0.328	0.328	0.331	0.335	0.336
	192	**0.364**	0.366	0.369	**0.364**	0.367	0.370	0.368	0.372
	336	0.400	0.402	**0.393**	0.404	0.411	0.409	0.405	0.404
	720	**0.466**	0.478	0.470	0.475	0.470	0.479	0.482	0.474
	Avg.	**0.389**	0.395	**0.389**	0.393	0.394	0.397	0.398	0.397

Table 12: Relative MSE/MAE reductions (%) after integrating KFM, computed from the dataset-averaged errors in Tables 3 and 4. Each entry reports MSE reduction / MAE reduction. Positive values denote lower error, while negative values denote degradation and are underlined.

Dataset	PatchTST	iTransformer	CATS	PDF	TimesNet	TSMixer
ETTh1	5.04 / 2.80	4.61 / 3.36	2.67 / 1.41	4.59 / 1.65	6.09 / 5.93	6.05 / 6.18
ETTh2	1.50 / 1.06	3.38 / 2.72	3.22 / 2.34	4.08 / 2.85	7.13 / 5.23	7.37 / 5.24
ETTm1	2.52 / 2.13	4.89 / 6.08	1.39 / 1.58	0.87 / 2.14	7.91 / 7.62	2.78 / 3.51
ETTm2	4.62 / 2.24	5.12 / 4.20	2.61 / 2.48	2.71 / 2.24	4.05 / 3.30	2.78 / 3.61
Traffic	1.46 / 5.32	$\underline{-2.82}$ / 3.56	0.97 / 1.50	$\underline{-0.25}$ / 5.08	3.52 / 2.41	4.75 / 10.68
Electricity	4.17 / 1.98	4.47 / 3.70	4.09 / 1.57	4.22 / 1.98	9.71 / 5.96	11.00 / 7.38
Weather	4.62 / 4.53	4.21 / 4.29	1.73 / 1.89	2.60 / 2.69	4.25 / 4.90	3.23 / 4.68
Exchange	8.98 / 5.72	1.34 / 1.21	13.98 / 5.16	8.56 / 4.54	2.70 / 4.11	$\underline{-0.27}$ / 0.49
Illness	6.98 / 5.20	28.26 / 21.85	37.32 / 26.05	22.60 / 19.22	9.11 / 10.75	44.40 / 31.22

relies heavily on the robustness of $\mathcal{F}_{\mathcal{W}}^{low}$. The frozen random mapping layers we introduce do not significantly compromise model stability.

I Supplementary Model Analysis

I.1 Complete Combination-wise Ablation

Table 16 evaluates all 2^3 combinations of soft gating, spectral fusion, and KOR using TimesNet as the low-frequency backbone. Across matched configurations that differ in only one component, KOR improves all 48/48 dataset–metric comparisons, with an average relative error reduction of 3.85%. Soft gating and spectral fusion improve 44/48 and 47/48 comparisons, with average reductions of 1.35% and 1.37%, respectively. These results identify KOR as the primary contributor while confirming the complementary gains of the other two components.

I.2 Module substitution Study

Table 17 presents the comparative experimental results after module replacement. In the frequency decomposition replacement experiments, we compare several approaches including Top-K hard thresholding, a FreDN-inspired method (An et al., 2026) that uses an MLP to learn amplitude masks on each frequency band with a decay prior, an Affirm-inspired approach (Wu et al., 2025b) where an MLP learns frequency threshold offsets, and a Gumbel-Softmax-based method (Jang et al., 2017) that extracts discrete high frequencies to be added to the low-frequency branch.

Among these, the learnable adaptive decomposition across all frequency bands exhibits the poorest performance. Every metric fails to surpass the Top-K hard thresholding baseline. The discrete full amplitude extraction based on Gumbel-Softmax also performs unsatisfactorily. It underperforms

Table 13: Full MSE results of baselines and their KFM variants across all prediction horizons. For each horizon (and the avg row), Avg. Δ (%) is the arithmetic mean of the six backbone-specific relative error reductions computed from the errors in that row. All other settings are identical to Table 3. The lower error within each baseline–KFM pair is bolded.

