PaCo: Evidence-Supported Phase Coordinates for Alignment-Aware Time Series Forecasting

Anonymous authors
Paper under double-blind review

Abstract

Long-term time series forecasting is challenging because recurring temporal patterns may reappear at different temporal positions across observation windows. Such alignment variation changes local observations while preserving the underlying pattern, making forecasting representations prone to mixing pattern content with alignment state. Existing temporal and spectral forecasting models have improved pattern extraction, and recent phase-aware methods further introduce spectral phase into temporal modeling. However, using phase is not equivalent to representing it as an alignment state. Spectral phase is a circular variable associated with temporal shifts rather than an unconstrained scalar, while the stability of an observed phase direction is related to the amplitude of the corresponding Fourier coefficient. In this work, we propose PaCo, an evidence-supported phase coordinate conditioning framework for time series forecasting. PaCo constructs spectral phase coordinates that preserve circular phase direction and incorporate amplitude-derived evidence for the observed direction. These coordinates serve as local alignment states to condition temporal representations and guide variable-level context modeling. Experiments on public long-term forecasting benchmarks and an industrial Wastewater Treatment Aeration (WWTA) dataset show that PaCo achieves competitive forecasting performance, while ablation and phase-response analyses support the importance of the proposed evidence-supported phase coordinate design. The code is available at `https://anonymous.4open.science/r/Anonymization-68A4/`.

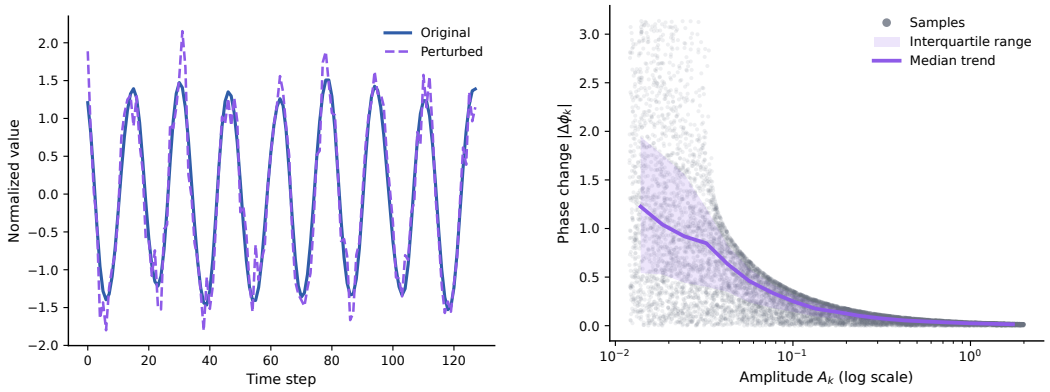
1 Introduction

Long-term time series forecasting aims to predict future values from historical observations (Li & Law, 2024; Kim et al., 2025; Houssein et al., 2025) and plays an important role in applications such as energy management (Javid et al., 2025; Zhao et al., 2025), traffic planning (Zhang et al., 2024; Tao, 2024), weather forecasting, and financial analysis (Zheng et al., 2024; Odumuwagun, 2023). A central difficulty is that recurring temporal patterns do not always appear at fixed positions across observation windows (Qiu et al., 2025). Daily peaks may appear earlier or later (Li et al., 2023), seasonal effects may drift (Seo & Lim, 2025), and local events may recur with temporal offsets (Sun et al., 2025b). As a result, the same pattern can produce different local observations under different alignments, making alignment variation not merely a temporal irregularity but a representation problem for learning predictive representations: shifted observations should remain related in content while being distinguishable in alignment state.

Recent forecasting models (Cao et al., 2026; Sun et al., 2025a; Li et al., 2024) have improved the extraction of temporal and spectral features, making them effective at modeling recurring structures and long-range dependencies. However, alignment variation remains a source of representation ambiguity. When a recurring pattern shifts in time, the observed values and local shapes change even though the pattern identity is preserved. A representation that mainly describes the observed content must then carry pattern identity and alignment state in a single feature state, which can make shifted instances of the same pattern appear unnecessarily different for prediction.

Spectral phase offers a useful way to describe the alignment state of recurring components (Oppenheim & Lim, 1981). In the spectral domain, amplitude and phase characterize different aspects of a temporal pattern: amplitude reflects the strength of a component, while phase is related to its

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.



(a) Signal before and after coefficient perturbation.

(b) Amplitude-dependent phase change.

Figure 1: Amplitude-dependent sensitivity of observed phase. We apply controlled perturbations to Fourier coefficients and measure the induced change in phase direction. (a) shows the original and perturbed signals under the selected perturbation scale. (b) shows that, under equal-magnitude coefficient perturbations, phase directions associated with lower amplitudes undergo larger angular deviations than those associated with higher amplitudes. This motivates the use of amplitude-derived evidence when constructing phase representations for forecasting.

temporal position. According to the Fourier shift theorem (Papoulis, 1977), temporal displacement is expressed as phase rotation while amplitude is preserved. This gives phase an alignment-related meaning: phase is not an unconstrained scalar attribute, but a circular direction associated with temporal shifts. Therefore, a phase-based forecasting representation should preserve the angular structure of phase when using it to describe local alignment.

However, the observed phase direction is not equally stable in practical forecasting. Local spectral coefficients are estimated from finite, noisy, and nonstationary windows (Liu et al., 2022; He et al., 2025). Since phase is the angular direction of a complex Fourier coefficient, this direction can change when the coefficient is perturbed. Under perturbations of comparable magnitude, coefficients with weaker amplitudes tend to exhibit larger angular deviations than coefficients with stronger amplitudes. Thus, amplitude provides evidence for the stability of an observed phase direction, rather than merely measuring spectral strength. We refer to this support as amplitude-derived evidence, which is derived from coefficient amplitude and supports the interpretation of the observed phase direction. Fig. 1 empirically illustrates the amplitude-dependent sensitivity of observed phase under controlled coefficient perturbations. Together with the shift-theoretic meaning of phase, this suggests two requirements for phase-based forecasting representations: *preserve circular phase direction and accompany the observed direction with amplitude-derived evidence.*

Based on these requirements, we propose PaCo, an evidence-supported phase coordinate conditioning framework for time series forecasting. PaCo represents local alignment using evidence-supported spectral phase coordinates constructed from circular phase direction and amplitude-derived evidence. These coordinates separate alignment-related information from temporal content and use it to condition temporal representations through residual affine transformations. In this way, temporal tokens encode local observed content, while phase coordinates provide alignment-related conditions that adjust the representation for forecasting. PaCo further reuses the evidence-supported phase coordinates to guide variable-level context mixing, so that cross-variable interactions can depend on both temporal content and spectral phase state.

We evaluate PaCo on standard long-term forecasting benchmarks. The results show that PaCo achieves strong forecasting performance compared with competitive baselines. Ablation studies verify the necessity of circular phase direction, amplitude-derived evidence, and phase-guided context mixing. Representation analyses further show that the learned phase-coordinate pathway preserves cyclic phase structure and maps phase direction together with the evidence indicator into conditioning signals. These results demonstrate that evidence-supported phase coordinate conditioning provides an effective representation principle for using spectral phase in time series forecasting.

Our contributions are summarized as follows:

- We formulate spectral phase as an evidence-supported coordinate for forecasting representations, combining its shift-induced circular direction with amplitude-derived evidence for the stability of observed phase directions.
- We propose PaCo, a phase coordinate conditioning framework that uses evidence-supported spectral phase coordinates to condition temporal representations and guide variable-level context mixing.
- We conduct forecasting experiments, ablation studies, and representation analyses to verify the effectiveness of PaCo and the necessity of its evidence-supported phase coordinate design.

2 Related Work

Frequency-domain time series forecasting. Frequency-domain forecasting has been explored because spectral representations provide a way to describe periodicity and long-range temporal structure. Representative methods introduce Fourier or temporal-frequency operations into forecasting architectures. FEDformer uses sparse Fourier-basis representations within a decomposed Transformer (Zhou et al., 2022), while ATFN adapts phase, frequency, and amplitude patterns between observation and prediction windows (Yang et al., 2022). Later methods further move learnable transformations into the spectral domain. FreTS applies MLPs to frequency-domain representations (Yi et al., 2023), FCNet develops fully complex-valued time-frequency networks (Zhu et al., 2024), and recent works extend spectral forecasting through Fourier basis mapping (Yang et al., 2024), wavelet-based multi-resolution mixing (Murad et al., 2025), frequency-domain objectives (Wang et al., 2025), learnable spectral decomposition (An et al., 2025), spectral-temporal graph modeling (Jin et al., 2025), and multi-level frequency enhancement (Wang et al., 2026). These methods demonstrate the effectiveness of spectral modeling for forecasting, but phase is usually handled as part of a complex-valued spectrum, a spectral transformation, or a reconstruction process. The alignment-related role of phase direction is therefore not the central object of representation design.

Phase-aware time series forecasting. Several recent methods have incorporated phase information into forecasting models. Quatformer uses learnable periods and phases in quaternion-based rotation attention to model phase shifts in periodic patterns (Chen et al., 2022). FITS learns amplitude scaling and phase shift during complex-domain interpolation (Xu et al., 2024). PhaseFormer introduces compact phase embeddings and cross-phase interaction (Niu et al., 2026), while PULSE models non-stationary forecasting through dynamical phase evolution and future phase generation (Liu et al., 2026b). Recent amplitude-phase methods further study frequency-domain non-stationarity through reconstruction and parameterization, such as APRNet (Liu et al., 2026a) and FreAPNet (Xu et al., 2026). These works show that spectral phase can benefit temporal modeling. Nevertheless, using phase in forecasting is not only a matter of adding phase-related signals to the model. Under local alignment variation, the observed phase direction should be represented as a circular alignment cue before it is used to condition forecasting representations. Moreover, because the stability of this direction depends on coefficient amplitude, phase direction should be accompanied by amplitude-derived evidence. This motivates the evidence-supported spectral phase coordinate used in PaCo.

3 Phase as an Evidence-Supported Alignment Signal

This section motivates the spectral phase coordinate used by PaCo. The derivation of phase sensitivity under coefficient perturbations is given in Appendix A.

