SFCATIME: SOFT SPECTRAL REPRESENTATION AND ADAPTIVE CROSS-MODAL FUSION FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time-series forecasting can benefit from complementary frequency-domain and language-derived representations, yet their effectiveness depends on how frequency structure is organized and how semantic information interacts with numerical features. A shared spectral projection leaves frequency specialization to downstream encoders, while standard cross-modal attention relies primarily on query–key affinity to determine numerical–semantic interactions. We propose **SFCATime**, an adaptive spectral–semantic forecasting framework that addresses both issues. First, SFCATime learns soft, overlapping assignments of frequency bins to latent spectral components and uses them to construct frequency-dependent corrections to a retained full-spectrum pathway. This preserves the complete magnitude spectrum while introducing spectral specialization without hard frequency boundaries. Second, **AdaptiveMask Cross-Modal Attention** augments base numerical–semantic affinity with a learnable signed relation correction before attention normalization, while adaptive branch aggregation and residual modulation separately control the contribution of the aligned representation. The pretrained language model remains frozen throughout forecasting training. Across eight widely used multivariate forecasting benchmarks, SFCATime achieves the best or tied-best average MAE on all eight datasets and the best average MSE on six. Compared with the most directly related multimodal baseline, it obtains lower average MSE and MAE on seven datasets each. Ablation studies and representation analyses further support the complementary effects of adaptive spectral representation and relation-aware cross-modal regulation.

1 INTRODUCTION

Multivariate time series forecasting is fundamental to applications such as electricity planning, weather monitoring, and economic analysis. Predictive structure may arise from local fluctuations, recurring patterns, long-range periodicity, and dependencies across variables. Accordingly, recent models exploit complementary views of historical observations: patch-based and variate-token methods emphasize temporal and cross-variable structure (Nie et al., 2023; Liu et al., 2024b), frequency-domain approaches expose periodic behavior (Zhou et al., 2022; Wu et al., 2023), and language-assisted methods introduce pretrained semantic representations (Jin et al., 2024; Liu et al., 2025a). The challenge is therefore not simply to combine more views, but to determine *how each view should be structured* and *how information should be selectively exchanged across them*.

The first challenge concerns spectral representation. A shared projection across Fourier magnitudes preserves the full spectrum but leaves frequency specialization to subsequent layers. Hard frequency partitions add structure, but impose fixed boundaries between neighboring frequencies. A more flexible representation should preserve all magnitude bins while allowing frequency regions to acquire distinct but smoothly varying representations before encoding.

The second challenge concerns cross-modal interaction. Prompt representations summarize the observed history in another feature space, but their relevance need not be uniform across numerical patterns. Standard cross-modal attention relies primarily on query–key affinity, although representational similarity does not necessarily imply forecasting relevance. Moreover, fusion involves two

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.

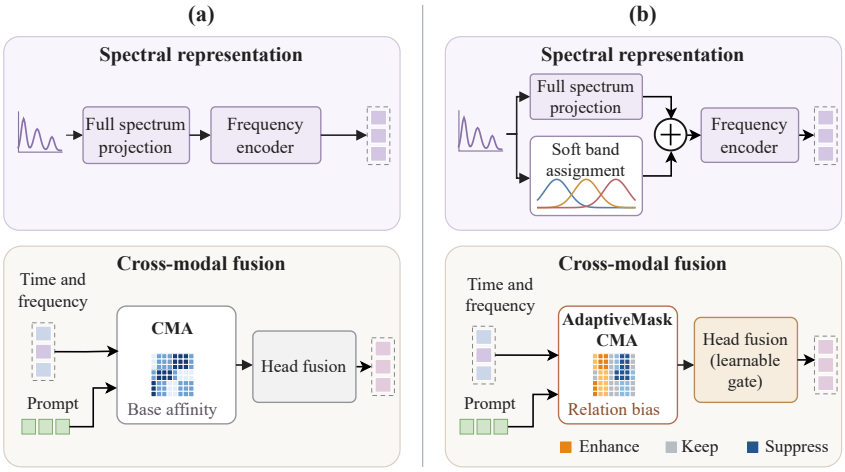

Figure 1: Motivation of SFCATime. (a) Conventional designs use a shared full-spectrum projection and base cross-modal affinity. (b) SFCATime introduces soft frequency-band assignment before spectral encoding and AdaptiveMask during cross-modal alignment. The former provides frequency-dependent structure without hard partitioning, while the latter applies signed corrections to numerical–semantic relation logits before attention normalization.

