PATTERNS PERSIST, EXPRESSIONS CHANGE: LONG-HORIZON TIME-SERIES FORECASTING VIA TEMPORAL PATTERN DEFORMATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-horizon forecasting implicitly relies on temporal patterns observed in history remaining informative beyond the observed window. Under structured non-stationarity, recurring patterns may remain informative even as their future expressions change in amplitude, phase, and cross-channel temporal alignment. We term this variation **temporal pattern deformation**. Existing methods model related phenomena through statistical shifts, spectral variation, latent dynamics, or future phase evolution. We study how recurring temporal structure can be represented together with time-varying amplitude and phase over the forecast horizon. We introduce **NSTime**, which applies a recursive moving-average decomposition to form a smoothed background and a hierarchy of detail residuals. Each detail residual is represented by an amplitude–phase component and a residual component. Continuous Circular Temporal Generators (CCTG) map historical and forecast time points onto shared differentiable cyclic coordinates, with learned channel offsets representing persistent channel-level temporal shifts. Along these coordinates, the AM–PM Non-Stationary Forecasting (NSF) module evaluates learned amplitude and phase fields over the historical and forecast intervals, while a residual path captures dynamics beyond the amplitude–phase formulation. Across 11 benchmarks, NSTime ranks first in 16 of 22 dataset–metric comparisons within the evaluated comparison pool and among the top two in 19.

1 INTRODUCTION

Time-series forecasting underpins many real-world applications, including energy systems Deb et al. (2017), traffic management Liu et al. (2025b), climate monitoring, and industrial control Fatima & Rahimi (2024), among others. Long-horizon forecasting implicitly relies on temporal patterns observed in history remaining informative beyond the input window. Under structured non-stationarity, recurring patterns may persist even as their expressions change in amplitude, phase, or cross-channel alignment. The forecasting challenge is therefore not always to discover a new pattern, but to predict how a recurring pattern will be re-expressed beyond the observed window.

Existing work addresses non-stationarity from several complementary perspectives. Statistical-alignment methods normalize instance statistics or estimate distribution shifts between input and output windows Kim et al. (2022); Fan et al. (2023). Latent-dynamics methods model time-varying evolution in learned state spaces Liu et al. (2023). Temporal–frequency methods stabilize temporal variation or emphasize spectral shifts Lu et al. (2026). Preservation-aware approaches restore or retain non-stationary information for dependency modeling Liu et al. (2022b; 2025a). These methods characterize non-stationarity through changing statistics, latent dy-

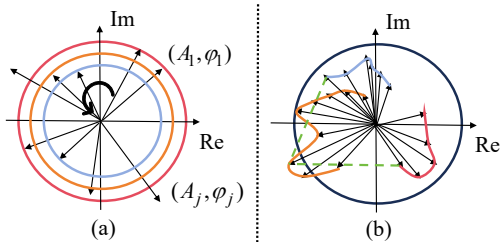


Figure 1: (a) An oscillatory pattern can be represented on a circular coordinate, with radius encoding amplitude and angular progression encoding phase. (b) Time-varying amplitude and phase change the expression of the pattern while retaining recurring cyclic organization.

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.

namics, spectral variation, or dependency structure. Recent work further models dynamical phase evolution and generates future phase trajectories Liu et al. (2026). NSTime complements these perspectives by modeling recurring temporal structure together with time-varying amplitude and phase over the forecast horizon.

From the complex-plane perspective illustrated in Fig. 1, an oscillatory component can be represented by a circular trajectory, with radius encoding amplitude and angular position encoding phase. When these quantities evolve over time, recurring cyclic organization may persist while its expression changes in amplitude, phase, or cross-channel alignment. We term this structured variation *temporal pattern deformation*, which conceptually distinguishes the recurring pattern from its time-varying expression. This distinction has a direct interpretation in real-world forecasting. In daily urban traffic, for example, the morning–evening rush-hour profile may remain recognizable across days, while the morning peak becomes higher on a rainy day, occurs later during a holiday, or appears with different delays across neighboring roads. Within this parameterization, these changes can be represented by amplitude deformation, phase deformation, and channel-specific phase variation, respectively.

Real-world series also contain slow background dynamics, variation at multiple temporal scales, and abrupt or non-cyclic changes. To model temporal pattern deformation, we apply amplitude–phase modulation (AM–PM) to the detail residuals produced by a recursive moving-average decomposition. Time-varying amplitude and phase model the future expression of components with recurring cyclic organization, while a residual neural predictor handles dynamics beyond the AM–PM prior. In this way, NSTime represents recurring temporal organization while extrapolating its time-varying future expression.

Based on this formulation, we propose **NSTime**, a hierarchical framework organized by decomposition level and modeling role. NSTime forms a smoothed background and detail residuals at successive decomposition levels, and represents each detail residual with an AM–PM component and a residual branch. Each detail branch combines an AM–PM Non-Stationary Forecasting (NSF) module with Continuous Circular Temporal Generators (CCTG), which place the historical and forecast intervals on shared differentiable cyclic coordinates with persistent channel-level offsets and fractional temporal shifts. CCTG specify the cyclic locations, NSF represents amplitude and phase variation along them, and the residual predictor captures the remaining dynamics.

Our main contributions are as follows:

- We characterize temporal pattern deformation as a forecasting regime where recurring temporal structure remains informative while its amplitude, phase, or cross-channel temporal alignment changes.
- We propose NSTime, which hierarchically models a slow background and detail residuals. NSF models time-varying amplitude and phase in each detail branch, while a residual predictor captures the remaining dynamics.
- We introduce CCTG, which map historical and forecast positions onto shared continuous cyclic coordinates through circular interpolation and persistent channel-level offsets.
- Extensive experiments on 11 multi-domain benchmarks demonstrate the effectiveness of NSTime.

2 Related Work

2.1 Decomposition for Time-Series Forecasting

Time-series forecasting has been approached with recurrent and gated models Cinar et al. (2018); Jia et al. (2024), Transformer architectures Zhang & Yan (2023); Li et al. (2019); Wu et al. (2020), and patch- or channel-centric representations such as PatchTST Nie et al. (2023), iTransformer Liu et al. (2024), and CSFormer Wang et al. (2025a). Despite these advances, dynamics at different temporal scales can remain entangled, motivating decomposition-based modeling Nematirad et al. (2025). Classical methods such as STL Ouyang et al. (2023) and MSTL Bandara et al. (2025) separate trend, seasonal components, and residual variation. Neural forecasters further incorporate progressive decomposition, exponential smoothing, frequency-domain representations, or basis expansion, includ-

ing Autoformer Wu et al. (2021), FEDformer Zhou et al. (2022b), ETSformer Woo et al. (2022), N-BEATS Oreshkin et al. (2020), and DLinear Zeng et al. (2023). TimeMixer Wang et al. (2024) mixes decomposed components across sampling scales, while TimeMixer++ Wang et al. (2025b) jointly explores temporal scales and frequency resolutions. These methods separate components, scales, or representation spaces, but typically leave how persistent patterns change beyond the observed window implicit. NSTime instead applies amplitude–phase forecasting to the detail residuals formed at successive stages of a recursive moving-average decomposition.

2.2 Forecasting under Non-Stationarity

Existing methods address non-stationarity by transforming the observed series or its representation. RevIN Kim et al. (2022) normalizes instance statistics and reverses the transformation at the output, while Dish-TS Fan et al. (2023) estimates input- and output-space distributions to alleviate distribution shift. Non-Stationary Transformers Liu et al. (2022b) combine series stationarization with de-stationary attention. FAN Ye et al. (2024) extracts dominant Fourier components and predicts their input–output discrepancy, whereas DTAF Lu et al. (2026) couples temporal stabilization with frequency differencing. Koopa Liu et al. (2023) separates time-varying and time-invariant dynamics through Koopman predictors, and TimeBridge Liu et al. (2025a) stabilizes short-term variation while preserving long-term non-stationarity for cross-variable cointegration. Recent work further models dynamical phase evolution and generates future phase trajectories Liu et al. (2026). NSTime models related non-stationary variation through amplitude and phase fields defined over recursively decomposed detail residuals.

2.3 Temporal Pattern Evolution in Forecasting

Existing work captures temporal structure through cyclic, phase-aligned, and spectral representations. CycleNet Lin et al. (2024) learns recurrent cycles, AlignTime Wang et al. (2026) aligns samples by dominant-period phase, and TQNet Lin et al. (2025) and TimesNet Wu et al. (2023) model periodic structure through shifted queries and period-specific representations. FEDformer Zhou et al. (2022b), FiLM Zhou et al. (2022a), FITS Xu et al. (2024), and FAN Ye et al. (2024) exploit spectral structure through decomposition, projection, interpolation, or dominant frequency extraction, while PULSE Liu et al. (2026) explicitly predicts future phase evolution. These methods provide complementary treatments of cyclic structure, phase variation, and spectral change. NSTime jointly models amplitude and phase variation over shared continuous coordinates for each detail residual.

3 Method

3.1 Modeling Temporal Pattern Deformation

Given a multivariate history $\mathbf{X} \in \mathbb{R}^{S\times C}$, we aim to forecast the future sequence $\widehat{\mathbf{Y}} \in \mathbb{R}^{P\times C}$. Motivated by the geometric amplitude–phase view in Fig. 1, we characterize temporal pattern deformation as a change in how a recurring temporal pattern is expressed rather than its disappearance. For a cyclic component, this distinction gives rise to two limiting regimes. The stable regime has time-invariant amplitude and phase. In the evolving regime, the same underlying oscillation is re-expressed through time-varying amplitude and phase, which can be formulated as:

$$\begin{aligned} x_c^{\text{stable}}(t) &= A_c \cos(\omega t + \phi_c) + \epsilon_c(t), \\ x_c^{\text{evolving}}(t) &= A_c(t) \cos\big(\omega t + \phi_c(t)\big) + \epsilon_c(t), \end{aligned} \tag{1}$$

where ω defines the reusable temporal structure of this component, while $A_c(t)$ and $\phi_c(t)$ determine its channel-specific expression over time. Because $\phi_c(t)$ is channel specific, relative phase differences across channels also encode changes in cross-channel alignment. The stable regime is recovered when $A_c(t) \equiv A_c$ and $\phi_c(t) \equiv \phi_c$, and $\epsilon_c(t)$ captures dynamics outside the amplitude–phase representation.

Time-varying modulation alone does not determine stationarity. Under the random-phase construction in Appendix B, constant amplitude with an affine phase displacement is a stationary special case, whereas modulation that induces time-dependent second-order statistics corresponds to a structured non-stationary regime.

Eq. 1 separates cyclic organization from its context-dependent expression for a single cyclic component. To accommodate real-world series with slow background evolution, cyclic variation at multiple decomposition levels, and non-cyclic dynamics, we extend this view as:

$$x_c(t) = g_c^{(B)}(t) + \sum_{j=1}^{J} A_{j,c}(t)\cos\big(\omega_j t + \phi_{j,c}(t)\big) + \epsilon_c(t), \tag{2}$$

where $g_c^{(B)}$ denotes the slowly varying background, each summand models cyclic structure associated with decomposition level j and its time-varying expression, and $\epsilon_c(t)$ absorbs residual dynamics such as abrupt or non-cyclic changes. The learned amplitude and phase are forecasting variables rather than uniquely identified physical quantities. Their ambiguity is constrained through the decomposition hierarchy, shared cyclic coordinates, and smooth generator fields.

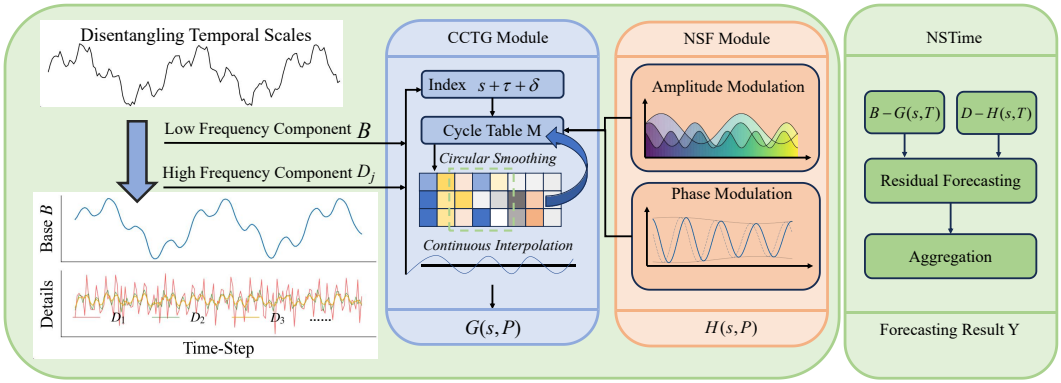


Figure 2: Overview of NSTime. The input is decomposed into a smoothed background and detail residuals at successive decomposition levels. Continuous Circular Temporal Generators establish shared differentiable cyclic coordinates, and the AM–PM forecasting module evaluates learned amplitude and phase fields for each detail residual.

NSTime Framework. Fig. 2 realizes this prior through three coupled components. Recursive moving-average decomposition forms a smoothed background and a hierarchy of detail residuals. Continuous Circular Temporal Generators (CCTG) place history and forecast on shared continuous cyclic coordinates, and the AM–PM Non-Stationary Forecasting (NSF) module represents amplitude and phase variation along them. A lightweight residual predictor captures variation outside the structured path, preventing the cyclic prior from being imposed where it is uninformative. The decomposition, CCTG interpolation, AM–PM component, and gated residual reconstruction correspond to Eqs. 3–5, Eqs. 6–10, Eq. 14, and Eqs. 19–20, respectively.

3.2 Recursive Decomposition

Starting from $\mathbf{T}_0 = \mathbf{X}$, NSTime recursively applies a channel-wise moving average:

$$\begin{aligned} \mathbf{T}_j &= \mathrm{MA}_k(\mathbf{T}_{j-1}), \\ \mathbf{D}_j &= \mathbf{T}_{j-1} - \mathbf{T}_j, \qquad j = 1, \ldots, J, \end{aligned} \tag{3}$$

where $\mathrm{MA}_k(\cdot)$ denotes a channel-wise moving average with window size k Wu et al. (2021). Repeated smoothing produces detail residuals at successive decomposition levels. The final smoothed component $\mathbf{T}_J$ forms the background component:

$$\mathbf{B} = \mathbf{T}_J. \tag{4}$$

Each $\mathbf{D}_j$ contains the variation removed at decomposition stage j. Because every detail component is defined residually, the decomposition is exact:

$$\mathbf{X} = \mathbf{B} + \sum_{j=1}^{J} \mathbf{D}_j. \tag{5}$$

This additive decomposition is exact, but it does not guarantee that the detail residuals are orthogonal or occupy disjoint frequency bands. At each decomposition level, a channel-wise gate controls the contribution of the AM–PM path, while a residual predictor captures the remaining dynamics.