Phase as an angular alignment cue. Let $X_k = A_k e^{i\phi_k}$ be the Fourier coefficient of a length-P segment at frequency k. Under a circular shift by τ, the Fourier shift theorem gives

$$\tilde{X}_k = X_k e^{-i2\pi k\tau/P}, \qquad |\tilde{X}_k| = A_k, \qquad \tilde{\phi}_k \equiv \phi_k - \frac{2\pi k\tau}{P} \pmod{2\pi}. \tag{1}$$

The phase rotation describes the shift while amplitude remains unchanged. As phase wraps at 2π, PaCo represents its direction by $[\sin\phi_k, \cos\phi_k]$ rather than by a scalar angle. Nearby directions therefore remain nearby across the wrapping boundary. For finite local patches, boundary effects and

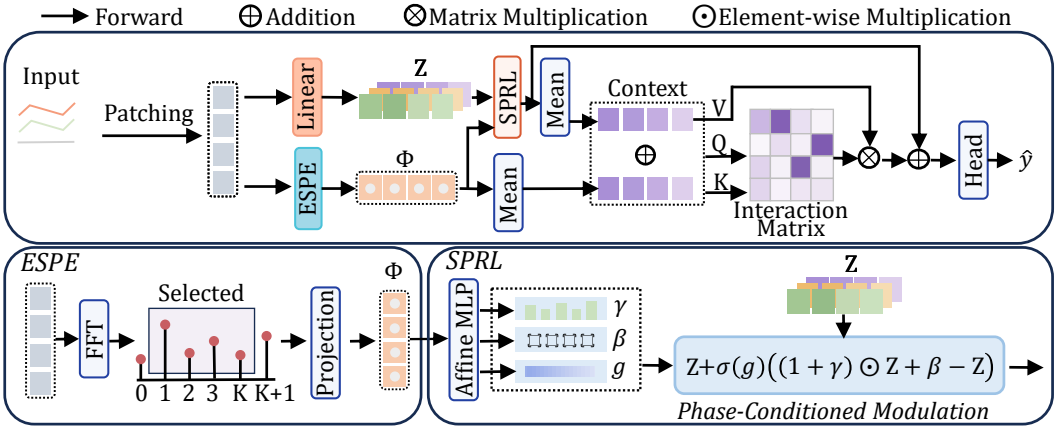


Figure 2: Overview of PaCo. Evidence-Supported Spectral Phase Encoding (ESPE) constructs phase coordinates $\mathbf{\Phi}$ from local spectral components, while temporal patch embedding produces tokens $\mathbf{Z}$. Spectral Phase-Conditioned Representation Learning (SPRL) modulates $\mathbf{Z}$ using $\mathbf{\Phi}$ through Phase-Conditioned Modulation. The same $\mathbf{\Phi}$ is reused for phase-guided contextual forecasting. Auxiliary dimension-alignment projections are omitted for clarity.

nonstationarity can also affect the observed coefficients, so this shift relation motivates an alignment cue rather than guaranteeing exact shift invariance for every patch.

Amplitude-dependent stability of observed phase. The shift relation does not make every estimated direction equally stable. Consider a nonzero coefficient $X = Ae^{i\phi}$ perturbed to $\widehat{X} = X + \eta$, and let $\Delta\phi$ be the smaller angular difference between their phases. Geometrically,

$$|\Delta\phi| \le \arcsin\left(\frac{|\eta|}{A}\right) \quad \text{when } |\eta| < A. \tag{2}$$

For perturbations of comparable magnitude, the direction of a weaker coefficient can therefore deviate more. Amplitude supplies *evidence* about the stability of the observed direction; it is neither a calibrated probability of correctness nor a reason to discard low-amplitude components. Fig. 1 illustrates this behavior, and Appendix A gives the local approximation and derivation.

Evidence-supported phase coordinate. For K selected non-DC components, PaCo couples each circular direction with logarithmic amplitude and a bounded indicator of relative amplitude:

$$\mathbf{d}_k = [\log(1 + A_k), \sin\phi_k, \cos\phi_k, \rho_k], \qquad \rho_k = \frac{A_k/\bar{A}}{A_k/\bar{A} + \alpha}, \qquad \bar{A} = \frac{1}{K}\sum_{j=1}^{K} A_j. \tag{3}$$

Here $\alpha > 0$ controls saturation, and ρ_k expresses amplitude-derived support relative to the selected components of the same segment. The evidence indicator accompanies the phase direction as a separate feature; it does not multiply or erase the sine and cosine terms. The descriptors are concatenated and projected into a learned phase coordinate, whose patch-level implementation is specified next.

4 Methodology

4.1 Problem Formulation and Framework

Given a multivariate time series $\mathbf{X} \in \mathbb{R}^{L\times C}$ with input length L and C variables, the task is to predict the future sequence $\mathbf{Y} \in \mathbb{R}^{H\times C}$ over horizon H. After reversible instance normalization (Kim et al., 2021), each variable is divided into N overlapping patches of length P and projected into temporal tokens $\mathbf{Z} \in \mathbb{R}^{C\times N\times D}$. From the same patches, PaCo extracts evidence-supported spectral phase coordinates $\mathbf{\Phi} \in \mathbb{R}^{C\times N\times D_\phi}$. Rather than adding $\mathbf{\Phi}$ to the tokens, PaCo uses it to generate gated affine transformations that produce phase-conditioned representations $\widetilde{\mathbf{Z}}$. These representations are

subsequently processed through phase-guided contextual interaction and temporal refinement, and then mapped to the prediction $\widehat{\mathbf{Y}}$. Fig. 2 illustrates the overall architecture of PaCo.

4.2 Evidence-Supported Spectral Phase Encoding

Alignment variation can differ across local regions of an observation window. PaCo therefore extracts spectral phase coordinates at the same patch resolution as the temporal tokens, allowing each token to receive a condition derived from its corresponding local observation. This establishes a direct correspondence between temporal content and spectral phase state while allowing the conditioning signal to vary across patches and variables.

Let $\mathbf{p}_{c,n} \in \mathbb{R}^P$ denote the n-th normalized patch of variable c. PaCo applies the discrete Fourier transform along the patch and retains the first K non-DC frequency components:

$$X_{c,n,k} = \sum_{t=0}^{P-1} p_{c,n,t} e^{-i2\pi kt/P} = A_{c,n,k} e^{i\phi_{c,n,k}}, \qquad k = 1, \ldots, K. \tag{4}$$

The DC component describes the patch mean and does not provide a shift-induced phase rotation, so it is excluded from the phase coordinate. The number of retained components K controls the spectral coverage and descriptor size of the conditioning pathway. Meanwhile, the temporal embedding operates on the full patch, preserving access to observations beyond those represented by the selected spectral components.

For each retained coefficient, we construct the descriptor introduced in Section 3, using logarithmic amplitude, sine and cosine of phase, and the evidence indicator $\rho_{c,n,k}$. Relative amplitudes are computed with respect to the mean amplitude of the selected components within the same patch. This gives the encoder both a compressed measure of component strength and a bounded representation of its strength relative to the local spectrum. These are complementary parameterizations of amplitude: the latter makes the within-patch comparison directly available when interpreting the corresponding phase direction.

The four feature groups are arranged into frequency-ordered blocks and concatenated into $\mathbf{s}_{c,n} \in \mathbb{R}^{4K}$. A multilayer perceptron then produces the spectral phase coordinate:

$$\mathbf{\Phi}_{c,n} = f_\phi(\mathbf{s}_{c,n}) \in \mathbb{R}^{D_\phi}. \tag{5}$$

Joint projection allows the encoder to learn dependencies among the selected directions and their amplitude-derived evidence before generating a condition for the temporal token. Keeping the evidence indicator as a separate input also retains access to the original circular direction, allowing its contribution to be learned from the forecasting objective. The resulting coordinates form a patch-aligned conditioning sequence that is used for both local representation adjustment and variable-level contextual modeling.

4.3 Spectral Phase-Conditioned Representation Learning

The temporal token $\mathbf{z}_{c,n} \in \mathbb{R}^D$ summarizes the observed content of patch $\mathbf{p}_{c,n}$. Under alignment variation, the relevance of its features to future prediction can depend on the local phase state. PaCo therefore uses $\mathbf{\Phi}_{c,n}$ to generate a transformation of the token, making its feature weighting and offset conditional on the spectral phase coordinate. This provides a direct mechanism for relating local alignment information to the temporal representation used for forecasting.

A conditioning MLP maps the phase coordinate to feature-wise scale and shift adjustments together with a patch-wise gating logit:

$$[\boldsymbol{\gamma}_{c,n}, \boldsymbol{\beta}_{c,n}, g_{c,n}] = f_{\text{cond}}(\mathbf{\Phi}_{c,n}), \qquad f_{\text{cond}} : \mathbb{R}^{D_\phi} \to \mathbb{R}^{2D+1}. \tag{6}$$

Here $\boldsymbol{\gamma}_{c,n}, \boldsymbol{\beta}_{c,n} \in \mathbb{R}^D$ and $g_{c,n} \in \mathbb{R}$. The resulting residual modulation is

$$\widetilde{\mathbf{z}}_{c,n} = \mathbf{z}_{c,n} + \sigma(g_{c,n}) \left(\boldsymbol{\gamma}_{c,n} \odot \mathbf{z}_{c,n} + \boldsymbol{\beta}_{c,n}\right), \tag{7}$$

where $\odot$ denotes element-wise multiplication and σ is the sigmoid function. This expression is equivalent to interpolating between the original token and the affine candidate $(1 + \boldsymbol{\gamma}_{c,n}) \odot \mathbf{z}_{c,n} + \boldsymbol{\beta}_{c,n}$.

The multiplicative term adjusts the contribution of existing temporal features according to the local phase coordinate, while the additive term supplies a phase-dependent offset. Together, they allow the same temporal feature to contribute differently under different local phase states. The scalar gate controls the overall strength of the patch-level adjustment, leaving the feature-wise affine parameters to determine its direction within the representation space.

Amplitude-derived evidence influences both the affine parameters and the gate through $\mathbf{\Phi}_{c,n}$, with their effects learned jointly from the forecasting objective. The residual path retains the original temporal token. Initializing the final layer of f_{cond} to zero makes the initial affine adjustment vanish, so the module begins as an identity mapping and learns phase-dependent adjustments during training.

4.4 Phase-Guided Contextual Forecasting

PaCo reuses the phase coordinates to make cross-variable relevance depend on both temporal content and spectral phase state.

For each variable c, we average its conditioned tokens and phase coordinates over the N patches to obtain the summaries $\bar{\mathbf{z}}_c$ and $\bar{\mathbf{\Phi}}_c$, respectively. Pooling places contextual interaction at the variable level, while the individual patch representations remain available in the temporal pathway. The phase summary is computed from learned coordinates, allowing the pooled representation to reflect information jointly encoded from phase directions and amplitude-derived evidence.

