FIRE: FOURIER-INFORMED REGIME ADAPTIVE ESTIMATION FOR NON-STATIONARY TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Non-stationarity remains a critical challenge in multivariate time series forecasting because data distributions and temporal dependencies continuously evolve over time. Recent frequency-domain models have demonstrated the value of spectral representations, yet most of them jointly model amplitude and phase within unified complex embeddings, overlooking their distinct physical roles and evolutionary behaviors. From a signal processing perspective, amplitude characterizes the strength of frequency components, whereas phase encodes their temporal alignment. These distinct physical roles and transformation properties motivate dedicated modeling of amplitude and phase under non-stationarity. We propose **FIRE** (Fourier Informed Regime Adaptive Estimation), a forecasting framework that explicitly decouples amplitude and phase into dedicated learning branches. To adapt these branches to evolving dynamics, FIRE introduces a signal-informed modulation mechanism driven by two historical window statistics: the Amplitude Non-stationarity Index (ANI) and the Phase Irregularity Index (PII). Across 12 diverse real-world benchmarks, FIRE achieves state-of-the-art forecasting performance against competitive baselines, securing the lowest average MSE and MAE on 9 and 10 datasets, respectively. Controlled ablations demonstrate the benefits of amplitude–phase decoupling and ANI/PII guided adaptive modulation, while channel-level analyses show larger forecasting gains for the evaluated channels with higher ANI and PII.

1 INTRODUCTION

Time series forecasting supports critical decision-making across diverse domains, including energy systems, transportation, finance, weather forecasting, and industrial operations. Despite significant progress, reliable forecasting remains challenging due to data non-stationarity, where the statistical properties and underlying distribution of time series continuously evolve over time.

To address this issue, recent research has explored several complementary modeling paradigms. Normalization and stationarization techniques aim to mitigate distribution shifts (Kim et al., 2022; Liu et al., 2022c), while time-domain architectures leverage decomposition-based linear models (Zeng et al., 2023), MLP mixers (Wang et al., 2024; Lu et al., 2024; Lin et al., 2024a), temporal convolutions (Wang et al., 2023; Luo & Wang, 2024), and Transformers (Zhou et al., 2021; Nie et al., 2023; Zhang & Yan, 2023; Liu et al., 2024) to capture temporal and multivariate dependencies. Concurrently, frequency-domain and hybrid methods offer compelling alternatives by transforming sequence patches into spectral representations (Yi et al., 2023b; Xu et al., 2024; Piao et al., 2024; Wang et al., 2025a; Zhou et al., 2022b; Wang et al., 2025b). Most Fourier-based forecasting models process complex spectra through unified representations, such as complex MLPs or real–imaginary mappings. Recent studies have also explored amplitude and phase information for frequency-domain modeling and non-stationarity adaptation (Zhuang et al., 2024; Hu et al., 2026). However, the distinct evolution of amplitude and phase under changing data distributions remains less explicitly modeled in forecasting architectures.

From a classical signal processing perspective, amplitude and phase inherently govern different dynamics and transformation laws: multiplicative energy fluctuations versus linear temporal shifts

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

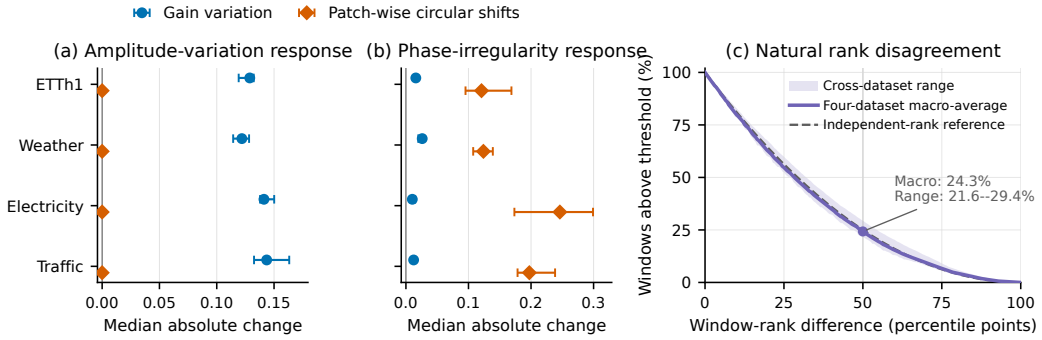

Figure 1: Empirical analysis of amplitude and phase variations. (a)–(b) Responses of diagnostic measures for amplitude and phase under patch-wise gain variation and circular shifts, evaluated over 7,141 training windows from four datasets. Error bars denote 95% bootstrap intervals across channels. (c) Distribution of the percentile rank difference between amplitude and phase variations within each channel over 29,566 unperturbed training windows. The solid line denotes the macro average across four datasets, the shaded area denotes the dataset range, and the dashed line denotes the independent rank reference. Details of the analysis are provided in Appendix A.3

(Oppenheim & Lim, 1981; Oppenheim & Schafer, 1989). Specifically, positive rescaling alters spectral amplitude while preserving phase, whereas temporal shifts leave amplitude invariant while inducing frequency-dependent phase rotations (Foroosh et al., 2002). Furthermore, their underlying mathematical properties are fundamentally mismatched: spectral amplitude is non-negative and describes signal strength, whereas phase is strictly periodic (bounded within $[-\pi, \pi)$) and encodes temporal alignment. By modeling these geometrically and physically distinct components within unified hidden states, standard complex-valued layers use a shared representation to capture their different transformation behaviors.

To examine whether these distinct behaviors also appear in data, we conduct a diagnostic analysis on four benchmark datasets, including ETTh1 (Zhou et al., 2021), Weather, Electricity, and Traffic (Wu et al., 2021). We use 7,141 training windows for controlled perturbations and 29,566 unperturbed training windows to analyze natural variations. As shown in Figure 1(a) and (b), amplitude and phase respond differently to the controlled perturbations. Gain variation applied to local patches produces a stronger amplitude response, while circular shifts produce a stronger phase response. On unperturbed sequences, Figure 1(c) further shows substantial rank disagreement between amplitude and phase variations within each channel. Windows with stronger amplitude variation do not necessarily exhibit stronger phase variation. Together, these observations suggest that amplitude and phase exhibit distinct variation patterns in real data, motivating their dedicated and decoupled modeling.

Motivated by these insights and empirical findings, we propose **FIRE** (**F**ourier **I**nformed **R**egime adaptive **E**stimation), a frequency-domain framework explicitly tailored for non-stationary forecasting. FIRE decouples spectral amplitude and phase into dedicated branches to avoid forcing their distinct dynamics into a shared modeling framework. Driven by two novel metrics, the *Amplitude Non-stationarity Index* (ANI) and *Phase Irregularity Index* (PII), FIRE dynamically modulates branch representations according to local spectral regimes before recombining them into complex spectra for final prediction.

Our main contributions are summarized as follows:

- We propose FIRE, a novel frequency-domain forecasting framework that explicitly separates spectral amplitude and phase into dedicated branches. This decoupled design aligns the model architecture with their distinct physical dynamics and avoids forcing both components into a shared representation.
- We introduce ANI and PII to respectively characterize energy non-stationarity and phase progression irregularities. These indices drive a branch-specific modulation mechanism, enabling adaptive feature recalibration.

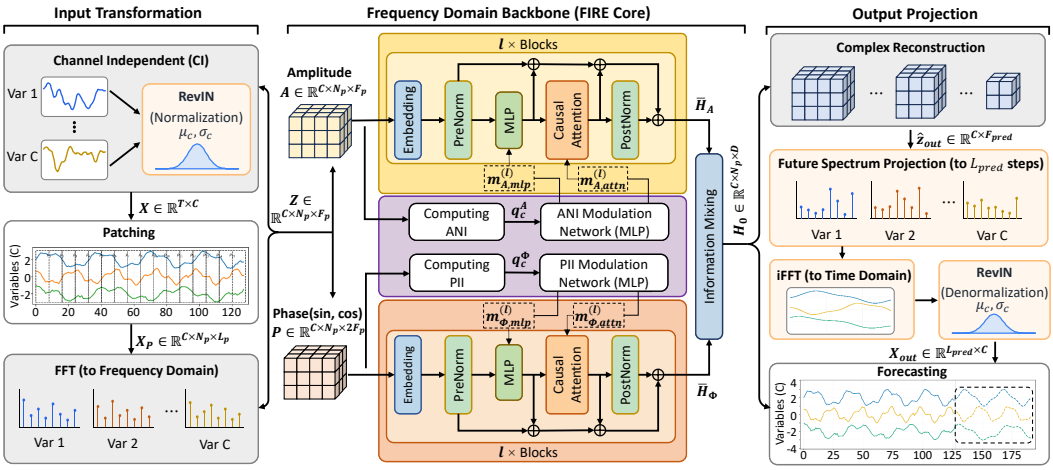

Figure 2: Overview of the FIRE architecture. **Left:** Sequence patches are normalized with RevIN and transformed into spectra via FFT. **Middle:** Amplitude and phase are modeled in dedicated branches, adaptively modulated by historical ANI and PII statistics. **Right:** After cross-variable attention, representations are recombined into complex spectra to generate the final forecasts via iFFT and denormalization.

- We evaluate FIRE across diverse benchmarks, consistently achieving state-of-the-art forecasting performance. Importantly, our diagnostic analysis demonstrates that FIRE's predictive advantage scales progressively with higher non-stationarity, validating the effectiveness of the proposed adaptive spectral modeling.

2 Method

2.1 Design Overview

Given an input series $\mathbf{X} \in \mathbb{R}^{T \times C}$ with C variables and length T, FIRE forecasts future horizons through three main components (Figure 2):

- **Input Transformation:** Channel-independently normalizes the input, extracts overlapping patches, and transforms them into frequency spectra via FFT.
- **Frequency-domain backbone:** Models amplitude and phase in separate branches, adaptively modulating them via ANI and PII to capture cross-patch energy fluctuations and phase irregularities.
- **Output projection:** Aggregates cross-variable information to predict the future spectrum, subsequently applying an inverse FFT and instance denormalization to generate the final time-domain output.

2.2 Input Transformation

FIRE processes each variable independently and successively applies normalization, patching, and frequency transformation:

$$\begin{aligned} \mathbf{X}_P &= \text{Patching}(\text{RevIN}(\mathbf{X})) \in \mathbb{R}^{C \times N_p \times L_p}, \\ \mathbf{Z} &= \text{FFT}(\mathbf{X}_P) \in \mathbb{C}^{C \times N_p \times F_p}. \end{aligned} \tag{1}$$

Here, N_p is the number of patches, L_p is the patch length, and $F_p = \lfloor L_p/2 \rfloor + 1$ is the number of one-sided frequency bins.

2.3 FREQUENCY-DOMAIN BACKBONE

The frequency-domain backbone implements FIRE's core design: decoupled modeling of amplitude and phase, with adaptive modulation driven by local spectral dynamics.