3.3 Continuous Circular Temporal Generators

For a sequence window starting at temporal coordinate s, CCTG maps each timestamp and channel to a continuous circular coordinate. At decomposition level j, for each field $q \in \{A, \phi\}$, we define:

$$\mathbf{U}_j^{(q)}(s,T)_{\tau,c} = s + \tau + \delta_{j,c}^{(q)}, \quad \tau = 0, \dots, T-1, \tag{6}$$

where T is the queried interval length and $\delta_{j,c}^{(q)} \in \mathbb{R}$ is a learnable channel-specific offset. The amplitude and phase generators use independently learned offsets, allowing both fields to be evaluated on their respective channel-aligned coordinates. The offsets are fixed after training and represent persistent channel-level coordinate shifts. Time-varying relative alignment is represented within the model by channel-specific phase fields evaluated along these coordinates. For each field type $q \in \{A, \phi\}$, CCTG maintains a learnable table defined over a length-L circular temporal domain:

$$\widetilde{\mathbf{M}}_j^{(q)} \in \mathbb{R}^{L \times C}, \quad q \in \{A, \phi\}. \tag{7}$$

Its temporal indices are interpreted modulo L, making positions k and $k+L$ equivalent and treating the two ends of the table as adjacent. The table therefore represents a closed temporal coordinate rather than a finite linear sequence. Given $u = \mathbf{U}_j^{(q)}(s,T)_{\tau,c}$, we wrap it onto $[0, L)$:

$$\bar{u} = u - L \left\lfloor \frac{u}{L} \right\rfloor. \tag{8}$$

Let

$$k_0 = \lfloor \bar{u} \rfloor, \quad k_1 = (k_0 + 1) \bmod L, \quad \alpha = \bar{u} - k_0. \tag{9}$$

The field value is obtained by circular linear interpolation:

$$G_j^{(q)}(s,T)_{\tau,c} = (1-\alpha)\widetilde{\mathbf{M}}_{j,k_0,c}^{(q)} + \alpha \widetilde{\mathbf{M}}_{j,k_1,c}^{(q)}. \tag{10}$$

CCTG coincides with discrete circular lookup at integer coordinates, while providing continuous, piecewise-differentiable access at fractional coordinates. It therefore represents fractional temporal shifts without abrupt index transitions. The channel-specific offsets represent persistent temporal shifts across variables. CCTG does not increase table capacity. It changes the access geometry from discrete lookup to continuous, channel-aligned retrieval. The resulting field is L-periodic rather than an arbitrary nonperiodic evolution. Residual branches handle departures from this cyclic prior. Factorization, smoothing, and the exact periodicity argument are provided in Appendix B.

3.4 AM–PM Non-Stationary Forecasting

At decomposition level j, NSF evaluates the amplitude field as:

$$\mathbf{A}_j(s,T) = \mathrm{Softplus}\Big(G_j^{(A)}(s,T)\Big), \tag{11}$$

which provides a smooth non-negative amplitude parameterization. Softplus is used for smooth gradients. The non-negativity convention avoids a redundant sign–phase representation, since a negative amplitude can be written as a positive amplitude with a π phase shift. ReLU provides an alternative non-negative parameterization but is non-differentiable at zero. The corresponding total phase is defined element-wise as:

$$\mathbf{\Phi}_j(s,T)_{\tau,c} = \omega_j(s+\tau) + G_j^{(\phi)}(s,T)_{\tau,c}, \tag{12}$$

$\tau = 0, \dots, T-1$, where

$$\omega_j = \frac{2\pi m_j}{L}, \qquad m_j \in \mathbb{N}, \tag{13}$$

is the carrier frequency associated with decomposition level j. The integer harmonic m_j makes the carrier compatible with the length-L circular coordinate. Here, $G_j^{(\phi)}$ represents an unwrapped phase-displacement field. Periodicity is imposed by the cosine operator rather than by explicitly wrapping the displacement to $[-\pi, \pi]$.

Table 1: **Comparison results.** Average forecasting performance over four horizons. Bold/underlined marks best/second-best. Per-horizon results are reported in Appendix C.4, Table 6.

	NSTime$_{\text{Ours}}$		DTAF$_{2026}$		TimeMosaic$_{2026}$		TimeBridge$_{2025}$		RAFT$_{2025}$		S-Mamba$_{2024}$		CycleNet$_{2024}$		iTransformer$_{2024}$		Koopa$_{2023}$		DLinear$_{2023}$		Crossformer$_{2023}$	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.419**	0.426	0.431	**0.422**	0.425	0.424	0.439	0.434	0.428	0.433	0.455	0.450	0.445	0.446	0.454	0.448	0.454	0.446	0.446	0.443	0.529	0.522
ETTh2	0.371	0.400	0.374	0.393	**0.363**	**0.387**	0.375	0.398	0.382	0.410	0.381	0.405	0.386	0.407	0.383	0.407	0.389	0.412	0.523	0.494	0.942	0.684
ETTm1	**0.366**	0.392	0.385	0.383	0.381	**0.381**	0.384	0.391	0.381	0.400	0.398	0.405	0.379	0.396	0.407	0.410	0.394	0.405	0.403	0.407	0.513	0.495
ETTm2	**0.267**	**0.316**	0.279	0.320	0.273	**0.316**	0.279	0.319	0.281	0.330	0.288	0.332	0.269	0.319	0.288	0.332	0.282	0.327	0.350	0.401	0.757	0.611
Electricity	**0.167**	**0.260**	0.195	0.278	0.187	0.279	0.175	0.265	0.175	0.272	0.170	0.265	0.169	0.262	0.178	0.270	0.190	0.284	0.212	0.300	0.244	0.334
Solar	**0.200**	0.254	0.271	0.277	0.240	0.270	0.233	**0.244**	0.301	0.303	0.240	0.273	0.210	0.261	0.233	0.262	0.255	0.287	0.330	0.401	0.641	0.639
Weather	**0.241**	0.270	0.256	0.272	0.251	**0.267**	0.254	0.271	0.270	0.309	0.251	0.276	0.243	0.271	0.258	0.278	0.245	0.272	0.265	0.317	0.259	0.315
PEMS03	**0.104**	**0.212**	0.224	0.308	0.175	0.275	0.167	0.267	0.144	0.230	0.122	0.228	0.125	0.233	0.113	0.222	0.180	0.278	0.262	0.357	0.169	0.282
PEMS04	**0.098**	**0.206**	0.297	0.355	0.215	0.304	0.185	0.287	0.104	0.210	0.103	0.211	0.128	0.242	0.111	0.221	0.167	0.275	0.261	0.351	0.209	0.314
PEMS07	**0.088**	**0.188**	0.238	0.308	0.169	0.266	0.181	0.271	0.094	0.193	0.089	**0.188**	0.113	0.214	0.101	0.204	0.143	0.250	0.329	0.396	0.235	0.315
PEMS08	**0.126**	**0.216**	0.299	0.342	0.220	0.297	0.226	0.299	0.151	0.234	0.148	0.224	0.150	0.246	0.150	0.226	0.192	0.292	0.379	0.416	0.268	0.307

The AM–PM realization at decomposition level j is:

$$\mathbf{H}_j(s,T) = \mathbf{A}_j(s,T) \odot \cos(\mathbf{\Phi}_j(s,T)), \tag{14}$$

where $\odot$ denotes element-wise multiplication. The historical and future components are evaluations of the same cyclic field over consecutive coordinate intervals:

$$\mathbf{H}_j^{\text{in}} = \mathbf{H}_j(s,S), \;\; \mathbf{H}_j^{\text{out}} = \mathbf{H}_j((s+S) \bmod L, P). \tag{15}$$

Thus, the forecast interval continues from the coordinate immediately following the observed interval rather than restarting from a local phase origin. CCTG specifies where each timestamp lies on a continuous, channel-aligned cycle, whereas NSF specifies how the pattern's amplitude and phase vary along that cycle.

3.5 Background and Residual Forecasting

The background branch follows the same history–forecast coordinate alignment. Let $G^{(B)}$ denote a background CCTG with its own learnable circular table and channel offsets. Its historical and future templates are:

$$\mathbf{C}_B^{\text{in}} = G^{(B)}(s,S), \;\; \mathbf{C}_B^{\text{out}} = G^{(B)}((s+S) \bmod L, P). \tag{16}$$

NSTime removes the historical template, forecasts its deviation, and restores the corresponding future template:

$$\widehat{\mathbf{B}} = F_B\Big[\big(\mathbf{B} - \mathbf{C}_B^{\text{in}}\big)^\top\Big]^\top + \mathbf{C}_B^{\text{out}}, \tag{17}$$

where $F_B : \mathbb{R}^{C\times S} \to \mathbb{R}^{C\times P}$ is a lightweight forecasting head. This residualized formulation prevents the predictor from repeatedly relearning the background template at every forecast origin. At each decomposition level, a learned channel-wise gate determines the global contribution of the structured AM–PM path:

$$\boldsymbol{\lambda}_j = \sigma(\mathbf{g}_j), \quad \boldsymbol{\lambda}_j \in (0,1)^C, \tag{18}$$

where $\boldsymbol{\lambda}_j$ is broadcast over the temporal dimension. The historical variation not explained by the gated AM–PM component is:

$$\mathbf{R}_j = \mathbf{D}_j - \boldsymbol{\lambda}_j \odot \mathbf{H}_j^{\text{in}}. \tag{19}$$

A lightweight residual predictor then extrapolates this unexplained component:

$$\widehat{\mathbf{D}}_j = \boldsymbol{\lambda}_j \odot \mathbf{H}_j^{\text{out}} + \rho_j F_{\text{R},j}\big(\mathbf{R}_j^\top\big)^\top, \tag{20}$$

where ρ_j is a learned scale controlling the global contribution of the residual predictor $F_{\text{R},j} : \mathbb{R}^{C\times S} \to \mathbb{R}^{C\times P}$. The gate attenuates the AM–PM path at decomposition levels and channels for which the structured trajectory is less informative, while the residual branch captures shocks, non-cyclic changes, and other dynamics outside the AM–PM prior. The final forecast is reconstructed by adding the predicted background and all predicted detail components following Eq. 5. Additional model-specification details are provided in Appendix B.

4 EXPERIMENTS

Datasets. We evaluate NSTime on 11 widely used multivariate forecasting benchmarks, including four ETT variants Zhou et al. (2021), Weather Wu et al. (2021), Electricity and Solar Lai et al. (2018), and four PEMS datasets Liu et al. (2022a). The input length is fixed to $S = 96$ for all experiments. We consider prediction horizons $P \in \{96, 192, 336, 720\}$ for ETT, Weather, Electricity, and Solar, and $P \in \{12, 24, 48, 96\}$ for PEMS.

Baselines. We compare NSTime with Koopa Liu et al. (2023), TimeBridge Liu et al. (2025a), DTAF Lu et al. (2026), TimeMosaic Ding et al. (2026), RAFT Han et al. (2025), S-Mamba Wang et al. (2025c), CycleNet Lin et al. (2024), iTransformer Liu et al. (2024), DLinear Zeng et al. (2023), and Crossformer Zhang & Yan (2023). Additional details on baseline selection are provided in Appendix C.2.

Additional Experimental Details. We present the main analyses in the paper, while complete results, few-shot results, experimental details, extended ablations, sensitivity, and efficiency analyses are provided in Appendix C–H.

Table 2: **Core ablation results.** Core variants at standard and longest horizons. The small subscript beside each ablation result reports its signed difference from Full, calculated as $\Delta = \text{Ablation} - \text{Full}$. Positive values indicate increased error, while negative values indicate decreased error. All ETTm2–720 variants use an identical predictor and differ only in the specified component.

		Full		w/o NSF		w/o Amp		w/o Phase		w/o CCTG		w/o Decomposition	
	Horizon	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Solar	96	**0.178**	**0.247**	$0.184_{+0.006}$	$0.249_{+0.002}$	$0.192_{+0.014}$	$0.246_{-0.001}$	$0.193_{+0.015}$	$0.245_{-0.002}$	$0.180_{+0.002}$	$0.250_{+0.003}$	$0.195_{+0.017}$	$0.248_{+0.001}$
	720	**0.214**	**0.257**	$0.222_{+0.008}$	$0.262_{+0.005}$	$0.239_{+0.025}$	$0.273_{+0.016}$	$0.216_{0.002}$	$0.258_{+0.001}$	$0.233_{+0.019}$	$0.272_{+0.015}$	$0.248_{+0.034}$	$0.279_{+0.022}$
ETTm2	96	**0.165**	**0.249**	**0.165**$_{0.000}$	$0.250_{+0.001}$	**0.165**$_{0.000}$	**0.249**$_{0.000}$	$0.166_{+0.001}$	$0.251_{+0.002}$	$0.167_{+0.002}$	$0.253_{+0.004}$	$0.167_{+0.002}$	$0.250_{+0.001}$
	720	**0.385**	**0.390**	$0.393_{+0.008}$	$0.392_{+0.002}$	$0.399_{+0.014}$	$0.395_{+0.005}$	$0.397_{+0.012}$	$0.396_{+0.006}$	$0.399_{+0.014}$	$0.396_{+0.006}$	$0.397_{+0.012}$	$0.394_{+0.004}$

4.1 COMPARATIVE RESULTS

Table 1 compares average performance across domains, resolutions, and forecast horizons within the evaluated comparison pool. NSTime is best in 16 of 22 dataset–metric comparisons, second-best in three, and attains the best MSE on 10 of 11 datasets (Table 1). It ranks in the top two on both metrics for eight datasets and is best on both for ETTm2, Electricity, PEMS03, PEMS04, PEMS07, and PEMS08 (including ties on ETTm2 and PEMS07). The gains extend beyond the ETT family: relative to the strongest competing MSE, NSTime reduces average error by approximately 8.0%, 4.9%, and 14.9% on PEMS03, PEMS04, and PEMS08, respectively, and by 4.8% on Solar. The result is not uniformly dominant: TimeMosaic is stronger on ETTh2, and NSTime is not top-two in MAE on ETTh1, ETTh2, or ETTm1. On PEMS, lower MSE is consistently accompanied by lower MAE, whereas Solar exhibits an MSE–MAE trade-off. These results show broad forecasting utility, but aggregate rankings alone cannot validate the proposed mechanism. We therefore next test the model components and the deformation regimes in which they matter.

4.2 ABLATION STUDY

Core Ablation. We ablate NSF, amplitude, phase, CCTG, and decomposition at $H = 96$ and $H = 720$ under matched forecasting capacity. Table 2 shows that the full model is most clearly favored at long horizons. On Solar at $H = 720$, removing amplitude, CCTG, or decomposition increases MSE by 11.7%, 8.9%, and 15.9%, respectively. On ETTm2, the variants are nearly tied at $H = 96$, whereas Full achieves the best result at $H = 720$, improving MSE by about 2.0%–3.5% over all ablations. These matched results quantify the effect of removing each tested component from the jointly trained model. The w/o-NSF variant retains the residual forecasting path while removing the structured AM–PM path, so its comparison with Full measures the incremental contribution of the structured path in the joint model. Additional ablations are provided in Appendix D, including the extended results in Table 8. The more pronounced performance separation at longer horizons further motivates us to examine whether these gains are associated with spectral drift that may intensify as the forecast horizon increases.

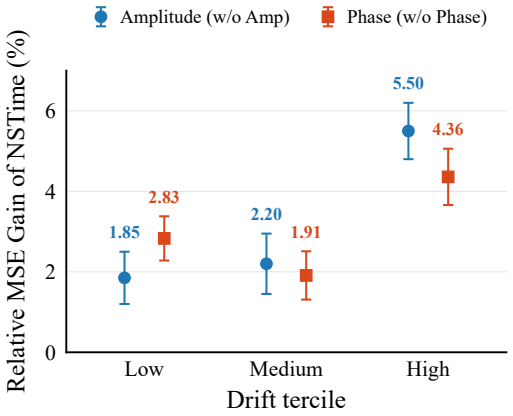


Figure 3: **Ablation on amplitude–phase drift.** NSF Gains across Spectral-Drift Regimes on ETTm2–720.

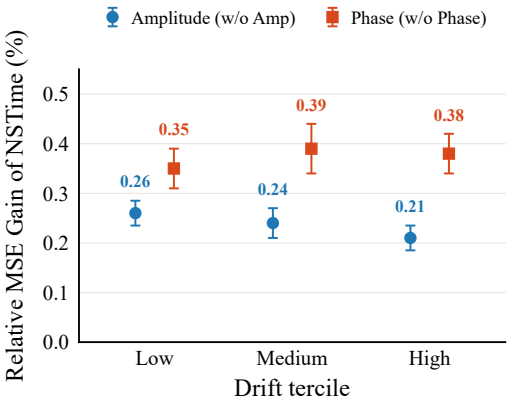


Figure 4: **Ablation on amplitude–phase drift.** NSF Gains across Spectral-Drift Regimes on Electricity–720.

Table 3: **CCTG module analysis.** Controlled RCF-to-CCTG replacement in CycleNet. The replacement jointly introduces circular interpolation and channel-specific offsets (best in bold). The complete per-horizon analysis is reported in Appendix F, Table 11.