Separate projections map these summaries into a shared D-dimensional space:

$$\mathbf{h}_c = f_{\text{tok}}(\bar{\mathbf{z}}_c) + f_{\text{ps}}(\bar{\mathbf{\Phi}}_c), \tag{8}$$

where $f_{\text{tok}} : \mathbb{R}^D \to \mathbb{R}^D$ and $f_{\text{ps}} : \mathbb{R}^{D_\phi} \to \mathbb{R}^D$. The separate projections accommodate the different dimensions and roles of temporal and phase summaries. Their combination provides a variable-level context from which both sources of information can influence contextual relevance.

Queries and keys are generated as $\mathbf{q}_c = W_q \mathbf{h}_c$ and $\mathbf{k}_c = W_k \mathbf{h}_c$, while values are generated as $\mathbf{v}_c = W_v \bar{\mathbf{z}}_c$, with $W_q, W_k, W_v \in \mathbb{R}^{D\times D}$. This assignment lets the joint temporal–phase context determine the interaction weights, while the exchanged values carry the summarized phase-conditioned temporal content. For target variable c, the contextual output is

$$a_{c,c'} = \frac{\exp(\mathbf{q}_c^\top \mathbf{k}_{c'}/\sqrt{D})}{\sum_{j=1}^{C} \exp(\mathbf{q}_c^\top \mathbf{k}_j/\sqrt{D})}, \qquad \mathbf{o}_c = \sum_{c'=1}^{C} a_{c,c'} \mathbf{v}_{c'}. \tag{9}$$

Here $a_{c,c'}$ determines the contribution of source variable c' to target variable c. Local modulation and contextual interaction thus use the same phase coordinate for two distinct decisions: how to adjust a patch representation, and how to weight information from other variables. Reusing the coordinate connects these decisions through a common spectral representation and requires no additional Fourier transform.

The contextual output $\mathbf{o}_c$ is broadcast to all patches of variable c and added residually, yielding $\mathbf{z}_{c,n}^{\text{ctx}} = \widetilde{\mathbf{z}}_{c,n} + \mathbf{o}_c$. This supplies a shared variable-level context while preserving differences among the locally conditioned patches. A refinement module f_{ref} then applies residual token-mixing and feature-mixing transformations: token mixing integrates information across patch positions, and feature mixing combines the resulting feature channels. The refined representation $\mathbf{Z}^{\text{ref}}$ therefore brings together local phase conditioning, cross-variable context, and temporal interactions before prediction.

For each variable, its N refined patch representations are concatenated and projected to the forecasting horizon:

$$\widehat{\mathbf{y}}_c = W_{\text{out}} \,\text{vec}\left(\mathbf{Z}^{\text{ref}}_{c,:,:}\right) + \mathbf{b}_{\text{out}} \in \mathbb{R}^H. \tag{10}$$

The per-variable predictions are assembled and transformed back to the original scale through inverse instance normalization, yielding $\widehat{\mathbf{Y}} \in \mathbb{R}^{H\times C}$. Given M training samples, all components are optimized jointly using mean squared error:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{MHC} \sum_{m=1}^{M} \sum_{h=1}^{H} \sum_{c=1}^{C} \left(\widehat{Y}_{m,h,c} - Y_{m,h,c}\right)^2. \tag{11}$$

Table 1: Average forecasting performance on each dataset across all prediction horizons. For each dataset, MSE and MAE are averaged over all evaluated forecasting horizons. Lower values indicate better performance. The **best** and second best results are marked in corresponding colors.

Model	PaCo (Ours)		PhaseFormer (2026)		PULSE (2026b)		FreDF (2025)		xPatch (2025)		iTransformer (2024)		PDF (2024)		TimeMixer (2024)		PatchTST (2023)	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.395**	**0.415**	0.414	0.418	0.422	0.425	0.436	0.433	0.471	0.460	0.457	0.449	0.408	0.430	0.450	0.445	0.413	0.434
ETTh2	0.333	**0.379**	0.352	0.394	0.361	0.390	0.372	0.397	0.432	0.438	0.383	0.406	0.341	0.387	0.377	0.404	**0.330**	0.381
ETTm1	**0.342**	**0.375**	0.349	**0.375**	0.362	0.387	0.394	0.399	0.360	0.377	0.407	0.411	0.349	0.377	0.386	0.400	0.352	0.382
ETTm2	**0.246**	**0.307**	0.261	0.320	0.262	0.314	0.278	0.319	0.271	0.316	0.291	0.335	0.252	0.316	0.274	0.324	0.256	0.317
Electricity	0.158	0.252	0.160	0.252	0.159	0.255	0.175	0.263	**0.157**	**0.247**	0.175	0.266	0.159	0.253	0.182	0.273	0.159	0.253
Weather	**0.217**	**0.252**	0.228	0.263	0.239	0.270	0.258	0.276	0.220	0.253	0.259	0.279	0.224	0.264	0.245	0.276	0.231	0.273
Exchange	**0.333**	**0.388**	0.471	0.471	0.418	0.429	0.387	0.424	0.382	0.416	0.367	0.415	0.426	0.434	0.395	0.421	0.341	0.395
WWTA	**0.371**	**0.276**	0.423	0.328	0.409	0.291	0.420	0.304	0.417	0.301	0.428	0.309	0.422	0.308	0.417	0.309	0.427	0.328

5 EXPERIMENTS

We evaluate PaCo on standard long-term time series forecasting benchmarks from multiple perspectives, including forecasting accuracy, component necessity, learned phase-coordinate behavior, parameter sensitivity, and computational efficiency.

5.1 EXPERIMENTAL SETUP

Datasets and baselines. We evaluate PaCo on widely used long-term forecasting benchmarks, including ETT (Zhou et al., 2021), Electricity, Weather (Wu et al., 2021), and Exchange. We also include the Wastewater Treatment Aeration (WWTA) dataset, collected from an operational wastewater treatment system, to evaluate PaCo in a real industrial forecasting scenario. We compare PaCo with recent competitive forecasting models, including PhaseFormer (Niu et al., 2026), PULSE (Liu et al., 2026b), FreDF (Wang et al., 2025), xPatch (Stitsyuk & Choi, 2025), iTransformer (Liu et al., 2024), PDF (Dai et al., 2024), TimeMixer (Wang et al., 2024), and PatchTST (Nie et al., 2023). Additional details are provided in Appendix B.

Forecasting protocol. For each dataset, the model observes a historical window of length L and predicts a future sequence of length H. We evaluate the standard forecasting horizons $H \in \{96, 192, 336, 720\}$. For all methods, we select the look-back length from $L \in \{96, 336, 512, 720\}$ according to validation performance under the same search space. For PaCo, each experiment is repeated with three random seeds, and we report the average test performance. All baseline results are reproduced under the same evaluation protocol. All datasets are split chronologically into training, validation, and test sets.

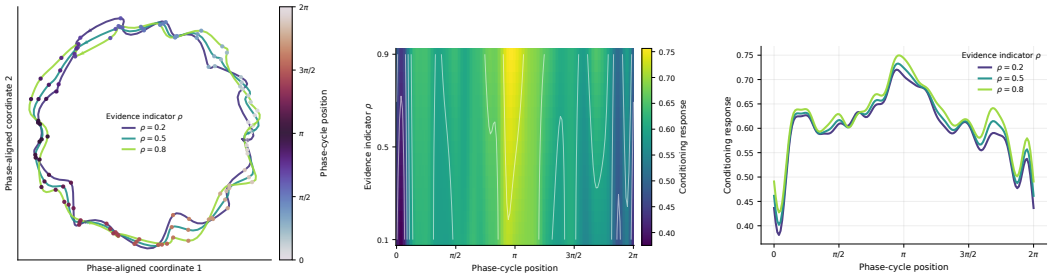
Metrics and implementation. We report mean squared error (MSE) and mean absolute error (MAE), where lower values indicate better performance. PaCo is implemented in PyTorch and trained with the Adam optimizer using early stopping based on validation loss. Hyperparameters are selected according to validation performance, and all reported results follow the same data preprocessing protocol as the baselines.

5.2 MAIN RESULTS

Tab. 1 reports the average forecasting performance across all prediction horizons on each dataset, with the complete results provided in Appendix B.3. PaCo achieves competitive overall performance and obtains the best or second-best results on most metrics. The advantage is more evident on several ETT subsets, Exchange, and the WWTA dataset, suggesting that phase-conditioned representations are useful when forecasting is affected by alignment-related temporal variation. On Electricity and Weather, PaCo remains among the top-performing methods, although the margin over strong baselines is smaller. This indicates that the benefit of phase conditioning depends on dataset characteristics, especially on whether local alignment variation is a major source of forecasting difficulty. Together with the ablation and phase-response analyses below, these results support the effectiveness of using evidence-supported spectral phase coordinates for temporal representation conditioning.

Table 2: Ablation study on the core components of PaCo. For each dataset, MSE and MAE are averaged over all evaluated forecasting horizons.

Model	ETTh1		ETTh2		ETTm1		ETTm2		Electricity		Weather		Exchange		Average		Δ (%)	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PaCo(Ours)	**0.395**	**0.415**	**0.333**	**0.379**	**0.342**	0.375	**0.246**	**0.307**	**0.158**	**0.252**	**0.217**	**0.252**	**0.333**	**0.388**	**0.289**	**0.338**	0.0%	0.0%
w/o Evidence	0.404	0.423	0.338	0.383	0.349	**0.374**	0.251	0.309	0.164	0.256	0.223	0.259	0.370	0.406	0.300	0.344	-3.7%	-1.7%
w/o Phase Direction	0.410	0.429	0.342	0.386	0.350	0.375	0.252	0.311	0.162	0.255	0.231	0.265	0.376	0.414	0.303	0.348	-4.6%	-2.9%
w/o Phase-Guided Context	0.410	0.432	0.356	0.385	0.351	0.376	0.253	0.310	0.168	0.259	0.225	0.261	0.397	0.419	0.309	0.349	-6.5%	-3.2%



(a) Phase-aligned coordinate (b) Response surface (c) Response profiles

Figure 3: Phase–evidence response analysis of the learned PaCo conditioning pathway on Weather with $H = 720$. (a) shows phase-aligned coordinate trajectories under different evidence indicators ρ. (b) shows the conditioning response surface over phase-cycle position and ρ. (c) shows representative response profiles. Details of the probing procedure are provided in Appendix C.

