ADAPTIVE PHASE ALIGNMENT: CORRECTING PHASE DRIFT IN MULTIVARIATE TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Phase drift is pervasive in real-world periodic time series, causing historical periodic structures to become misaligned with future observations and degrading multivariate forecasting performance. Existing methods for modeling periodic patterns often adopt fixed phase assignments within each cycle or establish correspondence using a shared set of phase indices. Such representations do not characterize frequency-specific phase variations or adapt periodic correspondence to their changes across observation windows, and may consequently preserve unreliable temporal information corrupted by phase drift and noise. To address these limitations, we characterize phase drift through the window-dependent variation of phase offsets across individual frequency components, providing a basis for adapting periodic representations to changing phase structures. Building on this formulation, we propose Adaptive Phase Alignment (APA), a lightweight and plug-and-play module for correcting phase drift. APA decomposes each variate into frequency components and applies a compact complex-valued affine mapping whose relative phase response depends on the current spectral state, allowing shared parameters to induce component- and window-dependent phase variations. It then converts these variations into representation-level phase alignment through temporal gating, which adaptively reweights periodic information across temporal positions. The resulting representation enhances reliable periodic cues while attenuating information corrupted by phase drift and noise. Extensive experiments on nine real-world benchmarks demonstrate state-of-the-art forecasting performance and consistent improvements across diverse forecasting backbones. Notably, APA achieves strong performance even with a shallow MLP backbone while introducing low computational overhead.

1 INTRODUCTION

Time series forecasting plays a central role in applications including energy management (Martinez Alvarez et al., 2011; Lai et al., 2018; Cheng et al., 2022), traffic planning (Cirstea et al., 2021; Wu et al., 2021b), weather forecasting (Wu et al., 2021a; Qureshi et al., 2025), and financial analysis (Huang et al., 2024; Koa et al., 2026). Real-world time series are often non-stationary, and their periodic structures may exhibit phase drift as recurring patterns evolve over time. As illustrated in Figure 1 (a), phase drift changes the phase offset of a periodic pattern, misaligning historical structures with future observations and reducing the reliability of periodic cues. Accurately characterizing and correcting such drift is therefore important for multivariate forecasting.

Existing methods establish periodic correspondence in two ways. The first uses fixed phase assignments within each cycle (Lin et al., 2024; 2025), which can become mismatched under phase drift, as illustrated in Figure 1 (b). The second uses globally shared phase indices across cycles (Wang et al., 2026a; Niu et al., 2026), which do not capture how phase correspondence varies across individual frequency components and observation windows and may therefore provide only partial alignment, as shown in Figure 1 (c) and (d). Thus, existing approaches rely on fixed or globally shared phase correspondence, raising a question: Can periodic correspondence be modeled at the level of individual frequency components while adapting to phase variations across forecasting windows?

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.

To address this limitation, we characterize phase drift through frequency-specific phase variations that change across observation windows. Based on this formulation, we propose Adaptive Phase Alignment (APA), a lightweight and plug-and-play module for correcting phase drift. APA applies a compact complex-valued affine mapping whose relative phase response depends on the spectral state, allowing shared parameters to induce phase variations across frequencies and observation windows. These variations are then converted into representation-level phase alignment through temporal gating, which adaptively reweights periodic information according to the current phase structure. APA can be integrated with different forecasting backbones at low computational overhead.

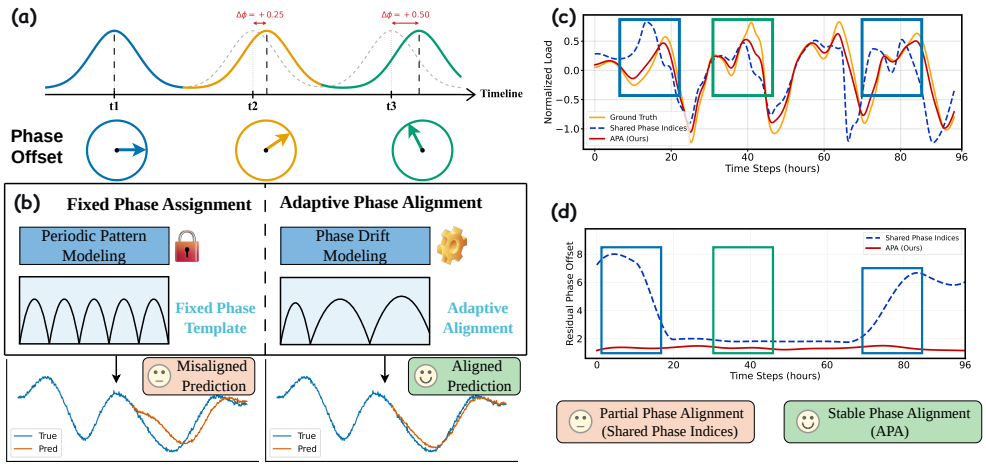


Figure 1: Motivation for Adaptive Phase Alignment (APA) under phase drift. (a) A periodic pattern can retain a similar structure while its phase offset evolves over time, causing historical patterns to become misaligned with future observations. (b) Fixed phase assignments reuse a fixed phase template and can lead to misaligned predictions under phase drift, whereas APA adapts the periodic representation to the current phase structure. (c) Forecasting results using shared phase indices and APA. Shared phase indices provide partial alignment over some temporal intervals but become mismatched as the phase changes, whereas APA more closely follows the ground truth. (d) Corresponding residual phase offsets. Shared phase indices exhibit larger and less stable residual phase offsets, whereas APA maintains lower and more stable residual phase offsets over time.

Our main contributions are summarized as follows:

- We characterize phase drift through the window-dependent variation of frequency-specific phase offsets, distinguishing it from fixed or globally shared phase correspondence.
- We propose Adaptive Phase Alignment (APA), a lightweight and plug-and-play module that induces frequency-specific and window-dependent phase responses with a shared complex-valued affine mapping and realizes representation-level phase alignment through adaptive temporal gating.
- Experiments on nine real-world multivariate datasets demonstrate state-of-the-art performance and consistent improvements across representative backbones. Controlled phase-drift experiments further validate phase drift modeling and alignment.

2 Related Work

2.1 Periodicity Modeling

Periodic structures are widely exploited in time series forecasting. Autoformer (Wu et al., 2021a) captures recurring dependencies through decomposition and Auto-Correlation, TimesNet (Wu et al., 2023) models multiple periods, while CycleNet (Lin et al., 2024) and MoFo (Ma et al., 2025) explicitly organize recurring patterns across cycles.

Frequency-domain methods expose spectral components underlying periodicity. FEDformer (Zhou et al., 2022) and FITS (Xu et al., 2024) use frequency representations for decomposition or prediction, while FAN (Ye et al., 2024) and FilterTS (Wang et al., 2025) adapt or filter frequency components. CFPT (Kou et al., 2025) and FreqCycle (Zhang et al., 2026a) model cross-frequency or multi-periodic structures. These methods enhance spectral representations, but how phase correspondence varies across frequency components and observation windows remains less characterized.

2.2 Phase Modeling and Alignment

Temporal variation has been studied through differentiable alignment and distortion modeling (Cuturi & Blondel, 2017; Guen & Thome, 2019; Tallec & Ollivier, 2018), as well as temporal-spectral and dynamic modeling (Yang & Hong, 2022; Yi et al., 2023; Liu et al., 2023). Phase-aware methods include temporal-spectral reconstruction (Zhang et al., 2024), phase-shift equivariance (Gui et al., 2024), phase-indexed forecasting (Wang et al., 2026a; Niu et al., 2026), phase-aligned periodic structures (Ma et al., 2026), and frequency-amplitude-phase modeling (Li et al., 2025). Recent studies further address periodic or phase drift more directly. Correctformer (Wang et al., 2026b) introduces periodic embedding and correction to mitigate attention-induced periodic drift, while AOSNET (Song et al., 2026) performs adaptive oscillatory-state alignment and POEM (Zhu et al., 2026) applies phase-aware latent feature rotation under periodicity drift. However, phase correspondence at the level of individual frequency components and its variation across observation windows remain less characterized. APA addresses this gap by inducing component- and window-dependent phase responses from the current spectral state and adapting periodic representations through temporal gating.

3 Method

3.1 Overview

At forecasting time t, given a historical observation window

$$\boldsymbol{X}^{(t)} = \{\boldsymbol{x}_{t-T+1}, \ldots, \boldsymbol{x}_t\} \in \mathbb{R}^{T\times N}, \tag{1}$$

multivariate time series forecasting aims to predict the subsequent S time steps

$$\boldsymbol{Y}^{(t)} = \{\boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_{t+S}\} \in \mathbb{R}^{S\times N}, \tag{2}$$

where T denotes the historical length, N is the number of variables, and S is the forecasting horizon.

APA models phase drift at the level of individual frequency components and adapts periodic representations to their window-dependent phase variations. It first applies a compact complex-valued affine mapping in the frequency domain, whose response depends on the spectral state of the current window, and maps the transformed representation back to the temporal domain as $\boldsymbol{R}^{(t)}$. A temporal gate derived from $\boldsymbol{R}^{(t)}$ then adaptively reweights a base periodic representation, yielding the aligned representation $\boldsymbol{H}^{(t)}$, which is used for forecasting. Figure 2 illustrates the two stages of APA.

As a lightweight and plug-and-play module, APA performs phase drift modeling and representation-level alignment before forecasting and can be integrated with different forecasting backbones. The complete computation is summarized in Appendix A.1.

3.2 Frequency-Specific Phase Drift Modeling

Periodic structures generally contain multiple frequency components whose phase states need not vary identically across observation windows. We therefore characterize phase drift through component-dependent phase responses that adapt to the spectral state of each window, rather than through a fixed or globally shared phase correspondence.