Dataset	Horizon	PatchTST		iTransformer		CATS		PDF		TimesNet		TSMixer		Avg.
		Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Δ (%)
ETTh1	96	0.378	**0.360**	0.391	**0.379**	0.375	**0.364**	0.361	**0.356**	0.389	**0.382**	0.404	**0.377**	3.44
	192	0.417	**0.398**	0.442	**0.430**	0.410	**0.404**	0.399	**0.396**	0.439	**0.425**	0.452	**0.429**	2.96
	336	0.428	**0.410**	0.490	**0.468**	0.422	**0.410**	0.423	**0.410**	0.494	**0.460**	0.489	**0.466**	4.37
	720	0.445	**0.416**	0.501	**0.464**	0.441	**0.426**	0.474	**0.418**	0.519	**0.459**	0.507	**0.468**	8.06
	avg	0.417	**0.396**	0.456	**0.435**	0.412	**0.401**	0.414	**0.395**	0.460	**0.432**	0.463	**0.435**	4.84
ETTh2	96	0.277	**0.273**	0.304	**0.289**	0.275	**0.273**	0.274	**0.272**	0.337	**0.303**	0.317	**0.290**	4.41
	192	0.343	**0.337**	0.381	**0.365**	0.335	**0.331**	0.339	**0.336**	0.405	**0.371**	0.400	**0.370**	3.99
	336	0.335	**0.332**	0.424	**0.412**	0.352	**0.330**	0.360	**0.325**	0.450	**0.425**	0.464	**0.417**	5.90
	720	0.380	**0.373**	0.430	**0.420**	0.404	**0.388**	0.397	**0.381**	0.434	**0.411**	0.447	**0.432**	3.47
	avg	0.334	**0.329**	0.385	**0.372**	0.342	**0.331**	0.343	**0.329**	0.407	**0.378**	0.407	**0.377**	4.44
ETTm1	96	0.298	**0.287**	0.342	**0.313**	0.299	**0.289**	0.284	**0.281**	0.336	**0.315**	0.325	**0.316**	4.27
	192	0.333	**0.326**	0.383	**0.366**	0.333	**0.330**	**0.320**	0.322	0.407	**0.362**	0.377	**0.363**	3.60
	336	0.372	**0.361**	0.420	**0.403**	0.371	**0.362**	0.358	**0.356**	0.416	**0.395**	0.407	**0.395**	3.00
	720	0.425	**0.416**	0.489	**0.472**	**0.433**	0.435	0.420	**0.411**	0.507	**0.464**	0.472	**0.463**	2.94
	avg	0.357	**0.348**	0.409	**0.389**	0.359	**0.354**	0.346	**0.343**	0.417	**0.384**	0.395	**0.384**	3.39
ETTm2	96	0.171	**0.162**	0.186	**0.172**	0.171	**0.167**	0.165	**0.162**	0.188	**0.172**	0.180	**0.172**	4.98
	192	0.229	**0.219**	0.254	**0.238**	0.235	**0.229**	0.224	**0.218**	0.252	**0.243**	0.249	**0.241**	3.78
	336	0.277	**0.266**	0.316	**0.299**	0.290	**0.279**	0.275	**0.271**	0.322	**0.311**	0.309	**0.299**	3.54
	720	0.362	**0.346**	0.415	**0.402**	0.377	**0.369**	0.366	**0.354**	0.420	**0.410**	0.414	**0.408**	2.80
	avg	0.260	**0.248**	0.293	**0.278**	0.268	**0.261**	0.258	**0.251**	0.296	**0.284**	0.288	**0.280**	3.65
Traffic	96	0.383	**0.377**	**0.398**	0.414	**0.372**	0.377	**0.360**	0.363	0.589	**0.560**	0.496	**0.484**	0.45
	192	0.395	**0.395**	**0.423**	0.424	0.400	**0.395**	**0.381**	0.383	0.617	**0.582**	0.506	**0.484**	1.75
	336	0.413	**0.404**	**0.424**	0.439	0.424	**0.411**	**0.389**	0.397	0.635	**0.595**	0.534	**0.493**	2.27
	720	0.447	**0.441**	**0.458**	0.473	0.461	**0.457**	0.441	**0.433**	**0.660**	0.674	0.568	**0.543**	0.50
	avg	0.410	**0.404**	**0.426**	0.438	0.414	**0.410**	**0.393**	0.394	0.625	**0.603**	0.526	**0.501**	1.27
Electricity	96	0.138	**0.130**	0.150	**0.141**	0.139	**0.130**	0.134	**0.128**	0.171	**0.158**	0.177	**0.150**	7.60
	192	0.154	**0.148**	0.166	**0.161**	0.158	**0.150**	0.151	**0.147**	0.187	**0.175**	0.194	**0.175**	5.14
	336	0.173	**0.164**	0.178	**0.175**	0.173	**0.168**	0.166	**0.162**	0.204	**0.195**	0.205	**0.176**	5.12
	720	0.208	**0.202**	0.220	**0.208**	0.213	**0.207**	0.212	**0.200**	0.262	**0.217**	0.261	**0.243**	6.81
	avg	0.168	**0.161**	0.179	**0.171**	0.171	**0.164**	0.166	**0.159**	0.206	**0.186**	0.209	**0.186**	6.28
Weather	96	0.161	**0.150**	0.176	**0.166**	0.154	**0.150**	0.153	**0.147**	0.169	**0.157**	0.161	**0.153**	5.18
	192	0.205	**0.194**	0.225	**0.214**	0.195	**0.194**	0.197	**0.191**	0.225	**0.208**	0.210	**0.202**	4.20
	336	0.259	**0.247**	0.283	**0.271**	0.252	**0.245**	0.248	**0.242**	0.282	**0.275**	0.272	**0.262**	3.37
	720	0.327	**0.317**	0.360	**0.349**	0.321	**0.319**	0.324	**0.319**	0.359	**0.353**	0.350	**0.343**	1.99
	avg	0.238	**0.227**	0.261	**0.250**	0.231	**0.227**	0.231	**0.225**	0.259	**0.248**	0.248	**0.240**	3.44
Exchange	96	0.100	**0.086**	0.087	**0.084**	**0.085**	0.087	0.092	**0.089**	0.105	**0.091**	0.086	**0.084**	5.67
	192	0.199	**0.183**	0.183	**0.179**	0.184	**0.172**	0.211	**0.196**	0.218	**0.185**	0.178	**0.177**	6.59
	336	0.395	**0.351**	**0.334**	0.336	0.342	**0.296**	0.360	**0.353**	0.363	**0.358**	**0.322**	0.328	4.24
	720	0.955	**0.879**	0.892	**0.877**	0.876	**0.723**	1.066	**0.941**	**0.940**	0.948	0.883	**0.881**	6.37
	avg	0.412	**0.375**	0.374	**0.369**	0.372	**0.320**	0.432	**0.395**	0.407	**0.396**	**0.367**	0.368	5.88
Illness	24	1.753	**1.510**	2.324	**1.651**	4.280	**1.745**	2.377	**1.552**	2.192	**1.683**	2.826	**1.366**	35.27
	36	1.390	**1.390**	2.253	**1.535**	4.515	**3.277**	2.189	**1.712**	2.191	**2.145**	2.817	**1.411**	22.18
	48	**1.678**	1.722	2.298	**1.714**	4.448	**4.174**	2.143	**1.827**	1.946	**1.750**	2.925	**1.667**	16.13
	60	2.170	**1.880**	2.397	**1.752**	4.600	**1.988**	2.351	**1.920**	2.231	**2.202**	3.115	**2.052**	25.14
	avg	1.748	**1.626**	2.318	**1.663**	4.461	**2.796**	2.265	**1.753**	2.140	**1.945**	2.921	**1.624**	24.78