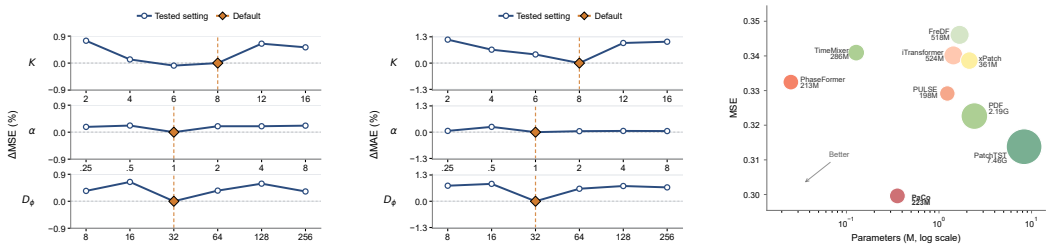
5.3 Ablation Study

To examine the contribution of each core design, we construct three ablated variants of PaCo. *w/o Evidence* removes the evidence indicator ρ from the phase descriptor while retaining logarithmic amplitude and circular phase direction. *w/o Phase Direction* removes the circular phase components and keeps only amplitude-related information. *w/o Phase-Guided Context* removes the use of phase coordinates in variable-level contextual interaction, so that contextual modeling is computed from temporal token summaries only.

Tab. 2 reports the ablation results. Removing any of these components leads to worse average performance, supporting the importance of the proposed phase-coordinate design. Among the variants, removing phase-guided contextual forecasting causes the largest performance drop, indicating that the extracted phase coordinate is useful not only for local representation conditioning but also for guiding variable-level interactions. Removing phase direction also clearly weakens the model, showing that circular phase direction carries important alignment-related information. Removing the evidence indicator results in a smaller but consistent drop, suggesting that amplitude-derived evidence is not the dominant predictive signal by itself, but helps the model interpret observed phase directions with amplitude-based support. Overall, the results show that circular phase direction, the evidence indicator, and phase-guided contextual modeling jointly contribute to effective phase-conditioned forecasting.

5.4 Learned Phase-Coordinate Response

To further examine whether the learned conditioning pathway captures the intended phase-coordinate behavior, we conduct a controlled phase–evidence response analysis on the trained PaCo model. Instead of perturbing the original input sequence, we construct controlled spectral descriptors by sweeping a shift-consistent phase pattern over one full cycle and varying the evidence indicator ρ while keeping the logarithmic amplitude term fixed. These descriptors are passed through the learned phase-coordinate encoder and affine-conditioning MLP to visualize both the phase-aligned coordinate trajectory and the resulting conditioning response. As shown in Fig. 3, the learned phase coordinates form smooth cyclic trajectories as the phase-cycle position changes, indicating that the coordinate preserves the circular structure of phase direction. Meanwhile, different values of ρ lead to distinguishable conditioning responses, showing that the evidence indicator affects the generated



(a) Relative MSE sensitivity. (b) Relative MAE sensitivity. (c) Accuracy–efficiency trade-off.

Figure 4: Parameter sensitivity and accuracy–efficiency analysis. (a) and (b) report relative MSE and MAE sensitivity on Weather with $H = 720$. (c) reports the accuracy–efficiency trade-off averaged over all datasets, with bubble area indicating computational cost. Detailed configurations and procedures are provided in Appendix D.

representation-conditioning signal. These results suggest that the learned conditioning pathway preserves phase-cycle variation while incorporating amplitude-derived evidence into phase-conditioned representation learning.

5.5 Parameter Sensitivity and Efficiency

We analyze the sensitivity of PaCo to K, the evidence saturation parameter α, and the phase-coordinate dimension D_ϕ. Fig. 4(a) and 4(b) report relative MSE and MAE changes on Weather with prediction horizon $H = 720$. Forecasting performance varies only modestly across the tested K settings, with the default $K = 8$ providing a compact representation and competitive accuracy. The configurations used in this analysis are detailed in Appendix D. The effect of α is milder because it only controls the saturation of the bounded evidence indicator and does not directly suppress phase components or alter the temporal backbone; within a moderate range, the relative amplitude-derived evidence is largely preserved. For D_ϕ, a small dimension may limit the capacity to represent the joint structure of circular phase direction and amplitude-derived evidence, while an excessively large dimension may introduce redundant degrees of freedom. The default setting therefore provides a balanced configuration.

Fig. 4(c) compares the accuracy–efficiency trade-off averaged over all datasets. PaCo achieves the lowest average MSE in this comparison while maintaining a moderate number of parameters and low computational cost. The additional overhead is limited because PaCo encodes only a compact set of selected spectral descriptors and uses the resulting phase coordinate to generate lightweight residual conditioning signals, rather than substantially enlarging the forecasting backbone.

6 Conclusion

This work studies long-term forecasting from an alignment-aware representation perspective. The results support representing recurring temporal patterns beyond their observed values, since their local alignment may vary across windows while the underlying pattern remains related. PaCo operationalizes this idea by representing spectral phase as a circular alignment coordinate and associating observed phase directions with amplitude-derived evidence. The resulting evidence-supported phase coordinate is used to calibrate temporal representations and guide context modeling.

Experiments, ablations, and phase-response analyses show that PaCo provides a useful representation signal for forecasting under local alignment variation. However, the benefit is not uniform across all datasets. On Electricity, where regular consumption patterns and variable interactions can already be effectively captured by strong forecasting models, PaCo shows a relatively limited margin. A key direction for future work is to develop adaptive phase-conditioning mechanisms that can identify when phase-based alignment cues should be emphasized and when they should be down-weighted.

AI USE STATEMENT

In this work, we used generative AI tools solely to improve the clarity, grammar, and readability of the manuscript. We did not use generative AI tools to generate research ideas, formulate the methodology, conduct experiments, analyze data, produce results, or draw conclusions. All AI-assisted revisions were carefully reviewed and verified by the authors to ensure that they preserved the intended meaning and technical accuracy. We take full responsibility for the final content of this work.

REFERENCES

Zhongde An, Jinhong You, Jiyanglin Li, Yiming Tang, Wen Li, Heming Du, and Shouguo Du. Fredn: Spectral disentanglement for time series forecasting via learnable frequency decomposition. In *AAAI Conference on Artificial Intelligence*, 2025. URL https://api.semanticscholar.org/CorpusID:283072770.

Fanpu Cao, Lu Dai, Jindong Han, and Hui Xiong. Enhancing multivariate time series forecasting with global temporal retrieval. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=QUJBPSfyui.

Weiqi Chen, Wenwei Wang, Bingqing Peng, Qingsong Wen, Tian Zhou, and Liang Sun. Learning to rotate: Quaternion transformer for complicated periodical time series forecasting. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 146–156, 2022.

Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Jigang Bao, Yong Jiang, and Shu-Tao Xia. Periodicity decoupling framework for long-term series forecasting. *International Conference on Learning Representations*, 2024.

Hui He, Qi Zhang, Kun Yi, Kaize Shi, Zhendong Niu, and Longbing Cao. Distributional drift adaptation with temporal conditional variational autoencoder for multivariate time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):7287–7301, 2025. doi: 10.1109/TNNLS.2024.3384842.

Essam H. Houssein, Meran Mohamed, Eman M. G. Younis, and Waleed M. Mohamed. Artificial intelligence and classical statistical models for time series forecasting: a comprehensive review. *Journal of Big Data*, 12, 2025. URL https://api.semanticscholar.org/CorpusID:283655625.

Hamid Reza Pilehvar Javid, Mahan Ahmadi Rahmatabadi, Behrooz Zaker, Hamid Reza Baghaee, and Gevork B. Gharehpetian. A comparative analysis of time-series deep learning forecasting models for peak electric load demand. In *2025 15th Smart Grid Conference (SGC)*, pp. 1–6, 2025. doi: 10.1109/SGC69320.2025.11372366.

Ming Jin, Guangsi Shi, Yuan-Fang Li, Bo Xiong, Tian Zhou, Flora D. Salim, Liang Zhao, Lingfei Wu, Qingsong Wen, and Shirui Pan. Towards expressive spectral-temporal graph neural networks for time series forecasting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47 (6):4926–4939, 2025. doi: 10.1109/TPAMI.2025.3545671.

Jongseon Kim, Hyungjoon Kim, HyunGi Kim, Dongjun Lee, and Sungroh Yoon. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. *Artificial Intelligence Review*, 58, 04 2025. doi: 10.1007/s10462-025-11223-9.

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 *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=cGDAkQo1C0p.

Fuxian Li, Huan Yan, Hongjie Sui, Deng Wang, Fan Zuo, Yue Liu, Yong Li, and Depeng Jin. Periodic shift and event-aware spatio-temporal graph convolutional network for traffic congestion prediction. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '23, New York, NY, USA, 2023. Association for Computing

Machinery. ISBN 9798400701689. doi: 10.1145/3589132.3625612. URL https://doi.org/10.1145/3589132.3625612.

Ruiqi Li, Maowei Jiang, Quangao Liu, Kai Wang, Kaiduo Feng, Yue Sun, and Xiufang Zhou. Faith: Frequency-domain attention in two horizons for time series forecasting. *Knowl. Based Syst.*, 309: 112790, 2024. URL https://api.semanticscholar.org/CorpusID:269983402.

Wenxiang Li and K. L. Eddie Law. Deep learning models for time series forecasting: A review. *IEEE Access*, 12:92306–92327, 2024. doi: 10.1109/ACCESS.2024.3422528.

Hao Liu, Xiaobin Zhu, Chun Yang, Lirui Deng, and Xucheng Yin. Amplitude-phase reconstruction for non-stationary time-series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–14, 2026a. doi: 10.1109/TKDE.2026.3701273.

Yangyou Liu, Zezhi Shao, Xinyu Chen, Hu Chen, Fei Wang, and Yuankai Wu. Pulse: Generative phase evolution for non-stationary time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026b.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 9881–9893. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4054556fcaa934b0bf76da52cf4f92cb-Paper-Conference.pdf.

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 B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 11116–11140, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/2ea18fdc667e0ef2ad82b2b4d65147ad-Paper-Conference.pdf.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. Wpmixer: efficient multi-resolution mixing for long-term 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.v39i18.34156. URL https://doi.org/10.1609/aaai.v39i18.34156.

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

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

Olanrewaju Odumuwagun. Time series-based quantitative risk models: Enhancing accuracy in forecasting and risk assessment. *International Journal of Computer Applications Technology and Research*, 12:29–41, 11 2023. doi: 10.7753/IJCATR1211.1006.

A.V. Oppenheim and J.S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5): 529–541, 1981. doi: 10.1109/PROC.1981.12022.

Athanasios Papoulis. *Signal Analysis*. McGraw-Hill, New York, 1977.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *SIGKDD*, pp. 1185–1196, 2025.

Hyunwoo Seo and Chiehyeon Lim. St-mtm: Masked time series modeling with seasonal-trend decomposition for time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, pp. 1209–1220, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400712456. doi: 10.1145/3690624.3709254. URL https://doi.org/10.1145/3690624.3709254.