distinct decisions: *which numerical–semantic relations should be emphasized* and *how strongly the aligned representation should modify the numerical pathway*.

We propose **SFCATime**, an adaptive spectral–semantic forecasting framework that addresses these two issues (Figure 1). Its **soft spectral representation** assigns each frequency bin learnable memberships over latent spectral components and adds the resulting frequency-dependent correction to a retained full-spectrum pathway, preserving spectral coverage without hard boundaries. Its **AdaptiveMask Cross-Modal Attention** augments base numerical–semantic affinity with a learned signed relation correction before softmax, while adaptive branch aggregation and residual modulation separately control the contribution of aligned semantic information. Our contributions are:

- We introduce a **soft spectral representation** that provides frequency-dependent structure before encoding while preserving the complete magnitude spectrum.
- We propose **AdaptiveMask Cross-Modal Attention**, which regulates numerical–semantic relations through a learnable signed correction and separates relation-level alignment from residual fusion strength.
- Across eight multivariate forecasting benchmarks, **SFCATime** achieves the best or tied-best average MAE on all eight datasets and the best average MSE on six; ablation and representation analyses support the complementary effects of the two proposed components.

2 Related Work

Temporal and multivariate forecasting. Time-series forecasting models capture long-range structure through efficient attention, decomposition, tokenization, and multiscale modeling. Representative approaches include Informer, Autoformer, DLinear, PatchTST, iTransformer, and TimeMixer (Zhou et al., 2021; Wu et al., 2021; Zeng et al., 2023; Nie et al., 2023; Liu et al., 2024b; Wang et al., 2024). SFCATime is complementary to these numerical architectures: rather than redesigning the temporal backbone, it focuses on spectral representation and semantic interaction.

Frequency-domain forecasting. Frequency information has been exploited through frequency-enhanced attention, dominant-period modeling, frequency-domain mappings, interpolation, and frequency-domain supervision (Zhou et al., 2022; Wu et al., 2023; Yi et al., 2023; Xu et al., 2024; Wang et al., 2025). SFCATime instead targets *pre-encoding spectral structure*: each frequency

bin receives soft memberships over latent components, producing frequency-dependent corrections while retaining the complete magnitude spectrum.

Language-assisted forecasting. Pretrained language models have been incorporated through backbone adaptation, input reprogramming, language-guided forecasting, and cross-modal alignment (Zhou et al., 2023; Jin et al., 2024; Liu et al., 2024a; 2025b;a). T3Time further combines spectral encoding with horizon-aware temporal–frequency fusion and adaptive head aggregation (Chowdhury et al., 2025). Related multimodal architectures also study directional attention and gated interactions (Tsai et al., 2019; Lin et al., 2025). SFCATime differs by explicitly modifying cross-modal relation scores through a signed bias and separately regulating the strength of the aligned residual.