Dataset	ETTh1				ETTh2				Dataset	PEMS03				PEMS04			
Methods	RCF		CCTG		RCF		CCTG		Methods	RCF		CCTG		RCF		CCTG	
Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	0.383	**0.393**	**0.380**	0.395	0.289	0.338	**0.288**	**0.337**	12	0.069	**0.176**	**0.068**	**0.176**	0.081	0.190	**0.077**	**0.184**
192	0.432	**0.421**	**0.428**	0.422	**0.374**	**0.393**	**0.374**	**0.393**	24	0.094	0.207	**0.092**	**0.204**	0.105	0.220	**0.099**	**0.213**
336	0.474	0.443	**0.464**	**0.439**	0.436	0.441	**0.424**	**0.435**	48	0.144	0.259	**0.141**	**0.255**	0.144	0.261	**0.137**	**0.255**
720	0.472	0.466	**0.470**	**0.460**	**0.457**	0.462	**0.457**	**0.461**	96	0.192	0.291	**0.190**	**0.290**	0.183	0.296	**0.171**	**0.287**
Avg	0.440	0.431	**0.436**	**0.429**	0.389	0.409	**0.386**	**0.407**	Avg	0.125	0.233	**0.123**	**0.231**	0.128	0.242	**0.121**	**0.235**

Performance under Amplitude–Phase Drift. We focus on ETTm2–720 and use amplitude and phase drift as two complementary measures of spectral drift. The statistics are computed post hoc from raw history–future pairs and are independent of NSTime's learned amplitude and phase fields. According to drift magnitude, we partition the forecasting windows into terciles and compare Full with the configuration-matched w/o-Amplitude and w/o-Phase variants. As shown in Fig. 3, Full has lower MSE than the corresponding ablation in all six groups, with the largest gaps occurring in the high-drift terciles, reaching 5.50% for amplitude drift and 4.36% for phase drift. Fig. 4 reports the same matched comparison on Electricity–720, where positive gaps are observed across all drift regimes. Across the evaluated bins, Full consistently outperforms the corresponding amplitude- or phase-removed variant. Detailed thresholds, bin-wise errors, and confidence intervals are reported in Appendix E, Table 10, with the Electricity replication shown in Fig. 6.

CCTG Module Analysis. To isolate the contribution of CCTG from the rest of NSTime, we replace CycleNet's discrete Residual Cycle Forecasting (RCF) lookup with CCTG while holding the backbone, data splits, training protocol, hyperparameters, and cycle-table capacity fixed. The replacement changes only temporal access: CCTG adds channel-specific offsets and circular interpolation, and reduces to RCF-style lookup at integer coordinates. CCTG improves MSE in 14 of 16 dataset–horizon settings and ties in two (Table 3). It improves MAE in 12 settings, ties in two, and is slightly worse in two. Improvements on both ETT and PEMS indicate that the combined continuous, channel-aligned access mechanism is useful beyond one horizon regime and the NSTime backbone. Because circular interpolation and channel-specific offsets are introduced together, this comparison measures their combined effect.

4.3 Relationship to Generic Statistical Non-Stationarity

The preceding analyses directly probe the proposed mechanism through component ablations, deformation-based stratification, and controlled replacement of cyclic indexing. Temporal pattern

Table 4: NSF gains on subsets diagnosed by each statistical non-stationarity criterion and on their complements. D and C denote the diagnosed subset and its complement, respectively. ADF∩KPSS requires ADF > 0.05 and KPSS ≤ 0.05 simultaneously.

ADF > 0.05				KPSS ≤ 0.05				ADF∩KPSS			
Dataset	Diagnosed Ratio	Gain (D)	Gain (C)	Dataset	Diagnosed Ratio	Gain (D)	Gain (C)	Dataset	Diagnosed Ratio	Gain (D)	Gain (C)
ETTh1	17.83%	1.40%	4.72%	ETTh1	15.98%	0.64%	4.79%	ETTh1	9.71%	0.12%	4.56%
ETTh2	13.97%	2.62%	0.98%	ETTh2	15.89%	2.96%	0.73%	ETTh2	10.34%	3.87%	0.80%
ETTm1	87.97%	3.37%	4.29%	ETTm1	58.56%	3.43%	3.54%	ETTm1	54.10%	3.32%	3.67%
ETTm2	64.10%	3.96%	2.13%	ETTm2	51.49%	4.61%	1.91%	ETTm2	38.37%	5.09%	2.18%
Weather	80.37%	3.70%	3.13%	Weather	71.99%	4.22%	1.97%	Weather	60.66%	4.38%	2.32%

deformation describes a regime in which recurring temporal organization is re-expressed through amplitude, phase, or relative cross-channel timing. Because ADF and KPSS characterize broader statistical properties rather than this specific structure, we next examine NSF gains across the subsets defined by these diagnostics. For each input window and variable, we apply the ADF and KPSS tests Dickey & Fuller (1979); Kwiatkowski et al. (1992). ADF > 0.05 indicates insufficient evidence to reject the unit-root null, while KPSS ≤ 0.05 rejects the null of stationarity. We also report their intersection, denoted by ADF∩KPSS, which requires ADF > 0.05 and KPSS ≤ 0.05 simultaneously. These tests characterize generic statistical non-stationarity rather than temporal pattern deformation itself. We therefore use them as diagnostic indicators rather than ground-truth labels of temporal pattern deformation.

Let $\mathrm{MSE}_{\mathrm{base}}^{(i)}$ and $\mathrm{MSE}_{\mathrm{NSF}}^{(i)}$ denote the forecasting errors of the model without and with NSF on sample i, respectively. For a diagnosed subset $\mathcal{S}$, we define the relative gain as:

$$\mathrm{Gain}(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \frac{\mathrm{MSE}_{\mathrm{base}}^{(i)} - \mathrm{MSE}_{\mathrm{NSF}}^{(i)}}{\mathrm{MSE}_{\mathrm{base}}^{(i)}} \times 100\%. \tag{21}$$

A positive value indicates that NSF reduces the forecasting error on the corresponding subset.

Results and Discussion. We evaluate the four ETT variants and Weather because they span distinct domains and temporal resolutions while supporting a uniform window-level stationarity analysis. Temporal pattern deformation describes changes in the expression of recurring temporal organization through amplitude, phase, or relative cross-channel timing. ADF and KPSS instead capture broader statistical properties of the input windows. Table 4 reports positive NSF gains on every diagnosed subset and on every complement, while their relative ordering varies across datasets. Thus, the selected ADF/KPSS criteria do not consistently determine where NSF is more beneficial.

Variable-Level Relationship. To assess whether the subset-level findings hold across variables and forecasting settings, we aggregate the window-level diagnostics into variable–setting pairs and compare them with the corresponding NSF gains in Fig. 7. Positive gains span broad ADF and KPSS ranges, but no consistent monotonic relation emerges. The selected diagnostics therefore do not consistently order NSF gains across variables and settings, and they do not identify temporal pattern deformation. Complete results are provided in Appendix F, Table 12, and Fig. 7.

5 Conclusion

NSTime casts long-horizon forecasting as temporal pattern deformation, where recurring temporal organization remains informative even as its amplitude, phase, and cross-channel temporal alignment evolve. It organizes the input into a smoothed background and detail residuals, maps historical and forecast positions onto shared continuous cyclic coordinates, and models each detail residual with learned amplitude–phase fields and complementary residual dynamics. Together, these components provide a structured approach to extrapolating recurring temporal organization while allowing its future expression to change. Extensive experiments on 11 multi-domain benchmarks demonstrate the effectiveness of NSTime.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation and to provide feedback on the clarity and consistency of methodological and experimental descriptions. We also used generative AI tools for language editing, improving readability, and checking consistency in the presentation of the paper. Generative AI tools were not used to generate experimental data or to independently determine the experimental results reported in this work. All AI-assisted text, code modifications, and technical suggestions were manually reviewed by the authors. The authors verified the relevant content against the actual code, experimental results, equations, citations, and claims presented in the paper.

ETHICS STATEMENT

This work studies general-purpose time-series forecasting using established public benchmark datasets. It does not involve human subjects, personally identifiable information, or the collection of new sensitive data. The proposed method is intended for general forecasting research. As with other forecasting systems, predictions may be imperfect and should not be used as the sole basis for safety-critical or high-stakes decisions without appropriate domain-specific validation and human oversight. We follow the ICLR Code of Ethics and the usage requirements of the datasets and baselines considered in this work.

REPRODUCIBILITY STATEMENT

We provide detailed information to facilitate reproduction of our results. The model formulation and implementation-relevant specifications are described in Section 3 and Appendix B. The datasets, forecasting settings, evaluation metrics, training protocol, and additional experimental details are documented in Section 4 and the corresponding experimental appendices. Complete per-horizon results, few-shot evaluations, extended ablations, analyses of temporal pattern deformation and statistical non-stationarity, sensitivity studies, and efficiency results are reported in Appendices C–H. These materials document the settings used for the reported experiments and provide the information necessary to reproduce the main empirical findings.

REFERENCES

Kasun Bandara, Rob J Hyndman, and Christoph Bergmeir. Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *International Journal of Operational Research*, 52(1):79–98, 2025.

Yagmur Gizem Cinar, Hamid Mirisaee, Parantapa Goswami, Eric Gaussier, and Ali Aït-Bachir. Period-aware content attention rnns for time series forecasting with missing values. *Neurocomputing*, 312:177–186, 2018.

Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 1979.

Kuiye Ding, Fanda Fan, Chunyi Hou, Zheya Wang, Lei Wang, Zhengxin Yang, and Jianfeng Zhan. TimeMosaic: Temporal heterogeneity guided time series forecasting via adaptive granularity patch and segment-wise decoding. In *AAAI Conference on Artificial Intelligence*, pp. 20790–20798, 2026.

Wei Fan, Pengyang Wang, Dongkun Wang, Dongjie Wang, Yuanchun Zhou, and Yanjie Fu. Dish-ts: A general paradigm for alleviating distribution shift in time series forecasting. In *AAAI Conference on Artificial Intelligence*, pp. 7522–7529, 2023.

Syeda Sitara Wishal Fatima and Afshin Rahimi. A review of time-series forecasting algorithms for industrial manufacturing systems. *Machines*, 12(6):380, 2024.

Sungwon Han, Seungeon Lee, Meeyoung Cha, Sercan Ö. Arik, and Jinsung Yoon. Retrieval augmented time series forecasting. In *International Conference on Machine Learning*, 2025.

Yuxin Jia, Youfang Lin, Jing Yu, Shuo Wang, Tianhao Liu, and Huaiyu Wan. Pgn: The rnn's new successor is effective for long-range time series forecasting. *Advances in Neural Information Processing Systems*, pp. 84139–84168, 2024.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022.

Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1–3):159–178, 1992.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *international ACM SIGIR conference on research & development in information retrieval*, pp. 95–104, 2018.

P. Langley. Crafting papers on machine learning. In Pat Langley (ed.), *International Conference on Machine Learning*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 2019.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37:106315–106345, 2024.

Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. In *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. *Advances in Neural Information Processing Systems*, 35:5816–5828, 2022a.

Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-Tao Xia. Timebridge: Non-stationarity matters for long-term time series forecasting. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *International Conference on Machine Learning*, 2025a.

Xin Liu, Lanqi Qin, Meng Xu, Yicheng Zhou, Bo Wang, Weiren Yu, and Wenxin Xiong. A comprehensive review of traffic flow forecasting based on deep learning. *Neurocomputing*, pp. 132269, 2025b.

Yangyou Liu, Zezhi Shao, Xinyu Chen, Hu Chen, Fei Wang, and Yuankai Wu. Pulse: Generative phase evolution for non-stationary time series forecasting. *arXiv preprint arXiv:2605.16793*, 2026.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in Neural Information Processing Systems*, pp. 9881–9893, 2022b.

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. *Advances in Neural Information Processing Systems*, 36:12271–12290, 2023.

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

Junkai Lu, Peng Chen, Chenjuan Guo, Yang Shu, Meng Wang, and Bin Yang. Towards non-stationary time series forecasting with temporal stabilization and frequency differencing. In Sven Koenig, Chad Jenkins, and Matthew E. Taylor (eds.), *AAAI Conference on Artificial Intelligence*, pp. 24070–24078, 2026.

Reza Nematirad, Anil Pahwa, and Balasubramaniam Natarajan. Times2d: Multi-period decomposition and derivative mapping for general time series forecasting. In *AAAI Conference on Artificial Intelligence*, volume 39, pp. 19651–19658, 2025.

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*, 2023.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*, 2020.

Zuokun Ouyang, Meryem Jabloun, and Philippe Ravier. Stlformer: Exploit STL decomposition and rank correlation for time series forecasting. In *European Signal Processing Conference*, pp. 1405–1409, 2023.

Haoxin Wang, Yipeng Mo, Kunlan Xiang, Nan Yin, Honghe Dai, Bixiong Li, and Songhai Fan. Csformer: Combining channel independence and mixing for robust multivariate time series forecasting. In *AAAI Conference on Artificial Intelligence*, pp. 21090–21098, 2025a.

Min Wang, Hua Wang, and Fan Zhang. Aligntime: Interperiodic phase alignment sampling for time-series forecasting. *Information Processing & Management*, 63(1):104296, 2026.

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*, 2024.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. Timemixer++: A general time series pattern machine for universal predictive analysis. In *International Conference on Learning Representations*, volume 2025, pp. 16980–17016, 2025b.

Zihan Wang, Fanheng Kong, Shi Feng, Ming Wang, Xiaocui Yang, Han Zhao, Daling Wang, and Yifei Zhang. Is mamba effective for time series forecasting? *Neurocomputing*, 619:129178, 2025c.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting. *arXiv preprint arXiv:2202.01381*, 2022.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, pp. 22419–22430, 2021.

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

Sifan Wu, Xi Xiao, Qianggang Ding, Peilin Zhao, Ying Wei, and Junzhou Huang. Adversarial sparse transformer for time series forecasting. *Advances in Neural Information Processing Systems*, pp. 17105–17115, 2020.

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

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. *Advances in Neural Information Processing Systems*, 37: 31350–31379, 2024.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *AAAI Conference on Artificial Intelligence*, pp. 11121–11128, 2023.

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

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 *AAAI conference on artificial intelligence*, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Xue Wang, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, and Rong Jin. FiLM: Frequency improved legendre memory model for long-term time series forecasting. In *Advances in Neural Information Processing Systems*, volume 35, pp. 12677–12690, 2022a.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, volume 162, pp. 27268–27286, 2022b.

A Appendix

Appendix B extends *Modeling Temporal Pattern Deformation* and records the complete NSTime computation. Appendix C expands the experimental protocol and per-horizon results. Appendix D extends the *Ablation Study*, and Appendix E details *Performance under Amplitude–Phase Drift*. Appendix F expands the *CCTG Module Analysis* and *Relationship to Generic Statistical Non-Stationarity*. Appendix G and Appendix H report robustness, efficiency, stability, and qualitative evidence. Throughout, temporal pattern deformation denotes a forecasting regime in which recurring organization is re-expressed through amplitude, phase, or relative cross-channel timing. When this variation induces time-dependent second-order statistics, it forms a structured non-stationary regime. The appendix analyses expose horizon-level exceptions, within-dataset reduced-supervision results, component- and parameterization-level interventions, backbone-augmentation reversals, run-to-run dispersion, and the distinction between post-hoc amplitude–phase drift diagnostics and generic statistical non-stationarity.

B Additional Details for Modeling Temporal Pattern Deformation

We first formalize the second-order motivation behind the main-paper amplitude–phase view, then specify how NSTime realizes that view. The argument motivates a forecasting prior. It neither identifies a unique physical decomposition nor claims that every non-stationary process is an AM–PM process.