2.3.1 SEPARATE AMPLITUDE AND PHASE MODELING

For the patch spectrum in Equation (1), amplitude is obtained as $\mathbf{A} = |\mathbf{Z}| \in \mathbb{R}^{C\times N_p\times F_p}$, capturing non-negative frequency strength. The corresponding phase is denoted by $\mathbf{\Phi} = \mathrm{angle}(\mathbf{Z}) \in [-\pi, \pi)^{C\times N_p\times F_p}$, encoding temporal alignment. Instead of feeding the wrapped angles directly into the model, FIRE uses their cosine and sine coordinates:

$$\mathbf{P} = [\cos\mathbf{\Phi}, \sin\mathbf{\Phi}] \in \mathbb{R}^{C\times N_p\times 2F_p}. \tag{2}$$

This representation preserves phase periodicity and avoids discontinuities across the $\pm\pi$ wrap-around boundary. For coefficients where amplitude falls below a small threshold ϵ_Φ, the phase is ill-conditioned and dominated by noise. Their coordinates are replaced with the fixed reference direction $[1, 0]$ to eliminate numerical instability caused by arbitrary phase rotations.

Amplitude and phase coordinates are projected into hidden spaces via linear embeddings:

$$\mathbf{H}_A^{(0)} = \mathbf{A}\mathbf{W}_A + \mathbf{b}_A \in \mathbb{R}^{C\times N_p\times D}, \tag{3}$$

$$\mathbf{H}_\Phi^{(0)} = \mathbf{P}\mathbf{W}_\Phi + \mathbf{b}_\Phi \in \mathbb{R}^{C\times N_p\times D}, \tag{4}$$

where $\mathbf{W}_A \in \mathbb{R}^{F_p\times D}$, $\mathbf{W}_\Phi \in \mathbb{R}^{2F_p\times D}$, and $\mathbf{b}_A, \mathbf{b}_\Phi \in \mathbb{R}^D$ are independent learnable parameters. Each branch contains L_b stacked blocks. For $r \in \{A, \Phi\}$, the $(\ell+1)$-th block update is formulated as:

$$\mathbf{H}_r^{(\ell+1)} = \mathrm{Block}_r^{(\ell)}\left(\mathbf{H}_r^{(\ell)}; \mathbf{m}_{r,\mathrm{mlp}}^{(\ell)}, \mathbf{m}_{r,\mathrm{attn}}^{(\ell)}\right) \in \mathbb{R}^{C\times N_p\times D}. \tag{5}$$

Within each block, an MLP transforms the feature dimension and causal self-attention models dependencies across ordered patches under a causal mask ($q \le p$). Inspired by the parallel-serial design of Branchformer (Peng et al., 2022), each block learns an interpolation between serial and parallel information flows between the MLP and attention sublayers, accompanied by LayerNorm and residual connections. While the two branches share this block topology, their parameters remain strictly disjoint, ensuring completely decoupled representations throughout the backbone.

2.3.2 SIGNAL-INFORMED ADAPTIVE MODULATION

Standard static networks struggle to accommodate sequences exhibiting different non-stationary regimes under fixed weights. To address this limitation, FIRE derives summary statistics from the observed patch spectra to dynamically adapt the MLP and attention representations in each branch.

For amplitude, the spectral strength of each valid patch is $s_{c,p} = \|\mathbf{A}_{c,p,:}\|_2 \in \mathbb{R}$, which quantifies the overall patch-level spectral energy. The ANI measures its relative variation across the $N_v = N_p - 1$ non-padded patches:

$$\mathrm{ANI}_c = \frac{\sqrt{\frac{1}{N_v}\sum_{p=1}^{N_v}(s_{c,p} - \bar{s}_c)^2}}{\max(\bar{s}_c, \epsilon_A)} \in \mathbb{R}, \tag{6}$$

where $\bar{s}_c = N_v^{-1}\sum_{p=1}^{N_v} s_{c,p}$ is the mean patch energy. A larger ANI reflects more severe cross-patch energy non-stationarity.

For phase, we consider adjacent-patch pairs whose amplitudes both exceed ϵ_Φ. Let $\Delta\Phi_{i,k} \in [-\pi, \pi)$ denote the principal phase increment of the i-th valid pair at the k-th retained frequency, and μ_k is the circular mean across all valid pairs. PII is defined as

$$\mathrm{PII}_c = \frac{1}{K}\sum_{k=1}^{K}\left\{\frac{1}{V_k}\sum_{i=1}^{V_k}\left[\mathrm{wrap}(\Delta\Phi_{i,k} - \mu_k)\right]^2\right\}^{1/2}, \tag{7}$$

where K is the number of frequencies with valid pairs, V_k is the number of valid pairs at frequency k, and $\mathrm{wrap}(\cdot)$ maps phase residuals to $[-\pi, \pi)$. A higher PII indicates greater irregularity in inter-patch phase evolution.

ANI and PII are mapped to the numerical ranges used by the two modulation networks:

$$q_c^A = \log(1 + \mathrm{ANI}_c), \qquad q_c^\Phi = \mathrm{PII}_c \,/\pi, \tag{8}$$

where $q^A, q^\Phi \in \mathbb{R}^C$. For branch $r \in \{A, \Phi\}$ and sublayer $u \in \{\mathrm{mlp}, \mathrm{attn}\}$, FIRE computes a positive feature-wise modulation vector $\mathbf{m}_{r,u}^{(\ell)}$ through MLP embedding. This modulation vector is broadcast over the patch dimension and multiplies the corresponding MLP or attention output before residual addition.

2.4 Output Projection

Before prediction, FIRE incorporates cross-variable information mixing into the amplitude and phase representations via branch-specific attention. The refined representations $\overline{\mathbf{H}}_A$ and $\overline{\mathbf{H}}_\Phi \in \mathbb{R}^{C \times N_p \times D}$ are mapped to positive amplitudes and unit-circle phase coordinates, respectively, and recombined into complex hidden features $\mathbf{H}_o \in \mathbb{C}^{C \times N_p \times D}$. An MLP-based projection head aggregates these representations across historical patches to predict the future spectrum:

$$\widehat{\mathbf{Z}}_{\mathrm{out}} = \mathrm{MLP}_{\mathrm{proj}}(\mathrm{Flat}(\mathbf{H}_o)) \in \mathbb{C}^{C \times F_{\mathrm{pred}}}, \tag{9}$$

where $F_{\mathrm{pred}} = \lfloor L_{\mathrm{pred}}/2 \rfloor + 1$ is the number of frequency bins in the predicted spectrum. Finally, the predicted spectrum is transformed back to the time domain via inverse FFT, followed by denormalization:

$$\mathbf{X}_{\mathrm{out}} = \mathrm{RevIN}^{-1}\left(\mathrm{iFFT}\left(\widehat{\mathbf{Z}}_{\mathrm{out}}, n = L_{\mathrm{pred}}\right)\right) \in \mathbb{R}^{L_{\mathrm{pred}} \times C}. \tag{10}$$

2.5 Loss Function

FIRE is trained with a composite loss balancing the time-domain prediction $\mathbf{X}_{\mathrm{out}}$ against the ground truth $\mathbf{X}_{\mathrm{true}} \in \mathbb{R}^{L_{\mathrm{pred}} \times C}$. It combines the Huber loss $\mathcal{L}_{\mathrm{Huber}}$ in the time domain with the frequency-domain loss $\mathcal{L}_{\mathrm{FFT}}$ adopted from FreDF (Wang et al., 2025a). The former penalizes point-wise forecasting errors while providing robustness against outliers, whereas the latter measures the mean absolute difference between the complex Fourier coefficients of the predicted and ground truth sequences. The overall objective is:

$$\mathcal{L} = \alpha \mathcal{L}_{\mathrm{Huber}} + \beta \mathcal{L}_{\mathrm{FFT}}, \tag{11}$$

where α and β control the relative weights of the time and frequency domain terms, respectively.

3 Experiments

We evaluate FIRE across 7 established forecasting benchmarks, comparing against SOTA models with an emphasis on frequency-domain baselines. Beyond standard benchmarks, we examine FIRE's broader applicability on extended datasets from diverse domains. To isolate our core architectural contributions, we provide detailed ablation and efficiency analyses. We further examine how the learned modulation responds to ANI/PII and how FIRE's gains vary across channels with different index levels. Additional results, including hyperparameter sensitivity, are detailed in Appendix A.5.

3.1 Datasets and Baselines

Datasets. We use 7 widely used multivariate long-term forecasting benchmarks: ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, and Traffic. To examine its applicability beyond these established benchmarks, we further consider 5 extended datasets: the Photovoltaic (PV) dataset (Pedro et al., 2019), the Leading to Cash (L2C) dataset (Saha et al., 2024), and three industry-level stock datasets constructed from HS 300 constituents in FinMultiTime (Xu et al., 2025), covering Media, Electronics, and Non-bank Financial sectors. Dataset statistics are provided in Appendix A.2.

Baselines.

We select representative models across major forecasting paradigms. Time-domain baselines include iTransformer (Liu et al., 2024) and TimeMixer (Wang et al., 2024). Given our spectral focus, we evaluate FreTS (Yi et al., 2023b), FITS (Xu et al., 2024), Fredformer (Piao et al., 2024),

Table 1: Full long-term forecasting results on seven widely used datasets. The input length is 96. Lower values are better. Best and second-best results are shown in **bold** and underlined, respectively. *Avg.* averages the four prediction horizons;