We decompose the historical observation window $\boldsymbol{X}^{(t)}$ along the temporal dimension:

$$\boldsymbol{Z}^{(t)} = \mathcal{F}\left(\boldsymbol{X}^{(t)}\right) \in \mathbb{C}^{K\times N}, \tag{3}$$

where $\mathcal{F}(\cdot)$ denotes the real Fourier transform and $K = \lfloor T/2 \rfloor + 1$. Each coefficient $Z^{(t)}_{k,n}$ encodes the magnitude and phase of frequency component k for variate n.

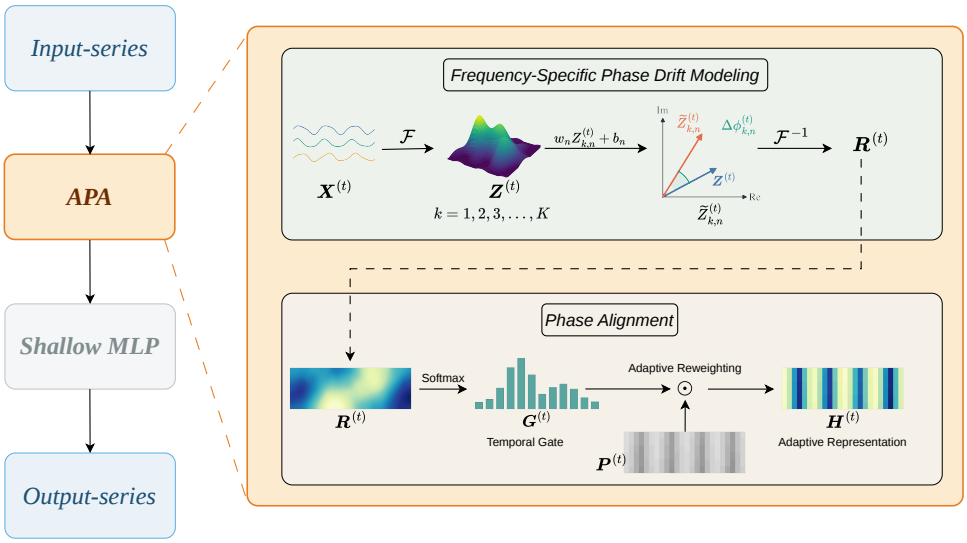


Figure 2: Overview of Adaptive Phase Alignment (APA). APA induces frequency-specific and window-dependent phase responses through a compact complex-valued affine mapping in the frequency domain and maps the transformed representation back to the temporal domain as $\boldsymbol{R}^{(t)}$. The resulting temporal gate $\boldsymbol{G}^{(t)}$ adapts the contribution of the base periodic representation $\boldsymbol{P}^{(t)}$ across temporal positions, yielding $\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}$. The relative phase variation $\Delta\phi_{k,n}^{(t)}$ characterizes the phase response induced for frequency component k of variate n.

To obtain adaptive phase responses with few parameters, we apply a compact complex-valued affine mapping:

$$\widetilde{Z}_{k,n}^{(t)} = w_n Z_{k,n}^{(t)} + b_n, \tag{4}$$

where $w_n, b_n \in \mathbb{C}$ are learnable parameters for variate n and are shared across its frequency components. Although the parameters are frequency-shared, the resulting response remains component-dependent because the mapping interacts with the spectral coefficient $Z_{k,n}^{(t)}$.

For nonzero original and transformed spectral coefficients, the relative phase variation induced by the mapping is

$$\Delta\phi_{k,n}^{(t)} = \arg\left(\frac{\widetilde{Z}_{k,n}^{(t)}}{Z_{k,n}^{(t)}}\right) = \arg\left(w_n + \frac{b_n}{Z_{k,n}^{(t)}}\right). \tag{5}$$

Equation (5) makes the source of adaptivity explicit. The same parameter pair (w_n, b_n) produces different phase responses because $Z_{k,n}^{(t)}$ varies across both frequency components and observation windows. Thus, frequency specificity and temporal adaptivity are induced by the current spectral state rather than by frequency-specific parameters or explicit time-dependent phase tracking.

This distinction can also be seen from the special case $b_n = 0$, for which

$$\Delta\phi_{k,n}^{(t)} = \arg(w_n), \tag{6}$$

corresponding to a fixed phase rotation for variate n. The affine term $b_n / Z_{k,n}^{(t)}$ therefore enables the response to vary with both the component and the observation window while requiring only one shared parameter pair per variate. The effect of this adaptive term is further examined with a *w/o Adaptive Bias* ablation in Appendix A.6.

The transformed spectral representation is mapped back to the temporal domain:

$$\boldsymbol{R}^{(t)} = \mathcal{F}^{-1}\left(\widetilde{\boldsymbol{Z}}^{(t)}\right) \in \mathbb{R}^{T \times N}, \tag{7}$$

where $\mathcal{F}^{-1}(\cdot)$ denotes the corresponding inverse real Fourier transform. The resulting $\boldsymbol{R}^{(t)}$ transfers the spectral-state-dependent response to the temporal domain and is subsequently used to construct adaptive alignment weights.

3.3 PHASE ALIGNMENT

APA realizes phase alignment at the representation level rather than by explicitly estimating a phase shift or temporally shifting or warping the input sequence. Specifically, it uses the spectral-state-dependent response of the current observation window to adapt the temporal contribution of a periodically indexed representation.

We introduce a learnable periodic template $\boldsymbol{C} \in \mathbb{R}^{L\times N}$, where L denotes the intrinsic period length of the recurring structure. The period length L and historical length T characterize different temporal scales: T determines the amount of history observed for each forecasting instance, whereas L specifies the temporal extent of the underlying periodic cycle. Therefore, L is not required to satisfy $L \leq T$. When $L > T$, each observation window covers only a local segment of a longer periodic cycle. The selection of L is detailed in Appendix A.2.

To preserve the position of each observation window within the periodic cycle, we associate the window ending at time t with a cyclic starting index

$$s^{(t)} = (t - T) \bmod L, \tag{8}$$

where $s^{(t)} \in \{0, \ldots, L-1\}$ denotes the periodic position immediately preceding the first observation in the current historical window. Accordingly, for the τ-th temporal position, the base periodic representation is defined as

$$P^{(t)}_{\tau,n} = C_{\left((s^{(t)}+\tau-1) \bmod L\right)+1,n}, \tag{9}$$

yielding $\boldsymbol{P}^{(t)} \in \mathbb{R}^{T\times N}$. Thus, different observation windows extract different but cycle-consistent segments from the same learnable template. In particular, when $L > T$, only the segment associated with the current periodic position is used, while windows separated by one complete period share the same template correspondence. This construction preserves dataset-level periodic correspondence without requiring a complete cycle to appear within every historical window.

The base representation $\boldsymbol{P}^{(t)}$ provides cycle-consistent periodic correspondence, but by itself cannot adapt to changes in the spectral phase structure across observation windows. To introduce window-dependent adaptation, we derive a temporal gate from $\boldsymbol{R}^{(t)}$:

$$\boldsymbol{G}^{(t)}_{:,n} = \mathrm{Softmax}\left(\boldsymbol{R}^{(t)}_{:,n}\right), \tag{10}$$

where the Softmax is applied along the temporal dimension and $\boldsymbol{G}^{(t)} \in \mathbb{R}^{T\times N}$ contains normalized temporal weights for each variate. Because $\boldsymbol{R}^{(t)}$ is obtained from the transformed spectrum of the current observation window, $\boldsymbol{G}^{(t)}$ varies across windows and adaptively changes the relative contribution of different temporal positions.

The adaptive periodic representation is then obtained as

$$\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}, \tag{11}$$

where $\odot$ denotes elementwise multiplication. The periodic template provides cycle-consistent correspondence, whereas the temporal gate adaptively modulates its effective contribution according to the current spectral state. Temporal positions assigned larger gate values contribute more strongly to the resulting representation, while positions receiving smaller weights are attenuated.

Importantly, APA does not treat $\boldsymbol{G}^{(t)}$ as an explicit estimator of a scalar phase offset. The complex-valued spectral transformation can affect both the magnitude and phase of the spectral coefficients, while Equation (5) characterizes the relative phase response induced for each frequency component. APA therefore uses these spectral-state-dependent responses to adapt the effective temporal emphasis of a periodically indexed representation, rather than explicitly translating the input sequence or the periodic template. We refer to this mechanism as *representation-level phase alignment*.

The two stages of APA are consequently coupled. The temporal gate $\boldsymbol{G}^{(t)}$ is derived from $\boldsymbol{R}^{(t)}$, whose frequency-specific and window-dependent phase response is characterized by Equation (5).

Frequency-specific phase drift modeling determines how the current spectral state modifies individual frequency components, while temporal gating converts the resulting response into adaptive representation-level alignment of the underlying periodic structure.

3.4 Forecasting with a Shallow MLP

Given the adaptive representation $\boldsymbol{H}^{(t)}$, we use a shallow MLP as the default forecasting backbone following the effectiveness of MLP architectures for time series forecasting (Das et al., 2023):

$$\boldsymbol{H}^{(t)}_{\mathrm{MLP}} = \mathrm{Linear}_2\left(\mathrm{GELU}\left(\mathrm{Linear}_1\left(\boldsymbol{H}^{(t)}\right)\right)\right) \in \mathbb{R}^{D\times N}, \tag{12}$$

where $\mathrm{Linear}_1 : \mathbb{R}^T \to \mathbb{R}^D$ and $\mathrm{Linear}_2 : \mathbb{R}^D \to \mathbb{R}^D$ operate along the temporal dimension.