B.1 Second-Order Motivation and Exact WSS Criterion

A real-valued process $\{h(t)\}_{t\in\mathbb{Z}}$ is wide-sense stationary (WSS) when its mean is constant and its autocovariance depends only on lag. Consider

$$h(t) = A(t)\cos\big(\omega t + \widetilde{\phi}(t) + \phi_0\big), \tag{22}$$

where $A(t) \geq 0$ and $\widetilde{\phi}(t)$ are deterministic and finite, and $\phi_0 \sim \mathrm{Unif}[0, 2\pi)$ is one random phase shared across all t. This global phase gives zero mean while retaining modulation in the covariance. All expectations below are taken with respect to ϕ_0. Finiteness of $A(t)$ guarantees finite second moments.

Because $\mathbb{E}_{\phi_0}[\cos(\alpha + \phi_0)] = 0$, $\mathbb{E}[h(t)] = 0$. For deterministic u and v, the product-to-sum identity gives

$$\mathbb{E}_{\phi_0}[\cos(u + \phi_0)\cos(v + \phi_0)] = \tfrac{1}{2}\cos(u - v). \tag{23}$$

Substituting the phases at t and $t+\tau$ yields

$$\mathrm{Cov}\big(h(t), h(t+\tau)\big) = \frac{1}{2}A(t)A(t+\tau)\cos\Big(\omega\tau + \Delta_\tau\widetilde{\phi}(t)\Big), \tag{24}$$

where $\Delta_\tau\widetilde{\phi}(t) = \widetilde{\phi}(t+\tau) - \widetilde{\phi}(t)$. Thus, this process is WSS exactly when the right-hand side of Eq. 24 is independent of t for every τ.

B.2 Stationary Special Case and Departures

Proposition 1. *If $A(t) \equiv A$ and $\widetilde{\phi}(t) = \beta t + \phi_c$ for constants $A \geq 0$, β, and ϕ_c, then the process in Eq. 22 is WSS.*

Proof. In this case $\Delta_\tau\widetilde{\phi}(t) = \beta\tau$, and hence

$$\mathrm{Cov}\big(h(t), h(t+\tau)\big) = \tfrac{1}{2}A^2\cos\big((\omega+\beta)\tau\big),$$

which depends only on τ. □

At zero lag,

$$\mathrm{Var}[h(t)] = \tfrac{1}{2}A(t)^2. \tag{25}$$

Thus, any nonconstant amplitude makes the variance time dependent. With constant amplitude, non-affine phase generally introduces time dependence through $\Delta_\tau\widetilde{\phi}(t)$. Special cosine symmetries remain possible, so Eq. 24 is the exact criterion. The proposition gives a sufficient stationary case, not a necessary characterization of all WSS realizations.

B.3 Connection to AM–PM Non-Stationary Forecasting

At decomposition level j, NSF constructs

$$\begin{aligned}[\mathbf{H}_j(\mathbf{s}, T)]_{b,\tau,c} = [\mathbf{A}_j(\mathbf{s}, T)]_{b,\tau,c} \\ \cdot \cos\Big(\omega_j(s_b+\tau) + [\boldsymbol{\phi}_j(\mathbf{s}, T)]_{b,\tau,c}\Big).\end{aligned} \tag{26}$$

CCTG parameterizes amplitude and phase on shared continuous cyclic coordinates. Under the random-phase construction above, constant amplitude and affine phase displacement give the WSS special case. Varying fields instead produce time-varying modulation. NSF itself is not a stochastic process model. It evaluates deterministic learned fields at the cyclic coordinates associated with each forecast origin, while the WSS construction provides a second-order motivation for the representation. Channel-specific phase differences describe cross-channel temporal alignment, and residual predictors handle variation outside this structured path.

B.4 Complete NSTime Specification

Given a multivariate series $\{\mathbf{x}_\ell\}_{\ell=1}^N$ and forecast origin t, a training pair is

$$\mathbf{X}_t = \mathbf{x}_{t-S+1:t} \in \mathbb{R}^{S\times C}, \qquad \mathbf{Y}_t = \mathbf{x}_{t+1:t+P} \in \mathbb{R}^{P\times C}. \tag{27}$$

NSTime models the structured recurrent component and assigns the remaining variation to residual predictors. For a batch element b, s_b denotes the cyclic coordinate of its first input point. The first forecast point is queried at $(s_b + S) \bmod L$. Thus the T used for a generic queried interval in the main paper equals S for the historical window and P for the forecast window.

Recursive Decomposition. Starting from $\mathbf{T}_0 = \mathbf{X}_t$, the hierarchical decomposition is:

$$\begin{aligned}\mathbf{T}_j &= \mathrm{MA}_k(\mathbf{T}_{j-1}), \quad \mathbf{D}_j = \mathbf{T}_{j-1} - \mathbf{T}_j, \\ \mathbf{B} &= \mathbf{T}_J, \qquad\qquad\quad j = 1, \ldots, J.\end{aligned} \tag{28}$$

This construction gives the exact identity $\mathbf{X}_t = \mathbf{B} + \sum_{j=1}^J \mathbf{D}_j$.

Continuous Circular Temporal Generators. Each CCTG maintains a circular table $\mathbf{M} \in \mathbb{R}^{L\times C}$. The full parameterization uses the low-rank form

$$\mathbf{M} = \mathbf{UV}, \qquad \mathbf{U} \in \mathbb{R}^{L\times r}, \quad \mathbf{V} \in \mathbb{R}^{r\times C}, \tag{29}$$

and circular smoothing $\widetilde{\mathbf{M}} = \mathcal{S}_{\mathrm{circ}}(\mathbf{M})$. For a normalized kernel $\{w_\ell\}_{\ell=-h}^{h}$, the operator has the form

$$\begin{aligned}[\mathcal{S}_{\mathrm{circ}}(\mathbf{M})]_{k,c} &= \sum_{\ell=-h}^{h} w_\ell \mathbf{M}_{(k+\ell) \bmod L, c}, \\ \sum_{\ell=-h}^{h} w_\ell &= 1.\end{aligned} \tag{30}$$

The corresponding ablations replace this factorization with a full table or replace circular smoothing with the identity operator.

For field $q \in \{A, \phi\}$ at decomposition level j, batch element b, step $\tau \in \{0, \dots, T-1\}$, and channel c, let $u_{b,\tau,c}^{(q)} = s_b + \tau + \delta_{j,c}^{(q)}$, $k_0 = \lfloor u_{b,\tau,c}^{(q)} \rfloor$, $k_1 = k_0 + 1$, and $\alpha = u_{b,\tau,c}^{(q)} - k_0$. Circular interpolation gives

$$G(\mathbf{s}, T)_{b,\tau,c} = (1-\alpha)\widetilde{\mathbf{M}}_{k_0 \bmod L, c} + \alpha \widetilde{\mathbf{M}}_{k_1 \bmod L, c}. \tag{31}$$

Amplitude and phase use independent tables and offsets. The background CCTG has its own table and offsets. Factorization, smoothing, and interpolation change parameterization and access geometry, not the number L of cycle positions. At integer coordinates, interpolation reduces exactly to discrete circular lookup.

Periodicity and Representation Scope. Modulo indexing makes every generated field periodic in its start coordinate, so $G_j^{(q)}(s + L, T) = G_j^{(q)}(s, T)$. The compatible carrier satisfies $\omega_j L = 2\pi m_j$. Therefore,

$$\mathbf{H}_j(s + L, T) = \mathbf{H}_j(s, T). \tag{32}$$

The structured path consequently represents an L-periodic modulation field. Under the random-phase construction, a nonconstant periodic amplitude or phase displacement is generally cyclostationary rather than WSS because its second-order statistics depend periodically on t. CCTG does not represent an arbitrary nonperiodic $A(t)$ or $\phi(t)$. Background and residual branches are responsible for departures from this bounded cyclic prior.

Background and Residual Forecasting. For the background branch,

$$\mathbf{C}_B^{\mathrm{in}} = G^{(B)}(\mathbf{s}, S), \qquad \mathbf{C}_B^{\mathrm{out}} = G^{(B)}((\mathbf{s} + S) \bmod L, P), \tag{33}$$

and

$$\widehat{\mathbf{B}} = F_B\big((\mathbf{B} - \mathbf{C}_B^{\mathrm{in}})^\top\big)^\top + \mathbf{C}_B^{\mathrm{out}}, \tag{34}$$

where $F_B : \mathbb{R}^{C\times S} \to \mathbb{R}^{C\times P}$.

AM–PM Non-Stationary Forecasting. At decomposition level j, CCTGs produce

$$\begin{aligned}\mathbf{A}_j(\mathbf{s}, T) &= \mathrm{Softplus}\Big(G_j^{(A)}(\mathbf{s}, T)\Big), \\ \boldsymbol{\phi}_j(\mathbf{s}, T) &= G_j^{(\phi)}(\mathbf{s}, T).\end{aligned} \tag{35}$$

The total phase and AM–PM component are

$$\begin{aligned}\big[\boldsymbol{\Phi}_j(\mathbf{s}, T)\big]_{b,\tau,c} &= \omega_j(s_b + \tau) + [\boldsymbol{\phi}_j(\mathbf{s}, T)]_{b,\tau,c}, \\ \mathbf{H}_j(\mathbf{s}, T) &= \mathbf{A}_j(\mathbf{s}, T) \odot \cos\big(\boldsymbol{\Phi}_j(\mathbf{s}, T)\big), \\ \omega_j &= \frac{2\pi m_j}{L}, \qquad m_j \in \mathbb{N}.\end{aligned} \tag{36}$$

History and forecast query consecutive intervals of the same cyclic field:

$$\mathbf{H}_j^{\mathrm{in}} = \mathbf{H}_j(\mathbf{s}, S), \qquad \mathbf{H}_j^{\mathrm{out}} = \mathbf{H}_j((\mathbf{s} + S) \bmod L, P). \tag{37}$$

A channel-wise gate and a bounded signed residual coefficient give

$$\begin{aligned} \boldsymbol{\lambda}_j &= \sigma(\mathbf{g}_j), \qquad \mathbf{R}_j = \mathbf{D}_j - \boldsymbol{\lambda}_j \odot \mathbf{H}_j^{\text{in}}, \\ \widehat{\mathbf{D}}_j &= \boldsymbol{\lambda}_j \odot \mathbf{H}_j^{\text{out}} + \tanh(a_j) F_{R,j}(\mathbf{R}_j^\top)^\top, \end{aligned} \tag{38}$$

where $\boldsymbol{\lambda}_j \in (0,1)^C$ is broadcast over time, $a_j \in \mathbb{R}$, $\rho_j \equiv \tanh(a_j)$ is the learned residual scale used in the main paper, and $F_{R,j} : \mathbb{R}^{C\times S} \to \mathbb{R}^{C\times P}$. The final forecast is

$$\widehat{\mathbf{Y}} = \widehat{\mathbf{B}} + \sum_{j=1}^{J} \widehat{\mathbf{D}}_j. \tag{39}$$

CCTG supplies aligned cyclic coordinates, NSF models amplitude and phase variation for each detail residual, and the residual branches capture dynamics not explained by this prior. All branches are optimized jointly with the MSE objective in the main paper.

Identifiability and scope. Amplitude and phase are forecasting variables, not uniquely identifiable physical quantities. Phase is periodic, amplitude–phase trade-offs can admit equivalent reconstructions, and the residual path can absorb structure outside the cyclic prior. NSTime constrains, but does not remove, this ambiguity through the recursive decomposition, shared cyclic coordinates, smooth low-rank fields, and channel-wise gates. The experiments evaluate a predictive claim and use matched interventions to assess the contribution of the structured path under the reported protocols. They do not establish causal or physically unique component recovery.

C Experimental Protocol and Full Benchmark Results

This section supplies the protocol and disaggregated evidence behind the main-paper averages. Unless an analysis explicitly states an intervention, all results use the same splits, horizons, and accuracy metrics.

C.1 Datasets and Forecasting Protocol

We evaluate 11 multivariate benchmarks covering energy, weather, and traffic at multiple temporal resolutions. We use $S = 96$ and $P \in \{96, 192, 336, 720\}$, except PEMS, which uses $P \in \{12, 24, 48, 96\}$. Data splits follow the main paper, and $J = 3$ unless stated otherwise. The model predicts all channels jointly from the same input window. Reported MSE and MAE aggregate over forecast steps and channels.

ETT datasets. The Electricity Transformer Temperature datasets contain seven load and temperature variables. We use ETTh1 and ETTh2 at hourly resolution and ETTm1 and ETTm2 at 15-minute resolution.

Weather dataset. Weather contains 21 meteorological variables, including temperature, humidity, wind, and pressure, sampled every 10 minutes.

Electricity dataset. Electricity records hourly consumption for 321 clients.

Solar dataset. Solar records generation from 137 photovoltaic plants at 10-minute resolution.

PEMS datasets. PEMS03, PEMS04, PEMS07, and PEMS08 record traffic flow from road sensors at 5-minute resolution. Their channel counts range from 170 to 883.

Dataset statistics. Table 5 summarizes length, dimensionality, and sampling resolution.

C.2 Baselines

We compare NSTime with Koopa Liu et al. (2023), TimeBridge Liu et al. (2025a), DTAF Lu et al. (2026), TimeMosaic Ding et al. (2026), RAFT Han et al. (2025), S-Mamba Wang et al. (2025c), CycleNet Lin et al. (2024), iTransformer Liu et al. (2024), DLinear Zeng et al. (2023), and Crossformer Zhang & Yan (2023). These methods provide representative comparisons across complementary forecasting paradigms, including latent dynamics, non-stationary and frequency-aware modeling, recurrent temporal patterns, state-space models, linear forecasting, and Transformer architectures. They also span different assumptions about temporal structure, cross-variable dependence,

and model complexity. All baselines are reported over the same datasets and prediction horizons, enabling a consistent comparison across methods and forecasting settings.

Table 5: Dataset statistics.

Dataset	Time Steps (T)	Variables (C)	Resolution
ETTh1	17,420	7	1 hour
ETTh2	17,420	7	1 hour
ETTm1	69,680	7	15 minutes
ETTm2	69,680	7	15 minutes
Weather	52,696	21	10 minutes
Electricity	26,304	321	1 hour
Solar	52,560	137	10 minutes
PEMS03	26,208	358	5 minutes
PEMS04	16,992	307	5 minutes
PEMS07	28,224	883	5 minutes
PEMS08	17,856	170	5 minutes

C.3 Evaluation Metrics

We report MSE and MAE for forecasting. ADF, KPSS, and the amplitude–phase drift statistics are post-hoc diagnostics. None enters training, model selection, or the forecasting loss.

C.3.1 Forecasting Accuracy Metrics

For target $\mathbf{Y}$ and prediction $\widehat{\mathbf{Y}}$ in $\mathbb{R}^{P\times C}$, both metrics average over horizons and channels.

Mean Squared Error (MSE).

$$\mathrm{MSE} = \frac{1}{PC}\sum_{\tau=1}^{P}\sum_{c=1}^{C}\left(Y_{\tau,c} - \widehat{Y}_{\tau,c}\right)^2. \tag{40}$$

Mean Absolute Error (MAE).

$$\mathrm{MAE} = \frac{1}{PC}\sum_{\tau=1}^{P}\sum_{c=1}^{C}\left|Y_{\tau,c} - \widehat{Y}_{\tau,c}\right|. \tag{41}$$

C.3.2 Non-Stationarity Diagnostics

For each input-window–variable pair, ADF>0.05 indicates insufficient evidence to reject a unit root, while KPSS≤ 0.05 rejects stationarity. Their intersection requires both criteria. These tests diagnose generic statistical non-stationarity rather than temporal pattern deformation, so they are used as descriptive indicators rather than ground-truth labels.

Augmented Dickey–Fuller (ADF) Test. The ADF regression is

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p}\delta_i \Delta x_{t-i} + \varepsilon_t, \tag{42}$$

where $\Delta x_t = x_t - x_{t-1}$. The unit-root null is $\gamma = 0$. More negative test statistics provide stronger evidence against this null.