Data	L_{pred}	**FIRE**		DistDF		Fredformer		APDNet		WPMixer		TimeMixer		iTransformer		FITS		FreTS		TimesNet	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.371**	**0.391**	0.373	0.393	0.373	0.392	0.376	**0.391**	0.375	0.393	0.375	0.400	0.386	0.405	0.384	0.394	0.392	0.406	0.384	0.402
	192	**0.415**	0.419	0.428	0.425	0.433	0.420	0.431	0.422	0.428	**0.417**	0.429	0.421	0.441	0.436	0.434	0.421	0.449	0.443	0.436	0.429
	336	**0.456**	0.439	0.466	0.445	0.470	**0.437**	0.477	0.445	0.477	0.439	0.484	0.458	0.487	0.458	0.474	0.444	0.499	0.470	0.638	0.469
	720	0.456	0.458	**0.453**	**0.453**	0.467	0.456	0.488	0.472	0.460	0.454	0.498	0.482	0.503	0.491	0.462	0.460	0.554	0.531	0.521	0.500
	Avg.	**0.424**	0.427	0.430	0.429	0.436	**0.426**	0.443	0.433	0.435	**0.426**	0.447	0.440	0.454	0.447	0.438	0.430	0.474	0.462	0.495	0.450
ETTh2	96	0.285	**0.334**	0.287	0.336	0.293	0.342	0.290	0.340	**0.283**	0.335	0.289	0.341	0.297	0.349	0.289	0.338	0.329	0.370	0.340	0.374
	192	**0.356**	0.382	0.358	**0.381**	0.371	0.389	0.369	0.390	0.364	0.391	0.372	0.392	0.380	0.400	0.375	0.390	0.426	0.441	0.402	0.414
	336	0.397	0.418	0.408	0.421	**0.382**	**0.409**	0.411	0.424	0.409	0.424	0.386	0.414	0.428	0.432	0.414	0.424	0.456	0.464	0.452	0.452
	720	**0.411**	**0.434**	0.416	0.435	0.415	**0.434**	0.421	0.439	0.429	0.443	0.412	**0.434**	0.427	0.445	0.416	0.435	0.587	0.545	0.462	0.468
	Avg.	**0.362**	**0.392**	0.367	0.393	0.365	0.394	0.373	0.398	0.371	0.398	0.364	0.395	0.383	0.407	0.374	0.397	0.449	0.455	0.414	0.427
ETTm1	96	**0.304**	**0.346**	0.316	0.357	0.326	0.361	0.315	0.354	0.316	0.352	0.320	0.357	0.334	0.368	0.353	0.374	0.341	0.377	0.338	0.375
	192	**0.356**	**0.375**	0.358	0.380	0.363	0.380	0.360	0.378	0.362	0.376	0.361	0.381	0.377	0.391	0.390	0.392	0.382	0.402	0.374	0.387
	336	**0.385**	**0.396**	0.392	0.404	0.395	0.403	0.392	0.401	0.387	**0.396**	0.390	0.404	0.426	0.420	0.422	0.412	0.416	0.422	0.410	0.411
	720	**0.440**	**0.431**	0.448	0.437	0.453	0.438	0.456	0.439	0.447	0.432	0.454	0.441	0.491	0.459	0.485	0.448	0.479	0.456	0.478	0.450
	Avg.	**0.371**	**0.387**	0.378	0.394	0.384	0.396	0.381	0.393	0.378	0.389	0.381	0.395	0.407	0.410	0.413	0.406	0.405	0.414	0.400	0.406
ETTm2	96	**0.171**	**0.252**	0.174	0.256	0.177	0.259	**0.171**	0.253	**0.171**	**0.252**	0.175	0.258	0.180	0.264	0.182	0.265	0.195	0.283	0.187	0.267
	192	0.236	0.296	0.239	0.298	0.243	0.301	0.235	0.295	**0.233**	**0.294**	0.237	0.299	0.250	0.309	0.246	0.304	0.270	0.341	0.249	0.309
	336	0.301	0.335	0.300	0.338	0.302	0.340	0.292	**0.333**	**0.290**	**0.333**	0.298	0.340	0.311	0.348	0.306	0.342	0.335	0.375	0.321	0.351
	720	0.398	0.392	0.397	0.394	0.397	0.396	0.388	**0.389**	**0.387**	0.390	0.391	0.396	0.412	0.407	0.407	0.397	0.550	0.507	0.408	0.403
	Avg.	0.276	0.319	0.277	0.321	0.280	0.324	0.272	0.318	**0.270**	**0.317**	0.275	0.323	0.288	0.332	0.285	0.327	0.337	0.376	0.291	0.333
Weather	96	**0.153**	**0.192**	0.164	0.209	0.163	0.207	0.159	0.205	0.162	0.204	0.163	0.209	0.174	0.214	0.163	0.208	0.170	0.233	0.172	0.220
	192	**0.206**	**0.243**	0.212	0.252	0.211	0.251	0.207	0.250	0.209	0.246	0.208	0.250	0.221	0.254	0.210	0.251	0.212	0.269	0.219	0.261
	336	0.263	**0.285**	0.270	0.295	0.267	0.292	0.263	0.291	0.263	0.287	0.251	0.287	0.278	0.296	0.264	0.290	0.263	0.307	**0.246**	0.337
	720	0.342	**0.338**	0.348	0.345	0.343	0.341	0.345	0.343	0.340	0.339	0.339	0.341	0.358	0.347	0.345	0.342	**0.324**	0.350	0.365	0.359
	Avg.	0.241	**0.265**	0.248	0.275	0.246	0.273	0.244	0.272	0.244	0.269	**0.240**	0.271	0.258	0.278	0.246	0.273	0.242	0.290	0.251	0.294
Traffic	96	0.405	**0.249**	**0.380**	0.262	0.406	0.277	**0.380**	0.265	0.465	0.286	0.462	0.285	0.395	0.268	0.649	0.387	0.498	0.310	0.593	0.321
	192	0.417	**0.260**	**0.407**	0.275	0.426	0.290	0.414	0.276	0.475	0.290	0.473	0.296	0.417	0.276	0.601	0.363	0.495	0.315	0.617	0.336
	336	0.432	**0.268**	**0.429**	0.284	0.432	0.281	0.432	0.283	0.489	0.296	0.498	0.296	0.433	0.283	0.608	0.365	0.529	0.324	0.629	0.336
	720	0.478	**0.286**	**0.452**	0.297	0.463	0.300	0.461	0.300	0.527	0.318	0.506	0.313	0.467	0.302	0.646	0.385	0.572	0.343	0.640	0.350
	Avg.	0.433	**0.266**	**0.417**	0.279	0.432	0.287	0.422	0.281	0.489	0.298	0.484	0.297	0.428	0.282	0.626	0.375	0.523	0.323	0.620	0.336
Electricity	96	**0.136**	**0.230**	0.137	0.235	0.147	0.241	0.141	0.239	0.150	0.241	0.153	0.247	0.148	0.240	0.198	0.275	0.164	0.257	0.168	0.272
	192	**0.155**	**0.246**	0.159	0.257	0.165	0.258	0.158	0.252	0.162	0.252	0.166	0.256	0.162	0.253	0.198	0.278	0.173	0.266	0.184	0.322
	336	**0.166**	**0.260**	0.178	0.272	0.177	0.273	0.174	0.271	0.179	0.270	0.185	0.277	0.178	0.269	0.213	0.293	0.192	0.284	0.198	0.300
	720	**0.193**	**0.284**	0.212	0.302	0.213	0.304	0.207	0.302	0.217	0.304	0.225	0.310	0.225	0.317	0.254	0.325	0.230	0.317	0.220	0.320
	Avg.	**0.162**	**0.255**	0.172	0.267	0.176	0.269	0.170	0.266	0.177	0.267	0.182	0.272	0.178	0.270	0.216	0.293	0.190	0.281	0.193	0.304
Best Count		**20**	**25**	6	2	1	4	2	3	6	7	1	1	0	0	0	0	1	0	1	0

and APDNet (Zhuang et al., 2024). We also include hybrid architectures TimesNet (Wu et al., 2023) and WPMixer (Murad et al., 2025), as well as DistDF (Wang et al., 2026a), a model-agnostic distribution-alignment training framework. Implementation details and configurations are provided in Appendix A.4.

3.2 Forecasting Results on Widely Used Datasets

Table 1 summarizes the long-term forecasting performance across 7 widely used benchmarks. Overall, **FIRE** achieves the best performance in **20 out of 35** cases for MSE and **25 out of 35** cases for MAE (surpassing the second-best model, WPMixer, which attains 6 and 7 best counts, respectively). The performance margin is especially noticeable on complex, large-scale systems such as Electricity, where FIRE reduces the average MSE and MAE by **4.71%** and **4.14%** over the competing baseline (APDNet). Remarkably, at the longest horizon ($L_{\text{pred}} = 720$), it achieves up to a **6.76%** relative MSE reduction. The gains vary across datasets and metrics, motivating further analysis of their relationship with the spectral non-stationarity using ANI and PII in Section 3.5.

Table 2: Average forecasting results on PV and L2C over $L_{\text{pred}} \in \{60, 240, 720\}$.

Data	**FIRE**		Fredformer		WPMixer	
	MSE	MAE	MSE	MAE	MSE	MAE
PV	**0.433**	**0.416**	0.515	0.458	0.518	0.450
L2C	**0.797**	**0.532**	0.895	0.570	0.882	0.566

Table 3: Average forecasting results on the HS 300 industry datasets over $L_{\text{pred}} \in \{5, 10, 20\}$.

Data	**FIRE**		Fredformer		WPMixer	
	MSE	MAE	MSE	MAE	MSE	MAE
Media	**0.0166**	**0.0756**	0.0174	0.0789	0.0173	0.0785
ELEC	**0.0235**	**0.0636**	0.0241	0.0671	0.0240	0.0655
NBF	**0.0374**	**0.0768**	0.0381	0.0798	0.0390	0.0810

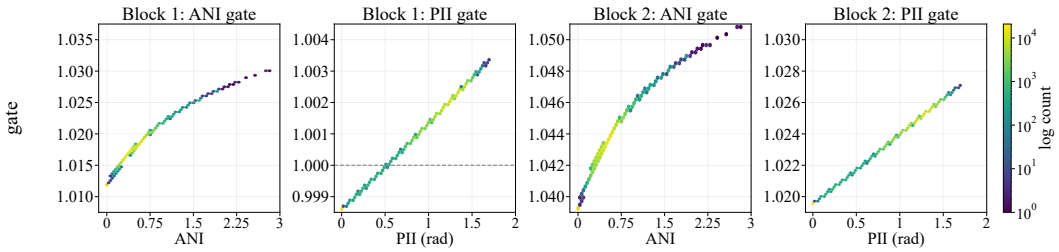

Figure 3: Learned ANI/PII-conditioned modulation responses on Weather ($L_{\text{pred}} = 720$). From left to right, panels display the branch-specific modulation gates conditioned on ANI and PII in Blocks 1 and 2, obtained by averaging the MLP and attention modulation vectors across sublayers and hidden dimensions. Colors denote sample counts on a logarithmic scale, and horizontal dashed lines mark the unscaled modulation baseline of one.

3.3 Forecasting Results on Extended Datasets

To evaluate FIRE beyond widely used forecasting benchmarks, we compare it with the reproduced competitive baselines Fredformer and WPMixer across five extended datasets. Tables 2 and 3 report the results averaged over three dataset-specific prediction horizons. Compared with SOTA models, FIRE reduces MSE by 15.92% and 9.64% on PV and L2C, respectively, with corresponding MAE reductions of 7.56% and 6.01%. It also yields consistent improvements across the three industry-level financial datasets, lowering MSE by 1.84%–4.05% and MAE by 2.90%–3.76%. patch wiserate FIRE's robust cross-domain generalization and adaptability. Full results are provided in Appendix A.5.1.

3.4 Ablation of Core Components

To isolate FIRE's core components, we evaluate two progressive variants in Table 4: *w/o Mod.* (removes ANI/PII-guided modulation while retaining separate branches) and *Joint Branch* (further merges the separated branches into a unified complex representation). The average results demonstrate that decoupling amplitude and phase provides the primary structural gain, with *w/o Mod.* outperforming *Joint Branch* on six of seven datasets in both MSE and MAE. Building upon this decoupled backbone, FIRE further improves all 14 dataset metric averages, confirming that ANI/PII-guided adaptive modulation yields substantial complementary gains. Full results and composite loss ablations are detailed in Appendices A.5.2 and A.5.3.

3.5 Analysis of ANI/PII-Guided Modulation

Learned modulation responses. Figure 3 visualizes the average modulation behavior on Weather ($L_{\text{pred}} = 720$). Conditioned on ANI and PII, the gates dynamically intensify update scales with increasing non-stationarity rather than applying static rescaling, realizing signal-informed modulation. Specifically, Block 1 transitions from slight phase attenuation (< 1.0) to amplification (> 1.0) under severe irregularity, while Block 2 consistently amplifies representations across both branches with higher baseline gains. These distinct patterns verify that FIRE executes layer-specific, instance-adaptive modulation driven by local spectral dynamics.

Table 4: Average module ablation results in 7 public datasets over $L_{\text{pred}} \in \{96, 192, 336, 720\}$.