Artyom Stitsyuk and Jaesik Choi. xpatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20601–20609, 2025.

Wenjie Sun, Jianhua Mei, Shengrui Liu, Chunhong Yuan, and Jiaxuan Zhao. Research on deep learning model for stock prediction by integrating frequency domain and time series features. *Scientific Reports*, 15, 2025a. URL https://api.semanticscholar.org/CorpusID:280690926.

Yanru Sun, Zongxia Xie, Emadeldeen Eldele, Dongyue Chen, Qinghua Hu, and Min Wu. Learning pattern-specific experts for time series forecasting under patch-level distribution shift. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, pp. 91810–91844. Curran Associates, Inc., 2025b. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/8491a7fcc218946b471b600a915c8b02-Paper-Conference.pdf.

Qingling Tao. Predictive analytics for traffic flow optimization in urban logistics: A transformer-based time series approach. *Science Progress*, 107(3):00368504241265196, 2024. doi: 10.1177/00368504241265196. URL https://doi.org/10.1177/00368504241265196. PMID: 39248169.

Guohong Wang, Xianhan Tan, Zengming Lin, Binli Luo, Shangjian Zhong, and Kele Xu. Freqevo: Enhancing time series forecasting with multi-level frequency domain feature extraction. *IEEE Transactions on Knowledge and Data Engineering*, 38(2):1099–1113, 2026. doi: 10.1109/TKDE.2025.3632365.

Hao Wang, Lichen Pan, Yuan Shen, Zhichao Chen, Degui Yang, Yifei Yang, Sen Zhang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. Fredf: Learning to forecast in the frequency domain. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 6893–6922, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/1457fb1e5d72cdc4ecd88bc10f916095-Paper-Conference.pdf.

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

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/bcc0d400288793e8bdcd7c19a8ac0c2b-Paper.pdf.

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

Nannan Xu, Kecheng Xu, Yuanyuan Cui, Guo Yang, and Meng Huang. Freapnet: Learnable amplitude–phase parameterization and dual-path frequency learning for multivariate time series forecasting. *Concurrency and Computation: Practice and Experience*, 38(7):e70675, 2026. doi: https://doi.org/10.1002/cpe.70675. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.70675.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with 10k parameters. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 26295–26318, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/701251e1db4a2e4dd2ef23f5265d5936-Paper-Conference.pdf.

Runze Yang, Longbing Cao, Jianxun Li, and Jie Yang. Rethinking fourier transform from a basis functions perspective for long-term time series forecasting. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 8515–8540. Curran Associates, Inc., 2024. doi: 10.52202/079017-0272. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/0fd4ce94d29be88a5a262a2c77a18f47-Paper-Conference.pdf.

Zhangjing Yang, Weiwu Yan, Xiaolin Huang, and Lin Mei. Adaptive temporal-frequency network for time-series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1576–1587, 2022. doi: 10.1109/TKDE.2020.3003420.

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 A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 76656–76679. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/f1d16af76939f476b5f040fd1398c0a3-Paper-Conference.pdf.

Weijia Zhang, Le Zhang, Jindong Han, Hao Liu, Yanjie Fu, Jingbo Zhou, Yu Mei, and Hui Xiong. Irregular traffic time series forecasting based on asynchronous spatio-temporal graph convolutional networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pp. 4302–4313, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400704901. doi: 10.1145/3637528.3671665. URL https://doi.org/10.1145/3637528.3671665.

Shuchen Zhao, Zhongjin Xu, Zhefan Zhu, Xiaoxiang Liang, Zecheng Zhang, and Ruxue Jiang. Short and long-term renewable electricity demand forecasting based on cnn-bi-gru model. *ICCK Transactions on Emerging Topics in Artificial Intelligence*, 2(1):1–15, 2025. ISSN 3068-6652. doi: 10.62762/TETAI.2024.532253. URL https://www.icck.org/article/abs/TETAI.2024.532253.

Haotian Zheng, Jiang Wu, Runze Song, Lingfeng Guo, and Zeqiu Xu. Predicting financial enterprise stocks and economic data trends using machine learning time series analysis. *Applied and Computational Engineering*, 87:26–32, 07 2024. doi: 10.54254/2755-2721/87/20241562.

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 *Proc. AAAI Conf. Artif. Intell.*, 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.

Xuanbing Zhu, Dunbin Shen, Hongyu Wang, and Yingguang Hao. Fcnet: Fully complex network for time series forecasting. *IEEE Internet of Things Journal*, 11(24):40127–40139, 2024. doi: 10.1109/JIOT.2024.3449903.

APPENDIX CONTENTS

A Derivation of Amplitude-Dependent Phase Stability **15**

B Additional Experimental Details **16**

B.1 Datasets and Baselines 16

B.2 Implementation Details 17

B.3 Complete Forecasting Results 18

C Details of Phase–Evidence Response Analysis **18**

D Details of Parameter Sensitivity and Efficiency Analysis **19**

E Notation **21**

F Additional Forecasting Examples **21**

A Derivation of Amplitude-Dependent Phase Stability

This appendix provides the derivation behind the amplitude-dependent stability of observed phase directions discussed in Section 3. Consider a nonzero Fourier coefficient

$$X = Ae^{i\phi}, \qquad A > 0, \tag{12}$$

and its perturbed observation

$$\widehat{X} = X + \eta, \tag{13}$$

where η is a complex perturbation. For $\widehat{X} \neq 0$, we define the wrapped phase error as

$$\Delta\phi = \arg\left(\frac{\widehat{X}}{X}\right) \in (-\pi, \pi]. \tag{14}$$

Here $\arg(\cdot)$ denotes the principal argument. Multiplying both coefficients by $e^{-i\phi}$ rotates the original coefficient to the positive real axis:

$$Xe^{-i\phi} = A, \qquad \widehat{X}e^{-i\phi} = A + \eta e^{-i\phi}. \tag{15}$$

Let the rotated perturbation be

$$\eta e^{-i\phi} = a + ib. \tag{16}$$

Then the phase error can be written exactly as

$$\Delta\phi = \arg(A + a + ib) = \operatorname{atan2}(b, A + a). \tag{17}$$

When the perturbation is small relative to the coefficient amplitude, i.e., $|\eta|/A$ is small, the phase error admits a first-order approximation. Since

$$\widehat{X} = X\left(1 + \frac{\eta}{X}\right), \tag{18}$$

we have

$$\Delta\phi = \arg\left(1 + \frac{\eta}{X}\right). \tag{19}$$

Using the expansion $\arg(1 + z) = \operatorname{Im}(z) + O(|z|^2)$ for small z, we obtain

$$\Delta\phi = \operatorname{Im}\left(\frac{\eta}{X}\right) + O\left(\frac{|\eta|^2}{A^2}\right). \tag{20}$$

Therefore,

$$|\Delta\phi| \lesssim \frac{|\eta|}{A}, \tag{21}$$

showing that, under perturbations of comparable magnitude, the phase direction of a smaller-amplitude coefficient is more sensitive.

A geometric bound leads to the same conclusion. If $|\eta| < A$, then the perturbed coefficient lies in a disk of radius $|\eta|$ centered at X, and the disk does not contain the origin. The maximum angular deviation from the original direction occurs along the tangent from the origin to this disk, yielding

$$|\Delta\phi| \leq \arcsin\left(\frac{|\eta|}{A}\right). \tag{22}$$

Thus, coefficient amplitude provides support for assessing the stability of an observed phase direction. This motivates the amplitude-derived evidence used in PaCo's evidence-supported phase coordinate.

Table 3: Dataset statistics.

Dataset	Var	Samples	Input Length	Horizon	Freq	Size	Domain
ETTh1	7	17420	{96,336,512,720}	{96,192,336,720}	1 hour	0.25M	Temperature
ETTh2	7	17420	{96,336,512,720}	{96,192,336,720}	1 hour	0.23M	Temperature
ETTm1	7	69680	{96,336,512,720}	{96,192,336,720}	15 min	1.00M	Temperature
ETTm2	7	69680	{96,336,512,720}	{96,192,336,720}	15 min	0.94M	Temperature
Electricity	321	26304	{96,336,512,720}	{96,192,336,720}	1 hour	93M	Electricity
Weather	21	52696	{96,336,512,720}	{96,192,336,720}	10 min	7M	Weather
Exchange	8	7588	{96,336,512,720}	{96,192,336,720}	1 day	0.6M	Finance
WWTA	6	44388	{96,336,512,720}	{96,192,336,720}	1 min	2.6M	Wastewater

B Additional Experimental Details

B.1 Datasets and Baselines

Datasets. We evaluate PaCo on both public long-term forecasting benchmarks and a real industrial dataset. The public benchmarks include ETT (Zhou et al., 2021), Electricity, Weather (Wu et al., 2021), and Exchange. ETT contains four subsets, namely ETTh1, ETTh2, ETTm1, and ETTm2. We obtain the public benchmark files from the data repository released with TimesNet (Wu et al., 2023)[1], and follow the commonly used preprocessing and chronological split protocol.

In addition to these public benchmarks, we use the Wastewater Treatment Aeration (WWTA) dataset to evaluate forecasting performance in a real industrial scenario. WWTA is collected from an operational wastewater treatment plant in a city from July 20, 2025 to August 20, 2025. The dataset contains six process variables related to wastewater treatment aeration, sampled at a one-minute frequency, resulting in 44,388 time steps. Compared with public benchmarks, WWTA reflects practical industrial operating conditions, where the monitored variables may exhibit strong temporal dynamics and process-dependent interactions. WWTA contains process-level measurements only and does not include personal data or facility-identifying information. Due to operational confidentiality, WWTA is not publicly released. To support reproducibility as much as possible, we report its collection period, sampling frequency, number of variables, number of time steps, and dataset statistics in Table 3. All datasets are divided chronologically into training, validation, and test sets. The detailed dataset statistics are summarized in Tab. 3.

Baselines. We compare PaCo with representative and recent competitive forecasting models from several categories. PhaseFormer (Niu et al., 2026) and PULSE (Liu et al., 2026b) are included as recent phase-aware forecasting methods. FreDF (Wang et al., 2025) and PDF (Dai et al., 2024) are included as frequency-domain or periodicity-aware forecasting methods. xPatch (Stitsyuk & Choi, 2025) and PatchTST (Nie et al., 2023) are included as strong patch-based forecasting models. iTransformer (Liu et al., 2024) and TimeMixer (Wang et al., 2024) are included as representative Transformer/Mixer-style methods.

Baseline details. We provide brief descriptions of the compared baselines below.