Table 14: Full MAE results of baselines and their KFM variants across all prediction horizons. For each horizon (and the avg row), Avg. Δ (%) is the arithmetic mean of the six backbone-specific relative error reductions computed from the errors in that row. All other settings are identical to Table 4. The lower error within each baseline–KFM pair is bolded.

Dataset	Horizon	PatchTST		iTransformer		CATS		PDF		TimesNet		TSMixer		Avg.
		Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Base	+KFM	Δ (%)
ETTh1	96	0.401	**0.387**	0.400	**0.394**	0.398	**0.391**	0.389	**0.386**	0.412	**0.396**	0.422	**0.387**	3.28
	192	0.419	**0.410**	0.435	**0.424**	0.420	**0.414**	0.413	**0.411**	0.442	**0.420**	0.442	**0.418**	2.83
	336	0.434	**0.422**	0.459	**0.446**	0.433	**0.423**	0.428	**0.422**	0.471	**0.438**	0.461	**0.436**	3.62
	720	0.460	**0.447**	0.492	**0.463**	0.449	**0.448**	0.464	**0.448**	0.494	**0.457**	0.487	**0.460**	4.24
	avg	0.429	**0.417**	0.447	**0.432**	0.425	**0.419**	0.424	**0.417**	0.455	**0.428**	0.453	**0.425**	3.56
ETTh2	96	0.335	**0.331**	0.347	**0.336**	0.334	**0.333**	0.335	**0.330**	0.371	**0.345**	0.359	**0.337**	3.22
	192	0.377	**0.373**	0.394	**0.385**	0.374	**0.369**	0.378	**0.375**	0.415	**0.388**	0.409	**0.388**	2.85
	336	0.381	**0.380**	0.432	**0.420**	0.390	**0.378**	0.399	**0.376**	0.448	**0.429**	0.456	**0.425**	3.82
	720	0.420	**0.413**	0.447	**0.436**	0.437	**0.421**	0.432	**0.418**	0.448	**0.435**	0.455	**0.441**	2.83
	avg	0.378	**0.374**	0.405	**0.394**	0.384	**0.375**	0.386	**0.375**	0.421	**0.399**	0.420	**0.398**	3.24
ETTm1	96	0.340	**0.328**	0.377	**0.343**	0.343	**0.333**	0.338	**0.330**	0.375	**0.350**	0.360	**0.346**	4.73
	192	0.365	**0.356**	0.396	**0.370**	0.368	**0.360**	0.361	**0.355**	0.413	**0.373**	0.385	**0.370**	4.41
	336	0.383	**0.378**	0.416	**0.395**	0.388	**0.381**	0.383	**0.372**	0.423	**0.397**	0.408	**0.392**	3.52
	720	0.417	**0.409**	0.456	**0.435**	**0.419**	0.421	0.415	**0.406**	0.470	**0.433**	0.444	**0.431**	3.17
	avg	0.376	**0.368**	0.411	**0.386**	0.380	**0.374**	0.374	**0.366**	0.420	**0.388**	0.399	**0.385**	3.84
ETTm2	96	0.250	**0.246**	0.269	**0.251**	0.256	**0.250**	0.255	**0.247**	0.268	**0.251**	0.263	**0.251**	4.11
	192	0.287	**0.285**	0.311	**0.295**	0.298	**0.292**	0.289	**0.284**	0.307	**0.297**	0.308	**0.297**	2.74
	336	0.327	**0.318**	0.348	**0.334**	0.337	**0.330**	0.325	**0.320**	0.349	**0.341**	0.347	**0.334**	2.74
	720	0.382	**0.371**	0.405	**0.394**	0.402	**0.387**	0.383	**0.374**	0.406	**0.399**	0.409	**0.399**	2.64
	avg	0.312	**0.305**	0.333	**0.319**	0.323	**0.315**	0.313	**0.306**	0.333	**0.322**	0.332	**0.320**	3.01
Traffic	96	0.248	**0.236**	**0.264**	0.268	0.252	**0.244**	0.240	**0.227**	0.315	**0.300**	0.321	**0.293**	4.23