Data	FIRE		w/o Mod.		Joint Branch	
	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.424**	**0.427**	0.429	0.427	0.478	0.456
ETTh2	**0.362**	**0.392**	0.366	0.394	0.377	0.403
ETTm1	**0.371**	**0.387**	0.372	0.389	0.375	0.390
ETTm2	**0.276**	**0.319**	0.278	0.320	0.282	0.321
Weather	**0.241**	**0.265**	0.243	0.266	0.246	0.269
Traffic	**0.433**	**0.266**	0.452	0.273	0.458	0.269
ELC	**0.162**	**0.255**	0.163	0.255	0.163	0.256
Best	**7/7**	**6/7**	1/7	1/7	0/7	1/7

Table 5: ANI/PII statistics and relative error reductions over WPMixer for selected Weather and ETTm1 channels.

Data	Channel	ANI	PII	Reduction (%)	
				MSE	MAE
Weather	Ch. 15	0.2683	0.0435	−0.94	0.11
	Ch. 13	0.3446	1.2977	8.85	5.52
	Ch. 12	0.3456	1.2996	16.78	12.21
	Ch. 21	0.5466	1.3109	**29.02**	**19.96**
ETTm1	Ch. 5	0.2775	1.1396	−3.05	−0.74
	Ch. 4	0.2864	1.1485	−0.82	0.11
	Ch. 3	0.3303	1.1982	1.94	1.06
	Ch. 1	0.3459	1.2065	2.79	2.32

Table 6: Computational efficiency on ETTh1 and Weather. FLOPs are measured per sample; memory and runtime are reported in MiB and seconds per iteration, respectively. Best and second-best results are shown in **bold** and underlined.

Dataset	Model	Complexity		Training		Inference	
		Params. (M)	FLOPs (G)	Mem. (MiB)	Time (s/iter)	Mem. (MiB)	Time (s/iter)
ETTh1	**FIRE**	1.231	0.170	939.12	0.02928	163.24	0.00921
	WPMixer	**0.507**	**0.043**	**276.76**	**0.00928**	**44.59**	**0.00160**
	Fredformer	48.787	0.761	1039.40	0.02678	264.37	0.00751
Weather	**FIRE**	1.231	0.509	2777.60	0.06568	451.00	0.02158
	WPMixer	**0.741**	**0.175**	**842.08**	**0.01248**	**120.30**	**0.00390**
	Fredformer	48.791	2.282	1784.87	0.06694	381.68	0.02093

Forecasting-gain alignment. Table 5 evaluates FIRE against WPMixer across four representative channels from Weather and ETTm1, ordered by increasing ANI and PII. We measure the relative error reduction, defined as $(E_{\text{WPMixer}} - E_{\text{FIRE}})/E_{\text{WPMixer}} \times 100\%$. Across both benchmarks, channels with minimal non-stationarity exhibit negligible improvements or slight degradations, whereas those characterized by elevated ANI and PII display substantial forecasting gains, reaching 29.02%/19.96% (MSE/MAE) on Weather (Ch. 21) and 2.79%/2.32% on ETTm1 (Ch. 1). These examples suggest that the larger performance gains of FIRE are associated with stronger energy fluctuations and phase irregularity.

Computational efficiency. Table 6 illustrates the computational profiles of FIRE, WPMixer, and Fredformer on a single NVIDIA A800 GPU under a strictly unified setting (batch size 128, input and prediction lengths of 96, patch length 16, stride 8, and hidden dimension 128). Across both benchmarks, FIRE achieves a 97.5% parameter compression (1.231M vs. ≈48.8M) and a 77.7% reduction in FLOPs per sample compared to the heavy spectral Transformer Fredformer. These architectural savings yield substantial hardware advantages: on ETTh1, FIRE reduces both training and inference memory consumptions over Fredformer while maintaining sub-30ms training iterations; on the high-dimensional Weather benchmark, it achieves comparable training latency and lower per-sample arithmetic load. While WPMixer requires less memory and runtime, FIRE uses substantially fewer parameters and FLOPs than Fredformer while maintaining competitive runtime and stronger forecasting accuracy.

4 Related Work

Time domain forecasting models. Recent time series forecasting architectures are predominantly designed in the time domain and can be broadly categorized by their underlying backbones. Linear models capture temporal patterns through structural decomposition (Zeng et al., 2023) or multi-period sparse prediction (Lin et al., 2024b). Concurrently, MLP-based architectures model se-

quences via hierarchical interpolation (Oreshkin et al., 2020; Challu et al., 2023), multiscale feature mixing (Ekambaram et al., 2023; Wang et al., 2024), multivariate channel fusion (Lu et al., 2024), or explicit cycle projections (Lin et al., 2024a). Convolutional architectures expand the temporal receptive field to capture both local and global dependencies (Liu et al., 2022a; Wang et al., 2023; Luo & Wang, 2024), while hybrid variants like DeepGLO integrate temporal convolutions with global matrix factorization (Sen et al., 2019). Transformer models address sequence representations through efficient or multiscale attention (Zhou et al., 2021; Liu et al., 2022b; Chen et al., 2024), internal decomposition (Wu et al., 2021), and patch or variable-level tokenization (Nie et al., 2023; Zhang & Yan, 2023; Liu et al., 2024). Although these models achieve strong empirical performance, their designs primarily focus on modeling dependencies across temporal steps and variables. Crucially, they rarely quantify non-stationarity as an explicit, measurable state to dynamically guide adaptive representation learning.

Forecasting with frequency and hybrid representations. Complementing temporal modeling, frequency-domain and hybrid approaches exploit spectral representations to characterize periodic structures and spectral dynamics. FEDformer combines decomposition with attention in the frequency domain, while FiLM and TimesNet use Fourier information to represent dependencies over long horizons or dominant periods (Zhou et al., 2022b;a; Wu et al., 2023). FreTS and FourierGNN learn mappings over complex spectra, whereas FITS interpolates complex frequency coefficients (Yi et al., 2023b;a; Xu et al., 2024). Fredformer mitigates frequency bias, FreDF introduces supervision in the frequency domain, and CFPT models interactions across frequencies (Piao et al., 2024; Wang et al., 2025a; Kou et al., 2025). TimeMixer++ and FACT further investigate multiresolution spectral patterns and detailed relations among variables (Wang et al., 2025b; 2026b). As an adjacent hybrid approach, WPMixer applies wavelet packet decomposition and mixes subband representations across resolutions (Murad et al., 2025).

Most of these methods model spectral information through unified representations. Recent studies have further explored amplitude and phase information from different perspectives. APDNet (Zhuang et al., 2024) learns temporal and cross-variable dependencies through Fourier interactions over real and imaginary spectral components. APRNet (Liu et al., 2026) explicitly models interactions between amplitude and phase from temporal and channel perspectives. TimeAPN (Hu et al., 2026) uses amplitude and phase information within a model-agnostic normalization framework to adapt to rapidly changing non-stationary factors and correct backbone forecasts. PAFNet (Luo et al., 2025) uses phase correlation to align temporal lags across manufacturing processes and models frequency-specific dependencies after spectral decomposition. These studies demonstrate different ways of using amplitude and phase information in spectral modeling and non-stationarity adaptation. In contrast, FIRE explicitly models amplitude and phase as separate forecasting representations through dedicated branches and uses ANI and PII to adapt the corresponding representations according to their spectral dynamics in the historical window.

5 Conclusion

Non-stationarity remains a critical challenge in real-world multivariate time series forecasting. To address this issue from a spectral perspective, we propose FIRE, a novel frequency-domain framework grounded in the fundamentally distinct physical dynamics of amplitude and phase. By explicitly decoupling these two components into dedicated learning branches and leveraging history-derived ANI and PII statistics, FIRE performs signal-informed adaptive modulation to dynamically align representations with evolving spectral regimes. Extensive evaluations across twelve diverse benchmarks (including seven standard and five extended datasets) demonstrate FIRE's state-of-the-art forecasting performance. Crucially, our ablations and modulation analyses support the benefits of separate amplitude and phase modeling together with adaptive modulation. Furthermore, learned-response visualizations and channel-level evaluations confirm that FIRE's predictive advantage scales progressively with the severity of historical non-stationarity. Building upon these spectral insights, future directions include exploring multiscale spectral representations and hybrid time-frequency architectures that adapt to non-stationarity across varying temporal resolutions. Further limitations and extended research directions are discussed in Appendix A.7.

AI USE STATEMENT

In this work, we have not used generative AI tools for any of the required disclosure tasks (including generating synthetic data sets, developing theoretical models or conceptual frameworks, formulating or proving mathematical claims, proposing hypotheses, designing methodology or experiments, implementing methods, cleaning datasets, assisting with translation, or interpreting results).

Additionally, we used generative AI tools occasionally for editing the research paper to improve readability (specifically, proposing new words, grammar, and spelling corrections).

We have reviewed all AI-assisted work. All technical ideas, experimental designs, analyses, conclusions, and the core writing were developed and carried out entirely by the authors. AI-suggested language edits were manually reviewed by the authors to ensure they accurately reflected our original intent without altering the scientific claims. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza Ramírez, Max Mergenthaler Canseco, and Artur Dubrawski. NHITS: neural hierarchical interpolation for time series forecasting. In Brian Williams, Yiling Chen, and Jennifer Neville (eds.), *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pp. 6989–6997. AAAI Press, 2023. doi: 10.1609/AAAI.V37I6.25854. URL https://doi.org/10.1609/aaai.v37i6.25854.

Peng Chen, Yingying Zhang, Yunyao Cheng, Yang Shu, Yihang Wang, Qingsong Wen, Bin Yang, and Chenjuan Guo. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=lJkOCMP2aW.

Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. In Ambuj K. Singh, Yizhou Sun, Leman Akoglu, Dimitrios Gunopulos, Xifeng Yan, Ravi Kumar, Fatma Ozcan, and Jieping Ye (eds.), *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, pp. 459–469. ACM, 2023. doi: 10.1145/3580305.3599533. URL https://doi.org/10.1145/3580305.3599533.

Hassan Foroosh, Josiane Zerubia, and Marc Berthod. Extension of phase correlation to subpixel registration. *IEEE Trans. Image Process.*, 11(3):188–200, 2002. doi: 10.1109/83.988953. URL https://doi.org/10.1109/83.988953.

Yue Hu, Jialiang Tang, Siwei Yu, Baosheng Yu, Jing Zhang, and Dacheng Tao. Timeapn: Adaptive amplitude-phase non-stationarity normalization for time series forecasting. *CoRR*, abs/2603.17436, 2026. doi: 10.48550/ARXIV.2603.17436. URL https://doi.org/10.48550/arXiv.2603.17436.

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 *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=cGDAkQo1C0p.

Feifei Kou, Jiahao Wang, Lei Shi, Yuhan Yao, Yawen Li, Suguo Zhu, Zhongbao Zhang, and Junping Du. CFPT: empowering time series forecasting through cross-frequency interaction and periodic-aware timestamp modeling. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Forty-second International Conference on Machine Learning, ICML 2025, Vancouver, BC, Canada, July 13-19, 2025*,

volume 267 of *Proceedings of Machine Learning Research*. PMLR / OpenReview.net, 2025. URL https://proceedings.mlr.press/v267/kou25b.html.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024a. URL http://papers.nips.cc/paper_files/paper/2024/hash/bfe7998398779dde03cad7a73b1f81b6-Abstract-Conference.html.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. Sparsetsf: Modeling long-term time series forecasting with *1k* parameters. In Ruslan Salakhutdinov, Zico Kolter, Katherine A. Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30211–30226. PMLR / OpenReview.net, 2024b. URL https://proceedings.mlr.press/v235/lin24n.html.