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test. KPSS decomposes the series as

$$x_t = d_t + r_t + \varepsilon_t, \qquad r_t = r_{t-1} + u_t, \tag{43}$$

where d_t is a constant or deterministic linear trend, ε_t is stationary, and u_t is a random-walk innovation. Under the stationarity null, $\mathrm{Var}(u_t) = 0$.

Let $\{\hat{\varepsilon}_t\}$ be the residuals after fitting d_t, and define

$$S_t = \sum_{i=1}^{t} \hat{\varepsilon}_i. \tag{44}$$

The KPSS statistic is then computed as

$$\mathrm{KPSS} = \frac{1}{T^2 \hat{\sigma}^2} \sum_{t=1}^{T} S_t^2, \tag{45}$$

where $\hat{\sigma}^2$ estimates the long-run variance of ε_t. Larger values provide stronger evidence against the stationarity null.

C.4 Full Per-Horizon Forecasting Results

Table 6 expands the main-paper averages to every horizon using the same runs and comparison pool. This is the table referenced as the main paper. It reports how the aggregate rankings vary across forecast lengths and makes horizon-specific exceptions explicit.

Table 6: Per-horizon forecasting results. Bold and underlined entries mark the best and second-best values. Ties are retained.

Dataset	P	NSTime		DTAF		TimeMosaic		TimeBridge		RAFT		S-Mamba		CycleNet		iTransformer		Koopa		DLinear		Crossformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.363**	0.388	0.376	**0.385**	0.369	0.389	0.375	0.390	0.376	0.401	0.386	0.405	0.388	0.398	0.386	0.405	0.385	0.408	0.381	0.393	0.423	0.448
	192	**0.406**	0.416	0.427	**0.414**	0.416	0.415	0.428	0.421	0.417	0.424	0.443	0.437	0.438	0.459	0.441	0.436	0.445	0.434	0.431	0.434	0.471	0.474
	336	**0.450**	0.438	0.461	**0.435**	0.454	**0.435**	0.470	0.443	0.454	0.441	0.489	0.468	0.476	0.456	0.487	0.458	0.489	0.460	0.475	0.450	0.570	0.546
	720	**0.456**	0.462	0.458	**0.453**	0.461	0.458	0.481	0.481	0.465	0.465	0.502	0.489	0.478	0.470	0.503	0.491	0.497	0.480	0.497	0.494	0.653	0.621
ETTh2	96	**0.281**	0.337	0.287	**0.332**	0.286	**0.332**	0.286	0.333	0.289	0.342	0.296	0.348	0.290	0.339	0.297	0.349	0.301	0.350	0.317	0.374	0.745	0.584
	192	**0.359**	0.389	0.369	0.383	0.362	**0.380**	0.364	0.385	0.382	0.400	0.376	0.396	0.376	0.394	0.380	0.400	0.375	0.398	0.412	0.438	0.877	0.656
	336	0.418	0.430	0.411	0.419	**0.409**	**0.416**	0.416	0.425	0.424	0.439	0.424	0.431	0.424	0.436	0.428	0.432	0.417	0.431	0.550	0.514	1.043	0.731
	720	0.425	0.442	0.428	0.438	**0.396**	**0.422**	0.435	0.448	0.433	0.457	0.426	0.444	0.455	0.460	0.427	0.445	0.464	0.467	0.813	0.648	1.104	0.763
ETTm1	96	**0.303**	0.352	0.314	0.342	0.308	**0.338**	0.312	0.344	0.325	0.369	0.333	0.368	0.319	0.360	0.334	0.368	0.333	0.370	0.345	0.372	0.404	0.426
	192	**0.344**	0.377	0.361	**0.367**	0.364	0.368	0.366	0.377	0.364	0.388	0.376	0.390	0.360	0.381	0.377	0.391	0.374	0.391	0.380	0.389	0.450	0.451
	336	**0.379**	0.402	0.395	**0.390**	0.394	**0.390**	0.403	0.404	0.391	0.407	0.408	0.413	0.389	0.403	0.426	0.420	0.403	0.410	0.413	0.413	0.532	0.515
	720	**0.436**	0.438	0.470	0.431	0.460	**0.428**	0.454	0.438	0.444	0.438	0.475	0.448	0.447	0.441	0.491	0.459	0.467	0.447	0.474	0.453	0.666	0.589
ETTm2	96	**0.165**	0.249	0.175	0.252	0.170	**0.248**	0.173	0.250	0.176	0.264	0.179	0.263	0.172	0.267	0.180	0.264	0.181	0.263	0.193	0.292	0.287	0.366
	192	**0.229**	0.293	0.241	0.297	0.236	0.291	0.243	0.297	0.241	0.306	0.250	0.309	**0.229**	**0.290**	0.250	0.309	0.241	0.303	0.284	0.362	0.414	0.492
	336	0.288	0.331	0.301	0.335	0.295	0.331	0.302	0.336	0.302	0.345	0.312	0.349	**0.284**	**0.327**	0.311	0.348	0.301	0.339	0.369	0.427	0.597	0.542
	720	**0.385**	0.390	0.400	0.394	0.392	**0.386**	0.399	0.394	0.404	0.405	0.411	0.406	0.389	0.391	0.412	0.407	0.404	0.403	0.554	0.522	1.730	1.042
Electricity	96	**0.135**	**0.230**	0.156	0.254	0.160	0.258	0.138	0.231	0.152	0.257	0.139	0.235	0.143	0.240	0.148	0.240	0.153	0.253	0.197	0.282	0.219	0.314
	192	**0.152**	0.245	0.185	0.262	0.174	0.269	0.158	0.250	0.156	0.257	0.159	0.255	**0.152**	**0.244**	0.162	0.253	0.183	0.277	0.196	0.285	0.231	0.322
	336	**0.169**	0.265	0.196	0.282	0.189	0.283	0.180	0.271	0.171	0.272	0.176	0.272	0.170	**0.264**	0.178	0.269	0.195	0.290	0.209	0.301	0.246	0.337
	720	0.210	0.301	0.244	0.314	0.225	0.304	0.223	0.308	0.221	0.300	**0.204**	**0.298**	0.212	0.299	0.225	0.317	0.230	0.317	0.245	0.333	0.280	0.363
Solar	96	**0.178**	0.247	0.232	0.255	0.209	0.248	0.196	**0.214**	0.278	0.278	0.205	0.244	0.190	0.247	0.203	0.237	0.219	0.265	0.290	0.378	0.310	0.331
	192	**0.198**	0.256	0.251	0.266	0.238	0.272	0.230	**0.239**	0.297	0.297	0.237	0.270	0.210	0.266	0.233	0.261	0.248	0.282	0.320	0.398	0.734	0.725
	336	**0.208**	**0.255**	0.312	0.295	0.255	0.277	0.250	0.258	0.311	0.311	0.258	0.288	0.217	0.266	0.248	0.273	0.272	0.296	0.353	0.415	0.750	0.735
	720	**0.214**	**0.257**	0.291	0.291	0.257	0.284	0.256	0.264	0.318	0.325	0.260	0.288	0.223	0.266	0.249	0.275	0.281	0.303	0.356	0.413	0.769	0.765
Weather	96	**0.156**	0.202	0.177	0.210	0.165	**0.198**	0.169	0.206	0.185	0.241	0.165	0.210	0.158	0.203	0.174	0.214	0.158	0.204	0.196	0.255	0.158	0.230
	192	**0.204**	0.246	0.220	0.250	0.215	**0.243**	0.219	0.249	0.237	0.287	0.214	0.252	0.207	0.247	0.221	0.254	0.208	0.250	0.237	0.296	0.206	0.277
	336	**0.261**	0.289	0.275	0.289	0.272	**0.286**	0.274	0.290	0.290	0.327	0.274	0.297	0.262	0.289	0.278	0.296	0.266	0.291	0.283	0.335	0.272	0.335
	720	**0.343**	0.344	0.353	**0.339**	0.353	0.340	0.352	0.341	0.367	0.380	0.350	0.345	0.344	0.344	0.358	0.349	0.348	0.345	0.345	0.381	0.398	0.418
PEMS03	12	**0.064**	**0.169**	0.090	0.199	0.078	0.186	0.076	0.188	0.071	0.175	0.065	**0.169**	0.069	0.176	0.071	0.174	0.069	0.177	0.103	0.218	0.090	0.203
	24	**0.083**	**0.194**	0.139	0.247	0.116	0.229	0.113	0.226	0.099	0.201	0.087	0.196	0.094	0.207	0.093	0.201	0.100	0.214	0.180	0.292	0.121	0.240
	48	**0.114**	**0.224**	0.251	0.337	0.190	0.296	0.191	0.292	0.153	0.243	0.133	0.243	0.144	0.259	0.125	0.236	0.207	0.310	0.316	0.409	0.202	0.317
	96	**0.156**	**0.261**	0.416	0.447	0.315	0.388	0.288	0.363	0.253	0.301	0.201	0.305	0.192	0.291	0.164	0.275	0.343	0.412	0.450	0.507	0.262	0.367
PEMS04	12	**0.069**	**0.171**	0.107	0.217	0.100	0.211	0.092	0.204	0.077	0.179	0.076	0.180	0.081	0.190	0.078	0.183	0.079	0.189	0.114	0.227	0.098	0.218
	24	**0.083**	**0.190**	0.176	0.281	0.140	0.248	0.128	0.243	0.091	0.195	0.084	0.193	0.105	0.220	0.095	0.205	0.098	0.212	0.187	0.296	0.131	0.256
	48	**0.106**	**0.217**	0.329	0.390	0.231	0.322	0.213	0.315	0.111	0.218	0.115	0.224	0.144	0.261	0.120	0.233	0.180	0.297	0.318	0.402	0.205	0.326
	96	**0.132**	**0.244**	0.577	0.530	0.388	0.434	0.307	0.384	0.137	0.247	0.137	0.248	0.183	0.296	0.150	0.262	0.311	0.402	0.424	0.479	0.402	0.457
PEMS07	12	**0.054**	**0.149**	0.082	0.187	0.074	0.181	0.073	0.184	0.062	0.159	0.063	0.159	0.062	0.162	0.067	0.165	0.065	0.167	0.115	0.242	0.094	0.200
	24	**0.069**	**0.169**	0.138	0.244	0.113	0.221	0.111	0.219	0.078	0.178	0.081	0.183	0.086	0.192	0.088	0.190	0.095	0.206	0.210	0.329	0.139	0.247
	48	0.097	0.196	0.271	0.342	0.197	0.292	0.237	0.328	0.102	0.203	**0.093**	**0.192**	0.128	0.234	0.110	0.215	0.156	0.270	0.398	0.458	0.311	0.369
	96	0.130	0.237	0.461	0.457	0.292	0.369	0.303	0.354	0.135	0.231	**0.117**	**0.217**	0.176	0.268	0.139	0.245	0.257	0.357	0.594	0.553	0.396	0.442
PEMS08	12	0.077	**0.172**	0.099	0.206	0.107	0.221	0.091	0.201	0.078	0.181	**0.076**	0.178	0.082	0.185	0.079	0.182	0.079	0.189	0.154	0.276	0.165	0.214
	24	**0.099**	**0.197**	0.165	0.266	0.124	0.224	0.137	0.246	0.102	0.205	0.104	0.209	0.117	0.226	0.115	0.219	0.108	0.224	0.248	0.353	0.215	0.260
	48	**0.141**	0.233	0.316	0.373	0.304	0.371	0.265	0.343	0.148	0.244	0.167	**0.228**	0.169	0.268	0.186	0.235	0.210	0.321	0.440	0.470	0.315	0.355
	96	**0.185**	**0.261**	0.615	0.522	0.345	0.371	0.410	0.407	0.277	0.307	0.245	0.280	0.233	0.306	0.221	0.267	0.369	0.433	0.674	0.565	0.377	0.397

The disaggregated results retain the broad main-paper pattern without implying uniform dominance. NSTime is competitive on Electricity, Solar, Weather, and PEMS, while TimeMosaic leads the longer ETTh2 settings and TimeBridge is often stronger on Solar MAE. Reporting these cells separately prevents the averages from obscuring horizon- and metric-specific exceptions.

NSTime's strongest results span datasets with different sampling rates and channel counts, so the empirical pattern is not confined to one benchmark family or dimensional regime. This breadth alone, however, does not identify the deformation prior as the cause of the gains. MSE and MAE also do not always move together. On Solar and several ETT settings, stronger MSE ranks coexist with weaker MAE ranks, whereas the two metrics improve together more often on PEMS. Because MSE weights large deviations more heavily, this difference may reflect non-uniform changes in the error distribution; without a direct error-tail or event-level analysis, it should not be attributed specifically to amplitude or timing errors.

NSTime remains competitive across multiple forecast lengths, although the strongest competitor changes with dataset, horizon, and metric. The table therefore supports empirical utility across the reported settings, not universal dominance or the claim that every dataset exhibits the same form or severity of temporal pattern deformation. The comparisons are primarily fixed-seed, so small leaderboard gaps should not be interpreted as paired multi-seed statistical significance.

C.5 Few-Shot Forecasting Results

The reduced-supervision experiment evaluates within-dataset performance when only 5% of each ETT training split is retained. We compare NSTime with forecasting, language-model, and vision–language baselines on the four ETT datasets. The test windows come from the same dataset as the retained training windows, so this protocol measures within-dataset generalization under limited supervision rather than cross-dataset, zero-shot, or pretrained transfer. ETTh1 and ETTh2 are reported at $P \in \{96, 192, 336\}$ under this protocol, so their averages cover three horizons, while ETTm1 and ETTm2 averages cover four. The corresponding results are reported in Table 7.

Table 7: Few-shot forecasting with 5% training data on the ETT benchmarks. Bold and underlined entries mark the best and second-best values across both panels, with ties retained. ETTh1 and ETTh2 averages cover three reported horizons.