- **PhaseFormer** (Niu et al., 2026) is a recent phase-aware forecasting model that rethinks time-series forecasting from a phase perspective. Instead of relying only on patch-level processing, PhaseFormer introduces phase-wise prediction to model periodic structure more efficiently. It uses compact phase embeddings and an efficient cross-phase interaction mechanism implemented with lightweight routing. We include PhaseFormer as a representative baseline that explicitly uses phase-level temporal modeling.
- **PULSE** (Liu et al., 2026b) is a recent framework for non-stationary time series forecasting based on generative phase evolution. The method argues that static historical assumptions may lead to a failure mode called phase amnesia, where the model becomes insensitive to evolving global context. PULSE adopts a Disentangle–Evolve–Simulate design, using phase-anchored disentanglement, a Phase Router for future trajectory generation, and

[1] https://drive.google.com/drive/folders/13Cg1KYOlzM5C7K8gK8NfC-F3EYxkM3D2

Statistic-Aware Mixup to improve robustness under distribution shifts. We include PULSE as a recent phase-related baseline designed for non-stationary forecasting.

- **FreDF** (Wang et al., 2025) is a frequency-enhanced direct forecasting method. It observes that many direct forecasting models focus on autocorrelation in the historical sequence but neglect autocorrelation within the future label sequence. To mitigate the estimation bias caused by label correlation, FreDF learns to forecast in the frequency domain. We include FreDF as a representative frequency-domain baseline that directly addresses the forecasting objective from a spectral perspective.
- **PDF** (Dai et al., 2024) is a Periodicity Decoupling Framework for long-term series forecasting. It exploits the intrinsic periodicity of time series by decoupling one-dimensional time series into periodic components and then modeling their two-dimensional temporal variations. PDF consists of a multi-periodic decoupling block, a dual variations modeling block, and a variations aggregation block. We include PDF as a representative periodicity-aware baseline.
- **xPatch** (Stitsyuk & Choi, 2025) is a non-Transformer patch-based forecasting model built on exponential seasonal-trend decomposition. It introduces a dual-stream architecture, where an MLP-based linear stream and a CNN-based nonlinear stream are used to capture different temporal patterns. The model also investigates the benefit of patching and channel independence outside the standard Transformer architecture. We include xPatch as a recent strong patch-based baseline with decomposition-based temporal modeling.
- **PatchTST** (Nie et al., 2023) is a strong Transformer-based baseline for long-term time series forecasting. It segments time series into subseries-level patches, which serve as input tokens to the Transformer. It also adopts channel independence, where each channel is treated as a separate univariate series while sharing the same embedding and Transformer weights. This design retains local semantic information, reduces the computational cost of attention, and enables the model to attend to longer historical contexts.
- **iTransformer** (Liu et al., 2024) is an inverted Transformer architecture for time series forecasting. Instead of treating time points as tokens, iTransformer embeds the historical sequence of each variable into a variate token. Self-attention is then applied across variate tokens to capture multivariate correlations, while the feed-forward network is used to learn nonlinear representations for each variate. We include iTransformer as a representative Transformer-style baseline with variable-centric representation learning.
- **TimeMixer** (Wang et al., 2024) is a fully MLP-based forecasting model based on decomposable multiscale mixing. It starts from the observation that time series may exhibit different patterns at different sampling scales. TimeMixer uses Past-Decomposable-Mixing to decompose and mix seasonal and trend components across scales, and Future-Multipredictor-Mixing to combine predictions from multiple scale-specific predictors. We include TimeMixer as a representative mixer-style baseline for multiscale temporal modeling.

Baseline implementation protocol. For each baseline, we use the official implementation when available and follow the recommended hyperparameter configuration provided by the authors. When multiple input lengths are supported, we evaluate the same candidate look-back lengths for all methods and report the best validation-selected setting. This protocol reduces the influence of input-length sensitivity and keeps the comparison consistent across different architectures. During testing, we disable `drop_last` for all models so that evaluation is conducted on the complete test set.

B.2 Implementation Details

PaCo is implemented in PyTorch. All experiments are conducted on a single NVIDIA GeForce RTX 5070 GPU with 12GB memory. We use the Adam optimizer and train each model with early stopping based on validation loss. The main architectural configuration and training hyperparameter selection strategy of PaCo are summarized in Tab. 4. Unless otherwise specified, all models follow the same preprocessing, chronological data split, and evaluation protocol. For a fair comparison, `drop_last` is disabled during testing for all methods so that evaluation is conducted on the complete test set.

Table 4: PaCo architectural configuration and training hyperparameter selection.

Hyperparameter	Setting
Selected phase frequencies K	8
Evidence saturation α	1.0
Phase-coordinate dimension D_ϕ	32
Patch length P	16
Patch stride	8
Token dimension D	64
Optimizer	Adam
Early stopping patience	12
Learning rate	0.0008
Dropout rate	selected on validation set
Batch size	selected on validation set

Table 5: Complete forecasting performance on all datasets and prediction horizons. For each dataset, we report MSE and MAE under $H \in \{96, 192, 336, 720\}$. Lower values indicate better performance. The **best** and second-best results are marked in corresponding colors.

Model		PaCo (ours)		PhaseFormer (2026)		PULSE (2026)		FreDF (2025)		xPatch (2025)		iTransformer (2024)		PDF (2024)		TimeMixer (2024)		PatchTST (2023)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.359**	0.388	0.366	**0.387**	0.365	0.392	0.377	0.397	0.384	0.401	0.386	0.404	0.360	0.393	0.385	0.401	0.372	0.402
	192	**0.394**	0.409	0.406	**0.407**	0.417	0.421	0.430	0.426	0.471	0.457	0.441	0.436	0.395	0.416	0.401	0.428	0.413	0.429
	336	**0.412**	**0.423**	0.444	0.432	0.456	0.436	0.465	0.442	0.511	0.473	0.490	0.462	0.427	0.440	0.513	0.470	0.421	0.439
	720	**0.416**	**0.442**	0.440	0.447	0.452	0.450	0.471	0.468	0.516	0.510	0.509	0.493	0.451	0.470	0.502	0.480	0.447	0.468
	AVG	**0.395**	**0.415**	0.414	0.418	0.422	0.425	0.436	0.433	0.471	0.460	0.457	0.449	0.408	0.430	0.450	0.445	0.413	0.434
ETTh2	96	**0.264**	**0.329**	0.281	0.343	0.277	**0.329**	0.286	0.335	0.326	0.366	0.300	0.349	0.278	0.340	0.290	0.342	0.273	0.337
	192	**0.327**	**0.369**	0.339	0.379	0.356	0.381	0.364	0.385	0.427	0.430	0.380	0.398	0.334	0.375	0.382	0.401	0.339	0.380
	336	0.355	0.394	0.379	0.412	0.398	0.415	0.419	0.427	0.467	0.459	0.423	0.432	0.358	0.399	0.416	0.431	**0.328**	**0.384**
	720	0.385	0.423	0.409	0.442	0.413	0.436	0.421	0.440	0.509	0.495	0.430	0.447	0.392	0.433	0.419	0.440	**0.379**	**0.422**
	AVG	0.333	**0.379**	0.352	0.394	0.361	0.390	0.372	0.397	0.432	0.438	0.383	0.406	0.341	0.387	0.377	0.404	**0.330**	0.381
ETTm1	96	**0.283**	**0.339**	0.301	0.345	0.302	0.346	0.324	0.361	0.293	0.340	0.341	0.376	0.302	0.350	0.322	0.361	0.290	0.344
	192	**0.326**	**0.363**	0.328	0.365	0.348	0.378	0.375	0.386	0.339	0.368	0.382	0.395	0.333	0.368	0.369	0.388	0.333	0.370
	336	0.353	0.383	0.356	0.379	0.374	0.397	0.405	0.406	0.371	**0.378**	0.418	0.418	**0.348**	0.379	0.390	0.402	0.369	0.392
	720	**0.408**	0.414	0.413	**0.412**	0.423	0.426	0.471	0.443	0.437	0.423	0.487	0.456	0.413	**0.412**	0.462	0.449	0.417	0.420
	AVG	**0.342**	**0.375**	0.349	**0.375**	0.362	0.387	0.394	0.399	0.360	0.377	0.407	0.411	0.349	0.377	0.386	0.400	0.352	0.382
ETTm2	96	**0.158**	**0.247**	0.173	0.259	0.162	**0.247**	0.174	0.252	0.170	0.250	0.185	0.272	0.163	0.255	0.174	0.257	0.166	0.256
	192	**0.211**	**0.285**	0.236	0.303	0.226	0.290	0.239	0.296	0.231	0.291	0.253	0.313	0.218	0.294	0.238	0.300	0.223	0.296
	336	**0.264**	**0.321**	0.280	0.335	0.283	0.329	0.301	0.336	0.285	0.327	0.315	0.350	0.272	0.330	0.295	0.339	0.273	0.329
	720	**0.352**	**0.377**	0.356	0.382	0.378	0.389	0.400	0.393	0.397	0.395	0.413	0.406	0.355	0.384	0.390	0.399	0.361	0.385
	AVG	**0.246**	**0.307**	0.261	0.320	0.262	0.314	0.278	0.319	0.271	0.316	0.291	0.335	0.252	0.316	0.274	0.324	0.256	0.317
Electricity	96	0.127	0.221	0.128	0.229	0.133	0.228	0.145	0.234	**0.126**	**0.218**	0.147	0.239	0.130	0.222	0.153	0.247	0.129	0.222
	192	0.147	0.241	0.147	0.237	0.150	0.243	0.160	0.248	**0.146**	**0.234**	0.166	0.257	0.149	0.243	0.166	0.256	0.147	0.240
	336	0.162	0.257	0.166	0.258	0.166	0.261	0.173	0.264	0.161	**0.252**	0.178	0.271	**0.159**	0.256	0.185	0.277	0.163	0.259
	720	0.196	0.290	0.198	0.284	**0.186**	0.288	0.224	0.306	0.193	**0.283**	0.208	0.297	0.198	0.291	0.225	0.310	0.197	0.290
	AVG	0.158	0.252	0.160	0.252	0.159	0.255	0.175	0.263	**0.157**	**0.247**	0.175	0.266	0.159	0.253	0.182	0.273	0.159	0.253
Weather	96	0.140	**0.184**	0.149	0.195	0.149	0.195	0.167	0.203	**0.137**	0.186	0.174	0.214	0.144	0.197	0.163	0.212	0.155	0.211
	192	0.182	0.227	0.194	0.240	0.200	0.244	0.216	0.250	**0.177**	**0.225**	0.225	0.257	0.192	0.245	0.212	0.255	0.208	0.259
	336	**0.235**	**0.272**	0.248	0.283	0.260	0.290	0.278	0.294	0.239	0.274	0.280	0.298	0.245	0.281	0.263	0.292	0.248	0.286
	720	**0.312**	**0.326**	0.319	0.333	0.347	0.353	0.371	0.359	0.327	0.327	0.358	0.349	0.316	0.333	0.343	0.346	0.313	0.336
	AVG	**0.217**	**0.252**	0.228	0.263	0.239	0.270	0.258	0.276	0.220	0.253	0.259	0.279	0.224	0.264	0.245	0.276	0.231	0.273
Exchange	96	**0.079**	**0.197**	0.109	0.237	0.084	0.202	0.102	0.228	0.085	0.202	0.110	0.237	0.093	0.216	0.089	0.207	0.088	0.207
	192	**0.173**	0.298	0.218	0.341	0.180	0.299	0.203	0.324	0.201	0.317	0.182	0.307	0.200	0.323	0.190	0.307	**0.173**	**0.296**
	336	**0.307**	**0.398**	0.441	0.493	0.354	0.431	0.351	0.430	0.370	0.441	0.343	0.426	0.346	0.426	0.385	0.451	0.312	0.406
	720	**0.772**	**0.658**	1.117	0.813	1.056	0.784	0.891	0.715	0.871	0.704	0.835	0.692	1.064	0.771	0.917	0.718	0.792	0.672
	AVG	**0.333**	**0.388**	0.471	0.471	0.418	0.429	0.387	0.424	0.382	0.416	0.367	0.415	0.426	0.434	0.395	0.421	0.341	0.395
WWTA	96	**0.202**	**0.149**	0.247	0.202	0.223	0.152	0.231	0.169	0.230	0.164	0.239	0.174	0.231	0.170	0.250	0.186	0.243	0.190
	192	**0.304**	**0.230**	0.361	0.299	0.326	0.235	0.336	0.250	0.336	0.247	0.344	0.258	0.341	0.257	0.347	0.255	0.356	0.295
	336	**0.399**	**0.303**	0.450	0.353	0.438	0.324	0.454	0.336	0.439	0.335	0.464	0.342	0.451	0.337	0.442	0.335	0.463	0.342
	720	**0.581**	**0.423**	0.636	0.458	0.649	0.454	0.659	0.463	0.664	0.458	0.667	0.462	0.664	0.470	0.630	0.462	0.648	0.484
	AVG	**0.371**	**0.276**	0.423	0.328	0.409	0.291	0.420	0.304	0.417	0.301	0.428	0.309	0.422	0.308	0.417	0.309	0.427	0.328
1st Count		29	27	0	4	1	2	0	0	5	7	0	0	2	1	0	0	4	3