Hao Liu, Xiaobin Zhu, Chun Yang, Lirui Deng, and Xucheng Yin. Amplitude-phase reconstruction for non-stationary time-series forecasting. *IEEE Trans. Knowl. Data Eng.*, 38(8):4868–4882, 2026. doi: 10.1109/TKDE.2026.3701273. URL https://doi.org/10.1109/TKDE.2026.3701273.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022a. URL http://papers.nips.cc/paper_files/paper/2022/hash/266983d0949aed78a16fa4782237dea7-Abstract-Conference.html.

Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X. Liu, and Schahram Dustdar. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022b. URL https://openreview.net/forum?id=0EXmFzUn5I.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022c. URL http://papers.nips.cc/paper_files/paper/2022/hash/4054556fcaa934b0bf76da52cf4f92cb-Abstract-Conference.html.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Han Lu, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. SOFTS: efficient multivariate time series forecasting with series-core fusion. In Amir Globersons, Lester Mackey, Danielle Belgrave, Angela Fan, Ulrich Paquet, Jakub M. Tomczak, and Cheng Zhang (eds.), *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024. URL http://papers.nips.cc/paper_files/paper/2024/hash/754612bde73a8b65ad8743f1f6d8ddf6-Abstract-Conference.html.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations, ICLR 2024,*

Vienna, Austria, May 7-11, 2024. OpenReview.net, 2024. URL https://openreview.net/forum?id=vpJMJerXHU.

Yang Luo, Haoyang Luan, Haoyun Pan, Yongquan Jia, Xiaofeng Gao, and Guihai Chen. Paf-net: Phase-aligned frequency decoupling network for multi-process manufacturing quality prediction. *CoRR*, abs/2507.22840, 2025. doi: 10.48550/ARXIV.2507.22840. URL https://doi.org/10.48550/arXiv.2507.22840.

Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. Wpmixer: Efficient multi-resolution mixing for long-term time series forecasting. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pp. 19581–19588. AAAI Press, 2025. doi: 10.1609/AAAI.V39I18.34156. URL https://doi.org/10.1609/aaai.v39i18.34156.

Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

A.V. Oppenheim and R.W. Schafer. *Discrete-time Signal Processing*. Prentice-Hall signal processing series. Prentice Hall, 1989. ISBN 9780132162920. URL https://books.google.co.jp/books?id=2CEqAQAAIAAJ.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL https://openreview.net/forum?id=r1ecqn4YwB.

Hugo Pedro, David Larson, and Carlos Coimbra. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. *Journal of Renewable and Sustainable Energy*, 11:036102, 05 2019. doi: 10.1063/1.5094494.

Yifan Peng, Siddharth Dalmia, Ian R. Lane, and Shinji Watanabe. Branchformer: Parallel mlp-attention architectures to capture local and global context for speech recognition and understanding. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17627–17643. PMLR, 2022. URL https://proceedings.mlr.press/v162/peng22a.html.

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In Ricardo Baeza-Yates and Francesco Bonchi (eds.), *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, pp. 2400–2410. ACM, 2024. doi: 10.1145/3637528.3671928. URL https://doi.org/10.1145/3637528.3671928.

Avirup Saha, Prerna Agarwal, Sambit Ghosh, Neelamadhav Gantayat, and Renuka Sindhgatta. Towards business process observability. In Sriraam Natarajan, Indrajit Bhattacharya, Richa Singh, Arun Kumar, Sayan Ranu, Kalika Bali, and Abinaya K. (eds.), *Proceedings of the 7th Joint International Conference on Data Science & Management of Data (11th ACM IKDD CODS and 29th COMAD), Bangalore, India, January 4-7, 2024*, pp. 257–265. ACM, 2024. doi: 10.1145/3632410.3632435. URL https://doi.org/10.1145/3632410.3632435.

Rajat Sen, Hsiang-Fu Yu, and Inderjit S. Dhillon. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (eds.),

Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, pp. 4838–4847, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/3a0844cee4fcf57de0c71e9ad3035478-Abstract.html.

Hao Wang, Lichen Pan, Yuan Shen, Zhichao Chen, Degui Yang, Yifei Yang, Sen Zhang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. Fredf: Learning to forecast in the frequency domain. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025a. URL https://openreview.net/forum?id=4A9IdSa1ul.

Hao Wang, Licheng Pan, Yuan Lu, Zhixuan Chu, Xiaoxi Li, Shuting He, Zi Ciu Chan, Qingsong Wen, Haoxuan Li, and Zhouchen Lin. DistDF: Time-series forecasting needs joint-distribution wasserstein alignment. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=VrdLwUmzBy.

Huiqiang Wang, Jian Peng, Feihu Huang, Jince Wang, Junhui Chen, and Yifei Xiao. MICN: multi-scale local and global context modeling for long-term series forecasting. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=zt53IDUR1U.

Huiqiang Wang, Jieming Shi, and Li Qing. FACT: Fine-grained across-variable convolution for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=j3gNYqrHtl.

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 *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=7oLshfEIC2.

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 *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025b. URL https://openreview.net/forum?id=1CLzLXSFNn.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 22419–22430, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Wenyan Xu, Dawei Xiang, Yue Liu, Xiyu Wang, Yanxiang Ma, Liang Zhang, Chang Xu, and Jiaheng Zhang. Finmultitime: A four-modal bilingual dataset for financial time-series analysis. *CoRR*, abs/2506.05019, 2025. doi: 10.48550/ARXIV.2506.05019. URL https://doi.org/10.48550/arXiv.2506.05019.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: modeling time series with 10k parameters. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=bWcnvZ3qMb.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023a. URL http://papers.nips.cc/paper_files/paper/2023/hash/dc1e32dd3eb381dbc71482f6a96cbf86-Abstract-Conference.html.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023b. URL http://papers.nips.cc/paper_files/paper/2023/hash/f1d16af76939f476b5f040fd1398c0a3-Abstract-Conference.html.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In Brian Williams, Yiling Chen, and Jennifer Neville (eds.), *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pp. 11121–11128. AAAI Press, 2023. doi: 10.1609/AAAI.V37I9.26317. URL https://doi.org/10.1609/aaai.v37i9.26317.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=vSVLM2j9eie.

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 *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pp. 11106–11115. AAAI Press, 2021. doi: 10.1609/AAAI.V35I12.17325. URL https://doi.org/10.1609/aaai.v35i12.17325.

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 Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022a. URL http://papers.nips.cc/paper_files/paper/2022/hash/524ef58c2bd075775861234266e5e020-Abstract-Conference.html.

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

Wei Zhuang, Jili Fan, Jiayu Fang, Wenxuan Fang, and Min Xia. Rethinking general time series analysis from a frequency domain perspective. *Knowl. Based Syst.*, 301:112281, 2024. doi: 10.1016/J.KNOSYS.2024.112281. URL https://doi.org/10.1016/j.knosys.2024.112281.

A APPENDIX

A.1 NOTATION

Table A1 summarizes the notation used throughout the paper.

Table A1: Notation list.

Symbol	Description
$\mathbf{X},\ X_{t,c},\ \mathbf{X}_{\text{out}},\ \mathbf{X}_{\text{true}}$	Input series and its entries, forecast, and ground truth
$T,\ L_p,\ S,\ L_{\text{pred}}$	Look back length, patch length, patch stride, and prediction length
$C,\ D,\ L_b$	Number of channels, hidden dimension, and number of blocks per branch
$N_p,\ N_v$	Number of patches and valid patches used for ANI
$F_p,\ F_{\text{pred}}$	Number of frequency bins in the input and predicted spectra
$\mathbf{X}_P,\ \mathbf{Z},\ \widehat{\mathbf{Z}}_{\text{out}}$	Input patches, input spectra, and predicted future spectrum
$\mathbf{A},\ \mathbf{\Phi},\ \mathbf{P}$	Spectral amplitude, phase, and phase coordinates $[\cos \mathbf{\Phi}, \sin \mathbf{\Phi}]$
$\mathbf{W}_r,\ \mathbf{b}_r$	Branch-specific embedding weights and biases
$\mathbf{H}_r^{(\ell)},\ \overline{\mathbf{H}}_A,\ \overline{\mathbf{H}}_\Phi,\ \mathbf{H}_o$	Branch representations, refined branch representations, and reconstructed complex features
$s_{c,p},\ \bar{s}_c$	Patch spectral strength and its mean over valid patches
$\text{ANI}_c,\ \text{PII}_c$	Amplitude Non-stationarity Index and Phase Irregularity Index
$\Delta\Phi_{i,k},\ \mu_k$	Phase increment of valid pair i and circular mean at frequency k
$K,\ V_k$	Number of frequencies with valid pairs and number of valid pairs at frequency k
$\epsilon_A,\ \epsilon_\Phi$	Numerical floor for ANI and amplitude threshold for reliable phase
$q_c^A,\ q_c^\Phi,\ \mathbf{m}_{r,u}^{(\ell)}$	Transformed ANI and PII inputs and modulation vectors
$r,\ \ell,\ u$	Branch, block, and sublayer indices
$\mathcal{L},\ \mathcal{L}_{\text{Huber}},\ \mathcal{L}_{\text{FFT}}$	Total loss, Huber loss, and spectral loss
$\alpha,\ \beta$	Weights of the Huber and FFT loss terms

A.2 DATASET STATISTICS

Table A2 summarizes the twelve datasets used in our experiments. We report the dataset size, number of variables, sampling interval, and prediction length. Dataset size refers to the number of records before constructing forecasting windows.

Widely used benchmarks. The seven standard benchmarks cover electricity transformer measurements (ETTh1, ETTh2, ETTm1, and ETTm2), weather observations (Weather), electricity consumption (Electricity), and road traffic measurements (Traffic) (Wu et al., 2021). Their sampling intervals range from 10 minutes to one hour, with 7 to 862 variables. We evaluate prediction lengths of 96, 192, 336, and 720 on all seven datasets.

Extended datasets. PV contains photovoltaic measurements sampled every minute (Pedro et al., 2019). We use prediction lengths of 60, 240, and 720, corresponding to 1, 4, and 12 hours. L2C contains six variables from the Leading to Cash business process sampled every five minutes (Saha et al., 2024). The same prediction lengths correspond to 5, 20, and 60 hours.

The three financial datasets are taken from FinMultiTime (Xu et al., 2025). They contain stock records from the Media, Electronics, and Non-bank Financial industries of the HS 300. Each dataset contains five numerical variables sampled once per trading day. We evaluate prediction lengths of 5, 10, and 20 trading days. PV is univariate, while the other eleven datasets are multivariate.

Table A2: Statistics of the twelve forecasting datasets. Records denote the dataset size, C denotes the number of variables, and prediction lengths are measured in time steps.