Dataset	P	NSTime		DTAF		Time-VLM		Time-LLM		GPT4TS		TimesNet		FEDformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.420	0.439	0.501	0.483	**0.417**	**0.435**	0.483	0.464	0.543	0.506	0.892	0.625	0.593	0.529
	192	0.460	**0.457**	0.584	0.529	**0.450**	0.458	0.629	0.540	0.748	0.580	0.940	0.665	0.652	0.563
	336	**0.453**	**0.458**	0.607	0.541	0.460	0.465	0.768	0.626	0.754	0.595	0.945	0.653	0.731	0.594
	Avg	0.444	**0.451**	0.564	0.518	**0.442**	0.453	0.627	0.543	0.682	0.560	0.926	0.648	0.659	0.562
ETTh2	96	0.315	0.373	0.382	0.419	**0.302**	**0.365**	0.336	0.397	0.376	0.421	0.409	0.420	0.390	0.424
	192	**0.345**	**0.390**	0.428	0.440	0.361	0.406	0.406	0.425	0.418	0.441	0.483	0.464	0.457	0.465
	336	**0.381**	**0.421**	0.443	0.458	0.398	0.434	0.405	0.432	0.408	0.439	0.499	0.479	0.477	0.483
	Avg	**0.347**	**0.395**	0.418	0.439	0.354	0.402	0.382	0.418	0.401	0.434	0.464	0.454	0.441	0.457
ETTm1	96	0.361	0.386	**0.312**	**0.350**	0.314	0.357	0.316	0.377	0.386	0.405	0.606	0.518	0.628	0.544
	192	0.358	0.385	0.351	0.375	**0.343**	**0.373**	0.450	0.464	0.440	0.438	0.681	0.539	0.666	0.566
	336	0.376	0.396	0.395	0.394	**0.373**	**0.391**	0.450	0.424	0.485	0.459	0.786	0.597	0.807	0.628
	720	0.427	0.427	0.427	**0.418**	**0.425**	0.420	0.483	0.471	0.577	0.499	0.796	0.593	0.822	0.633
	Avg	0.381	0.399	0.371	**0.384**	**0.364**	0.385	0.425	0.434	0.472	0.450	0.717	0.562	0.731	0.593
ETTm2	96	0.174	0.267	0.181	0.270	**0.169**	**0.260**	0.174	0.261	0.199	0.280	0.220	0.299	0.229	0.320
	192	0.238	0.310	0.247	0.312	0.224	0.298	**0.215**	**0.287**	0.256	0.316	0.311	0.361	0.394	0.361
	336	0.302	0.349	0.304	0.353	0.282	0.338	**0.273**	**0.330**	0.318	0.353	0.338	0.366	0.378	0.427
	720	0.383	0.400	0.389	0.399	**0.375**	**0.397**	0.433	0.412	0.460	0.436	0.509	0.465	0.523	0.510
	Avg	0.274	0.332	0.280	0.334	**0.263**	**0.323**	0.274	**0.323**	0.308	0.346	0.345	0.373	0.381	0.405

Dataset	P	NSTime		Autoformer		Stationary		ETSformer		LightTS		Informer		Reformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.420	0.439	0.681	0.570	0.952	0.650	1.169	0.832	1.483	0.910	1.225	0.812	1.198	0.795
	192	0.460	**0.457**	0.725	0.602	0.943	0.645	1.221	0.853	1.525	0.930	1.249	0.828	1.273	0.853
	336	**0.453**	**0.458**	0.761	0.624	0.935	0.644	1.179	0.832	1.347	0.870	1.202	0.811	1.254	0.857
	Avg	0.444	**0.451**	0.722	0.599	0.943	0.646	1.190	0.839	1.452	0.903	1.225	0.817	1.242	0.835
ETTh2	96	0.315	0.373	0.428	0.468	0.408	0.423	0.678	0.619	2.022	1.006	3.837	1.508	3.753	1.518
	192	**0.345**	**0.390**	0.496	0.504	0.497	0.468	0.845	0.697	3.534	1.348	3.975	1.933	3.516	1.473
	336	**0.381**	**0.421**	0.486	0.496	0.507	0.481	0.905	0.727	4.063	1.451	3.956	1.520	3.312	1.427
	Avg	**0.347**	**0.395**	0.470	0.489	0.471	0.457	0.809	0.681	3.206	1.268	3.923	1.654	3.527	1.473
ETTm1	96	0.361	0.386	0.726	0.578	0.823	0.587	1.031	0.747	1.048	0.733	1.130	0.775	1.234	0.798
	192	0.358	0.385	0.750	0.591	0.844	0.591	1.087	0.766	1.097	0.756	1.150	0.788	1.287	0.839
	336	0.376	0.396	0.851	0.659	0.870	0.603	1.138	0.787	1.147	0.775	1.198	0.809	1.288	0.842
	720	0.427	0.427	0.857	0.655	0.893	0.611	1.245	0.831	1.200	0.799	1.175	0.794	1.247	0.828
	Avg	0.381	0.399	0.796	0.621	0.857	0.598	1.125	0.783	1.123	0.766	1.163	0.792	1.264	0.827
ETTm2	96	0.174	0.267	0.232	0.322	0.238	0.316	0.404	0.485	1.108	0.772	3.599	1.478	3.883	1.545
	192	0.238	0.310	0.291	0.357	0.298	0.349	0.479	0.521	1.317	0.850	3.578	1.475	3.553	1.484
	336	0.302	0.349	0.478	0.517	0.353	0.380	0.552	0.555	1.415	0.879	3.561	1.473	3.446	1.460
	720	0.383	0.400	0.553	0.538	0.475	0.445	0.701	0.627	1.822	0.984	3.896	1.533	3.445	1.460
	Avg	0.274	0.332	0.388	0.434	0.341	0.372	0.534	0.547	1.415	0.871	3.659	1.490	3.582	1.487

The central observation is competitive within-dataset performance rather than uniform dominance. Trained from scratch with 5% of the data, NSTime ranks within the top three on all eight dataset–metric averages, first in three and second in three. It leads both ETTh2 metrics and ETTh1 MAE without external language or vision pretraining. These results provide evidence that NSTime can remain competitive under the reported reduced-data protocol. They are compatible with the hypoth-

esis that the structured prior constrains the history-to-future mapping in a useful way, but they do not directly establish a reusable lower-dimensional deformation rule or isolate that explanation from model capacity and optimization effects.

Pretrained Time-VLM or Time-LLM remains stronger on several ETTm settings, so the structural prior does not replace external pretraining in every reduced-data regime. The supported claim is limited to generalization from a smaller same-dataset training subset to held-out windows at the reported horizons. Cross-dataset or zero-shot transfer would require training on one collection and evaluating on an unseen collection, which is outside this experiment.

D Ablation Study

The main paper isolates NSF, amplitude, phase, CCTG, and decomposition at standard and longest horizons. Here we extend the same component-level analysis across horizons, then test decomposition depth and residual-head choice. Each table should be read as a matched intervention under the stated configuration. Comparisons are made within a row against its corresponding Full reference, not across independently tuned tables.

D.1 Core Ablation of NSF and CCTG Parameterization

Table 8 extends the component analysis to every horizon on ETTh1, ETTm1, and Weather. Removing NSF increases MSE in every reported row, while removing amplitude or phase usually increases error. These matched comparisons provide setting-specific evidence that the structured path and its two fields contribute to forecasting accuracy. The w/o-NSF configuration retains the residual predictors and serves as the matched reference for the incremental effect of the structured AM–PM path. Internal CCTG choices have smaller and less uniform effects, including ties and reversals. The discrete-index intervention retains the same table and changes its access rule; the other columns should be interpreted only with respect to the component explicitly altered in each ablation.

Table 8: Extended component and CCTG-parameterization ablations across the four standard prediction horizons. Full NSTime is the corresponding per-horizon result in Table 6, and every intervention is compared with its matched Full row. Boldface identifies this reference rather than the best entry in every row. Ties and reversals are retained.

Dataset	P	Full NSTime		w/o NSF		w/o A		w/o ϕ		w/ Discrete Index		w/o offset δ		w/o Smoothing		w/o low-rank	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.363**	**0.388**	0.391	0.402	0.372	0.392	0.371	0.391	0.365	0.389	0.369	0.391	0.365	0.389	0.366	0.389
	192	**0.406**	**0.416**	0.430	0.433	0.419	0.422	0.418	0.422	0.407	0.417	0.416	0.422	0.408	0.417	0.416	0.420
	336	**0.450**	**0.438**	0.470	0.456	0.466	0.449	0.461	0.444	0.453	0.441	0.457	0.446	0.452	0.441	0.459	0.445
	720	**0.456**	**0.462**	0.489	0.479	0.476	0.465	0.483	0.469	0.463	0.462	0.477	0.470	0.454	0.459	0.476	0.473
ETTm1	96	**0.303**	**0.352**	0.326	0.363	0.305	0.353	0.306	0.353	0.307	0.354	0.315	0.356	0.308	0.355	0.314	0.355
	192	**0.344**	**0.377**	0.366	0.385	0.348	0.378	0.347	0.378	0.347	0.388	0.355	0.379	0.349	0.379	0.357	0.379
	336	**0.379**	**0.402**	0.405	0.411	0.387	0.405	0.385	0.404	0.385	0.405	0.392	0.405	0.386	0.404	0.396	0.406
	720	**0.436**	**0.438**	0.464	0.448	0.447	0.443	0.447	0.445	0.447	0.446	0.454	0.442	0.441	0.441	0.456	0.443
Weather	96	**0.156**	**0.202**	0.175	0.217	0.162	0.210	0.161	0.209	0.158	0.204	0.158	0.204	0.157	0.203	0.160	0.205
	192	**0.204**	**0.246**	0.216	0.256	0.211	0.251	0.211	0.251	0.210	0.248	0.211	0.249	0.207	0.247	0.209	0.248
	336	**0.261**	**0.289**	0.272	0.297	0.271	0.296	0.271	0.295	0.267	0.290	0.268	0.291	0.265	0.291	0.267	0.292
	720	**0.343**	**0.344**	0.354	0.356	0.350	0.348	0.350	0.348	0.346	0.342	0.347	0.342	0.346	0.344	0.345	0.343

The extended table reveals different effect sizes among the design choices. Across the 12 dataset–horizon rows, removing NSF increases MSE every time, by 6.50% on average. Removing amplitude and phase increases MSE by 2.77% and 2.67%, respectively. Their similar average penalties are compatible with both fields contributing, but do not uniquely determine how the model uses either field. Within CCTG, discrete indexing is worse in all 12 rows (a 1.39% mean MSE increase), while removing channel offsets or the low-rank parameterization increases MSE by 2.80% and 2.84% on average. These results provide measurable evidence for the tested access and parameterization choices under the reported configurations.

D.2 Decomposition and Residual Predictor Analysis

Fig. 5 shows that $J = 3$ is consistently competitive. All other experiments therefore use three levels. Table 9 compares four residual heads. MLP is strongest or tied in 11 of 12 MSE settings and 10 of

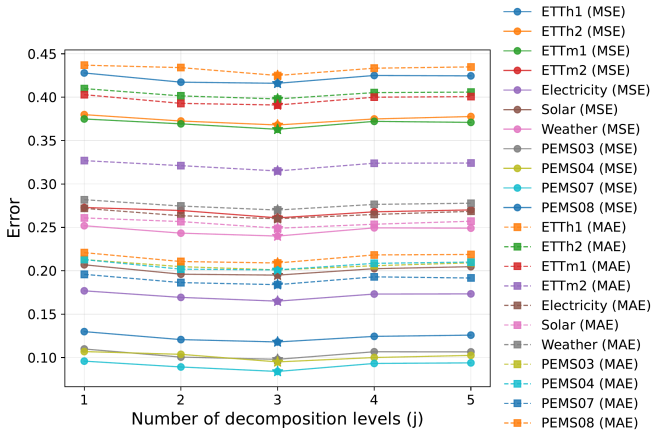


Figure 5: Sensitivity to decomposition depth J.

Table 9: Residual-predictor ablation. Within each dataset, rows follow $P = 96, 192, 336, 720$. The MLP column is the same Full reference used in Table 8.

	MLP		Linear		Convolution		KAN Linear	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.363	0.388	0.380	0.392	0.369	0.390	0.381	0.395
	0.406	0.416	0.428	0.421	0.411	0.416	0.429	0.424
	0.450	0.438	0.469	0.442	0.455	0.440	0.473	0.446
	0.456	0.462	0.474	0.464	0.465	0.465	0.474	0.466
ETTm1	0.303	0.352	0.310	0.354	0.311	0.354	0.311	0.355
	0.344	0.377	0.348	0.377	0.347	0.377	0.346	0.376
	0.379	0.402	0.382	0.404	0.381	0.402	0.381	0.402
	0.436	0.438	0.439	0.439	0.438	0.438	0.439	0.438
Weather	0.156	0.202	0.156	0.203	0.156	0.203	0.156	0.204
	0.204	0.246	0.204	0.245	0.206	0.247	0.203	0.246
	0.261	0.289	0.263	0.291	0.262	0.291	0.264	0.291
	0.343	0.344	0.344	0.344	0.346	0.345	0.344	0.345

12 MAE settings, and is used elsewhere. Alternative heads occasionally win: KAN Linear is best on Weather–192 MSE, while Linear is best on the corresponding MAE. The generally small gaps show that performance in these three datasets is not uniquely dependent on one residual-head choice. This comparison supports robustness to the tested heads; it does not identify a unique functional division between the residual and structured paths.

E Performance under Amplitude–Phase Drift

The component ablations quantify average Full–ablation differences but do not show how those differences vary across forecasting windows. We therefore stratify two long-horizon test settings using post-hoc deformation statistics computed from raw history–future pairs and compare Full only with its component-matched ablation.

E.1 Directed Drift Analysis

E.1.1 Targeted Component Test on ETTm2–720

The main-paper ablation shows near ties among ETTm2 variants at $P = 96$ but a clear Full advantage at $P = 720$. We therefore analyze ETTm2–720 to examine whether the Full–ablation gap differs across windows with lower or higher amplitude or phase drift. We partition its 10,801 test windows into empirical terciles using drift computed from each raw history–future pair at the dominant historical frequency, independently of the learned fields. The amplitude thresholds are

$0.5967/0.7781$ and the phase thresholds $0.1648/0.2314$, producing $3{,}601/3{,}600/3{,}600$ windows per tercile. All variants share the predictor, hidden dimensions, activation, and training protocol, and each ablation removes only its matched component.

For a bin $\mathcal{S}$ and its matched ablation, we report

$$\mathrm{Gain}(\mathcal{S}) = \frac{\mathrm{MSE}_{\mathrm{ablation}}(\mathcal{S}) - \mathrm{MSE}_{\mathrm{Full}}(\mathcal{S})}{\mathrm{MSE}_{\mathrm{ablation}}(\mathcal{S})} \times 100\%. \tag{46}$$

Confidence intervals are obtained by paired bootstrap resampling of windows within each bin, preserving the Full–ablation pairing. The drift statistics and tercile assignment are computed from raw history–future pairs and do not use NSTime's learned amplitude or phase fields.

Table 10: ETTm2–720 amplitude- and phase-drift analysis. Gain is the relative MSE reduction of Full over the matched ablation within each empirical tercile. Brackets report 95% paired-bootstrap confidence intervals.

Drift	Bin	N	Full MSE	Compared variant MSE	Gain (%)	95% CI
Amplitude	Low	3601	0.3498	0.3563 (w/o Amp)	1.85	[1.15, 2.52]
Amplitude	Medium	3600	0.3672	0.3754 (w/o Amp)	2.20	[1.43, 2.90]
Amplitude	High	3600	0.4383	0.4637 (w/o Amp)	5.50	[4.79, 6.22]
Phase	Low	3601	0.3137	0.3228 (w/o Phase)	2.83	[2.21, 3.47]
Phase	Medium	3600	0.3953	0.4029 (w/o Phase)	1.91	[1.24, 2.52]
Phase	High	3600	0.4463	0.4666 (w/o Phase)	4.36	[3.60, 5.09]

Interpretation. As summarized in Fig. 3 of the main paper and detailed in Table 10, Full has lower MSE than its matched ablation in all six empirical terciles, with every paired-bootstrap confidence interval excluding zero. On ETTm2–720, the largest average gaps occur in the high-drift terciles, reaching 5.50% for amplitude and 4.36% for phase. However, the intermediate bins are not ordered monotonically, and the window-level rank correlations are weak ($\rho = 0.053/0.034$). Because drift is computed post hoc from history–future pairs, this analysis describes an association within the evaluated test set; it neither provides a prospective selector nor establishes a causal dose–response relation.

E.1.2 Second-Dataset Replication on Electricity–720

The main paper additionally repeats the matched tercile analysis on Electricity–720. Fig. 6 reproduces that result: the gains remain positive in all six bins, but are smaller and non-monotone (0.21–0.26% for amplitude and 0.35–0.39% for phase). This replication provides sign-consistent Full–ablation differences on a second dataset, but it does not imply that larger drift must produce proportionally larger gains.

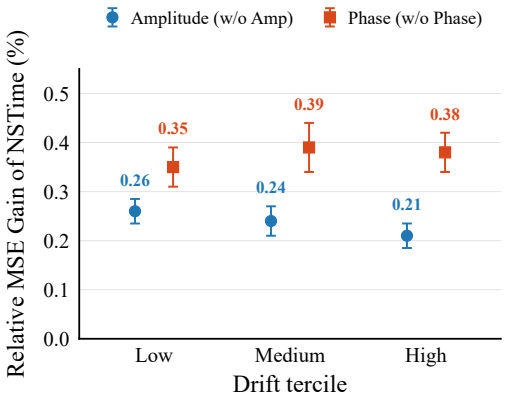


Figure 6: Electricity–720 replication of the amplitude- and phase-drift analysis. Markers report the relative MSE gain of Full over the matched ablation. Bars show paired-bootstrap 95% confidence intervals.

Taken together, the two datasets support a narrower conclusion than a generic "more drift, more gain" claim. ETTm2 has the largest matched gaps in the high-drift terciles, but its intermediate bins are not monotone. Electricity shows positive but much smaller and flatter differences. The repeatable observation is the positive average Full–ablation difference across the reported bins; its magnitude is dataset- and regime-dependent. The analysis does not show that drift magnitude is a universal predictor of component benefit or that the same ordering will hold on unseen datasets.