B.3 Complete Forecasting Results

Tab. 5 provides the complete long-term forecasting results on all datasets and prediction horizons. For each dataset, we report MSE and MAE under horizons $H \in \{96, 192, 336, 720\}$. These detailed results correspond to the averaged performance summarized in Tab. 1 of the main text.

C Details of Phase–Evidence Response Analysis

This section provides the detailed probing procedure used in Section 5.4. The goal is to inspect the learned phase-coordinate conditioning pathway of a trained PaCo model, rather than to evaluate forecasting accuracy under additional input perturbations. Therefore, the original input sequence is

not modified. Instead, we construct controlled spectral descriptors and directly feed them into the learned phase-coordinate encoder and affine-conditioning MLP.

Let K denote the number of selected non-DC frequency components used by the phase encoder. We sweep a normalized phase-cycle position $\tau \in [0, 1]$ and assign the phase of the k-th component as

$$\phi_k(\tau) = -2\pi k\tau, \qquad k = 1, \dots, K. \tag{23}$$

This construction follows the shift-consistent phase pattern induced by temporal displacement, where different frequency components rotate at rates proportional to their frequency indices. For each τ, we construct the circular phase representation using $\sin \phi_k(\tau)$ and $\cos \phi_k(\tau)$. The logarithmic amplitude term is fixed to a constant value for all components, so that the probing focuses on the effects of phase direction and the evidence indicator.

For a given value of the evidence indicator ρ, the controlled descriptor is constructed in the same form as the phase descriptor used by PaCo:

$$\mathbf{s}(\tau, \rho) = \left[\{\log(1 + A)\}_{k=1}^{K}, \{\sin \phi_k(\tau)\}_{k=1}^{K}, \{\cos \phi_k(\tau)\}_{k=1}^{K}, \{\rho\}_{k=1}^{K}\right], \tag{24}$$

where A is fixed across all phase-cycle positions and all values of ρ. We sweep τ over one full cycle and evaluate multiple values of ρ. The descriptors are passed through the learned phase-coordinate encoder,

$$\mathbf{\Phi}(\tau, \rho) = f_\phi(\mathbf{s}(\tau, \rho)), \tag{25}$$

and then through the learned affine-conditioning MLP,

$$[\boldsymbol{\gamma}(\tau, \rho), \boldsymbol{\beta}(\tau, \rho), g(\tau, \rho)] = f_{\text{cond}}(\mathbf{\Phi}(\tau, \rho)). \tag{26}$$

To summarize the magnitude of the generated conditioning signal, we define the conditioning response as

$$R(\tau, \rho) = \sigma(g(\tau, \rho)) \frac{\|\boldsymbol{\gamma}(\tau, \rho)\|_2 + \|\boldsymbol{\beta}(\tau, \rho)\|_2}{\sqrt{D}}, \tag{27}$$

where D is the token dimension and $\sigma(\cdot)$ is the sigmoid function. This response measures the gated strength of the residual affine conditioning generated from the controlled phase coordinate.

For visualization, we also compute a phase-aligned projection of the learned coordinates. Specifically, we fit a linear projection from $\mathbf{\Phi}(\tau, \rho)$ to the two-dimensional target

$$[\cos(2\pi\tau), \sin(2\pi\tau)]. \tag{28}$$

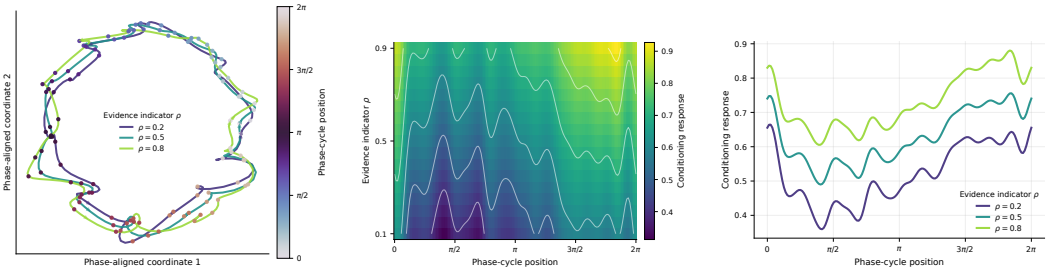
This projection is used only for visualization and does not affect the trained model. It allows us to inspect whether the learned phase coordinates preserve the cyclic structure induced by the controlled phase sweep.

In Fig. 3, we apply this probing procedure to the trained PaCo model on Weather with prediction horizon $H = 720$. The phase-aligned trajectories visualize how the learned coordinates vary across phase-cycle positions and values of ρ, while the response surface and response profiles show how the learned conditioning pathway changes its generated residual affine signal. This analysis complements the ablation study by directly inspecting the behavior of the learned phase-coordinate conditioning pathway.

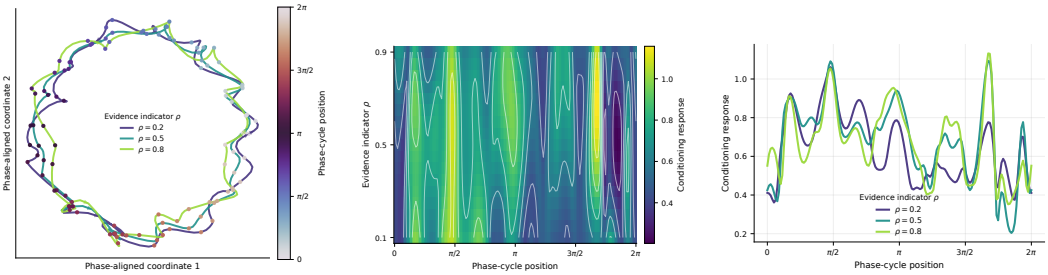
Additional results on other datasets. To further examine whether the learned phase–evidence response behavior is consistent beyond Weather, we provide the same analysis on ETTm1, Exchange, and Electricity. Fig. 5 shows the phase-aligned coordinate trajectories, conditioning response surfaces, and response profiles on these datasets. Although the detailed response shapes differ across datasets, the learned conditioning pathway generally preserves cyclic variation along the controlled phase-cycle position and produces distinguishable responses under different evidence indicators. These results provide additional evidence that the phase-coordinate pathway captures structured responses associated with circular phase direction and amplitude-derived evidence.

D Details of Parameter Sensitivity and Efficiency Analysis

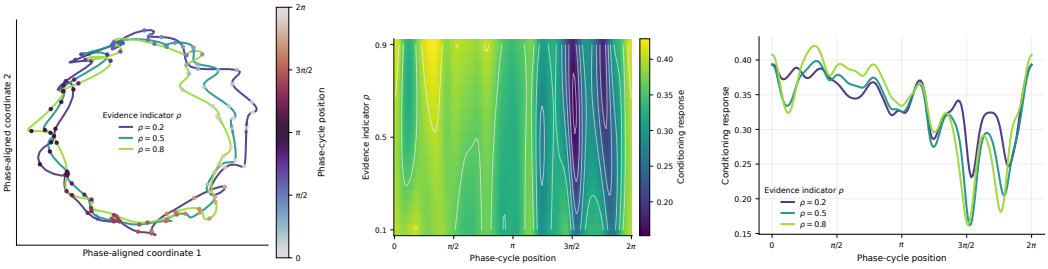
Parameter sensitivity. We conduct parameter sensitivity analysis on the Weather dataset with prediction horizon $H = 720$. The default configuration is $K = 8$, $\alpha = 1.0$, and $D_\phi = 32$.



(a) ETTm1: Phase-aligned coordinate (b) ETTm1: Response surface (c) ETTm1: Response profiles



(d) Exchange: Phase-aligned coordinate (e) Exchange: Response surface (f) Exchange: Response profiles



(g) Electricity: Phase-aligned coordinate (h) Electricity: Response surface (i) Electricity: Response profiles

Figure 5: Additional phase–evidence response analysis on ETTm1, Exchange, and Electricity. Each row corresponds to one dataset, and the three columns show the phase-aligned coordinate, conditioning response surface, and representative response profiles, respectively.