The prediction is obtained by

$$\hat{\boldsymbol{Y}}^{(t)} = \mathrm{Linear}_{\mathrm{out}}\left(\mathrm{Dropout}\left(\boldsymbol{H}^{(t)}_{\mathrm{MLP}}\right)\right) \in \mathbb{R}^{S\times N}, \tag{13}$$

where $\mathrm{Linear}_{\mathrm{out}} : \mathbb{R}^D \to \mathbb{R}^S$. The shallow backbone keeps the forecasting stage simple, while APA itself is backbone-agnostic and can be integrated with other forecasting architectures.

4 Experiment

4.1 Experiment Setup

Datasets. We evaluate APA on nine multivariate forecasting benchmarks: ETTh1 and ETTm1 (Zhou et al., 2021), PEMS03, PEMS04, PEMS07, and PEMS08 (Liu et al., 2022), and Electricity, Solar, and Weather (Wu et al., 2021a). We use forecasting horizons $S \in \{12, 24, 48, 96\}$ for PEMS and $S \in \{96, 192, 336, 720\}$ for the other datasets, with the historical length T fixed at 96.

Baselines. We compare APA with representative forecasting methods, including PhaseFormer (Niu et al., 2026), Time-TK (Zhang et al., 2026b), TQNet (Lin et al., 2025), CMoS (Si et al., 2025), CycleNet (Lin et al., 2024), iTransformer (Liu et al., 2024), PatchTST (Nie et al., 2023), and DLinear (Zeng et al., 2023).

Implementation Details. We use mean squared error (MSE) and mean absolute error (MAE) as evaluation metrics. All experiments are implemented in PyTorch (Paszke et al., 2019) and conducted on a single NVIDIA RTX 3090 24GB GPU. We optimize models using Adam (Kingma & Ba, 2015). Detailed experimental settings, dataset-specific hyperparameters, and the selection of the period length L are provided in Appendix A.2.

4.2 Multivariate Forecasting Results

We evaluate APA in terms of its applicability across forecasting backbones and its performance against representative state-of-the-art methods.

General applicability of APA. We integrate APA into eight representative forecasting backbones. As shown in Table 1, APA consistently improves their average performance, with MSE/MAE reductions of 40.1%/30.9% on DLinear and 23.5%/14.9% on CMoS. APA also improves PhaseFormer, which already incorporates phase modeling, indicating that frequency-specific and window-dependent phase responses provide complementary benefits.

Comparison with representative methods. We compare APA with representative forecasting methods on nine benchmark datasets. As reported in Table 2, APA achieves the best performance in 16 of 18 cases. Compared with methods based on fixed periodic representations, such as CycleNet and TQNet, APA adapts periodic information according to window-dependent phase variations. It also improves upon phase modeling methods such as PhaseFormer, suggesting that frequency-specific and window-dependent phase responses provide complementary information for representation-level phase alignment. These results demonstrate the effectiveness of frequency-specific phase drift modeling and representation-level alignment for multivariate forecasting. We additionally report results over five random initializations in Appendix A.5 to examine the robustness of APA to random initialization.

Table 1: General applicability of APA across different forecasting architectures. The reported results are averaged across all four forecasting horizons. The historical length T is fixed at 96. The cells highlighted in gray indicate the results after integrating APA. The last row reports the average relative improvement over the corresponding backbone.

Dataset	Metric	DLinear (2023)		PatchTST (2023)		iTransformer (2024)		CycleNet (2024)		CMoS (2025)		TQNet (2025)		Time-TK (2026b)		PhaseFormer (2026)	
		Base	+APA	Base	+APA	Base	+APA	Base	+APA	Base	+APA	Base	+APA	Base	+APA	Base	+APA
ETTh1	MSE	0.456	**0.429** (↑5.9%)	0.469	**0.430** (↑8.3%)	0.454	**0.428** (↑5.7%)	0.457	**0.433** (↑5.3%)	0.448	**0.434** (↑3.1%)	0.441	**0.429** (↑2.7%)	0.432	**0.429** (↑0.7%)	0.430	**0.429** (↑0.2%)
	MAE	0.452	**0.429** (↑5.1%)	0.455	**0.430** (↑5.5%)	0.448	**0.431** (↑3.8%)	0.441	**0.430** (↑2.5%)	0.442	**0.432** (↑2.3%)	0.434	**0.429** (↑1.2%)	0.430	**0.430** (↑0.0%)	0.430	**0.430** (↑0.0%)
ETTm1	MSE	0.403	**0.370** (↑8.2%)	0.387	**0.373** (↑3.6%)	0.407	**0.373** (↑8.4%)	0.379	**0.368** (↑2.9%)	0.412	**0.369** (↑10.4%)	0.377	**0.370** (↑1.9%)	0.379	**0.367** (↑3.2%)	0.376	**0.370** (↑1.6%)
	MAE	0.407	**0.394** (↑3.2%)	0.400	**0.394** (↑1.5%)	0.410	**0.394** (↑3.9%)	0.396	**0.391** (↑1.3%)	0.410	**0.392** (↑4.4%)	0.393	**0.392** (↑0.3%)	0.393	**0.390** (↑0.8%)	0.395	**0.394** (↑0.3%)
Electricity	MSE	0.212	**0.166** (↑21.7%)	0.205	**0.167** (↑18.5%)	0.178	**0.170** (↑4.5%)	0.168	**0.167** (↑0.6%)	0.204	**0.166** (↑18.6%)	0.164	**0.162** (↑1.2%)	0.174	**0.166** (↑4.6%)	0.170	**0.166** (↑2.4%)
	MAE	0.300	**0.259** (↑13.7%)	0.290	**0.262** (↑9.7%)	0.270	**0.264** (↑2.2%)	0.259	**0.259** (↑0.0%)	0.284	**0.258** (↑9.2%)	0.259	**0.256** (↑1.2%)	0.265	**0.258** (↑2.6%)	0.262	**0.258** (↑1.5%)
Solar	MSE	0.330	**0.202** (↑38.8%)	0.270	**0.231** (↑14.4%)	0.233	**0.202** (↑13.3%)	0.210	**0.198** (↑5.7%)	0.332	**0.201** (↑39.5%)	0.198	**0.189** (↑4.5%)	0.205	**0.191** (↑6.8%)	0.200	**0.194** (↑3.0%)
	MAE	0.401	**0.249** (↑37.9%)	0.307	**0.282** (↑8.1%)	0.262	**0.251** (↑4.2%)	0.261	**0.245** (↑6.1%)	0.322	**0.248** (↑23.0%)	0.256	**0.249** (↑2.7%)	0.257	**0.244** (↑5.1%)	0.247	**0.241** (↑2.4%)
Weather	MSE	0.265	**0.239** (↑9.8%)	0.259	**0.239** (↑7.7%)	0.258	**0.240** (↑7.0%)	0.243	**0.239** (↑1.6%)	0.251	**0.239** (↑4.8%)	0.242	**0.240** (↑0.8%)	0.256	**0.239** (↑6.6%)	0.236	**0.228** (↑3.4%)
	MAE	0.317	**0.267** (↑15.8%)	0.273	**0.268** (↑1.8%)	0.278	**0.269** (↑3.2%)	0.271	**0.267** (↑1.5%)	0.278	**0.267** (↑4.0%)	0.269	**0.267** (↑0.7%)	0.278	**0.266** (↑4.3%)	0.270	**0.264** (↑2.2%)
PEMS03	MSE	0.278	**0.089** (↑68.0%)	0.180	**0.091** (↑49.4%)	0.113	**0.086** (↑23.9%)	0.118	**0.090** (↑23.7%)	0.147	**0.090** (↑38.8%)	0.097	**0.086** (↑11.3%)	0.112	**0.079** (↑29.5%)	0.101	**0.086** (↑14.9%)
	MAE	0.375	**0.190** (↑49.3%)	0.291	**0.195** (↑33.0%)	0.222	**0.191** (↑14.0%)	0.226	**0.190** (↑15.9%)	0.253	**0.189** (↑25.3%)	0.203	**0.188** (↑7.4%)	0.219	**0.181** (↑17.4%)	0.203	**0.186** (↑8.4%)
PEMS04	MSE	0.295	**0.089** (↑69.8%)	0.195	**0.094** (↑51.8%)	0.111	**0.081** (↑27.0%)	0.119	**0.085** (↑28.6%)	0.124	**0.086** (↑30.6%)	0.091	**0.078** (↑14.3%)	0.109	**0.078** (↑28.4%)	0.090	**0.077** (↑14.4%)
	MAE	0.388	**0.193** (↑50.3%)	0.307	**0.202** (↑34.2%)	0.221	**0.184** (↑16.7%)	0.232	**0.186** (↑19.8%)	0.249	**0.187** (↑24.9%)	0.197	**0.179** (↑9.1%)	0.218	**0.178** (↑18.3%)	0.195	**0.178** (↑8.7%)
PEMS07	MSE	0.329	**0.075** (↑77.2%)	0.211	**0.081** (↑61.6%)	0.101	**0.084** (↑16.8%)	0.113	**0.076** (↑32.7%)	0.154	**0.075** (↑51.3%)	0.075	**0.073** (↑2.7%)	0.093	**0.066** (↑29.0%)	0.085	**0.073** (↑14.1%)
	MAE	0.396	**0.167** (↑57.8%)	0.303	**0.177** (↑41.6%)	0.204	**0.179** (↑12.3%)	0.214	**0.166** (↑22.4%)	0.247	**0.167** (↑32.4%)	0.171	**0.166** (↑2.9%)	0.195	**0.157** (↑19.5%)	0.190	**0.172** (↑9.5%)
PEMS08	MSE	0.379	**0.146** (↑61.5%)	0.280	**0.117** (↑58.2%)	0.150	**0.121** (↑19.3%)	0.150	**0.134** (↑10.7%)	0.176	**0.151** (↑14.2%)	0.142	**0.134** (↑5.6%)	0.145	**0.116** (↑20.0%)	0.141	**0.127** (↑9.9%)
	MAE	0.416	**0.230** (↑44.7%)	0.321	**0.210** (↑34.6%)	0.226	**0.207** (↑8.4%)	0.246	**0.218** (↑11.4%)	0.255	**0.232** (↑9.0%)	0.229	**0.216** (↑5.7%)	0.224	**0.201** (↑10.3%)	0.228	**0.217** (↑4.8%)
Avg. Impv.		40.1% / 30.9%		30.4% / 18.9%		14.0% / 7.6%		12.4% / 9.0%		23.5% / 14.9%		5.0% / 3.5%		14.3% / 8.7%		7.1% / 4.2%	

Table 2: Multivariate forecasting results on nine real-world datasets. The reported results are averaged across all four forecasting horizons, with the historical length T fixed at 96. The best results are highlighted in **bold**, while the second-best results are underlined. *Count* reports the number of best results across datasets and metrics.