F CCTG Module Analysis and Statistical Non-Stationarity Diagnostics

This section mirrors the two analyses that follow the main-paper drift study. First, we distinguish a controlled CycleNet replacement from broader CCTG augmentation experiments. Second, we use ADF and KPSS post hoc to relate NSF gains to broader statistical non-stationarity without treating either test as a measurement of temporal pattern deformation.

F.1 CCTG Module Analysis and Compatibility Beyond NSTime

The main paper uses a parameter- and protocol-matched RCF-to-CCTG replacement in CycleNet to evaluate the combined change from discrete lookup to continuous, channel-aligned cyclic access. RCF reads a recurrent-cycle table at discrete integer indices and has neither fractional cyclic access nor learnable channel-specific coordinate offsets. CCTG keeps the same table length and capacity but jointly adds circular interpolation and channel-specific offsets. Because these two changes are introduced together, the comparison attributes results to the combined access mechanism rather than to either element alone. We complement this controlled test by inserting CCTG into eight heterogeneous forecasting backbones. Table 11 reports every horizon and evaluates empirical compatibility beyond NSTime.

Table 11: Per-horizon results before and after CCTG augmentation. Within each adjacent backbone and augmented pair, bold marks the lower MSE or MAE for the same dataset and horizon. Both values are bold when tied.

Dataset	P	CycleNet+RCF		CycleNet+CCTG		TQNet+TQ&MHA		TQNet+CCTG		PatchTST		PatchTST+CCTG		DLinear		DLinear+CCTG	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.383	**0.393**	**0.380**	0.395	0.373	**0.393**	**0.372**	**0.393**	0.377	0.396	**0.373**	**0.393**	0.381	0.393	**0.369**	**0.391**
ETTh1	192	0.432	**0.421**	**0.428**	0.422	0.430	0.427	**0.426**	**0.422**	**0.432**	**0.423**	**0.432**	0.424	0.431	0.434	**0.411**	**0.419**
ETTh1	336	0.474	0.443	**0.464**	**0.439**	0.476	0.446	**0.466**	**0.442**	**0.481**	0.452	**0.481**	**0.439**	0.475	0.450	**0.446**	**0.441**
ETTh1	720	0.472	0.466	**0.470**	**0.460**	0.486	0.475	**0.479**	**0.466**	0.489	0.477	**0.478**	**0.471**	0.497	0.494	**0.467**	**0.481**
ETTh1	Avg	0.440	0.431	**0.436**	**0.429**	0.441	0.435	**0.436**	**0.431**	0.445	0.437	**0.441**	**0.432**	0.446	0.443	**0.423**	**0.433**
ETTh2	96	0.289	0.338	**0.288**	**0.337**	0.290	0.342	**0.289**	**0.341**	**0.283**	**0.336**	**0.283**	**0.336**	0.317	0.374	**0.308**	**0.366**
ETTh2	192	**0.374**	**0.393**	**0.374**	**0.393**	0.369	0.392	**0.360**	**0.387**	0.367	0.389	**0.364**	**0.388**	0.412	0.438	**0.388**	**0.412**
ETTh2	336	0.436	0.441	**0.424**	**0.435**	**0.412**	**0.428**	**0.412**	0.431	**0.413**	**0.426**	**0.413**	**0.426**	0.550	0.514	**0.547**	**0.510**
ETTh2	720	**0.457**	0.462	**0.457**	**0.461**	0.426	**0.444**	**0.425**	**0.444**	0.425	0.443	**0.424**	**0.442**	0.813	0.648	**0.647**	**0.571**
ETTh2	Avg	0.389	0.409	**0.386**	**0.407**	0.374	0.402	**0.372**	**0.401**	0.372	0.399	**0.371**	**0.398**	0.523	0.494	**0.473**	**0.465**
PEMS03	12	0.069	**0.176**	**0.068**	**0.176**	0.063	0.169	**0.061**	**0.164**	0.085	0.196	**0.074**	**0.185**	0.103	0.218	**0.074**	**0.182**
PEMS03	24	0.094	0.207	**0.092**	**0.204**	0.084	0.190	**0.083**	**0.188**	0.127	0.241	**0.106**	**0.223**	0.180	0.292	**0.105**	**0.213**
PEMS03	48	0.144	0.259	**0.141**	**0.255**	0.131	0.238	**0.118**	**0.224**	0.195	0.302	**0.152**	**0.265**	0.316	0.409	**0.155**	**0.257**
PEMS03	96	0.192	0.291	**0.190**	**0.290**	0.160	0.264	**0.158**	**0.257**	0.250	0.350	**0.205**	**0.303**	0.450	0.507	**0.199**	**0.292**
PEMS03	Avg	0.125	0.233	**0.123**	**0.231**	0.110	0.215	**0.105**	**0.208**	0.164	0.272	**0.134**	**0.244**	0.262	0.357	**0.133**	**0.236**
PEMS04	12	0.081	0.190	**0.077**	**0.184**	0.072	0.176	**0.070**	**0.171**	0.099	0.212	**0.088**	**0.200**	0.114	0.227	**0.087**	**0.198**
PEMS04	24	0.105	0.220	**0.099**	**0.213**	0.089	0.197	**0.084**	**0.187**	0.143	0.259	**0.118**	**0.234**	0.187	0.296	**0.111**	**0.222**
PEMS04	48	0.144	0.261	**0.137**	**0.255**	0.111	0.222	**0.097**	**0.206**	0.216	0.323	**0.163**	**0.281**	0.318	0.402	**0.153**	**0.262**
PEMS04	96	0.183	0.296	**0.171**	**0.287**	0.139	0.251	**0.109**	**0.221**	0.271	0.368	**0.200**	**0.311**	0.424	0.479	**0.189**	**0.293**
PEMS04	Avg	0.128	0.242	**0.121**	**0.235**	0.103	0.212	**0.090**	**0.196**	0.182	0.291	**0.142**	**0.257**	0.261	0.351	**0.135**	**0.244**

Dataset	P	TimeBridge		TimeBridge+CCTG		FilterNet		FilterNet+CCTG		Autoformer		Autoformer+CCTG		iTransformer		iTransformer+CCTG	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.383	0.389	**0.382**	**0.388**	0.377	0.397	**0.375**	**0.395**	0.376	0.423	**0.365**	**0.417**	0.386	**0.401**	**0.384**	**0.401**
ETTh1	192	0.436	0.422	**0.431**	**0.419**	0.444	0.434	**0.432**	**0.429**	0.438	0.450	**0.405**	**0.438**	0.438	**0.430**	**0.434**	**0.430**
ETTh1	336	0.484	0.450	**0.471**	**0.439**	0.515	0.468	**0.492**	**0.459**	0.544	0.515	**0.523**	**0.510**	0.478	**0.450**	**0.476**	0.451
ETTh1	720	0.480	0.471	**0.479**	**0.469**	0.527	0.504	**0.469**	**0.463**	0.537	0.530	**0.500**	**0.511**	0.475	**0.469**	**0.472**	**0.469**
ETTh1	Avg	0.446	0.433	**0.441**	**0.429**	0.466	0.451	**0.442**	**0.437**	0.474	0.480	**0.448**	**0.469**	0.444	**0.438**	**0.442**	**0.438**
ETTh2	96	0.289	0.334	**0.288**	**0.333**	0.293	0.343	**0.288**	**0.338**	0.264	0.346	**0.261**	**0.345**	0.301	**0.350**	**0.300**	**0.350**
ETTh2	192	0.366	0.384	**0.365**	**0.383**	0.370	0.394	**0.367**	**0.391**	0.307	0.373	**0.302**	**0.370**	0.373	0.397	**0.372**	**0.396**
ETTh2	336	0.415	0.421	**0.414**	**0.420**	0.421	0.435	**0.415**	**0.430**	0.458	0.466	**0.444**	**0.458**	**0.417**	**0.428**	**0.417**	**0.428**
ETTh2	720	**0.416**	**0.434**	0.421	0.436	**0.435**	**0.449**	0.436	0.450	0.445	0.469	**0.441**	**0.446**	**0.422**	**0.443**	**0.422**	**0.443**
ETTh2	Avg	**0.372**	**0.393**	**0.372**	**0.393**	0.380	0.405	**0.377**	**0.402**	0.369	0.414	**0.362**	**0.405**	**0.378**	0.405	**0.378**	**0.404**
ETTm1	96	0.312	0.339	**0.302**	**0.337**	0.318	0.358	**0.311**	**0.356**	0.429	0.442	**0.396**	**0.425**	0.337	0.369	**0.330**	**0.366**
ETTm1	192	0.370	0.371	**0.359**	**0.370**	0.362	0.384	**0.354**	**0.380**	0.643	0.627	**0.508**	**0.469**	0.380	0.391	**0.372**	**0.386**
ETTm1	336	0.400	**0.393**	**0.398**	0.394	0.394	0.405	**0.392**	**0.401**	0.653	0.544	**0.488**	**0.479**	0.414	0.422	**0.405**	**0.408**
ETTm1	720	0.472	0.435	**0.466**	**0.433**	**0.454**	0.441	0.455	**0.434**	0.710	0.567	**0.506**	**0.488**	0.482	0.451	**0.469**	**0.445**
ETTm1	Avg	0.389	0.385	**0.381**	**0.384**	0.382	0.397	**0.378**	**0.393**	0.609	0.545	**0.475**	**0.465**	0.403	0.408	**0.394**	**0.401**
ETTm2	96	0.175	0.254	**0.173**	**0.252**	0.175	0.257	**0.162**	**0.245**	0.231	0.315	**0.203**	**0.284**	0.184	0.269	**0.176**	**0.260**
ETTm2	192	0.240	0.298	**0.238**	**0.297**	0.240	0.301	**0.226**	**0.288**	0.284	0.339	**0.256**	**0.315**	0.254	0.314	**0.243**	**0.305**
ETTm2	336	0.302	0.338	**0.300**	**0.337**	0.295	0.337	**0.284**	**0.327**	0.338	0.370	**0.309**	**0.348**	0.319	0.354	**0.309**	**0.345**
ETTm2	720	0.402	0.395	**0.398**	**0.392**	0.392	0.394	**0.386**	**0.389**	0.433	0.420	**0.404**	**0.401**	0.419	0.410	**0.410**	**0.403**
ETTm2	Avg	0.280	0.321	**0.277**	**0.320**	0.276	0.322	**0.265**	**0.312**	0.322	0.361	**0.293**	**0.337**	0.294	0.337	**0.285**	**0.328**

In the controlled CycleNet replacement, CCTG improves MSE in 14 of 16 dataset–horizon settings and ties in two. For MAE, it improves 12 settings, ties two, and is slightly worse in two. Because RCF and CCTG share the same table length and capacity and coincide at integer coordinates, this comparison provides the most direct evidence for the combined continuous, channel-aligned access change under the tested CycleNet protocol. It does not separate interpolation from channel offsets, and optimization may still differ between the two access rules.

The broader experiment examines augmentation across recurrent-cycle, query-based, patch-based, linear, filtering, and Transformer backbones. CCTG improves MSE in 116 of 128 backbone–dataset–horizon comparisons (nine ties, three reversals) and MAE in 108 of 128 (12 ties, eight reversals), with improvements represented across all eight backbones. This provides broad compatibility evidence in the tested implementations, not a guarantee of architecture-independent improvement.

F.2 Relationship to Generic Statistical Non-Stationarity

Temporal pattern deformation describes a regime in which recurring temporal organization may be re-expressed through amplitude, phase, or relative cross-channel timing. We use the four ETT variants and Weather to examine whether NSF gains remain positive across subsets defined by broader statistical diagnostics. This is a scope and specificity check, not a proxy-label experiment. ADF and KPSS partition input-window–variable pairs by generic statistical properties but do not identify temporal pattern deformation itself.

Diagnostic criteria. ADF>0.05 indicates insufficient evidence to reject the unit-root null, while KPSS≤ 0.05 rejects the stationarity null. We report each criterion and their intersection. Section C.3.2 gives the full definitions.

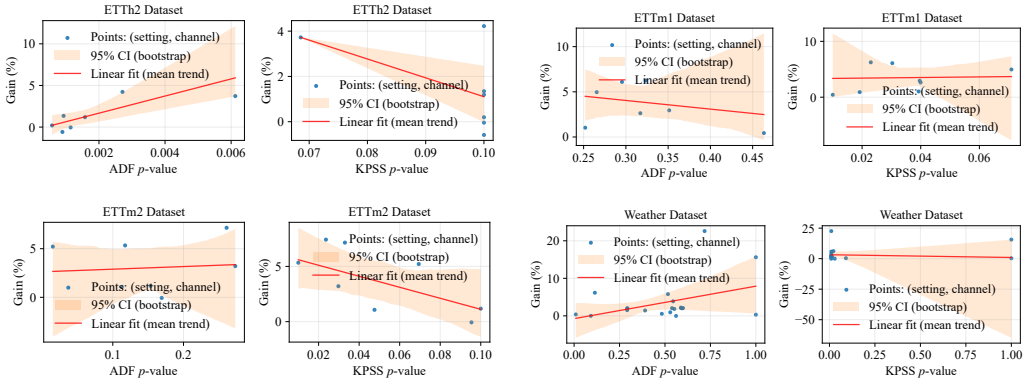
Gain. Let $\mathrm{MSE}^{(i)}_{\mathrm{base}}$ and $\mathrm{MSE}^{(i)}_{\mathrm{NSF}}$ denote errors without and with NSF on sample i. For subset $\mathcal{S}$,

$$\mathrm{Gain}(\mathcal{S}) = \frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} \frac{\mathrm{MSE}^{(i)}_{\mathrm{base}} - \mathrm{MSE}^{(i)}_{\mathrm{NSF}}}{\mathrm{MSE}^{(i)}_{\mathrm{base}}} \times 100\%. \tag{47}$$

Positive values indicate lower error with NSF.

Table 12: NSF gains on subsets diagnosed by each statistical non-stationarity criterion and on their complements. D and C denote the diagnosed subset and its complement, respectively. ADF∩KPSS requires both criteria.

	ADF>0.05			KPSS≤ 0.05			ADF∩KPSS		
Dataset	Diagnosed Ratio	Gain (D)	Gain (C)	Diagnosed Ratio	Gain (D)	Gain (C)	Diagnosed Ratio	Gain (D)	Gain (C)
ETTh1	17.83%	1.40%	4.72%	15.98%	0.64%	4.79%	9.71%	0.12%	4.56%
ETTh2	13.97%	2.62%	0.98%	15.89%	2.96%	0.73%	10.34%	3.87%	0.80%
ETTm1	87.97%	3.37%	4.29%	58.56%	3.43%	3.54%	54.10%	3.32%	3.67%
ETTm2	64.10%	3.96%	2.13%	51.49%	4.61%	1.91%	38.37%	5.09%	2.18%
Weather	80.37%	3.70%	3.13%	71.99%	4.22%	1.97%	60.66%	4.38%	2.32%



(a) (b)

Figure 7: Variable–setting-level associations between statistical non-stationarity diagnostics and NSF forecasting gains, with trend lines summarizing empirical rather than causal relationships.

Diagnostic–gain relationship. Table 12 reports positive NSF gains on every diagnosed subset and on every complement. Under ADF∩KPSS, diagnosed-subset gains exceed complement gains on

Table 13: Residual-MLP hidden dimensions used in each setting.

Horizon	96	192	336	720
ETTh1	64	256	256	64
ETTh2	128	128	32	32
ETTm1	128	256	256	256
ETTm2	128	128	256	256
Weather	2048	2048	2048	2048
Electricity	512	512	512	512
Solar	2048	2048	2048	2048
Horizon	**12**	**24**	**48**	**96**
PEMS03	1024	1024	1024	1024
PEMS04	1024	1024	1024	1024
PEMS07	1024	1024	1024	1024
PEMS08	1024	1024	1024	1024

ETTh2, ETTm2, and Weather, whereas ETTh1 and ETTm1 reverse that ordering. Fig. 7 likewise reports positive gains across broad ADF and KPSS ranges without a consistent monotonic association. The selected binary diagnostics therefore do not consistently order when NSF is more beneficial.