3 METHOD

3.1 OVERVIEW

Given a multivariate history $\mathbf{X} \in \mathbb{R}^{L\times N}$, our goal is to predict $\mathbf{Y} \in \mathbb{R}^{P\times N}$, where L, P, and N denote the input length, forecasting horizon, and number of variables, respectively. We omit the batch dimension for clarity. We first apply reversible instance normalization (RevIN) (Kim et al., 2022) and transpose the normalized history to $\widetilde{\mathbf{X}} \in \mathbb{R}^{N\times L}$. A linear history projection followed by a Transformer encoder (Vaswani et al., 2017) produces the temporal representation $\mathbf{Z}_t \in \mathbb{R}^{N\times C}$. Following the prompt construction protocol of TimeCMA (Liu et al., 2025a), prompts are generated exclusively from the observed history. A frozen language model provides cached embeddings, which are refined by a trainable prompt encoder to obtain $\mathbf{Z}_p \in \mathbb{R}^{N\times E}$. No future observations are provided to the language branch. We retain the temporal encoder, horizon-conditioned temporal–spectral fusion, and adaptive alignment aggregation of the underlying multimodal forecasting backbone (Chowdhury et al., 2025). SFCATime modifies two stages of information flow: it introduces frequency-dependent structure *before* spectral encoding and explicitly regulates numerical–semantic relations *during* cross-modal alignment.

3.2 SOFT SPECTRAL REPRESENTATION

Full-spectrum pathway. We transform each normalized variable into the frequency domain and retain its magnitude spectrum:

$$\mathbf{F} = \left|\mathrm{RFFT}(\widetilde{\mathbf{X}})\right| \in \mathbb{R}^{N\times L_f}, \qquad L_f = \left\lfloor \frac{L}{2} \right\rfloor + 1. \tag{1}$$

The full-spectrum pathway retains every magnitude bin. A shared affine projection $\phi_{\mathrm{full}} : \mathbb{R} \to \mathbb{R}^C$ maps each magnitude into the latent space.

Soft frequency-band assignment. Although the shared projection preserves spectral coverage, it treats frequency bins identically before encoding. To introduce frequency-dependent structure without imposing hard boundaries, each bin ℓ is assigned soft memberships over R latent spectral components:

$$s_{\ell r} = \frac{\exp(A_{\ell r})}{\sum_{j=1}^{R} \exp(A_{\ell j})}, \qquad v_{n\ell r} = F_{n\ell} s_{\ell r}, \tag{2}$$

where $\mathbf{A} \in \mathbb{R}^{L_f\times R}$ contains learnable assignment logits. We use $R = 3$. The assignments are shared across samples and variables, while input-specific spectral variation is carried by $F_{n\ell}$. Since $\sum_r s_{\ell r} = 1$, each magnitude is distributed across the latent components rather than removed.

We initialize the assignments with broad low-, middle-, and high-frequency priors and optimize them jointly with the forecasting objective. The component-weighted magnitudes are projected into the latent space and added as a correction to the retained full-spectrum token:

$$\mathbf{U}_{n\ell} = \phi_{\mathrm{full}}(F_{n\ell}) + \lambda_f \mathbf{v}_{n\ell} \mathbf{W}_b, \qquad \lambda_f = \sigma(\eta_f), \tag{3}$$

where $\mathbf{v}_{n\ell} \in \mathbb{R}^R$ and $\mathbf{W}_b \in \mathbb{R}^{R\times C}$. We initialize $\mathbf{W}_b = \mathbf{0}$, so training starts from the full-spectrum pathway and gradually learns frequency-dependent corrections.