When one hyperparameter is varied, the other two are fixed to their default values. We test $K \in \{2, 4, 6, 8, 12, 16\}$, $\alpha \in \{0.25, 0.5, 1, 2, 4, 8\}$, and $D_\phi \in \{8, 16, 32, 64, 128, 256\}$. With $P = 16$, the settings $K > 8$ retain all eight available non-DC rFFT coefficients and pad the remaining positions with zero-valued complex coefficients before descriptor computation. Padding affects both the descriptor dimension and the stabilized mean amplitude used to compute ρ, while the retained frequency components remain unchanged. For each setting, we report relative MSE and MAE changes with respect to the default configuration, e.g., $\Delta\text{MSE} = (\text{MSE}(p) - \text{MSE}(p_{\text{default}}))/\text{MSE}(p_{\text{default}}) \times 100\%$.

Accuracy–efficiency trade-off. For efficiency analysis, we compare PaCo with representative baselines using the average MSE over all evaluated datasets and forecasting horizons. We also report the number of trainable parameters and the computational cost measured in MFLOPs per sample. The computational cost is computed as

$$\text{MFLOPs/sample} = \frac{\text{FLOPs per batch}}{\text{batch size} \times 10^6}.$$

Table 6: Accuracy–efficiency statistics used in Fig. 4(c). Average MSE is computed over all evaluated datasets and forecasting horizons.

Model	MSE	Params	FLOPs/sample
PaCo	0.2996	0.36M	222.52M
PhaseFormer	0.3324	0.03M	213.39M
PULSE	0.3292	1.23M	198.00M
FreDF	0.3401	1.43M	518.08M
xPatch	0.3386	2.13M	361.24M
iTransformer	0.3461	1.67M	523.88M
PDF	0.3226	2.41M	2.19G
TimeMixer	0.3410	0.13M	286.40M
PatchTST	0.3138	8.31M	7.46G

For consistency, we count one multiply–accumulate operation as two floating-point operations, i.e., 1 MAC $= 1$ multiplication $+1$ addition $= 2$ FLOPs. In Fig. 4(c), the x-axis denotes parameters, the y-axis denotes average MSE, and the bubble area indicates MFLOPs per sample. The second line of each label reports FLOPs, using M for MFLOPs and G for GFLOPs.

The corresponding statistics are reported in Tab. 6.

E Notation

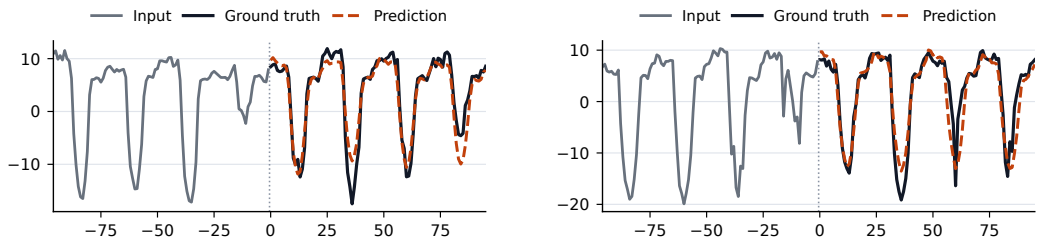
Table 7 summarizes the main symbols used in the PaCo framework and the accompanying analyses. Scalar quantities are listed with size 1, and tensor sizes omit the batch dimension. The indices c, n, and k denote variables, patches, and frequency components, respectively.

F Additional Forecasting Examples

We provide additional qualitative forecasting examples to illustrate the prediction behavior of PaCo on representative datasets. Fig. 6 shows two test examples for each dataset, including ETTh1, ETTm2, Electricity, and Weather, with prediction length $H = 96$. Each example compares the ground-truth future sequence with the prediction generated by PaCo. These visualizations provide a direct view of how PaCo follows temporal trends and local variations across different data domains.

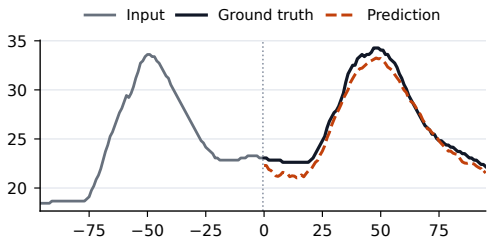
Table 7: Main notation used in PaCo. Tensor sizes omit the batch dimension.

Symbol	Size	Meaning
$L,\ H$	1	Input length and forecasting horizon.
C	1	Number of variables.
$P,\ N$	1	Patch length and number of patches per variable.
$D,\ D_\phi$	1	Temporal-token and phase-coordinate dimensions.
$K,\ M$	1	Numbers of selected frequencies and training samples.
$\mathbf{X}$	$L \times C$	Input multivariate sequence.
$\mathbf{Y},\ \widehat{\mathbf{Y}}$	$H \times C$	Ground-truth and predicted future sequences.
$\mathbf{p}_{c,n}$	P	Normalized patch n of variable c.
$X_{c,n,k}$	1	Complex Fourier coefficient at frequency k.
$A_{c,n,k},\ \bar{A}_{c,n}$	1	Component amplitude and mean selected amplitude.
$\phi_{c,n,k}$	1	Phase angle of the corresponding coefficient.
$\rho_{c,n,k}$	1	Relative-amplitude evidence indicator.
α	1	Evidence saturation parameter.
$\mathbf{d}_k$	4	Descriptor of a single frequency component.
$\mathbf{s}_{c,n}$	$4K$	Concatenated patch-level spectral descriptor.
$\mathbf{\Phi}_{c,n}$	D_ϕ	Learned phase coordinate of patch (c, n).
$\mathbf{\Phi}$	$C \times N \times D_\phi$	Phase coordinates across variables and patches.
$\mathbf{z}_{c,n},\ \widetilde{\mathbf{z}}_{c,n}$	D	Original and phase-conditioned temporal tokens.
$\mathbf{Z},\ \widetilde{\mathbf{Z}}$	$C \times N \times D$	Original and phase-conditioned token tensors.
$\boldsymbol{\gamma}_{c,n},\ \boldsymbol{\beta}_{c,n}$	D	Feature-wise scale adjustment and shift.
$g_{c,n}$	1	Patch-wise gating logit.
$\bar{\mathbf{z}}_c$	D	Mean phase-conditioned token over patches.
$\bar{\mathbf{\Phi}}_c$	D_ϕ	Mean phase coordinate over patches.
$\mathbf{h}_c$	D	Joint temporal–phase context.
$\mathbf{q}_c,\ \mathbf{k}_c,\ \mathbf{v}_c$	D	Variable-level query, key, and value vectors.
$W_q,\ W_k,\ W_v$	$D \times D$	Query, key, and value projection matrices.
$a_{c,c'}$	1	Attention weight from target c to source c'.
$\mathbf{o}_c$	D	Aggregated cross-variable context.
$\mathbf{z}^{\mathrm{ctx}}_{c,n}$	D	Token after residual context addition.
$\mathbf{Z}^{\mathrm{ctx}},\ \mathbf{Z}^{\mathrm{ref}}$	$C \times N \times D$	Context-enhanced and refined token tensors.
W_{out}	$H \times ND$	Forecasting-head weight matrix.
$\mathbf{b}_{\mathrm{out}}$	H	Forecasting-head bias.
$\widehat{\mathbf{y}}_c$	H	Per-variable forecast before inverse normalization.
$\mathcal{L}_{\mathrm{MSE}}$	1	Mean squared forecasting error.
f_ϕ	–	Spectral descriptor encoder: $4K \to D_\phi$.
f_{cond}	–	Conditioning MLP: $D_\phi \to 2D + 1$.
$f_{\mathrm{tok}},\ f_{\mathrm{ps}}$	–	Token-summary and phase-summary projections.
f_{ref}	–	Residual token-mixing and feature-mixing module.
$X,\ \widehat{X},\ \eta$	1	Reference coefficient, perturbed coefficient, and perturbation.
$\Delta\phi$	1	Angular change under coefficient perturbation.
$a,\ b$	1	Real and imaginary parts of the rotated perturbation.
τ	1	Temporal shift; normalized cycle position in response analysis.
$R(\tau, \rho)$	1	Gated affine-parameter response statistic.

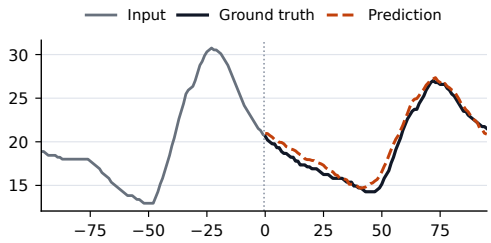


(a) ETTh1, Example 1

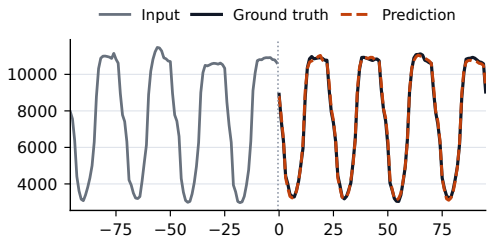
(b) ETTh1, Example 2



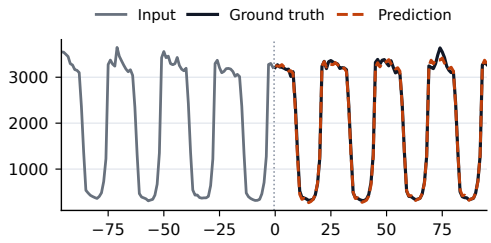
(c) ETTm2, Example 1



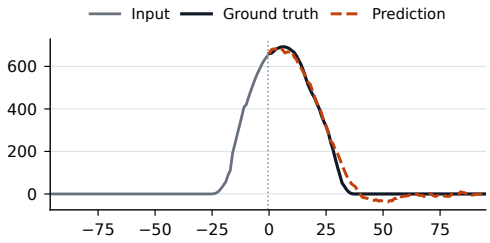
(d) ETTm2, Example 2



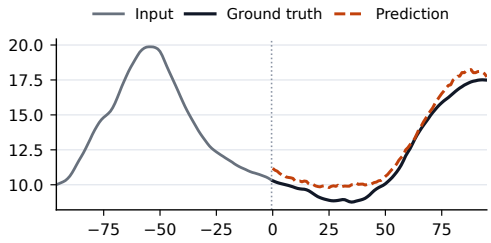
(e) Electricity, Example 1



(f) Electricity, Example 2



(g) Weather, Example 1



(h) Weather, Example 2

Figure 6: Additional forecasting examples of PaCo on representative datasets with prediction length $H = 96$. Each row corresponds to one dataset, and each column shows a different test example. The plots compare the ground-truth future sequence with the prediction generated by PaCo.