This result complements the post-hoc amplitude–phase stratification in Appendix E. ADF/KPSS grouping and amplitude–phase stratification describe different empirical partitions: the former captures broader statistical properties, whereas the latter measures changes between raw history–future pairs. Neither analysis identifies a unique latent data-generating mechanism.

G Robustness Analysis

These analyses ask whether the main result depends on one capacity setting or one context length. They complement, but do not replace, the matched component interventions in the main paper.

G.1 Configuration and Input-Length Sensitivity

Hidden dimension. Table 13 records the residual-MLP width used in each setting for reproducibility. The values are not a cross-dataset capacity ranking.

Input length. Fig. 8 varies Electricity context from 24 to 720. NSTime has the lowest MSE at every tested length, so its advantage is not explained solely by longer history. This experiment does not compare preferred lengths across datasets. The separation is visually largest at the shortest context ($S = 24$) and persists after the curves flatten from moderate to long contexts. This result shows that NSTime's Electricity performance is robust across the evaluated input lengths. The larger short-context gap is descriptive: without a controlled analysis of the number and quality of observed cycles, it does not establish that the deformation prior is specifically more useful when fewer cycles are available or that additional context cannot substitute for the structured path. Because the sweep covers only Electricity, no general context-length law is inferred. Likewise, Table 13 documents the capacities used but is not a controlled width sweep and should not be read as evidence that larger hidden dimensions cause better performance.

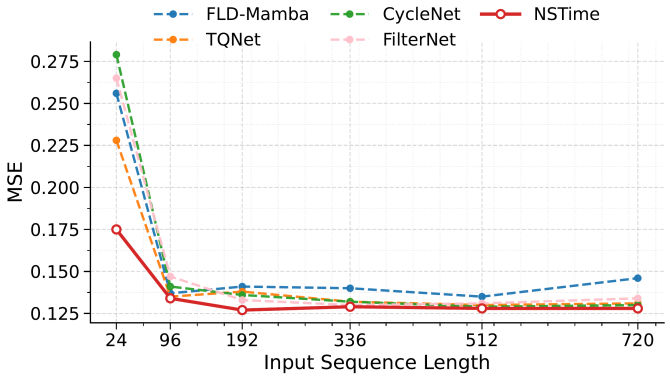


Figure 8: MSE on Electricity across input sequence lengths.

Table 14: Efficiency on ETTh1 ($S = 512$, batch size 1, RTX 4090). Latency is end-to-end time per batch.

Model	Params (M)	FLOPs (G)	Latency (ms)
DLinear	0.098	0.0014	0.165 ± 0.028
Autoformer	1.465	1.1020	16.74 ± 4.235
FEDformer	1.467	1.1040	29.183 ± 0.392
iTransformer	0.277	0.0060	1.028 ± 0.242
TimeBridge	0.542	0.1000	2.900 ± 0.690
CycleNet	0.049	0.0007	0.259 ± 0.023
TQNet	1.888	0.0117	0.765 ± 0.267
NSTime	0.181	0.0024	1.046 ± 0.354

H Efficiency, Stability, and Qualitative Analysis

We close with three complementary checks. They cover computational cost, multi-seed dispersion, and the AM–PM components generated at successive decomposition levels.

H.1 Efficiency Analysis

We measure ETTh1 inference with $S = 512$, batch size 1, and one NVIDIA RTX 4090. Latency uses CUDA events after 50 warm-up and 200 timed runs. Table 14 reports parameters, FLOPs, and latency.

NSTime uses 0.181M parameters and 0.0024G FLOPs. Its 1.046-ms latency is close to iTransformer and substantially below the attention-heavy Autoformer and FEDformer measurements, while remaining slower than DLinear and CycleNet. The deformation prior therefore adds a modest, nonzero runtime cost.

The comparison exposes the relevant efficiency trade-off. Relative to iTransformer, NSTime uses about 35% fewer parameters and 60% fewer FLOPs at essentially the same measured latency. Relative to Autoformer and FEDformer, it is more than an order of magnitude cheaper in both FLOPs and latency. The cost is not negligible: NSTime is roughly four times slower than CycleNet and over six times slower than DLinear in this batch-1 measurement. The supported conclusion is limited to this implementation, hardware, sequence length, and batch setting: the structured path remains lightweight relative to the tested attention-heavy baselines, but it is not a zero-cost addition.

H.2 Stability Analysis

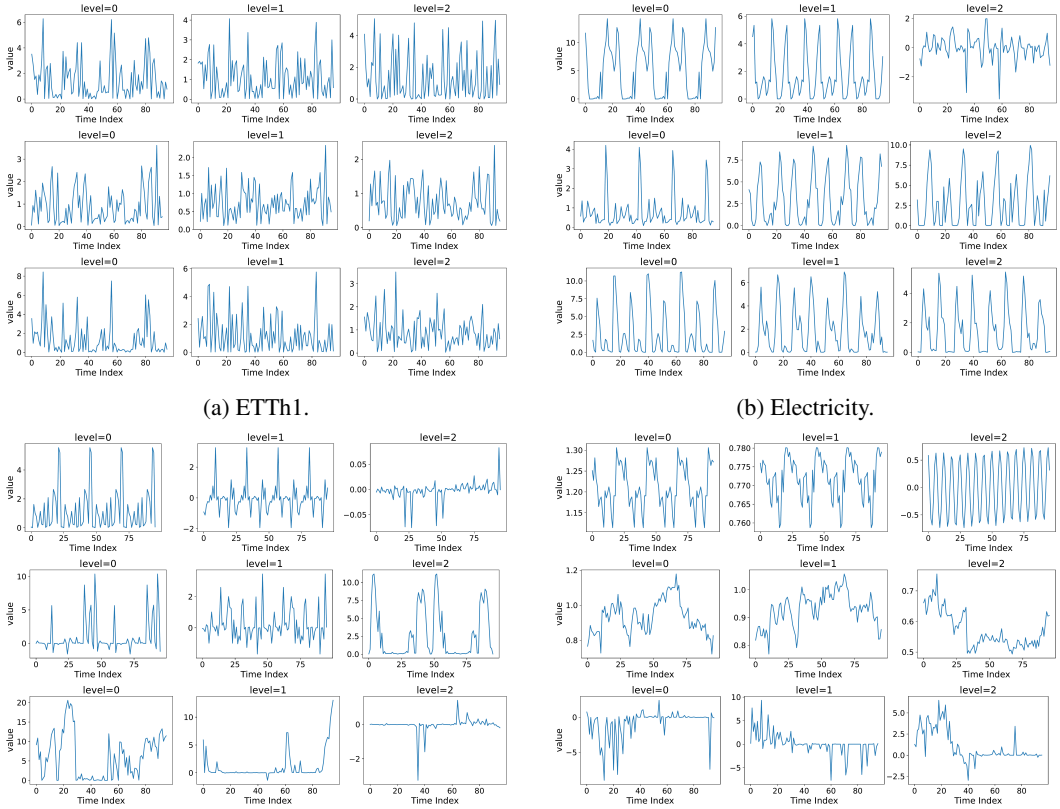
Table 15 reports three seeds. Seed 2024 matches the main comparison. Dispersion is small in most settings, with larger variation at several long horizons. We therefore treat the main table as a fixed-seed comparison and use Table 15 to expose run-to-run spread.

At the dataset-average level, NSTime's MSE standard deviations are at most 0.0050. The largest horizon-level MSE deviation is 0.0136 on PEMS08–96. Thus the reported NSTime averages have limited dispersion in most settings, with larger variation in several individual cells. This table does not provide multi-seed results for every baseline, so it cannot establish paired statistical significance for leaderboard differences or show that rankings are seed-invariant. Its narrower role is to expose NSTime's run-to-run spread and to identify cells for which rounded single-seed differences should be treated cautiously.

Table 15: Three-seed stability. The final columns report mean ± sample standard deviation.

		2024		2025		2026		mean ± std	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.363	0.388	0.361	0.388	0.361	0.386	0.3617 ± 0.0012	0.3873 ± 0.0012
	192	0.406	0.416	0.402	0.414	0.406	0.413	0.4047 ± 0.0023	0.4143 ± 0.0015
	336	0.450	0.438	0.444	0.437	0.448	0.438	0.4473 ± 0.0031	0.4377 ± 0.0006
	720	0.456	0.462	0.469	0.466	0.467	0.465	0.4640 ± 0.0070	0.4643 ± 0.0021
	Avg	0.419	0.426	0.419	0.426	0.421	0.426	0.4197 ± 0.0012	0.4260 ± 0.0000
ETTh2	96	0.281	0.337	0.278	0.334	0.280	0.334	0.2797 ± 0.0015	0.3350 ± 0.0017
	192	0.359	0.389	0.360	0.388	0.359	0.388	0.3593 ± 0.0006	0.3883 ± 0.0006
	336	0.418	0.430	0.411	0.428	0.409	0.428	0.4127 ± 0.0047	0.4287 ± 0.0012
	720	0.425	0.442	0.427	0.442	0.425	0.442	0.4257 ± 0.0012	0.4420 ± 0.0000
	Avg	0.371	0.400	0.369	0.398	0.368	0.401	0.3693 ± 0.0015	0.3997 ± 0.0015
ETTm1	96	0.303	0.352	0.305	0.358	0.297	0.348	0.3017 ± 0.0042	0.3527 ± 0.0050
	192	0.344	0.377	0.342	0.381	0.350	0.381	0.3453 ± 0.0042	0.3797 ± 0.0023
	336	0.379	0.402	0.379	0.397	0.372	0.396	0.3767 ± 0.0040	0.3983 ± 0.0032
	720	0.436	0.438	0.429	0.438	0.434	0.432	0.4330 ± 0.0036	0.4360 ± 0.0035
	Avg	0.366	0.392	0.364	0.393	0.363	0.389	0.3643 ± 0.0015	0.3913 ± 0.0021
ETTm2	96	0.165	0.249	0.164	0.247	0.164	0.247	0.1643 ± 0.0006	0.2477 ± 0.0012
	192	0.229	0.293	0.228	0.296	0.226	0.287	0.2277 ± 0.0015	0.2920 ± 0.0046
	336	0.288	0.331	0.285	0.329	0.290	0.334	0.2877 ± 0.0025	0.3313 ± 0.0025
	720	0.385	0.390	0.384	0.392	0.389	0.391	0.3860 ± 0.0026	0.3910 ± 0.0010
	Avg	0.267	0.316	0.265	0.316	0.267	0.315	0.2663 ± 0.0012	0.3157 ± 0.0006
Weather	96	0.156	0.202	0.158	0.205	0.150	0.198	0.1547 ± 0.0042	0.2017 ± 0.0035
	192	0.204	0.246	0.202	0.247	0.209	0.243	0.2050 ± 0.0036	0.2453 ± 0.0021
	336	0.261	0.289	0.256	0.292	0.262	0.292	0.2597 ± 0.0032	0.2910 ± 0.0017
	720	0.343	0.344	0.337	0.342	0.335	0.341	0.3383 ± 0.0042	0.3423 ± 0.0015
	Avg	0.241	0.270	0.238	0.271	0.239	0.268	0.2393 ± 0.0015	0.2697 ± 0.0015
Electricity	96	0.135	0.230	0.129	0.222	0.138	0.224	0.1340 ± 0.0046	0.2253 ± 0.0042
	192	0.152	0.245	0.155	0.248	0.154	0.248	0.1537 ± 0.0015	0.2470 ± 0.0017
	336	0.169	0.265	0.172	0.262	0.170	0.263	0.1703 ± 0.0015	0.2633 ± 0.0015
	720	0.210	0.301	0.207	0.307	0.213	0.301	0.2100 ± 0.0030	0.3030 ± 0.0035
	Avg	0.167	0.260	0.166	0.260	0.169	0.262	0.1673 ± 0.0015	0.2607 ± 0.0012
Solar	96	0.178	0.247	0.173	0.243	0.173	0.242	0.1747 ± 0.0029	0.2440 ± 0.0026
	192	0.198	0.256	0.198	0.246	0.193	0.253	0.1963 ± 0.0029	0.2517 ± 0.0051
	336	0.208	0.255	0.196	0.249	0.204	0.256	0.2027 ± 0.0061	0.2533 ± 0.0038
	720	0.214	0.257	0.213	0.256	0.214	0.257	0.2137 ± 0.0006	0.2567 ± 0.0006
	Avg	0.200	0.254	0.195	0.248	0.196	0.252	0.1970 ± 0.0026	0.2513 ± 0.0031
PEMS03	12	0.064	0.169	0.063	0.157	0.058	0.159	0.0617 ± 0.0032	0.1617 ± 0.0064
	24	0.083	0.194	0.079	0.180	0.079	0.183	0.0803 ± 0.0023	0.1857 ± 0.0074
	48	0.114	0.224	0.113	0.205	0.111	0.210	0.1127 ± 0.0015	0.2130 ± 0.0098
	96	0.156	0.261	0.147	0.251	0.147	0.247	0.1500 ± 0.0052	0.2530 ± 0.0072
	Avg	0.104	0.212	0.101	0.198	0.099	0.200	0.1013 ± 0.0025	0.2033 ± 0.0076
PEMS04	12	0.069	0.171	0.066	0.165	0.072	0.175	0.0690 ± 0.0030	0.1703 ± 0.0050
	24	0.083	0.190	0.088	0.193	0.084	0.184	0.0850 ± 0.0026	0.1890 ± 0.0046
	48	0.106	0.217	0.104	0.213	0.104	0.214	0.1047 ± 0.0012	0.2147 ± 0.0021
	96	0.132	0.244	0.129	0.236	0.130	0.229	0.1303 ± 0.0015	0.2363 ± 0.0075
	Avg	0.098	0.206	0.097	0.202	0.098	0.201	0.0977 ± 0.0006	0.2030 ± 0.0026
PEMS07	12	0.054	0.149	0.056	0.147	0.059	0.151	0.0563 ± 0.0025	0.1490 ± 0.0020
	24	0.069	0.169	0.072	0.166	0.067	0.172	0.0693 ± 0.0025	0.1690 ± 0.0030
	48	0.097	0.196	0.090	0.198	0.089	0.197	0.0920 ± 0.0044	0.1970 ± 0.0010
	96	0.130	0.237	0.130	0.223	0.133	0.228	0.1310 ± 0.0017	0.2293 ± 0.0071
	Avg	0.088	0.188	0.087	0.184	0.087	0.187	0.0873 ± 0.0006	0.1863 ± 0.0021
PEMS08	12	0.077	0.172	0.079	0.167	0.072	0.178	0.0760 ± 0.0036	0.1723 ± 0.0055
	24	0.099	0.197	0.102	0.193	0.098	0.194	0.0997 ± 0.0021	0.1947 ± 0.0021
	48	0.141	0.233	0.134	0.218	0.134	0.224	0.1363 ± 0.0040	0.2250 ± 0.0075
	96	0.185	0.261	0.162	0.242	0.161	0.254	0.1693 ± 0.0136	0.2523 ± 0.0096
	Avg	0.126	0.216	0.120	0.205	0.116	0.213	0.1207 ± 0.0050	0.2113 ± 0.0057

H.3 VISUALIZATION



(a) ETTh1. (b) Electricity.

(c) ETTm1. (d) Weather.

Figure 9: Components produced by the NSF structured path at three decomposition levels on four representative datasets. The selected examples show recurring-looking variation together with changes in envelope and timing across decomposition levels.

Fig. 9 visualizes components assigned to the structured NSF path at three decomposition levels. In the selected examples, several components contain recurring-looking motifs together with changes in envelope, timing, and variation across decomposition levels. In these examples, earlier levels show relatively faster local fluctuations, whereas later levels show relatively slower variation.