Model	APA (Ours)		PhaseFormer (2026)		Time-TK (2026b)		TQNet (2025)		CMoS (2025)		CycleNet (2024)		iTransformer (2024)		PatchTST (2023)		DLinear (2023)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.422**	**0.426**	0.430	0.430	0.432	0.430	0.441	0.434	0.448	0.442	0.457	0.441	0.454	0.448	0.469	0.455	0.456	0.452
ETTm1	**0.367**	**0.391**	0.376	0.395	0.379	0.393	0.377	0.393	0.412	0.410	0.379	0.396	0.407	0.410	0.387	0.400	0.403	0.407
Electricity	0.166	**0.258**	0.170	0.262	0.174	0.265	**0.164**	0.259	0.204	0.284	0.168	0.259	0.178	0.270	0.205	0.290	0.212	0.300
Solar	**0.197**	**0.245**	0.200	0.247	0.205	0.257	0.198	0.256	0.332	0.322	0.210	0.261	0.233	0.262	0.270	0.307	0.330	0.401
Weather	0.239	**0.267**	**0.236**	0.270	0.256	0.278	0.242	0.269	0.251	0.278	0.243	0.271	0.258	0.278	0.259	0.273	0.265	0.317
PEMS03	**0.087**	**0.188**	0.101	0.203	0.112	0.219	0.097	0.203	0.147	0.253	0.118	0.226	0.113	0.222	0.180	0.291	0.278	0.375
PEMS04	**0.085**	**0.187**	0.090	0.195	0.109	0.218	0.091	0.197	0.124	0.249	0.119	0.232	0.111	0.221	0.195	0.307	0.295	0.388
PEMS07	**0.075**	**0.167**	0.085	0.190	0.093	0.195	0.075	0.171	0.154	0.247	0.113	0.214	0.101	0.204	0.211	0.303	0.329	0.396
PEMS08	**0.133**	**0.217**	0.141	0.228	0.145	0.224	0.142	0.229	0.176	0.255	0.150	0.246	0.150	0.226	0.280	0.321	0.379	0.416
Count	**16**		1		0		1		0		0		0		0		0	

4.3 Visualization of Phase Modeling and Alignment

Figure 3 compares periodic representations on PEMS03 for a representative test window. APA uses the aligned representation $\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}$, while the fixed-phase baseline retains its periodic representation without adaptive alignment. All representations are shown over the same temporal positions. Fixed phase assignments remain anchored to predefined positions and become misaligned as the phase changes, whereas APA adapts the periodic representation to the current phase structure.

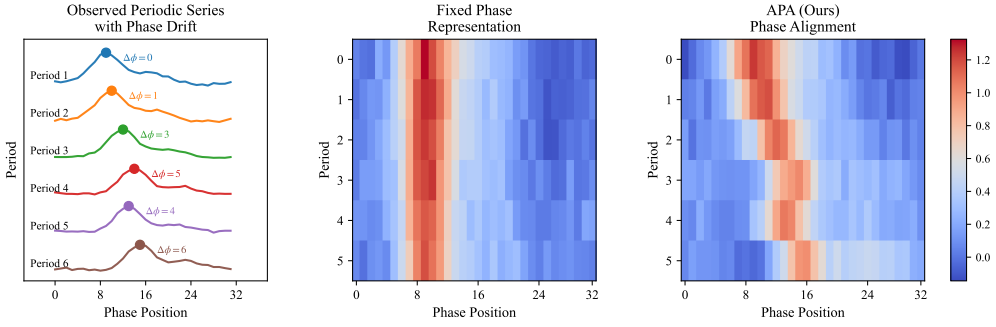


Figure 3: Periodic representations on PEMS03 under fixed phase assignments (TQNet (2025)) and Adaptive Phase Alignment (APA).

Figure 4 further examines phase alignment through multivariate correlation structures on four real-world benchmarks. For each representation, we compute pairwise Pearson correlations between variates along the temporal dimension using the same test window and set of variates. Shared phase indices show correlation structures that become less consistent with the ground truth as the phase changes, whereas APA better reflects the phase structure of the current observation window.

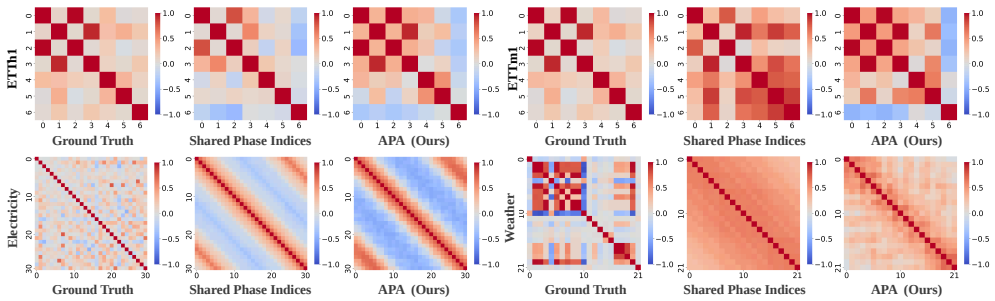


Figure 4: Phase alignment visualization on four real-world benchmarks comparing ground truth, shared phase indices (PhaseFormer (2026)), and APA.

4.4 Ablation and Phase Drift Analysis

Component Ablation. To assess the contributions of the two core components of APA, we construct two ablated variants by removing each stage in turn. Specifically, *w/o Drift Modeling* removes the frequency-specific phase drift modeling stage, while *w/o Alignment* removes the subsequent representation-level alignment based on temporal gating.

As shown in Table 3, removing either component consistently degrades performance. The average MSE increases from 0.197 for APA to 0.224 without phase drift modeling and 0.222 without phase alignment. These results indicate that the two stages play complementary roles: phase drift modeling produces frequency-specific and window-dependent responses, while temporal gating converts them into adaptive weighting of periodic representations. The Shallow MLP performs worst overall, further confirming the contribution of APA beyond the forecasting backbone.

Table 3: Ablation study of phase drift modeling and alignment in APA.

Model	ETTh1		ETTm1		Electricity		Solar		Weather		PEMS03		PEMS04		PEMS07		PEMS08		Average	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
APA (Ours)	**0.422**	**0.426**	**0.367**	**0.391**	**0.166**	**0.258**	**0.197**	**0.245**	**0.239**	**0.267**	**0.087**	**0.188**	**0.085**	**0.187**	**0.075**	**0.167**	**0.133**	**0.217**	**0.197**	**0.261**
w/o Drift Modeling	0.465	0.447	0.382	0.399	0.193	0.283	0.205	0.261	0.249	0.276	0.101	0.201	0.105	0.206	0.093	0.187	0.220	0.291	0.224	0.283
w/o Alignment	0.443	0.432	0.383	0.397	0.179	0.268	0.206	0.251	0.246	0.274	0.115	0.217	0.117	0.222	0.111	0.202	0.195	0.262	0.222	0.281
Shallow MLP	0.445	0.434	0.384	0.397	0.189	0.275	0.212	0.258	0.258	0.278	0.134	0.240	0.150	0.256	0.145	0.239	0.234	0.294	0.239	0.297

Adaptive Phase Response. Adaptive Phase Response. Setting $b_n = 0$ in Equation (5) reduces the adaptive response to the fixed phase rotation $\arg(w_n)$, removing its dependence on the current spectral coefficient. The resulting *w/o Adaptive Bias* variant consistently underperforms APA on Electricity, Weather, and PEMS03, supporting the benefit of spectral-state-dependent adaptation. Detailed results and additional controlled variants are provided in Appendices A.6 and A.7.

Controlled Phase Drift Analysis. We evaluate APA under controlled phase drift on Electricity by applying a smooth temporal displacement $\delta_\alpha(t)$. For frequency ω_k, this induces an additional phase offset $\Delta\phi_k^{\text{extra}}(t) = \omega_k \delta_\alpha(t)$, yielding frequency-dependent and time-varying perturbations controlled by α, with $\alpha = 0$ recovering the original series. Construction details are provided in Appendix A.4.

As shown in Figure 5 (left), APA degrades less than the ablated variants as controlled phase drift increases, with a widening performance gap under stronger perturbations.

Figure 5 (right) illustrates how the temporal gate $\boldsymbol{G}^{(t)}$, derived from $\boldsymbol{R}^{(t)}$, adaptively reweights the base periodic representation $\boldsymbol{P}^{(t)}$ for a representative test window.

Their elementwise combination,

$$\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)},$$

produces a window-adaptive periodic representation, illustrating how the spectral-state-dependent response is converted into adaptive temporal weighting of periodic information.