Dataset	Records	C	Interval	L_{pred}
ETTh1	17,420	7	1 hour	96, 192, 336, 720
ETTh2	17,420	7	1 hour	96, 192, 336, 720
ETTm1	69,680	7	15 min	96, 192, 336, 720
ETTm2	69,680	7	15 min	96, 192, 336, 720
Weather	52,696	21	10 min	96, 192, 336, 720
Electricity	26,304	321	1 hour	96, 192, 336, 720
Traffic	17,544	862	1 hour	96, 192, 336, 720
PV	293,500	1	1 min	60, 240, 720
L2C	31,968	6	5 min	60, 240, 720
Media	103,061	5	1 trading day	5, 10, 20
Electronics	129,264	5	1 trading day	5, 10, 20
Non-bank Financial	156,183	5	1 trading day	5, 10, 20

A.3 Details of Empirical Analysis

We provide additional details for the empirical analysis in Figure 1. The analysis uses the training splits of ETTh1, Weather, Electricity, and Traffic. Each 96-point window is divided into six nonoverlapping patches of length 16. For each patch p, let $X_{p,k} = A_{p,k}e^{j\phi_{p,k}}$ denote its Fourier coefficient at frequency k. We first compute the spectral strength of each patch as:

$$S_p = \left(\sum_{k \in \mathcal{K}_A} A_{p,k}^2 \right)^{1/2}, \tag{A1}$$

where $\mathcal{K}_A$ contains the nonzero frequency bins. Amplitude variation across patches is measured by:

$$D_A = \frac{\mathrm{Std}_p(S_p)}{\mathrm{Mean}_p(S_p) + \epsilon}, \tag{A2}$$

where ϵ is a small constant for numerical stability. For phase, we compute the phase increment between adjacent patches as:

$$\Delta\phi_{p,k} = \mathrm{angle}\left(X_{p+1,k} X_{p,k}^*\right). \tag{A3}$$

For each frequency, its circular dispersion is computed from the phase increments and weighted by the spectral amplitudes of the adjacent patches. The resulting phase variation is

$$D_P = \frac{\sum_{k \in \mathcal{K}_P} W_k \left(1 - |R_k|\right)}{\sum_{k \in \mathcal{K}_P} W_k}, \tag{A4}$$

where

$$R_k = \frac{\sum_p w_{p,k} e^{j\Delta\phi_{p,k}}}{\sum_p w_{p,k}}, \qquad w_{p,k} = \sqrt{A_{p+1,k} A_{p,k}}, \qquad W_k = \sum_p w_{p,k}. \tag{A5}$$

Controlled analysis. We separately apply gain variation and circular shifts to local patches. Positive gain changes spectral amplitude while preserving phase, whereas a circular shift preserves spectral amplitude while changing phase. Each patch is shifted by at most four samples. For each transformation, we compute the absolute changes $|D'_A - D_A|$ and $|D'_P - D_P|$ relative to the original window. Figure 1(a) and (b) summarize the responses over $7{,}141$ training windows. Error bars denote 95% bootstrap intervals across channels.

Natural variation analysis. We further compute D_A and D_P on $29{,}566$ unperturbed training windows. Within each channel, windows are ranked separately according to the two measures. We define their percentile rank difference as

$$D_R = |r_A - r_P|, \tag{A6}$$

where r_A and r_P are the percentile ranks of D_A and D_P, respectively. Figure 1(c) reports the fraction of windows with D_R above each threshold. The solid curve denotes the macro average across the four datasets, and the shaded region denotes the dataset range. The dashed curve provides a reference for two independent percentile ranks, given by $P(D_R \geq \tau) = (1 - \tau)^2$.

A.4 Implementation Details

We use a fixed look-back window of 96 time points for all datasets and evaluate the dataset-specific forecasting horizons reported in Table A2. Unless otherwise specified, all experiments follow the same data preprocessing and evaluation protocol. Patch lengths are selected from 8 to 48, and training uses mini-batches of 32 to 256 samples. We optimize all models using Adam with the composite Huber–FFT objective defined in Equation 11. Early stopping is applied when the validation loss does not improve for eight consecutive epochs. The implementation is based on PyTorch, and all experiments are conducted on a single NVIDIA A800 GPU with 80 GB of memory. Forecasting performance is evaluated using mean squared error (MSE) and mean absolute error (MAE).

For baseline methods, we follow their original experimental settings whenever applicable. In particular, Fredformer and WPMixer are reproduced using their released implementations and published configurations. Model selection is performed on the validation set, and the test set is used only for final evaluation. The main hyperparameter search space for FIRE is summarized below:

- **Look-back length:** $L = 96$;
- **Prediction length:** $T \in \{96, 192, 336, 720\}$, according to the dataset setting;
- **Patch length:** $P \in \{8, 12, 16, 24, 32, 48\}$;
- **Batch size:** $B \in \{32, 64, 128, 256\}$;
- **Optimizer:** Adam;
- **Learning rate:** $\eta \in \{1 \times 10^{-4}, 3 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}\}$;
- **Early stopping patience:** 8 epochs.

The complete forecasting procedure of FIRE is provided in Algorithm A1. The algorithm follows the notation summarized in Table A1. Given an input window, FIRE first transforms the historical sequence into the frequency domain and separates its amplitude and phase components. The two components are processed by their corresponding branches and modulated using the Amplitude Non-stationarity Index (ANI) and Phase Irregularity Index (PII), respectively. The modulated representations are then combined to reconstruct the forecast.

Algorithm A1 Pseudocode of the FIRE forecasting procedure.

Require: Historical series $\mathbf{X} \in \mathbb{R}^{T \times C}$; prediction length L_{pred}; patch length L_p and stride S; hidden dimension D; blocks per branch L_b; model parameters.
Ensure: Forecast $\mathbf{X}_{\text{out}} \in \mathbb{R}^{L_{\text{pred}} \times C}$.

▷ **Input transformation**
$\widetilde{\mathbf{X}} \leftarrow \text{RevIN}(\mathbf{X})$; retain the normalization statistics
$\mathbf{X}_P \leftarrow \text{Patching}(\widetilde{\mathbf{X}}; L_p, S)$ ▷ $\mathbb{R}^{C \times N_p \times L_p}$
$\mathbf{Z} \leftarrow \text{rFFT}(\mathbf{X}_P)$ ▷ $\mathbb{C}^{C \times N_p \times F_p}$
▷ **Amplitude–phase separation and embedding**
$\mathbf{A} \leftarrow |\mathbf{Z}|, \quad \mathbf{\Phi} \leftarrow \text{angle}(\mathbf{Z})$
$\mathbf{P} \leftarrow [\cos \mathbf{\Phi}, \sin \mathbf{\Phi}]$; replace coordinates with $[1, 0]$ where $\mathbf{A} \leq \epsilon_\Phi$
$\mathbf{H}_A^{(0)} \leftarrow \mathbf{A}\mathbf{W}_A + \mathbf{b}_A, \quad \mathbf{H}_\Phi^{(0)} \leftarrow \mathbf{P}\mathbf{W}_\Phi + \mathbf{b}_\Phi$ ▷ $\mathbb{R}^{C \times N_p \times D}$
▷ **History-derived modulation inputs**
Compute ANI_c and PII_c from the first N_v patches, for each channel c
$q_c^A \leftarrow \log(1 + \text{ANI}_c), \quad q_c^\Phi \leftarrow \text{PII}_c / \pi$
▷ **Separate branch updates**
for $\ell = 0$ **to** $L_b - 1$ **do**
  for $r \in \{A, \Phi\}$ **do**
    Generate $\mathbf{m}_{r,\text{mlp}}^{(\ell)}$ and $\mathbf{m}_{r,\text{attn}}^{(\ell)}$ from q^r using separate modulation networks
    $\mathbf{H}_r^{(\ell+1)} \leftarrow \text{Block}_r^{(\ell)}(\mathbf{H}_r^{(\ell)}; \mathbf{m}_{r,\text{mlp}}^{(\ell)}, \mathbf{m}_{r,\text{attn}}^{(\ell)})$
  end for
end for
▷ **Output projection**
$\overline{\mathbf{H}}_r \leftarrow \text{CrossVarAttn}_r(\mathbf{H}_r^{(L_b)})$ for $r \in \{A, \Phi\}$ ▷ Includes residual update
$\mathbf{H}_o \leftarrow \text{Recombine}(\overline{\mathbf{H}}_A, \overline{\mathbf{H}}_\Phi)$ ▷ Positive amplitude $\times$ unit phase
$\widehat{\mathbf{Z}}_{\text{out}} \leftarrow \text{MLP}_{\text{proj}}(\text{Flat}(\mathbf{H}_o))$ ▷ $\mathbb{C}^{C \times F_{\text{pred}}}$
$\mathbf{X}_{\text{out}} \leftarrow \text{RevIN}^{-1}(\text{irFFT}(\widehat{\mathbf{Z}}_{\text{out}}, n = L_{\text{pred}}))$
return $\mathbf{X}_{\text{out}}$

A.5 Full Results

The main paper focuses on the key experimental results and findings. For completeness, we provide the full results and further analyses across extended datasets, core component ablations, loss ablations, and hyperparameter sensitivity analyses.

A.5.1 Forecasting Results on Extended Datasets

Table A3 provides the full horizon-specific results corresponding to the averaged comparisons reported in Tables 2 and 3. PV and L2C are evaluated at prediction lengths of 60, 240, and 720, while the three HS 300 industry datasets are evaluated at horizons of 5, 10, and 20.

Across all five extended datasets, FIRE achieves the best MSE at every reported prediction horizon. It also achieves the best MAE at all horizons except $L_{\text{pred}} = 60$ on L2C, where WPMixer and Fredformer obtain better results. On PV and L2C, the MSE differences between FIRE and the baselines become more pronounced at the two longer prediction horizons. On the three HS 300 industry datasets, FIRE achieves the best MSE and MAE at all three horizons, consistent with the averaged results reported in the main paper.

Table A3: Full forecasting results on the five extended datasets. *Avg.* denotes the average over the three prediction lengths for each dataset.

Dataset	L_{pred}	**FIRE**		Fredformer		WPMixer	
		MSE	MAE	MSE	MAE	MSE	MAE
PV	60	**0.189**	**0.229**	0.199	0.244	0.199	0.243
	240	**0.449**	**0.426**	0.565	0.490	0.578	0.473
	720	**0.659**	**0.592**	0.781	0.640	0.778	0.633
	Avg.	**0.433**	**0.416**	0.515	0.458	0.518	0.450
L2C	60	**0.211**	0.231	0.219	0.230	0.213	**0.227**
	240	**0.726**	**0.512**	0.839	0.555	0.832	0.550
	720	**1.454**	**0.852**	1.626	0.925	1.600	0.922
	Avg.	**0.797**	**0.532**	0.895	0.570	0.882	0.566
Media	5	**0.0108**	**0.0566**	0.0119	0.0598	0.0119	0.0601
	10	**0.0156**	**0.0731**	0.0162	0.0769	0.0163	0.0765
	20	**0.0234**	**0.0971**	0.0240	0.0999	0.0237	0.0990
	Avg.	**0.0166**	**0.0756**	0.0174	0.0789	0.0173	0.0785
Electronics	5	**0.0191**	**0.0483**	0.0196	0.0522	0.0196	0.0512
	10	**0.0233**	**0.0627**	0.0238	0.0658	0.0237	0.0638
	20	**0.0280**	**0.0798**	0.0288	0.0834	0.0287	0.0815
	Avg.	**0.0235**	**0.0636**	0.0241	0.0671	0.0240	0.0655
Non-bank Financial	5	**0.0278**	**0.0581**	0.0290	0.0630	0.0302	0.0630
	10	**0.0363**	**0.0752**	0.0371	0.0770	0.0382	0.0797
	20	**0.0481**	**0.0971**	0.0483	0.0995	0.0486	0.1002
	Avg.	**0.0374**	**0.0768**	0.0381	0.0798	0.0390	0.0810

A.5.2 Ablation of Core Components

Table 4 provides the full horizon-specific results of the core component ablations, complementing the averaged results reported in the main paper.