	192	0.260	**0.242**	0.275	**0.266**	0.261	**0.253**	0.246	**0.237**	0.324	**0.317**	0.324	**0.291**	4.88
	336	0.259	**0.246**	0.283	**0.271**	**0.260**	0.262	0.257	**0.243**	0.338	**0.330**	0.346	**0.297**	5.08
	720	0.283	**0.270**	0.300	**0.277**	**0.289**	0.290	0.280	**0.266**	**0.349**	0.350	0.358	**0.324**	4.35
	avg	0.263	**0.249**	0.281	**0.271**	0.266	**0.262**	0.256	**0.243**	0.332	**0.324**	0.337	**0.301**	4.76
Electricity	96	0.222	**0.217**	0.240	**0.231**	0.225	**0.220**	0.218	**0.216**	0.270	**0.258**	0.268	**0.245**	3.69
	192	0.238	**0.234**	0.258	**0.251**	0.240	**0.237**	0.240	**0.234**	0.288	**0.274**	0.283	**0.265**	3.23
	336	0.260	**0.252**	0.271	**0.265**	0.261	**0.257**	0.258	**0.251**	0.305	**0.293**	0.297	**0.270**	3.76
	720	0.289	**0.285**	0.310	**0.291**	0.292	**0.290**	0.293	**0.285**	0.346	**0.310**	0.343	**0.322**	4.58
	avg	0.252	**0.247**	0.270	**0.260**	0.255	**0.251**	0.252	**0.247**	0.302	**0.284**	0.298	**0.276**	3.76
Weather	96	0.198	**0.187**	0.216	**0.200**	0.202	**0.191**	0.193	**0.187**	0.219	**0.199**	0.210	**0.194**	6.38
	192	0.246	**0.230**	0.257	**0.245**	0.242	**0.235**	0.237	**0.229**	0.265	**0.245**	0.253	**0.239**	5.09
	336	0.283	**0.271**	0.297	**0.288**	0.279	**0.275**	0.280	**0.270**	0.304	**0.296**	0.298	**0.286**	3.16
	720	0.332	**0.323**	0.348	**0.340**	0.333	**0.333**	0.328	**0.324**	0.354	**0.347**	0.349	**0.340**	1.80
	avg	0.265	**0.253**	0.280	**0.268**	0.264	**0.259**	0.260	**0.253**	0.286	**0.272**	0.278	**0.265**	3.83
Exchange	96	0.224	**0.206**	0.208	**0.204**	**0.204**	0.208	0.212	**0.211**	0.235	**0.214**	0.207	**0.203**	3.22
	192	0.322	**0.306**	0.307	**0.301**	0.305	**0.299**	0.329	**0.316**	0.341	**0.303**	0.303	**0.299**	4.22
	336	0.463	**0.428**	**0.419**	0.419	0.423	**0.395**	0.437	**0.429**	0.438	**0.429**	**0.411**	0.414	2.89
	720	0.739	**0.707**	0.717	**0.706**	0.697	**0.642**	0.784	**0.727**	0.739	**0.735**	0.711	**0.708**	3.66
	avg	0.437	**0.412**	0.413	**0.408**	0.407	**0.386**	0.441	**0.421**	0.438	**0.420**	0.408	**0.406**	3.54
Illness	24	0.842	**0.780**	1.038	**0.794**	1.477	**0.838**	1.105	**0.774**	0.995	**0.796**	1.180	**0.720**	27.18
	36	0.756	**0.739**	1.022	**0.773**	1.499	**1.217**	1.014	**0.850**	0.962	**0.917**	1.171	**0.754**	16.98
	48	0.866	**0.855**	1.035	**0.837**	1.490	**1.457**	1.011	**0.873**	0.952	**0.865**	1.201	**0.812**	12.97
	60	0.999	**0.909**	1.062	**0.844**	1.522	**0.917**	1.075	**0.897**	1.034	**0.941**	1.124	**0.929**	18.70
	avg	0.866	**0.821**	1.039	**0.812**	1.497	**1.107**	1.051	**0.849**	0.986	**0.880**	1.169	**0.804**	19.05

Table 15: Robustness performance of the baseline models and their integration within KFM framework. Results represent the mean and standard deviation across five random seeds with standard deviations exceeding 0.003 highlighted in bold.