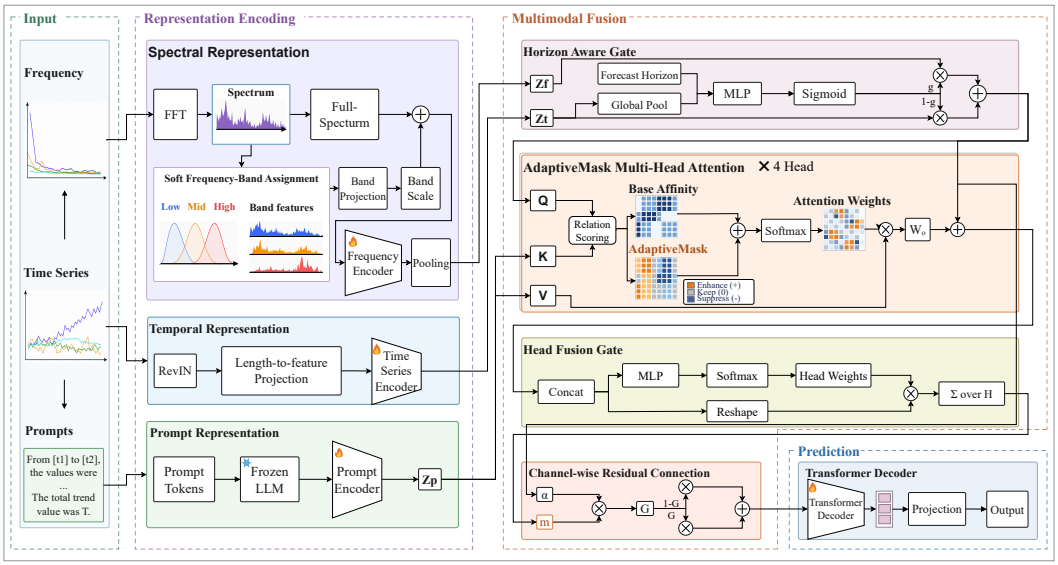

Figure 2: Architecture of SFCATime. Temporal, spectral, and prompt branches encode the observed history. Soft frequency-band assignment introduces frequency-dependent corrections before spectral encoding. A horizon-conditioned gate combines temporal and spectral representations, which query prompt features through parallel AdaptiveMask cross-modal alignment branches. Dynamic branch aggregation and residual modulation control the aligned semantic contribution before forecasting. The pretrained language model remains frozen.

A shared Transformer encoder processes the resulting frequency tokens. Attention pooling then aggregates the encoded bins into $\mathbf{Z}_f \in \mathbb{R}^{N\times C}$. The two operations serve different roles: soft assignment determines *how a frequency is represented before encoding*, whereas attention pooling determines *how encoded frequencies contribute to the final spectral representation*.

Frequency distinguishability. The soft correction also addresses a structural symmetry of a position-agnostic spectral pathway. Consider a shared scalar projection followed by a Transformer without positional information and permutation-invariant pooling. For the corresponding mapping $\mathcal{G}_0$ and any permutation $\mathbf{\Pi}$ over frequency bins,

$$\mathcal{G}_0(\mathbf{\Pi f}) = \mathcal{G}_0(\mathbf{f}). \tag{4}$$

The shared projection and Transformer are permutation equivariant, whereas the final pooling is permutation invariant. Such a pathway therefore cannot distinguish spectra solely from the locations at which identical magnitudes occur.

Writing $\phi_{\text{full}}(f) = f\mathbf{w}_{\text{full}} + \mathbf{b}_{\text{full}}$, Equation 3 becomes

$$\mathbf{U}_{n\ell} = F_{n\ell}\mathbf{w}_\ell^{\text{eff}} + \mathbf{b}_{\text{full}}, \qquad \mathbf{w}_\ell^{\text{eff}} = \mathbf{w}_{\text{full}} + \lambda_f \sum_{r=1}^{R} s_{\ell r}\mathbf{W}_{b,r:}. \tag{5}$$

Because $\mathbf{w}_\ell^{\text{eff}}$ depends on the frequency index through the learned memberships, SFCATime can break this permutation symmetry whenever different bins acquire different effective projection directions. Frequency location is therefore explicitly available before the shared spectral encoder. This is an architectural property rather than a guarantee that every pair of spectra produces distinct pooled representations.

3.3 Adaptive Cross-Modal Fusion

Temporal–spectral conditioning. We transpose the temporal and spectral representations to $\mathbf{H}_t, \mathbf{H}_f \in \mathbb{R}^{C\times N}$. Following the backbone, they are combined by a horizon-conditioned gate:

$$\mathbf{H}_g = \mathbf{g} \odot \mathbf{H}_f + (1-\mathbf{g}) \odot \mathbf{H}_t, \qquad \mathbf{g} = \sigma(f_g[\text{mean}_n(\mathbf{H}_t); P/1000]), \tag{6}$$

where f_g is a two-layer MLP and $\mathbf{g} \in \mathbb{R}^C$ is broadcast over variables. Because each forecasting configuration is trained with a fixed P, the horizon term provides configuration-level conditioning rather than sample-wise variation.