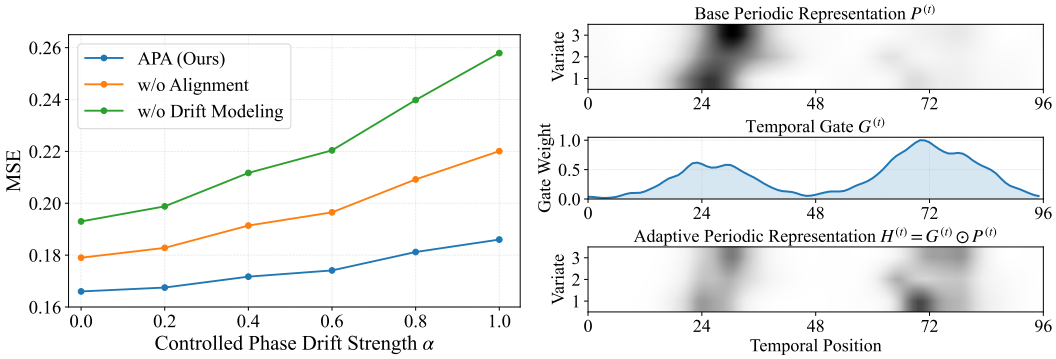


Figure 5: Controlled phase drift analysis on the Electricity dataset. Left: forecasting MSE as the phase drift strength α increases, where $\alpha = 0$ denotes the original series and larger values indicate stronger phase drift. Right: representation-level alignment for a representative test window, showing the base periodic representation $\boldsymbol{P}^{(t)}$, temporal gate $\boldsymbol{G}^{(t)}$, and adaptive representation $\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}$.

4.5 Efficiency Analysis

Figure 6 shows that APA achieves low MSE on Weather ($S = 96$) with relatively short training time and few parameters, while maintaining lower peak GPU memory than several representative methods on PEMS07. These results support the computational efficiency and lightweight design of APA.

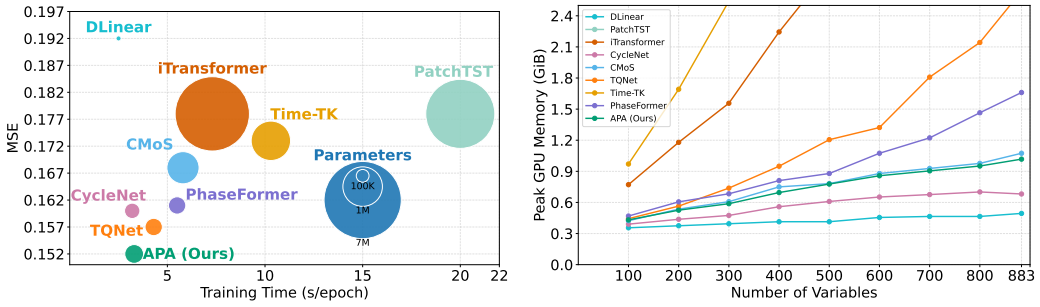


Figure 6: Computational efficiency comparison of APA with representative forecasting methods. Left: MSE versus training time on the Weather dataset with forecasting horizon $S = 96$, where bubble size indicates the parameter count. Right: peak GPU memory usage with increasing numbers of variables on the PEMS07 dataset. PatchTST is omitted from the right panel because its peak GPU memory exceeds the displayed range at all evaluated variable counts.

5 Conclusion

We propose Adaptive Phase Alignment (APA) for correcting phase drift in multivariate time series forecasting. APA combines frequency-specific phase modeling with adaptive temporal gating, achieving consistent improvements across nine benchmarks with low overhead.

AI USE STATEMENT

In this work, generative AI tools were used only to aid and polish the writing, including improving grammar, clarity, and linguistic expression. We did not use generative AI tools for other tasks requiring disclosure. All AI-assisted edits were reviewed by the authors. We take full responsibility for the final content of this work.

ETHICS STATEMENT

This work focuses on methodological research in time series forecasting and does not involve human subjects, sensitive personal data, or other apparent ethical concerns.

REPRODUCIBILITY STATEMENT

The model architecture and learning procedure are described in the main text with explicit equations. Additional implementation details, including dataset descriptions, evaluation metrics, hyperparameter settings, and experimental configurations, are provided in the Appendix.

REFERENCES

Xu Cheng, Fan Shi, Xiufeng Liu, Meng Zhao, and Shengyong Chen. A novel deep class-imbalanced semisupervised model for wind turbine blade icing detection. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2558–2570, 2022. doi: 10.1109/TNNLS.2021.3102514.

Razvan-Gabriel Cirstea, Tung Kieu, Chenjuan Guo, Bin Yang, and Sinno Jialin Pan. Enhancenet: Plugin neural networks for enhancing correlated time series forecasting. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp. 1739–1750, 2021. doi: 10.1109/ICDE51399.2021.00153.

Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 894–903. JMLR.org, 2017.

Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan K Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tiDE: Time-series dense encoder. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=pCbC3aQB5W.

Vincent Le Guen and Nicolas Thome. Shape and time distortion loss for training deep time series forecasting models. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/466accbac9a66b805ba50e42ad715740-Paper.pdf.

Haokun Gui, Xiucheng Li, and Xinyang Chen. Vector quantization pretraining for EEG time series with random projection and phase alignment. In *Forty-first International Conference on Machine Learning*, 2024. URL https://openreview.net/forum?id=7uwLvFvpis.

Hongbin Huang, Minghua Chen, and Xiao Qiao. Generative learning for financial time series with irregular and scale-invariant patterns. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=CdjnzWsQax.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference for Learning Representations*, 2015.

Kelvin J.L. Koa, Jan Chen, Yunshan Ma, Zheng Huanhuan, and Tat-Seng Chua. Reasoning on time-series for financial technical analysis. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=PcjIe5xNaY.

Feifei Kou, Jiahao Wang, Lei Shi, Yuhan Yao, Yawen Li, Suguo Zhu, Zhongbao Zhang, and Junping Du. CFPT: Empowering time series forecasting through cross-frequency interaction and periodic-aware timestamp modeling. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=hHYjiOJFum.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '18, pp. 95–104, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450356572. doi: 10.1145/3209978.3210006. URL https://doi.org/10.1145/3209978.3210006.

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 *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=gqoeQPhQcE.

Shengsheng Lin, Weiwei Lin, Xinyi HU, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=clBiQUgj4w.

Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia LAI, Lingna Ma, and Qiang Xu. SCINet: Time series modeling and forecasting with sample convolution and interaction. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=AyajSjTAzmg.

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=A4zzxu82a7.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Jiaming Ma, Binwu Wang, Qihe Huang, Guanjun Wang, Pengkun Wang, Zhengyang Zhou, and Yang Wang. Mofo: Empowering long-term time series forecasting with periodic pattern modeling. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=sbvLts2HqR.

Jiaming Ma, Qihe Huang, Haofeng Ma, Guanjun Wang, Sheng Huang, Zhengyang Zhou, Pengkun Wang, Xu Wang, Binwu Wang, and Yang Wang. PHAT: Modeling period heterogeneity for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=lr4RlISR6x.

Henrik Madsen. *Time series analysis*. CRC Press, 2007.

Francisco Martinez Alvarez, Alicia Troncoso, Jose C. Riquelme, and Jesus S. Aguilar Ruiz. Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8):1230–1243, 2011. doi: 10.1109/TKDE.2010.227.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Yiming Niu, Jinliang Deng, and Yongxin Tong. Phaseformer: From patches to phases for efficient and effective time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Lk9SqMQzhX.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Md. Mahin Uddin Qureshi, Amrin Binte Ahmed, Adisha Dulmini, Mohammad Mahboob Hussain Khan, and Rumana Rois. Developing a seasonal-adjusted machine-learning-based hybrid time-series model to forecast heatwave warning. *Scientific Reports*, 15(1):8699, March 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-93227-7. URL https://doi.org/10.1038/s41598-025-93227-7.

Haotian Si, Changhua Pei, Jianhui li, Dan Pei, and Gaogang Xie. CMos: Rethinking time series prediction through the lens of chunk-wise spatial correlations. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=eyjrms4HHm.

Zhangyao Song, Chaofeng Qu, Chao Zha, Xiaoyu Zhao, Yinfei Xu, and Tao Guo. Adaptive oscillatory-state alignment for time series forecasting, 2026. URL https://arxiv.org/abs/2606.06010.

Corentin Tallec and Yann Ollivier. Can recurrent neural networks warp time? In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJcKhk-Ab.

Min Wang, Hua Wang, and Fan Zhang. Aligntime: Interperiodic phase alignment sampling for time-series forecasting. *Information Processing & Management*, 63(1):104296, 2026a. ISSN 0306-4573. doi: https://doi.org/10.1016/j.ipm.2025.104296. URL https://www.sciencedirect.com/science/article/pii/S0306457325002377.

Min Wang, Hua Wang, and Fan Zhang. Correctformer: A transformer architecture for correcting periodic drift in time-series forecasting. *Neural Networks*, 196:108375, 2026b. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2025.108375. URL https://www.sciencedirect.com/science/article/pii/S0893608025012560.

Yulong Wang, Yushuo Liu, Xiaoyi Duan, and Kai Wang. Filterts: comprehensive frequency filtering for multivariate time series forecasting. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. ISBN 978-1-57735-897-8. doi: 10.1609/aaai.v39i20.35438. URL https://doi.org/10.1609/aaai.v39i20.35438.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021a. URL https://openreview.net/forum?id=J4gRj6d5Qm.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Xinle Wu, Dalin Zhang, Chenjuan Guo, Chaoyang He, Bin Yang, and Christian S. Jensen. Autocts: automated correlated time series forecasting. *Proc. VLDB Endow.*, 15(4):971–983, December 2021b. ISSN 2150-8097. doi: 10.14778/3503585.3503604. URL https://doi.org/10.14778/3503585.3503604.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with \$10k\$ parameters. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=bWcnvZ3qMb.