Model		iTransformer		iTransformer+KFM		PatchTST		PatchTST+KFM	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.391±0.001	0.400±0.003	0.379±0.003	0.394±0.003	0.378±0.002	0.401±0.003	0.360±0.002	0.387±0.002
	192	0.442±0.003	0.435±0.002	0.430±0.002	0.424±0.002	0.417±0.002	0.419±0.003	0.398±0.003	0.410±0.001
	336	0.490±0.002	0.459±0.001	0.468±0.003	0.446±0.002	0.428±0.002	0.434±0.001	0.410±0.003	0.422±0.002
	720	0.501±0.003	0.492±0.002	0.464±0.003	**0.463±0.004**	0.445±0.002	0.460±0.002	0.416±0.002	0.447±0.002
	avg	0.456±0.002	0.447±0.002	0.435±0.001	0.432±0.001	0.417±0.002	0.429±0.002	0.396±0.001	0.417±0.001
ETTh2	96	0.304±0.002	0.347±0.001	0.289±0.002	0.336±0.001	0.277±0.002	0.335±0.002	0.273±0.002	0.331±0.001
	192	0.381±0.001	0.394±0.001	0.365±0.002	0.385±0.001	**0.343±0.004**	0.377±0.003	0.337±0.002	0.373±0.002
	336	0.424±0.002	0.432±0.001	0.412±0.001	0.420±0.002	**0.335±0.004**	**0.381±0.004**	**0.332±0.005**	**0.380±0.004**
	720	**0.430±0.004**	0.447±0.002	0.420±0.002	0.436±0.002	**0.380±0.004**	0.420±0.002	0.373±0.003	0.413±0.002
	avg	0.385±0.002	0.405±0.001	0.372±0.001	0.394±0.001	**0.334±0.004**	0.378±0.003	0.329±0.003	0.374±0.002
ETTm1	96	0.342±0.002	0.377±0.003	0.313±0.002	0.343±0.002	0.298±0.003	0.340±0.003	0.287±0.003	0.328±0.002
	192	0.383±0.002	0.396±0.001	0.366±0.003	0.370±0.003	0.333±0.002	0.365±0.002	0.326±0.001	0.356±0.001
	336	0.420±0.003	0.416±0.001	0.403±0.002	0.395±0.003	0.372±0.003	0.383±0.002	**0.361±0.004**	0.378±0.003
	720	0.489±0.003	0.456±0.002	0.472±0.002	0.435±0.002	**0.425±0.005**	0.417±0.002	0.416±0.003	0.409±0.003
	avg	0.409±0.003	0.411±0.002	0.389±0.002	0.386±0.002	0.357±0.003	0.376±0.002	0.348±0.003	0.368±0.002
ETTm2	96	0.186±0.002	0.269±0.002	0.172±0.003	0.251±0.002	0.171±0.002	0.250±0.001	0.162±0.003	0.246±0.002
	192	0.254±0.003	0.311±0.001	0.238±0.003	0.295±0.002	0.229±0.003	0.287±0.001	0.219±0.003	0.285±0.002
	336	0.316±0.003	0.348±0.003	**0.299±0.004**	**0.334±0.004**	0.277±0.003	0.327±0.003	**0.266±0.004**	0.318±0.003
	720	0.415±0.002	0.405±0.003	0.402±0.002	0.394±0.002	0.362±0.002	0.382±0.003	0.346±0.002	0.371±0.002
	avg	0.293±0.003	0.333±0.002	0.278±0.002	0.319±0.002	0.260±0.003	0.312±0.002	0.248±0.003	0.305±0.002
Traffic	96	**0.398±0.004**	**0.264±0.004**	**0.414±0.005**	**0.268±0.006**	0.383±0.003	**0.248±0.004**	**0.377±0.004**	0.236±0.001
	192	0.423±0.003	0.275±0.002	**0.424±0.004**	**0.266±0.005**	**0.395±0.005**	0.260±0.003	0.395±0.003	0.242±0.002
	336	0.424±0.003	**0.283±0.004**	**0.439±0.005**	**0.271±0.005**	0.413±0.003	0.259±0.003	**0.404±0.006**	0.246±0.003
	720	**0.458±0.006**	0.300±0.003	**0.473±0.007**	**0.277±0.006**	**0.447±0.005**	0.283±0.002	**0.441±0.007**	0.270±0.002
	avg	**0.426±0.004**	0.281±0.003	**0.438±0.005**	**0.271±0.006**	**0.410±0.004**	0.263±0.003	**0.404±0.005**	0.249±0.002
Weather	96	0.176±0.001	0.216±0.001	0.166±0.002	0.200±0.003	0.161±0.003	0.198±0.003	0.150±0.003	0.187±0.003
	192	0.225±0.002	0.257±0.001	0.214±0.002	**0.245±0.004**	0.205±0.001	0.246±0.002	0.194±0.002	0.230±0.002
	336	0.283±0.002	0.297±0.001	0.271±0.002	0.288±0.002	0.259±0.001	0.283±0.002	0.247±0.003	0.271±0.001
	720	0.360±0.001	0.348±0.001	0.349±0.003	0.340±0.002	0.327±0.002	0.332±0.002	0.317±0.003	0.323±0.003
	avg	0.261±0.002	0.280±0.001	0.250±0.002	0.268±0.003	0.238±0.003	0.265±0.002	0.227±0.003	**0.253±0.004**
Electricity	96	0.150±0.001	0.240±0.002	0.141±0.002	0.231±0.002	0.138±0.002	0.222±0.002	0.130±0.003	0.217±0.003
	192	0.166±0.002	0.258±0.002	0.161±0.002	0.251±0.001	0.154±0.003	0.238±0.002	0.148±0.002	0.234±0.003
	336	0.178±0.001	0.271±0.002	**0.175±0.004**	0.265±0.003	0.173±0.003	0.260±0.003	**0.164±0.004**	0.252±0.003
	720	**0.220±0.006**	**0.310±0.005**	**0.208±0.004**	0.291±0.003	0.208±0.002	**0.289±0.004**	**0.202±0.005**	0.285±0.002
	avg	0.179±0.003	0.270±0.003	0.171±0.003	0.260±0.002	**0.168±0.004**	**0.252±0.004**	0.161±0.002	0.247±0.003

Table 16: Complete combination-wise ablation of the three KFM components. SG and SF denote soft gating and spectral fusion, respectively. +/− indicates that a component is enabled/disabled. All configurations use TimesNet as $\mathcal{F}_{\mathcal{W}}^{low}$. The best results are highlighted in bold.