The prompt representation is transposed to $\mathbf{H}_p = \mathbf{Z}_p^\top \in \mathbb{R}^{E\times N}$. Numerical latent channels serve as queries, while prompt latent channels provide keys and values, using their profiles across the N variables as the shared feature dimension.

AdaptiveMask relation modulation. We use M independently parameterized cross-modal alignment branches, with $M = 4$ in our implementation. These are parallel alignment modules rather than the internal heads of a conventional multi-head attention layer. For branch $h \in \{1, \dots, M\}$, let

$$\mathbf{Q}_h \in \mathbb{R}^{C\times d}, \qquad \mathbf{K}_h, \mathbf{V}_h \in \mathbb{R}^{E\times d}, \qquad d = N. \tag{7}$$

Standard cross-modal attention forms the base affinity

$$\mathbf{S}_h = \frac{\mathbf{Q}_h \mathbf{K}_h^\top}{\sqrt{d}}. \tag{8}$$

AdaptiveMask augments this affinity with a learned relation-specific correction:

$$\mathbf{B}_h = \beta_h \tanh\left(\frac{(\mathbf{Q}_h \mathbf{R}_{Q,h})(\mathbf{K}_h \mathbf{R}_{K,h})^\top}{\sqrt{d}}\right), \qquad \mathbf{P}_h = \mathrm{softmax}_{\mathrm{key}}(\mathbf{S}_h + \mathbf{B}_h), \tag{9}$$

where $\mathbf{R}_{Q,h}, \mathbf{R}_{K,h} \in \mathbb{R}^{d\times d}$ are bias-free relation projections and β_h is a learnable scalar. We initialize $\beta_h = 0$, such that every branch starts from standard cross-modal attention.

AdaptiveMask is a soft relation modulator rather than a binary mask. Positive and negative entries of $\mathbf{B}_h$ represent corrections to the attention logits; they do not necessarily imply an absolute increase or decrease in the corresponding attention probability because softmax couples all keys within each query.

Relative attention control. Let $\mathbf{P}_h^0 = \mathrm{softmax}_{\mathrm{key}}(\mathbf{S}_h)$ denote the base attention. For query i and semantic keys j and k,

$$\log \frac{P_{h,ij}}{P_{h,ik}} - \log \frac{P^0_{h,ij}}{P^0_{h,ik}} = B_{h,ij} - B_{h,ik}. \tag{10}$$

Thus, AdaptiveMask explicitly changes the relative preference between semantic relations through differences in their relation corrections. A row-constant bias has no effect because it cancels under softmax.

Branch aggregation and residual modulation. After value aggregation, output projection, and residual transformations, branch h produces $\mathbf{O}_h \in \mathbb{R}^{C\times N}$. A learned softmax network assigns variable-specific branch weights $\pi_{n,h}$, yielding

$$\mathbf{\Lambda}_{:,n} = \sum_{h=1}^{M} \pi_{n,h} \mathbf{O}_{h,:,n}. \tag{11}$$

Relation selection and fusion strength are modeled separately. AdaptiveMask controls how prompt information is combined within each alignment branch, whereas a subsequent residual modulation determines how strongly the aggregated representation modifies the numerical pathway:

$$\mathbf{\Theta} = \mathbf{H}_g + (m\boldsymbol{\alpha}) \odot (\mathbf{\Lambda} - \mathbf{H}_g), \qquad m = \exp[-\mathrm{softplus}(\rho)] = \sigma(-\rho). \tag{12}$$

Here $\boldsymbol{\alpha} \in \mathbb{R}^C$ contains unconstrained channel-wise residual coefficients and $m \in (0, 1)$ provides a learned global modulation factor. Consequently, $m\boldsymbol{\alpha}$ should be interpreted as a residual coefficient rather than a convex gating probability.