Ling Yang and Shenda Hong. Unsupervised time-series representation learning with iterative bilinear temporal-spectral fusion. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 25038–25054. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/yang22e.html.

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=T0axIflVDD.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=iif9mGCTfy.

A. Zeng, M. Chen, L. Zhang, and Q. Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023.

B. Zhang, S. Yin, H. Zhu, and X. He. Freqcycle: A multi-scale time-frequency analysis method for time series forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(33): 28159–28167, 2026a.

Fan Zhang, Shiming Fan, and Hua Wang. Time-tk: A multi-offset temporal interaction framework combining transformer and kolmogorov-arnold networks for time series forecasting. In *Proceedings of the ACM Web Conference 2026*, WWW '26, pp. 7495–7506, New York, NY, USA, 2026b. Association for Computing Machinery. ISBN 9798400723070. doi: 10.1145/3774904.3792618. URL https://doi.org/10.1145/3774904.3792618.

Wenrui Zhang, Ling Yang, Shijia Geng, and Shenda Hong. Self-supervised time series representation learning via cross reconstruction transformer. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):16129–16138, 2024. doi: 10.1109/TNNLS.2023.3292066.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Jiawen Zhu, Shuhan Liu, Shengxuan Li, Qiming Shi, and Di Weng. Poem: Phase-aware $\mathrm{SO}(2)$ feature rotation for time series forecasting under periodicity drift, 2026. URL https://arxiv.org/abs/2608.03630.

A Implementation Details

A.1 More Details of APA

Algorithm 1 summarizes the overall computation of Adaptive Phase Alignment (APA) for a historical observation window ending at time t. The procedure consists of two stages: frequency-specific phase drift modeling and representation-level phase alignment.

Algorithm 1 Overall Pseudocode of Adaptive Phase Alignment (APA)

Require: Historical window $\boldsymbol{X}^{(t)} \in \mathbb{R}^{T\times N}$ ending at time t, period length L
Ensure: Adaptive periodic representation $\boldsymbol{H}^{(t)} \in \mathbb{R}^{T\times N}$
Initialize learnable periodic template $\boldsymbol{C} \in \mathbb{R}^{L\times N}$
Initialize complex affine parameters $w_n, b_n \in \mathbb{C}$, $n = 1, \ldots, N$
{Frequency-specific phase drift modeling}
$\boldsymbol{Z}^{(t)} \leftarrow \mathcal{F}\big(\boldsymbol{X}^{(t)}\big)$, where $K = \lfloor T/2 \rfloor + 1$
for $n = 1, \ldots, N$ **do**
  for $k = 1, \ldots, K$ **do**
    $\widetilde{Z}^{(t)}_{k,n} \leftarrow w_n Z^{(t)}_{k,n} + b_n$
  end for
end for
$\boldsymbol{R}^{(t)} \leftarrow \mathcal{F}^{-1}\Big(\widetilde{\boldsymbol{Z}}^{(t)}\Big)$
{Representation-level phase alignment}
for $n = 1, \ldots, N$ **do**
  $\boldsymbol{G}^{(t)}_{:,n} \leftarrow \mathrm{Softmax}\Big(\boldsymbol{R}^{(t)}_{:,n}\Big)$ {along the temporal dimension}
end for
$s^{(t)} \leftarrow (t - T) \bmod L$
for $\tau = 1, \ldots, T$ **do**
  for $n = 1, \ldots, N$ **do**
    $P^{(t)}_{\tau,n} \leftarrow C_{((s^{(t)}+\tau-1) \bmod L)+1,n}$
  end for
end for
$\boldsymbol{H}^{(t)} \leftarrow \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}$
return $\boldsymbol{H}^{(t)}$

Given a historical observation window $\boldsymbol{X}^{(t)}$, APA first applies a real Fourier transform along the temporal dimension to obtain $\boldsymbol{Z}^{(t)} \in \mathbb{C}^{K\times N}$, where $K = \lfloor T/2 \rfloor + 1$. For each variate, a shared complex-valued affine transformation is applied to all frequency components:

$$\widetilde{Z}^{(t)}_{k,n} = w_n Z^{(t)}_{k,n} + b_n.$$

Although w_n and b_n are shared across frequency components, the induced response remains frequency-specific and window-dependent because it interacts with the component- and window-dependent spectral coefficient $Z^{(t)}_{k,n}$. The transformed spectrum is then mapped back to the temporal domain to obtain $\boldsymbol{R}^{(t)}$, which carries the spectral-state-dependent response of the current observation window.

APA next converts $\boldsymbol{R}^{(t)}$ into a temporal gate $\boldsymbol{G}^{(t)}$ by applying Softmax independently to each variate along the temporal dimension. The gate therefore assigns normalized, window-dependent weights to different temporal positions.

In parallel, APA constructs a base periodic representation from the learnable template $\boldsymbol{C} \in \mathbb{R}^{L\times N}$. Rather than always starting from the first position of the template, the current historical window is associated with the cyclic starting index

$$s^{(t)} = (t - T) \bmod L.$$

The corresponding periodic representation is

$$P^{(t)}_{\tau,n} = C_{((s^{(t)}+\tau-1) \bmod L)+1,n},$$

which preserves the position of the current observation window within the underlying periodic cycle. Consequently, different observation windows extract different but cycle-consistent segments from the same learnable template. This construction also applies when $L > T$, in which case each observation window covers only a local segment of the longer periodic cycle.

Finally, APA combines the window-dependent temporal gate with the cycle-consistent periodic representation:

$$\boldsymbol{H}^{(t)} = \boldsymbol{G}^{(t)} \odot \boldsymbol{P}^{(t)}.$$

The resulting representation adaptively modulates the contribution of periodic information across temporal positions according to the spectral state of the current observation window. It can then be supplied to an arbitrary forecasting backbone, enabling APA to perform representation-level phase alignment before prediction.

A.2 Experiment Details

All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 3090 24GB GPU. We optimize all models using Adam. The historical input length is fixed at $T = 96$ for all datasets. For each forecasting backbone, we follow its original implementation and recommended hyperparameter settings whenever available. When integrating APA into a backbone, we keep the backbone configuration unchanged and introduce only the APA module.

The main dataset-specific hyperparameter of APA is the period length L, which determines the length of the learnable periodic template $\boldsymbol{C} \in \mathbb{R}^{L \times N}$. The corresponding base periodic representation is constructed by periodically repeating this template along the temporal dimension, as defined in Equation (9). The values of L used for different datasets are summarized in Table 4.

Table 4: Hyperparameter settings for different datasets. "lr" denotes the learning rate, and L denotes the period length of the learnable periodic template in APA.

Dataset	Period (L)	lr	Batch size	Epochs	patience
ETTh1	24	1×10^{-3}	256	30	5
ETTm1	96	1×10^{-3}	256	30	5
Electricity	168	3×10^{-3}	32	30	5
Weather	144	1×10^{-3}	64	30	5
Solar	144	3×10^{-3}	64	30	5
PEMS03	288	3×10^{-3}	32	30	5
PEMS04	288	3×10^{-3}	32	30	5
PEMS07	288	3×10^{-3}	32	30	5
PEMS08	288	3×10^{-3}	32	30	5

Following the periodicity setting in CycleNet (Lin et al., 2024), we determine the period length L according to the dominant periodic structure of each dataset. Specifically, the dominant period is identified using the autocorrelation function (ACF) (Madsen, 2007), while the sampling interval of each dataset determines the corresponding number of temporal positions. Accordingly, L represents the number of sampled time steps covered by one dominant period of the base periodic representation $\boldsymbol{P}^{(t)}$.

For example, ETTh1 is sampled hourly and exhibits a dominant daily cycle, corresponding to $L = 24$, whereas Electricity is also sampled hourly but exhibits a dominant weekly cycle, giving $L = 24 \times 7 = 168$. ETTm1 is sampled every 15 minutes, yielding $L = 96$ for a daily cycle, while the PEMS datasets are sampled every 5 minutes and use $L = 288$ for their dominant daily periodicity. The remaining dataset-specific period lengths are determined in the same manner and are reported in Table 4.

For a discrete time series $\{y_t\}_{t=1}^{T}$ with sample mean $\bar{y}$, the sample autocorrelation at lag k is defined as

$$r_k = \frac{\sum_{t=1}^{T-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}. \tag{14}$$

A prominent peak in the ACF at lag k indicates a strong recurring structure at that lag. We therefore select the dominant periodic lag as the period length, i.e., $L = k$. Consequently, the period lengths used in APA are determined from the intrinsic periodic structure of each dataset rather than being independently tuned for different forecasting horizons.

A.3 Impact of the Period Length L

The period length L determines the length of the learnable periodic template $\boldsymbol{C}$ in APA and therefore controls the periodic structure represented by the base periodic representation $\boldsymbol{P}^{(t)}$. In principle, L should be consistent with the dominant periodicity of the dataset. To investigate the influence of this hyperparameter, we conduct a sensitivity analysis on the Electricity dataset with the forecasting horizon fixed at $S = 720$.

As shown in Figure 7, APA achieves the best performance when $L = 168$, with an MSE of 0.208 and an MAE of 0.298. Since the Electricity dataset is sampled hourly, $L = 168$ corresponds to a weekly period of 24×7 time steps. In comparison, setting $L = 24$, which corresponds only to the daily periodicity, yields higher errors (0.219 MSE and 0.308 MAE), indicating that the longer weekly cycle provides a more suitable periodic prior for this dataset.