The full FIRE model achieves the best performance across most prediction horizons. Removing the amplitude or phase modeling components generally degrades forecasting accuracy, while removing ANI or PII modulation also leads to performance degradation across different settings. The magnitude of these changes varies across datasets and prediction horizons, showing that the contribution of each component depends on the forecasting setting. These horizon-specific results are consistent with the averaged trends reported in the main paper.

Table A4: Full ablation of the core components on seven datasets. *Avg.* averages four prediction lengths.

Dataset	L_{pred}	**FIRE**		w/o Mod.		Joint Branch	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.371**	**0.391**	0.374	**0.391**	0.381	0.396
	192	**0.415**	**0.419**	0.423	0.420	0.434	0.428
	336	**0.456**	**0.439**	0.458	**0.439**	0.547	0.489
	720	**0.456**	**0.458**	0.459	0.459	0.551	0.511
	Avg.	**0.424**	**0.427**	0.429	**0.427**	0.478	0.456
ETTh2	96	**0.285**	**0.334**	**0.285**	**0.334**	0.301	0.348
	192	**0.356**	**0.382**	0.364	0.386	0.379	0.399
	336	**0.397**	**0.418**	0.398	0.419	0.406	0.424
	720	**0.411**	**0.434**	0.417	0.437	0.423	0.443
	Avg.	**0.362**	**0.392**	0.366	0.394	0.377	0.403
ETTm1	96	**0.304**	**0.346**	0.308	0.349	0.310	0.351
	192	**0.356**	**0.375**	**0.356**	**0.375**	0.361	0.379
	336	**0.385**	**0.396**	**0.385**	0.397	0.388	0.397
	720	0.440	**0.431**	**0.439**	0.434	0.442	0.434
	Avg.	**0.371**	**0.387**	0.372	0.389	0.375	0.390
ETTm2	96	**0.171**	**0.252**	0.173	0.255	0.176	0.255
	192	**0.236**	**0.296**	0.240	0.298	0.241	0.299
	336	**0.301**	**0.335**	**0.301**	0.336	0.303	0.337
	720	**0.398**	**0.392**	0.399	**0.392**	0.407	0.395
	Avg.	**0.276**	**0.319**	0.278	0.320	0.282	0.321
Weather	96	**0.153**	**0.192**	0.156	0.196	0.158	0.197
	192	**0.206**	**0.243**	0.207	0.244	0.209	0.246
	336	**0.263**	**0.285**	0.264	**0.285**	0.270	0.290
	720	**0.342**	**0.338**	0.344	0.340	0.347	0.343
	Avg.	**0.241**	**0.265**	0.243	0.266	0.246	0.269
Traffic	96	**0.405**	**0.249**	0.412	0.256	0.414	0.251
	192	**0.417**	**0.260**	0.439	0.270	0.446	0.262
	336	**0.432**	**0.268**	0.464	0.273	0.467	0.274
	720	**0.478**	**0.286**	0.492	0.292	0.506	0.290
	Avg.	**0.433**	**0.266**	0.452	0.273	0.458	0.269
Electricity	96	**0.136**	0.230	**0.136**	**0.229**	**0.136**	**0.229**
	192	0.155	**0.246**	**0.154**	0.247	**0.154**	**0.246**
	336	**0.166**	**0.260**	**0.166**	**0.260**	0.169	0.263
	720	**0.193**	**0.284**	0.195	0.285	**0.193**	0.285
	Avg.	**0.162**	**0.255**	0.163	**0.255**	0.163	0.256

A.5.3 LOSS ABLATION

Table A5 provides the full horizon-specific results of the loss ablation, complementing the averaged results reported in the main paper.

The composite Huber + FFT objective achieves the best performance across most prediction horizons. Removing either loss component generally degrades forecasting accuracy, although the magnitude of the degradation varies across datasets and horizons.

Table A5: Loss ablation on three datasets. *Avg.* averages four prediction lengths.

Dataset	L_{pred}	**Huber+FFT**		Huber only		MSE only	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.371**	**0.391**	0.385	**0.391**	0.382	0.400
	192	**0.415**	**0.419**	0.434	0.423	0.430	0.428
	336	**0.456**	**0.439**	0.479	0.452	0.478	0.449
	720	**0.456**	**0.458**	0.467	0.462	0.477	0.468
	Avg.	**0.424**	**0.427**	0.441	0.432	0.442	0.436
ETTh2	96	**0.285**	**0.334**	0.290	0.342	0.296	0.348
	192	**0.356**	**0.382**	0.367	0.393	0.377	0.401
	336	**0.397**	**0.418**	0.410	0.428	0.431	0.438
	720	**0.411**	**0.434**	0.424	0.442	0.429	0.448
	Avg.	**0.362**	**0.392**	0.373	0.401	0.383	0.409
Weather	96	**0.153**	**0.192**	0.160	0.200	0.160	0.205
	192	**0.206**	**0.243**	0.209	0.246	0.211	0.251
	336	**0.263**	**0.285**	0.267	0.286	0.269	0.293
	720	**0.342**	**0.338**	0.347	0.341	0.348	0.346
	Avg.	**0.241**	**0.265**	0.245	0.268	0.247	0.274

A.5.4 HYPERPARAMETER SENSITIVITY ANALYSIS

We examine the sensitivity of FIRE to representation dimension, backbone depth, patch length and stride, and look-back window length on ETTh1, ETTm1, Weather, and Traffic. For each hyperparameter, we vary its configuration while keeping the remaining settings fixed. Tables A6–A9 report the results over all four prediction horizons and their averages.

Representation dimension. Table A6 shows the results for representation dimension $D \in \{32, 64, 128, 256\}$. Increasing D has the clearest effect on Traffic, where the average MSE decreases from 0.467 to 0.449 and the average MAE from 0.301 to 0.271 as D increases from 32 to 256. Weather shows only small changes across the evaluated dimensions. On ETTh1 and ETTm1, the best results occur at different intermediate dimensions depending on the metric. Thus, increasing the representation dimension does not consistently improve forecasting performance across datasets.

Backbone layers. Table A7 demonstrates the results for the number of backbone blocks per branch, $L_b \in \{1, 3, 6, 9\}$. Increasing backbone depth does not consistently improve forecasting performance. One or three layers achieve the best dataset-level averages across the four datasets. For example, Weather achieves its best average MSE and MAE with three layers, while ETTh1 achieves its best average MSE with one layer. Increasing the depth to nine layers results in higher average errors than the shallower configurations across the evaluated datasets.

Patch length and stride. Table A8 illustrates the results for five patch/stride pairs: $4/2$, $8/4$, $16/8$, $32/16$, and $48/24$, each maintaining 50% overlap. The best-performing configuration varies across datasets. The shortest pair, $4/2$, achieves the best average MSE and MAE on ETTh1 and Traffic, while intermediate patch/stride pairs perform best on ETTm1. Weather shows relatively small variations across the evaluated configurations.

Table A6: Sensitivity to representation dimension D. Bold marks the lowest error for each dataset, prediction length, and metric.

Dataset	L_{pred}	$D = 32$		$D = 64$		$D = 128$		$D = 256$	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.387	0.396	**0.374**	**0.391**	0.377	0.397	**0.374**	0.395
	192	0.433	0.424	**0.423**	**0.420**	0.433	0.426	0.435	0.426
	336	0.470	0.446	0.467	0.442	**0.457**	**0.439**	0.460	0.442
	720	0.471	0.472	0.469	0.463	0.462	0.461	**0.458**	**0.459**
	Avg.	0.440	0.434	0.433	**0.429**	**0.432**	0.431	**0.432**	0.430
ETTm1	96	0.308	0.350	**0.307**	**0.348**	0.314	0.350	0.310	**0.348**
	192	0.364	**0.375**	**0.357**	**0.375**	0.360	0.380	0.361	0.385
	336	0.409	0.402	0.395	**0.395**	**0.385**	0.397	0.387	0.403
	720	0.457	0.432	0.445	**0.430**	**0.440**	0.434	0.452	0.445
	Avg.	0.385	0.390	0.376	**0.387**	**0.375**	0.390	0.378	0.395
Weather	96	0.158	0.198	0.160	0.200	0.158	0.196	**0.156**	**0.195**
	192	0.210	0.246	**0.208**	**0.245**	0.210	0.246	0.211	0.246
	336	**0.266**	0.288	**0.266**	**0.286**	0.269	0.289	0.267	0.289
	720	**0.346**	0.342	**0.346**	**0.341**	**0.346**	**0.341**	0.347	0.342
	Avg.	**0.245**	0.269	**0.245**	**0.268**	**0.245**	**0.268**	**0.245**	**0.268**
Traffic	96	0.439	0.288	0.422	0.270	0.417	0.261	**0.409**	**0.255**
	192	0.453	0.290	0.443	0.278	0.442	0.272	**0.438**	**0.266**
	336	0.470	0.299	**0.454**	0.283	0.456	0.277	0.456	**0.273**
	720	0.507	0.326	0.490	0.309	**0.485**	0.300	0.495	**0.292**
	Avg.	0.467	0.301	0.452	0.285	0.450	0.277	**0.449**	**0.271**

Table A7: Sensitivity to backbone layers L_b. Bold marks the lowest error for each dataset, prediction length, and metric.

Dataset	L_{pred}	$L_b = 1$		$L_b = 3$		$L_b = 6$		$L_b = 9$	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.378	**0.391**	**0.376**	0.392	0.377	0.395	0.378	0.395
	192	0.429	**0.422**	**0.428**	**0.422**	0.432	0.424	0.436	0.426
	336	**0.460**	0.441	0.464	**0.440**	0.461	0.443	0.463	0.442
	720	**0.465**	**0.461**	**0.465**	0.463	0.478	0.469	0.486	0.472
	Avg.	**0.433**	**0.429**	0.434	**0.429**	0.437	0.433	0.441	0.434
ETTm1	96	**0.305**	**0.347**	0.312	0.348	0.315	0.352	0.316	0.353
	192	0.358	**0.375**	**0.357**	0.376	0.363	0.379	0.364	0.380
	336	0.388	**0.395**	**0.385**	0.397	0.386	0.399	0.390	0.400
	720	**0.443**	**0.432**	**0.443**	0.437	0.447	0.440	0.457	0.444
	Avg.	**0.374**	**0.387**	**0.374**	0.389	0.378	0.392	0.382	0.394
Weather	96	0.159	0.198	**0.155**	**0.195**	0.157	0.198	0.163	0.206
	192	0.209	0.245	0.209	0.245	0.208	0.245	**0.207**	**0.244**
	336	0.267	0.288	**0.266**	**0.286**	0.268	0.289	0.271	0.293
	720	0.347	**0.341**	**0.346**	0.342	**0.346**	0.342	0.348	**0.341**
	Avg.	0.245	0.268	**0.244**	**0.267**	0.245	0.269	0.247	0.271
Traffic	96	0.410	0.262	0.416	0.255	0.426	**0.253**	**0.409**	0.260
	192	0.438	0.269	0.440	**0.264**	0.448	0.267	**0.437**	0.278
	336	0.459	0.274	0.457	**0.272**	**0.452**	0.282	0.456	0.281
	720	**0.488**	0.296	0.490	0.293	0.499	**0.292**	0.498	0.308
	Avg.	**0.449**	0.275	0.451	**0.271**	0.456	0.273	0.450	0.282

Table A8: Sensitivity to patch length and stride. Column headings show L_p/S. Bold marks the lowest error for each dataset, prediction length, and metric.