Forecasting objective. Finally, $\mathbf{\Theta}^\top \in \mathbb{R}^{N \times C}$ is passed to a Transformer decoder and a linear forecasting head. Inverse RevIN restores the original scale, producing $\widehat{\mathbf{Y}} \in \mathbb{R}^{P \times N}$. All trainable forecasting components are optimized using

$$\mathcal{L} = \frac{1}{PN} \left\| \widehat{\mathbf{Y}} - \mathbf{Y} \right\|_F^2, \tag{13}$$

averaged over the batch. The pretrained language model and cached prompt embeddings remain fixed, and no auxiliary supervision is required.

4 Experiments

4.1 Experimental Setup

Datasets and metrics. We evaluate SFCATime on eight widely used multivariate forecasting benchmarks spanning energy, meteorology, finance, and public health. ETTh1/ETTh2 and ETTm1/ETTm2 contain seven electricity-transformer variables sampled hourly and every 15 minutes, respectively. ECL records hourly electricity consumption from 321 clients, Weather contains 21 meteorological variables sampled every 10 minutes, Exchange consists of eight daily exchange-rate series, and ILI contains seven weekly influenza-like-illness indicators. The input length is 96 for all datasets except ILI, where it is 36. Forecast horizons are $\{96, 192, 336, 720\}$ and $\{24, 36, 48, 60\}$ for ILI. We report mean squared error (MSE) and mean absolute error (MAE), with lower values indicating better forecasts.

Baselines and implementation. We compare with numerical forecasting methods DLinear, TimesNet, PatchTST, and iTransformer, together with language-assisted approaches GPT4TS/OFA, Time-LLM, UniTime, TimeCMA, and T3Time, introduced in Section 2. SFCATime uses cached representations from a frozen GPT-2 (Radford et al., 2019); only the downstream forecasting modules are optimized. Hyperparameters are selected separately for each dataset and forecasting horizon using Optuna (Akiba et al., 2019). Further implementation details and search spaces are provided in the **Appendix**.

4.2 Overall Forecasting Performance

Table 1 shows that SFCATime achieves the best or tied-best average MAE on all eight datasets and the best average MSE on six. The two exceptions in average MSE are ETTm2 and ECL, where TimeCMA and T3Time, respectively, obtain lower errors. Relative to T3Time, the most directly related multimodal baseline, SFCATime achieves lower average MSE on seven of eight datasets and lower average MAE on seven, with a tie on ETTm2. The magnitude of the improvement varies across datasets. On ETTh1, average MSE decreases from 0.4180 to 0.4070, a relative reduction of 2.6%. The gain is larger on ILI, where average MSE decreases from 1.7050 to 1.5440 (9.4%). In contrast, improvements on ETTm1, ETTm2, and Weather are comparatively modest. This pattern suggests that the proposed modifications do not simply shift performance on a single benchmark family, but their benefit depends on the spectral and cross-modal structure of the forecasting task. We next separate the two proposed components and examine the representations they learn.

4.3 Component Ablations

We evaluate the two proposed components on ETTh1, ETTh2, Weather, Exchange, and ILI across all four forecasting horizons. *w/o SSC* removes the soft spectral correction while retaining the full-spectrum encoder and cross-modal pathway. *w/o AM* removes the AdaptiveMask relation correction together with its associated global residual modulation, while retaining base cross-modal attention, adaptive branch aggregation, and channel-wise residual fusion. As shown in Table 2, SFCATime obtains lower dataset-level average MSE and MAE than both ablated variants on all five datasets. Removing SSC degrades the average results even though the full-spectrum pathway remains intact. This indicates that explicitly introducing frequency-dependent structure before the shared spectral encoder contributes beyond the underlying frequency representation itself. Removing the AdaptiveMask block also worsens the dataset-level averages while preserving base cross-modal attention and adaptive branch aggregation. Because this variant removes both the signed relation correction

Table 1: Multivariate forecasting results on eight benchmarks. H denotes the forecasting horizon and Avg averages the four horizons within each dataset. Lower MSE and MAE are better; bold denotes the best result, including ties.