Component			ETTh1		ETTm1		Weather		Electricity		Traffic		Illness	
SG	SF	KOR	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
−	−	−	0.460	0.455	0.417	0.420	0.259	0.286	0.206	0.302	0.625	0.332	2.140	0.986
+	−	−	0.455	0.449	0.414	0.416	0.256	0.283	0.203	0.297	0.623	0.330	2.107	0.970
−	+	−	0.457	0.451	0.413	0.415	0.257	0.283	0.201	0.296	0.620	0.331	2.083	0.977
−	−	+	0.439	0.437	0.394	0.398	0.253	0.278	0.193	0.292	0.611	0.328	2.000	0.927
+	+	−	0.447	0.445	0.408	0.410	0.253	0.281	0.196	0.294	0.618	0.328	2.001	0.932
+	−	+	0.433	0.432	0.394	0.395	0.249	0.274	0.194	0.291	0.609	0.328	1.965	0.914
−	+	+	0.437	0.435	**0.383**	0.389	0.252	0.276	0.190	0.289	0.607	0.326	1.997	0.948
+	+	+	**0.432**	**0.428**	0.384	**0.388**	**0.248**	**0.272**	**0.186**	**0.284**	**0.603**	**0.324**	**1.945**	**0.880**

Table 17: We conduct replacement experiments on the frequency decomposition module and KOR within our TSF framework, using TimesNet as $\mathcal{F}_{\mathcal{W}}^{low}$. The best results are highlighted in bold, and the second best are underlined.

Dataset (→)		ETTh1		ETTm1		Weather		Electricity		Traffic		Illness	
Origin (↓)	Variant (↓)	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Top-K Soft Gating	Top-K Hard Threshold	0.437	0.435	**0.383**	0.389	0.252	0.276	0.190	0.289	0.607	0.326	1.997	0.948
	Full Spectrum Learnable Mask	0.450	0.447	0.390	0.395	0.257	0.284	0.206	0.303	0.628	0.334	2.111	0.964
	Gumbel Softmax Gating	0.440	0.438	0.385	0.391	0.251	0.275	0.191	0.291	0.626	0.335	1.980	0.917
	Adaptive Shiftable Threshold	0.435	0.432	0.385	0.390	0.251	0.276	0.188	0.288	0.619	0.330	1.988	0.939
KOR	-	0.447	0.445	0.408	0.410	0.253	0.281	0.196	0.294	0.618	0.328	2.001	0.932
	Time Domain MLP	0.440	0.441	0.402	0.408	0.253	0.282	0.198	0.295	0.616	0.330	2.103	0.957
	Freq. Domain MLP	0.435	0.438	0.399	0.401	0.252	0.282	0.196	0.290	0.609	0.326	1.990	0.936
	Time Domain 1D CNN	0.435	0.440	0.392	0.397	0.252	0.280	0.191	0.287	0.610	0.326	1.982	0.925
	Learnable Projection in KOR	0.433	**0.425**	0.384	**0.388**	**0.247**	0.273	0.188	0.288	0.606	0.325	1.953	0.891
Full Model		**0.432**	0.428	0.384	**0.388**	0.248	**0.272**	**0.186**	**0.284**	**0.603**	**0.324**	**1.945**	**0.880**
TimesNet (Baseline)		0.460	0.455	0.417	0.420	0.259	0.286	0.206	0.302	0.625	0.332	2.140	0.986

Top-K hard thresholding by approximately 3% on the Traffic dataset. These results suggest that the energy of introduced high frequencies must be carefully controlled during frequency decomposition. This prevents high-frequency patterns from becoming dominant and interfering with the low-frequency module in capturing periodicity. The shiftable threshold approach achieves performance closest to our soft-gating method. However, a performance gap of 1% to 2% remains. These results support Top-K truncation as a stable decomposition prior rather than establishing hard thresholding as universally optimal. The remaining gains arise from minor and controllable data-dependent modulation beyond this prior.

In the module replacement experiments for the high-frequency processing module, we compare KOR with a time-domain MLP, a frequency-domain MLP, and an xPatch-inspired 1D CNN with residual connections (Stitsyuk & Choi, 2025). The remaining KFM pipeline is kept unchanged so that the experiment isolates the choice of high-frequency encoder. KOR consistently outperforms these neural alternatives across the evaluated datasets and metrics, indicating that its gains are not explained solely by adding another high-frequency processing branch.

Making the projection layers in KOR learnable does not improve the average results and increases MAE by approximately 0.5%. This suggests that frozen projections are sufficient for exposing the oscillator responses while avoiding additional trainable parameters and gradient-storage overhead.