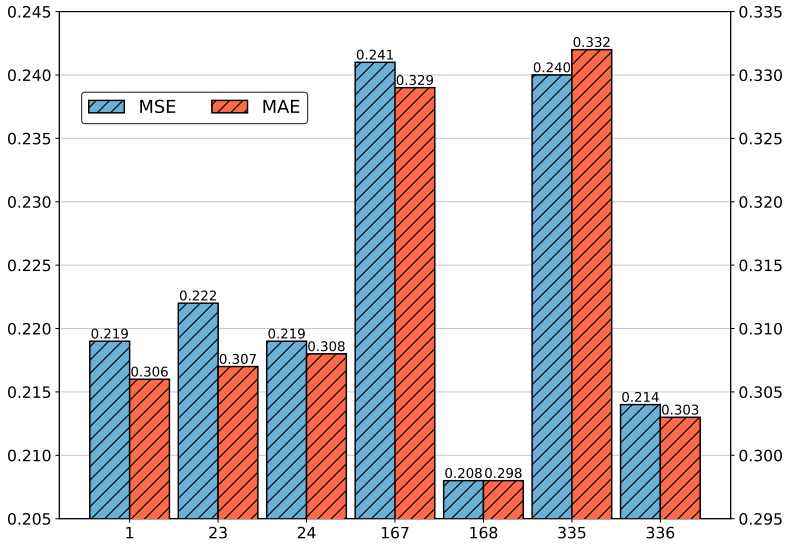


Figure 7: Performance of APA on the Electricity dataset with varying period length L under the forecasting horizon $S = 720$. Lower values indicate better forecasting performance.

Interestingly, setting $L = 336 = 2 \times 168$ still achieves competitive performance, with an MSE of 0.214 and an MAE of 0.303. This suggests that an integer multiple of the dominant period can preserve useful periodic structure. In contrast, period lengths that are close to but misaligned with the dominant cycle lead to substantial degradation. For example, $L = 167$ results in an MSE of 0.241 and an MAE of 0.329, while $L = 335$ yields an MSE of 0.240 and an MAE of 0.332.

Overall, these results show that APA is most effective when L is aligned with the intrinsic periodic structure of the data. They also support our use of the dominant periodicity identified by the ACF to determine the dataset-specific period length L.

A.4 Controlled Phase Drift Experiment

We provide additional details of the controlled phase drift experiment used to examine model behavior under systematically increasing phase variation. The experiment is conducted on the Electricity dataset by introducing a smooth, time-varying temporal displacement into the original series.

We use a dimensionless coefficient $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ to control the perturbation strength. The coefficient α is used only to construct the controlled datasets and is not a hyperparameter of APA. The setting $\alpha = 0$ leaves the original series unchanged, whereas $\alpha = 1$ corresponds to the predefined maximum displacement.

Controlled Temporal Displacement. Let $\boldsymbol{x}_t \in \mathbb{R}^N$ denote the multivariate observation at time step t. We define a smooth temporal displacement as

$$\delta_\alpha(t) = \alpha\delta_{\max} \sin\left(\frac{2\pi t}{P_d}\right), \tag{15}$$

where $\delta_{\max}$ determines the maximum displacement magnitude and P_d controls the temporal scale over which the displacement varies. We set $P_d = 4T = 384$ with historical length $T = 96$, and use $\delta_{\max} = 6$ time steps. Since Electricity is sampled hourly, $\alpha = 1$ corresponds to a maximum temporal displacement of six hours. The evaluated values of α therefore correspond to maximum displacement magnitudes of $0, 1.2, 2.4, 3.6, 4.8$, and 6.0 time steps.

For each original temporal index t, we sample the series at the displaced coordinate

$$u_\alpha(t) = t + \delta_\alpha(t), \tag{16}$$

using linear interpolation when $u_\alpha(t)$ falls between two discrete observations. The sinusoidal displacement changes gradually over time rather than producing an abrupt shift.

Under our parameterization, the temporal mapping remains monotonic. Indeed,

$$\frac{\mathrm{d}u_\alpha(t)}{\mathrm{d}t} = 1 + \alpha\delta_{\max}\frac{2\pi}{P_d}\cos\left(\frac{2\pi t}{P_d}\right), \tag{17}$$

whose minimum value is

$$1 - \alpha\delta_{\max}\frac{2\pi}{P_d} \geq 1 - \frac{12\pi}{384} > 0.$$

Thus, the controlled transformation modifies temporal correspondence while preserving chronological ordering.

Algorithm 2 Construction of the Controlled Phase-Drift Series

Require: Time series $\boldsymbol{X} = \{\boldsymbol{x}_t\}_{t=1}^M$, drift strength α, $T = 96$, $\delta_{\max} = 6$
Ensure: Controlled series $\widetilde{\boldsymbol{X}}^{(\alpha)}$

$P_d \leftarrow 4T$
for $t = 1, \ldots, M$ **do**
  $\delta \leftarrow \alpha\delta_{\max}\sin(2\pi t/P_d)$
  $u \leftarrow t + \delta$
  if u permits interpolation within the current data partition **then**
    $j \leftarrow \lfloor u \rfloor$
    $\lambda \leftarrow u - j$
    $\widetilde{\boldsymbol{x}}_t^{(\alpha)} \leftarrow (1-\lambda)\boldsymbol{x}_j + \lambda\boldsymbol{x}_{j+1}$
  else
    $u \leftarrow \min(\max(u, 1), M)$
    $j \leftarrow \min(\lfloor u \rfloor, M-1)$
    $\lambda \leftarrow u - j$
    $\widetilde{\boldsymbol{x}}_t^{(\alpha)} \leftarrow (1-\lambda)\boldsymbol{x}_j + \lambda\boldsymbol{x}_{j+1}$
  end if
end for
return $\widetilde{\boldsymbol{X}}^{(\alpha)}$

The same temporal displacement is applied to all variables at a given time step. This preserves their cross-variate synchronization while perturbing their temporal correspondence with the original time axis. No circular shifting is used. The original chronological training, validation, and test partitions are retained, and interpolation is restricted to observations from the same partition so that the controlled transformation does not introduce cross-partition information leakage.

Relation to Phase Drift. The controlled temporal displacement has a direct interpretation for an individual periodic component. Consider a component with angular frequency ω_k. Evaluating this component at the displaced coordinate $t + \delta_\alpha(t)$ changes its instantaneous phase relative to the undisplaced coordinate by

$$\Delta\phi_k^{\mathrm{extra}}(t) = \omega_k\delta_\alpha(t) = \alpha\omega_k\delta_{\max}\sin\left(\frac{2\pi t}{P_d}\right). \tag{18}$$

Equation (18) provides the motivation for the controlled perturbation used in Figure 5. Increasing α scales the magnitude of the induced phase variation while leaving its temporal profile unchanged. Moreover, because the induced offset is proportional to ω_k, the same temporal displacement produces different phase variations for different frequency components. The sinusoidal displacement further causes these variations to evolve smoothly over time. The construction therefore provides a controlled setting with both frequency-dependent and time-varying phase perturbations, matching the two properties targeted by APA.

When $\alpha = 0$, $\delta_\alpha(t) = 0$ for all t, and the controlled series exactly reduces to the original Electricity dataset.

Experimental Protocol. For each $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1.0\}$, we independently train APA, *w/o Alignment*, and *w/o Drift Modeling* on the corresponding controlled dataset. The historical length, optimization procedure, training configuration, and model hyperparameters are kept unchanged across perturbation strengths.

Following the main forecasting experiments, we evaluate horizons $S \in \{96, 192, 336, 720\}$. For each value of α, the MSE reported in Figure 5 is averaged across the four forecasting horizons. Because $\alpha = 0$ exactly recovers the original Electricity dataset, the first point of each curve corresponds to the associated setting in the main ablation study.

Importantly, models are retrained independently for each perturbation strength. Therefore, this experiment is not intended to measure zero-shot robustness to an unseen phase shift. Instead, it provides a controlled comparison of how forecasting performance changes as the amount of phase variation present in the data is systematically increased. The relative degradation of APA and its ablated variants can consequently be used to assess whether the proposed phase-drift modeling and representation-level alignment mechanisms become more beneficial as the controlled phase variation grows.

A.5 Robustness to Random Initialization

To evaluate the robustness of APA to random initialization, we conduct five independent runs with random seeds $\{2026, 2027, 2028, 2029, 2030\}$. Table 5 reports the performance of each run together with the mean and standard deviation across the five runs.

A.6 Effect of Adaptive Phase Response

To further examine the source of frequency-specific and window-dependent adaptation in APA, we construct a variant termed *w/o Adaptive Bias*, in which the complex bias term is removed by setting $b_n = 0$ for all variates. Under this setting, the complex affine transformation becomes

$$\widetilde{Z}_{k,n}^{(t)} = w_n Z_{k,n}^{(t)}, \tag{19}$$

and the induced relative phase variation reduces to

$$\Delta\phi_{k,n}^{(t)} = \arg(w_n). \tag{20}$$

Therefore, for each variate, the induced phase response becomes fixed and no longer depends on the spectral coefficient $Z_{k,n}^{(t)}$, which varies across frequency components and observation windows. This variant directly tests the role of the additive term b_n, which enables the spectral-state-dependent phase response characterized in Equation (5).

As shown in Table 6, removing the adaptive bias term consistently degrades forecasting performance across the three evaluated datasets. When $b_n = 0$, the relative phase response changes from

$$\arg\left(w_n + \frac{b_n}{Z_{k,n}^{(t)}}\right)$$

to the fixed response $\arg(w_n)$, eliminating its dependence on the frequency- and window-dependent spectral coefficient $Z_{k,n}^{(t)}$. The consistent performance degradation indicates that the additive term b_n, and the resulting spectral-state-dependent response, contributes to the effectiveness of APA. Together with the component ablations, this result provides further evidence that adapting the response across frequency components and observation windows is beneficial for forecasting.