Dataset	H	SFCATime		T3Time		TimeCMA		Time-LLM		UniTime		TimesNet		DLinear		iTrans.		PatchTST		GPT4TS	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	**0.3050**	**0.3510**	0.3080	0.3540	0.3120	**0.3510**	0.3590	0.3810	0.3220	0.3630	0.3380	0.3750	0.3450	0.3720	0.3340	0.3680	0.3440	0.3730	0.3350	0.3690
	192	**0.3530**	0.3790	0.3570	0.3810	0.3610	**0.3780**	0.3830	0.3930	0.3660	0.3870	0.3740	0.3870	0.3800	0.3890	0.3770	0.3910	0.3670	0.3860	0.3740	0.3850
	336	**0.3800**	**0.3990**	0.3820	0.4000	0.3920	0.4010	0.4160	0.4140	0.3980	0.4070	0.4100	0.4110	0.4130	0.4130	0.4260	0.4200	0.3920	0.4070	0.4070	0.4060
	720	**0.4410**	0.4380	0.4420	**0.4370**	0.4530	0.4380	0.4830	0.4490	0.4540	0.4400	0.4780	0.4500	0.4740	0.4530	0.4910	0.4590	0.4640	0.4420	0.4690	0.4420
	Avg	**0.3690**	**0.3910**	0.3720	0.3930	0.3800	0.3920	0.4100	0.4090	0.3850	0.3990	0.4000	0.4060	0.4030	0.4070	0.4070	0.4100	0.3920	0.4020	0.3960	0.4010
ETTm2	96	**0.1720**	0.2550	**0.1720**	**0.2540**	0.1730	0.2580	0.1930	0.2800	0.1830	0.2660	0.1870	0.2670	0.1930	0.2920	0.1800	0.2640	0.1770	0.2600	0.1900	0.2750
	192	0.2410	**0.2980**	**0.2370**	0.3000	0.2380	0.3010	0.2570	0.3180	0.2510	0.3100	0.2490	0.3090	0.2840	0.3620	0.2500	0.3090	0.2460	0.3050	0.2530	0.3130
	336	0.2970	0.3370	0.3060	0.3370	0.2970	0.3380	0.3170	0.3530	**0.1830**	**0.2660**	0.3190	0.3510	0.3690	0.4270	0.3110	0.3480	0.3050	0.3430	0.3210	0.3600
	720	0.3970	0.3980	0.4000	0.3980	**0.3930**	**0.3940**	0.4190	0.4110	0.4200	0.4100	0.4080	0.4030	0.5540	0.5220	0.4120	0.4070	0.4100	0.4050	0.4110	0.4060
	Avg	0.2770	**0.3220**	0.2790	**0.3220**	**0.2750**	0.3230	0.2960	0.3400	0.2930	0.3340	0.2910	0.3330	0.3500	0.4010	0.2880	0.3320	0.2850	0.3280	0.2940	0.3390
ETTh1	96	**0.3660**	0.3950	0.3710	0.3970	0.3730	**0.3910**	0.3980	0.4100	0.3970	0.4180	0.3840	0.4020	0.3860	0.4000	0.3860	0.4050	0.4040	0.4130	0.3980	0.4240
	192	**0.4000**	**0.4170**	0.4110	0.4210	0.4270	0.4210	0.4510	0.4400	0.4340	0.4390	0.4340	0.4290	0.4370	0.4320	0.4410	0.4360	0.4540	0.4300	0.4490	0.4270
	336	**0.4334**	**0.4368**	0.4480	0.4410	0.4580	0.4480	0.4750	0.4510	0.4680	0.4570	0.4910	0.4690	0.4810	0.4590	0.4870	0.4580	0.4970	0.4620	0.4920	0.4660
	720	**0.4280**	**0.4538**	0.4410	0.4600	0.4490	0.4600	0.4690	0.4700	0.4690	0.4770	0.5210	0.5000	0.5190	0.5160	0.5030	0.4910	0.4960	0.4810	0.4870	0.4830