Dataset	L_{pred}	4/2		8/4		16/8		32/16		48/24	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.372**	**0.390**	0.375	0.391	0.374	0.391	0.377	0.393	0.379	0.395
	192	**0.417**	**0.418**	0.426	**0.418**	0.432	0.422	0.430	0.422	0.429	0.424
	336	0.469	**0.438**	**0.461**	0.442	**0.461**	**0.438**	0.483	0.452	0.470	0.452
	720	**0.464**	**0.464**	0.468	0.465	0.491	0.478	0.467	**0.464**	0.473	0.468
	Avg.	**0.431**	**0.428**	0.432	0.429	0.440	0.432	0.439	0.433	0.438	0.435
ETTm1	96	0.313	**0.345**	0.312	0.350	**0.310**	0.351	0.312	0.350	0.320	0.351
	192	0.359	0.378	**0.356**	0.377	**0.356**	**0.375**	0.363	0.376	0.365	**0.375**
	336	**0.385**	0.398	**0.385**	**0.397**	**0.385**	**0.397**	0.393	**0.397**	0.398	0.398
	720	0.442	0.436	0.440	0.435	**0.439**	0.434	0.448	**0.432**	0.455	0.434
	Avg.	0.375	**0.389**	**0.373**	0.390	**0.373**	**0.389**	0.379	**0.389**	0.384	0.390
Weather	96	**0.155**	0.198	0.157	0.199	0.157	**0.197**	0.158	0.198	0.162	0.202
	192	**0.207**	0.247	0.210	0.248	0.209	**0.246**	0.209	**0.246**	0.210	0.247
	336	0.271	0.292	0.271	0.292	**0.269**	**0.290**	**0.269**	**0.290**	0.271	0.291
	720	0.348	0.343	0.346	0.342	**0.345**	**0.341**	0.350	0.343	0.348	0.344
	Avg.	**0.245**	0.270	0.246	0.270	**0.245**	**0.269**	0.247	**0.269**	0.248	0.271
Traffic	96	**0.405**	**0.254**	0.410	0.262	0.427	0.255	0.415	0.258	0.422	0.262
	192	**0.426**	0.268	0.431	0.268	0.442	**0.266**	0.433	0.268	0.434	0.270
	336	**0.437**	**0.269**	0.455	0.271	0.459	0.274	0.451	0.276	0.451	0.277
	720	**0.494**	**0.287**	0.495	0.288	0.502	0.291	0.496	0.296	0.499	0.299
	Avg.	**0.440**	**0.269**	0.448	0.272	0.457	0.272	0.449	0.275	0.452	0.277

Look-back window length. Table A9 presents the results for input length $T \in \{96, 192, 288, 384, 512\}$. Longer histories improve forecasting performance in several settings, but the trend varies across datasets and metrics. On Weather, the average MSE decreases from 0.244 at $T = 96$ to 0.227 at $T = 512$. ETTm1 achieves its best average results at intermediate window lengths. Traffic achieves its best average MSE at $T = 384$ and its best average MAE at $T = 192$. On ETTh1, longer windows improve the average MSE, while $T = 96$ achieves the best average MAE. Overall, increasing the look-back window does not consistently improve both metrics across datasets.

Table A9: Sensitivity to look-back window length T. Bold marks the lowest error for each dataset, prediction length, and metric.

Dataset	L_{pred}	$T = 96$		$T = 192$		$T = 288$		$T = 384$		$T = 512$	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.374**	**0.391**	0.385	0.401	0.398	0.408	0.394	0.408	0.400	0.418
	192	0.423	**0.420**	0.427	0.424	0.428	0.428	**0.414**	**0.420**	0.421	0.436
	336	0.456	0.439	0.445	0.436	0.437	**0.432**	0.426	**0.432**	**0.417**	0.434
	720	0.458	**0.460**	0.458	0.464	0.451	0.464	0.444	0.466	**0.443**	0.461
	Avg.	0.428	**0.427**	0.429	0.431	0.429	0.433	**0.420**	0.432	**0.420**	0.437
ETTm1	96	0.307	0.348	0.299	**0.341**	0.295	0.343	**0.289**	**0.341**	0.294	0.347
	192	0.357	0.375	0.333	0.366	0.325	0.363	**0.322**	**0.361**	0.333	0.367
	336	0.385	0.397	0.363	0.386	0.361	**0.385**	**0.359**	**0.385**	0.362	0.387
	720	0.440	0.434	0.421	0.426	**0.409**	**0.417**	0.421	0.421	0.423	0.420
	Avg.	0.372	0.388	0.354	0.380	**0.348**	**0.377**	**0.348**	**0.377**	0.353	0.380
Weather	96	0.156	0.195	0.152	0.195	0.148	**0.190**	0.149	0.195	**0.147**	0.192
	192	0.207	0.245	0.199	0.240	0.200	0.240	**0.192**	**0.237**	0.193	**0.237**
	336	0.267	0.288	0.254	0.281	0.253	0.283	0.248	0.280	**0.246**	**0.277**
	720	0.345	0.341	0.333	0.335	0.327	0.333	0.325	**0.332**	**0.322**	0.334
	Avg.	0.244	0.267	0.234	0.263	0.232	0.261	0.229	0.261	**0.227**	**0.260**
Traffic	96	0.410	0.259	0.382	0.249	0.386	0.261	0.383	0.256	**0.377**	**0.248**
	192	0.442	0.271	0.425	0.259	0.399	**0.254**	**0.385**	0.262	0.403	0.268
	336	0.457	0.274	0.432	**0.264**	0.414	0.267	0.413	0.266	**0.405**	0.268
	720	0.486	0.291	0.472	**0.285**	0.461	0.292	**0.449**	0.286	**0.449**	0.290
	Avg.	0.449	0.274	0.428	**0.264**	0.415	0.268	**0.408**	0.268	0.409	0.268

A.6 Visual Analysis

Figure A1 presents forecasting examples from seven benchmark datasets, with three channels shown for each dataset. Across diverse temporal patterns, FIRE captures recurring patterns, overall trends, and local variations in the forecast windows. The predictions generally follow the corresponding ground-truth trajectories across datasets and channels. These visualizations complement the aggregate forecasting results reported in Table 1.

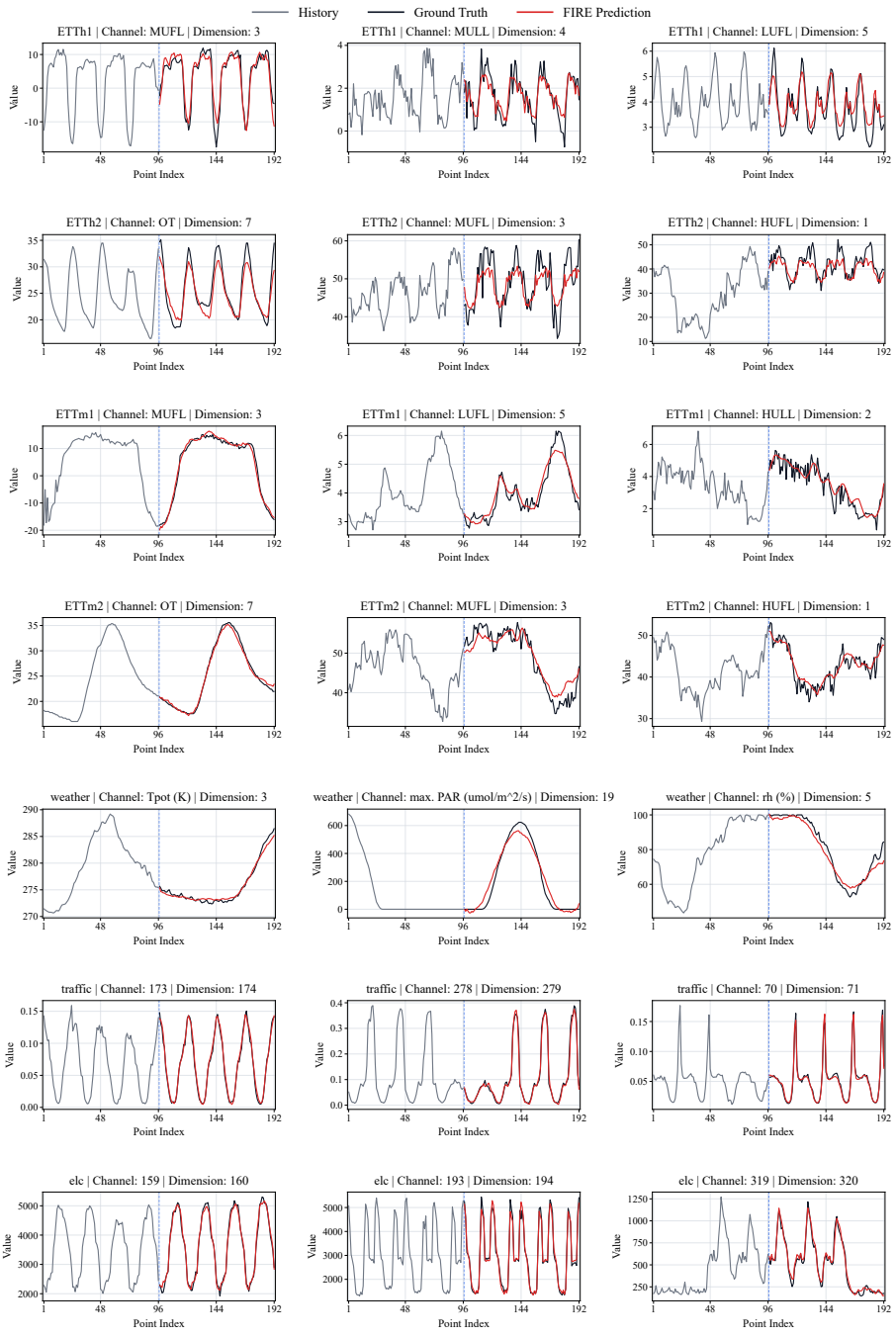

Figure A1: Forecasting examples on ETTh1, ETTh2, ETTm1, ETTm2, Weather, Traffic, and Electricity (top to bottom). Gray denotes the 96-point history; black and red show the next 96 ground-truth values and FIRE predictions. The dashed line marks the forecast boundary.

A.7 Limitations and Future Work

FIRE models amplitude and phase separately and adapts their representations to spectral changes observed in the historical window. The current formulation operates at a fixed temporal resolution. Spectral dynamics may vary across temporal scales, and amplitude and phase may respond differently to these changes. Future work can extend FIRE to multiscale spectral modeling and combine information from different temporal resolutions. This may provide a more complete view of spectral evolution while retaining the decoupled modeling of amplitude and phase.

Another direction is to extend FIRE to general forecasting models trained on heterogeneous time series. Such models need to capture shared temporal patterns while adapting to different datasets, sampling intervals, and forecasting tasks. The decoupled modeling of amplitude and phase provides a natural way to study which spectral patterns can be shared and which should be adapted to specific data characteristics. Future work can explore this principle in time series foundation models and other shared forecasting frameworks.