I.3 Frequency Decomposition Visualization

On the ETTm1 dataset ($T = 96$ forecasting task) using TSMixer as $\mathcal{F}_{\mathcal{W}}^{low}$, we visualize the frequency decomposition mask matrix M (formulated in Equation 1) for a specific test sample during inference. We observe that initializing the gating MLP bias with a large β significantly enhances the sharpness of M. Furthermore, the mask weights exhibit substantial channel-wise variance, indicating the effectiveness of the channel-independent design. For instance, the mask values for channels 1, 3, and 7 are predominantly below 0.2. In contrast, for the remaining channels, the mask values generally exceed 0.6 beyond the 23rd frequency bin, with multiple high-frequency components in channels 4 and 6 approaching 1.0.

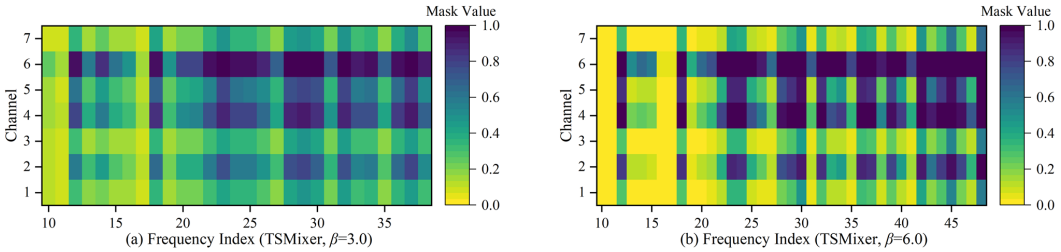


Figure 12: Heatmap of the frequency extraction mask $\mathbf{M}$ for a randomly picked test sample on the ETTm1 dataset.

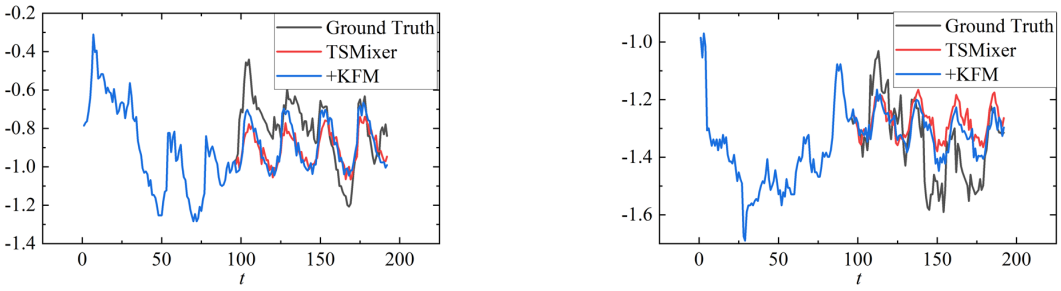
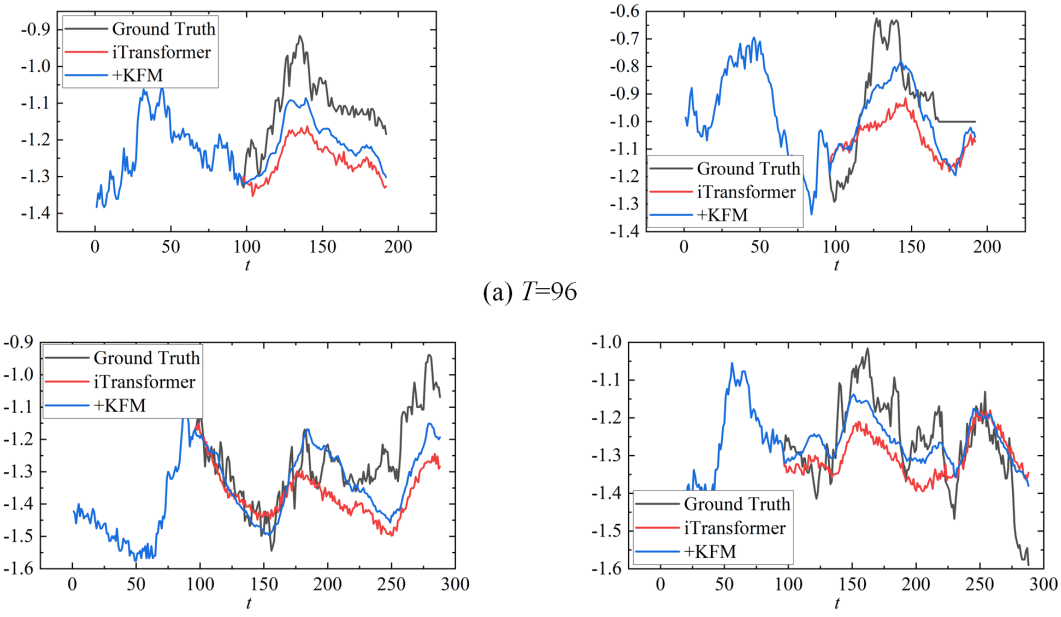


Figure 13: Comparison between ground truth and forecasting results on ETTh1 dataset configuring $\mathcal{F}_{\mathcal{W}}^{low}$ as TSMixer with a look-back window $L = 96$ and target window $T = 96$.

Notably, this phenomenon contradicts the conventional intuition that frequency components closer to the Top-K threshold are more likely to be gated into the low-frequency branch. Taking channel 4 as an example, the average spectral energy of frequencies beyond the 23rd bin is merely 19.4% of the energy within the 10th to 23rd bins. This indicates an inverse energy-weighting mechanism: low-energy, high-frequency components further from the threshold are preferentially assigned higher mask values to modulate the low-frequency representations. Ultimately, the following modules can leverage these entropy-rich components to accurately forecast future non-stationary dynamics without disrupting the global smooth trend and primary periodicity. This also explains the underperformance of the shiftable threshold in Table 17, where a monotonic cutoff cannot select distant, channel-dependent high-frequency bins.