Table 5: Forecasting performance of APA across datasets and forecasting horizons over five independent random seeds. The mean and standard deviation of MSE and MAE are reported.

Dataset	Length	Random Seed													
		2026		2027		2028		2029		2030		Mean		Std	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.370	0.393	0.367	0.392	0.364	0.391	0.368	0.392	0.366	0.392	0.367	0.392	0.002	0.001
	192	0.414	0.417	0.419	0.420	0.416	0.419	0.411	0.416	0.415	0.418	0.415	0.418	0.003	0.002
	336	0.455	0.440	0.449	0.434	0.457	0.442	0.453	0.438	0.451	0.436	0.453	0.438	0.003	0.003
	720	0.453	0.454	0.450	0.452	0.446	0.450	0.460	0.458	0.456	0.456	0.453	0.454	0.005	0.003
ETTm1	96	0.307	0.352	0.310	0.353	0.308	0.353	0.311	0.354	0.309	0.353	0.309	0.353	0.002	0.001
	192	0.351	0.379	0.345	0.375	0.348	0.377	0.347	0.376	0.349	0.378	0.348	0.377	0.002	0.002
	336	0.381	0.400	0.378	0.398	0.375	0.396	0.379	0.399	0.377	0.397	0.378	0.398	0.002	0.002
	720	0.432	0.432	0.438	0.438	0.436	0.436	0.430	0.430	0.434	0.434	0.434	0.434	0.003	0.003
Electricity	96	0.133	0.226	0.130	0.225	0.134	0.227	0.132	0.226	0.131	0.226	0.132	0.226	0.002	0.001
	192	0.152	0.244	0.151	0.244	0.150	0.243	0.154	0.245	0.153	0.244	0.152	0.244	0.002	0.001
	336	0.168	0.262	0.172	0.263	0.170	0.263	0.174	0.264	0.171	0.263	0.171	0.263	0.002	0.001
	720	0.211	0.300	0.205	0.296	0.208	0.298	0.207	0.297	0.209	0.299	0.208	0.298	0.002	0.002
Solar	96	0.183	0.233	0.182	0.232	0.181	0.231	0.182	0.232	0.182	0.232	0.182	0.232	0.001	0.001
	192	0.193	0.245	0.196	0.246	0.195	0.245	0.192	0.244	0.194	0.245	0.194	0.245	0.002	0.001
	336	0.205	0.248	0.202	0.247	0.206	0.249	0.204	0.248	0.203	0.248	0.204	0.248	0.002	0.001
	720	0.209	0.253	0.208	0.253	0.207	0.252	0.211	0.254	0.210	0.253	0.209	0.253	0.002	0.001
Weather	96	0.151	0.196	0.152	0.197	0.152	0.197	0.153	0.198	0.152	0.197	0.152	0.197	0.001	0.001
	192	0.204	0.244	0.200	0.242	0.202	0.243	0.201	0.243	0.203	0.243	0.202	0.243	0.002	0.001
	336	0.260	0.286	0.258	0.285	0.256	0.284	0.259	0.285	0.257	0.285	0.258	0.285	0.002	0.001
	720	0.341	0.341	0.345	0.345	0.343	0.343	0.339	0.339	0.342	0.342	0.342	0.342	0.002	0.002
PEMS03	12	0.057	0.156	0.056	0.155	0.058	0.157	0.057	0.156	0.057	0.156	0.057	0.156	0.001	0.001
	24	0.071	0.173	0.071	0.173	0.070	0.172	0.072	0.174	0.071	0.173	0.071	0.173	0.001	0.001
	48	0.094	0.197	0.097	0.198	0.095	0.198	0.098	0.199	0.096	0.198	0.096	0.198	0.002	0.001
	96	0.129	0.228	0.121	0.224	0.125	0.226	0.123	0.225	0.127	0.227	0.125	0.226	0.003	0.002
PEMS04	12	0.068	0.167	0.067	0.166	0.066	0.165	0.067	0.166	0.067	0.166	0.067	0.166	0.001	0.001
	24	0.074	0.177	0.077	0.178	0.076	0.177	0.073	0.176	0.075	0.177	0.075	0.177	0.002	0.001
	48	0.091	0.194	0.087	0.191	0.093	0.195	0.090	0.193	0.089	0.192	0.090	0.193	0.002	0.002
	96	0.109	0.213	0.107	0.212	0.105	0.210	0.113	0.216	0.111	0.214	0.109	0.213	0.003	0.002
PEMS07	12	0.050	0.142	0.051	0.143	0.051	0.143	0.052	0.144	0.051	0.143	0.051	0.143	0.001	0.001
	24	0.063	0.156	0.061	0.154	0.062	0.155	0.062	0.155	0.062	0.155	0.062	0.155	0.001	0.001
	48	0.082	0.174	0.080	0.173	0.078	0.172	0.081	0.173	0.079	0.173	0.080	0.173	0.002	0.001
	96	0.104	0.194	0.113	0.200	0.110	0.197	0.101	0.193	0.107	0.196	0.107	0.196	0.005	0.003
PEMS08	12	0.074	0.170	0.071	0.169	0.075	0.171	0.073	0.170	0.072	0.170	0.073	0.170	0.002	0.001
	24	0.097	0.192	0.096	0.191	0.094	0.190	0.100	0.194	0.098	0.193	0.097	0.192	0.002	0.002
	48	0.135	0.223	0.147	0.230	0.139	0.226	0.151	0.233	0.143	0.228	0.143	0.228	0.006	0.004
	96	0.230	0.283	0.210	0.271	0.220	0.277	0.215	0.274	0.225	0.280	0.220	0.277	0.008	0.005

Table 6: Ablation study of the adaptive bias term in APA. *w/o Adaptive Bias* sets $b_n = 0$, reducing the induced relative phase response to the fixed rotation $\Delta\phi_{k,n}^{(t)} = \arg(w_n)$. Results are averaged over the four dataset-specific forecasting horizons. Lower values indicate better forecasting performance.

Model	Electricity		Weather		PEMS03		Average	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
APA (Ours)	**0.166**	**0.258**	**0.239**	**0.267**	**0.087**	**0.188**	**0.164**	**0.238**
w/o Adaptive Bias ($b_n = 0$)	0.175	0.269	0.249	0.279	0.095	0.196	0.173	0.248

Table 7: Controlled analysis of phase-aware adaptation. Results on Electricity and Weather are averaged over forecasting horizons $\{96, 192, 336, 720\}$, while results on PEMS03 are averaged over $\{12, 24, 48, 96\}$. The last two columns report the average across the three datasets. Lower values are better.

Variant	Electricity		Weather		PEMS03		Average	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Periodic Template Only	0.206	0.293	0.256	0.283	0.119	0.229	0.194	0.268
Generic Temporal Gate	0.193	0.288	0.247	0.276	0.105	0.201	0.182	0.255
Magnitude-Only	0.178	0.271	0.252	0.281	0.098	0.197	0.176	0.250
Full APA	**0.166**	**0.258**	**0.239**	**0.267**	**0.087**	**0.188**	**0.164**	**0.238**

A.7 Controlled Analysis of Phase-Aware Adaptation

To disentangle the contribution of phase-aware adaptation from that of the periodic template and generic temporal reweighting, we construct three controlled variants of APA. These variants progressively isolate whether the performance gains arise from the periodic template itself, generic input-dependent temporal gating, or frequency-domain adaptation without phase modification.

Periodic Template Only retains the cycle-indexed periodic representation $P^{(t)}$ but removes adaptive gating, directly feeding the periodic representation into the forecasting backbone. This variant isolates the contribution of the learnable periodic template without window-dependent adaptation.

Generic Temporal Gate retains the same periodic representation and adaptive reweighting mechanism as APA, but replaces the frequency-domain phase modeling module with a parameter-matched time-domain gating module. This variant controls for gains that may arise from generic input-dependent temporal reweighting rather than phase-aware spectral adaptation.

Magnitude-Only retains frequency-domain adaptation while preventing additional phase modification. Specifically, the spectral magnitude is adaptively transformed while the original spectral phase is preserved, thereby removing the induced relative phase response.

Full APA uses the complete complex-valued spectral transformation together with temporal gating, allowing the spectral response to adapt in both magnitude and phase according to the current observation window. All variants use the same forecasting backbone and training protocol.

Table 7 reports the results on Electricity, Weather, and PEMS03. For each dataset, MSE and MAE are averaged over the four forecasting horizons used in the main experiments. Electricity and Weather use horizons $\{96, 192, 336, 720\}$, while PEMS03 uses $\{12, 24, 48, 96\}$. The final two columns report the average across the three datasets.

As shown in Table 7, the periodic template alone provides useful periodic structure, but consistently underperforms variants with window-dependent adaptation. Introducing a generic temporal gate reduces the average MSE from 0.194 to 0.182, indicating that adaptive temporal reweighting itself contributes to forecasting performance. Frequency-domain magnitude adaptation further reduces the average MSE to 0.176, although its advantage over generic temporal gating is not uniform across all datasets.

Full APA achieves the best performance on all three datasets and both metrics, with an average MSE/MAE of $0.164/0.238$. Compared with the strongest controlled variant, Magnitude-Only, Full APA reduces the average MSE and MAE by 6.8% and 4.8%, respectively. The improvement is also consistent across individual datasets: relative to Magnitude-Only, Full APA reduces MSE from 0.178 to 0.166 on Electricity, from 0.252 to 0.239 on Weather, and from 0.098 to 0.087 on PEMS03. These results indicate that neither the periodic template, generic temporal gating, nor spectral magnitude adaptation alone fully accounts for the gains of APA. Allowing the spectral response to adapt its phase according to the current window provides additional and consistent benefits.