J Test Results Visualization

We randomly select test samples from the ETTh1, ETTm1, Electricity, and Traffic datasets to contrast the ground truth against the predictions generated by the original baseline models and the KFM framework in Figures 13 through 16, respectively. As observed in Figure 14, the iTransformer attempts to fit micro-fluctuations and noise components, making its prediction curve to exhibit jagged artifacts that deviate significantly from the true distribution and ultimately fail to capture macroscopic trends. Conversely, applying KFM successfully extracts the structured envelope and notably rectifies this flaw. For extended forecasting windows T as illustrated in Figures 15 and 16, KFM effectively calibrates prediction errors in terminal regions distant from the look-back window while achieving significantly higher precision in fitting peak and valley extrema.



(a) T=96

(b) T=192

Figure 14: Comparison between ground truth and forecasting results on ETTm1 dataset configuring $\mathcal{F}_{\mathcal{W}}^{low}$ as iTransformer with a look-back window $L = 96$.

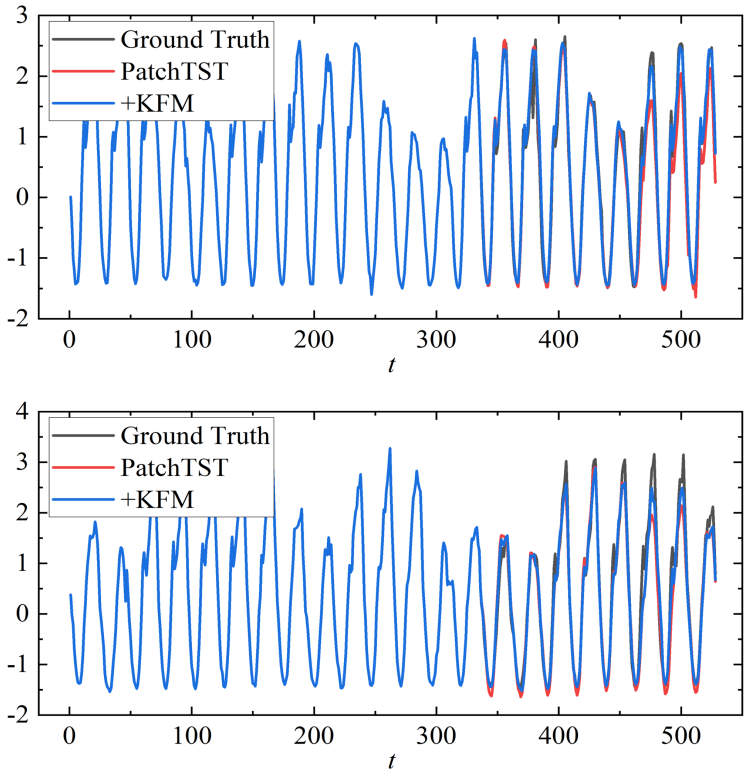


Figure 15: Comparison between ground truth and forecasting results on Traffic dataset configuring $\mathcal{F}_{\mathcal{W}}^{low}$ as PatchTST with a look-back window $L = 96$ and target window $T = 192$.

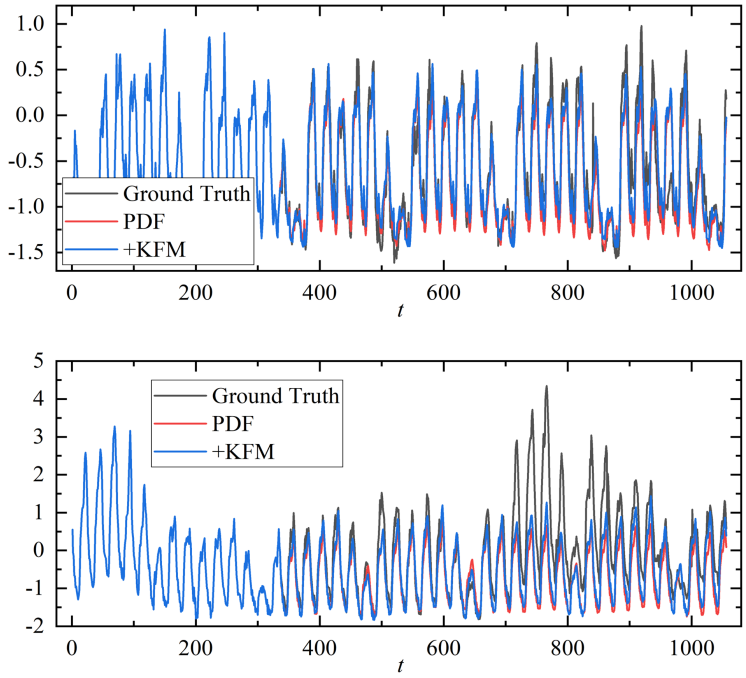


Figure 16: Comparison between ground truth and forecasting results on Electricity dataset configuring $\mathcal{F}_{\mathcal{W}}^{low}$ as PDF with a look-back window $L = 336$ and target window $T = 720$.