	Avg	**0.4070**	**0.4250**	0.4180	0.4300	0.4230	0.4310	0.4480	0.4430	0.4420	0.4480	0.4580	0.4500	0.4560	0.4520	0.4540	0.4470	0.4630	0.4490	0.4570	0.4500
ETTh2	96	**0.2774**	0.3398	0.2780	0.3380	0.2860	**0.3360**	0.2950	0.3450	0.2960	0.3450	0.3400	0.3740	0.3330	0.3870	0.2970	0.3490	0.3120	0.3580	0.3120	0.3600
	192	**0.3460**	0.3890	0.3510	0.3890	0.3630	**0.3870**	0.3860	0.3990	0.3740	0.3940	0.4020	0.4140	0.4770	0.4760	0.3800	0.4000	0.3970	0.4080	0.3870	0.4050
	336	**0.3483**	**0.3938**	0.3580	0.3980	0.4060	0.4210	0.4190	0.4290	0.4150	0.4270	0.4520	0.4520	0.5940	0.5410	0.4280	0.4320	0.4350	0.4400	0.4240	0.4370
	720	**0.3970**	**0.4300**	0.4040	0.4330	0.4170	0.4380	0.4250	0.4420	0.4250	0.4440	0.4620	0.4680	0.8310	0.6570	0.4270	0.4450	0.4360	0.4490	0.4330	0.4530
	Avg	**0.3420**	**0.3880**	0.3480	0.3900	0.3720	0.3970	0.3810	0.4040	0.3780	0.4030	0.4140	0.4270	0.5590	0.5150	0.3830	0.4070	0.3950	0.4140	0.3890	0.4140
ECL	96	0.1420	**0.2290**	**0.1380**	0.2330	0.1430	0.2380	0.1720	0.2650	0.1960	0.2870	0.1680	0.2720	0.1970	0.2820	0.1480	0.2400	0.1860	0.2690	0.1970	0.2900
	192	0.1560	**0.2440**	**0.1550**	0.2500	0.1610	0.2590	0.1820	0.2790	0.1990	0.2910	0.1840	0.2890	0.1960	0.2850	0.1620	0.2530	0.1900	0.2730	0.2010	0.2920
	336	0.1840	0.2640	**0.1680**	0.2650	0.1690	**0.2610**	0.1950	0.2880	0.2140	0.3050	0.1980	0.3000	0.2090	0.3010	0.1780	0.2690	0.2060	0.2900	0.2170	0.3090
	720	0.2420	0.3180	**0.2180**	**0.3140**	0.2190	0.3150	0.2330	0.3200	0.2540	0.3350	0.2200	0.3200	0.2450	0.3330	0.2250	0.3170	0.2470	0.3220	0.2530	0.3390
	Avg	0.1810	**0.2640**	**0.1700**	0.2660	0.1740	0.2690	0.1950	0.2880	0.2160	0.3060	0.1920	0.2950	0.2120	0.3000	0.1780	0.2700	0.2070	0.2890	0.2170	0.3080
Weather	96	**0.1590**	**0.2050**	0.1620	0.2100	0.1670	0.2110	0.1980	0.2350	0.1710	0.2140	0.1720	0.2200	0.1960	0.2550	0.1740	0.2140	0.1770	0.2180	0.2030	0.2440
	192	**0.2072**	**0.2522**	0.2110	0.2530	0.2120	0.2530	0.2400	0.2690	0.2170	0.2540	0.2190	0.2610	0.2370	0.2960	0.2210	0.2540	0.2220	0.2590	0.2470	0.2770
	336	**0.2660**	0.2937	0.2670	0.2930	0.2700	**0.2920**	0.2950	0.3080	0.2740	0.2930	0.2800	0.3060	0.2830	0.3350	0.2780	0.2960	0.2770	0.2970	0.2970	0.3110
	720	0.3407	0.3432	**0.3350**	0.3460	0.3500	0.3480	0.3680	0.3530	0.3510	**0.3430**	0.3650	0.3590	0.3450	0.3810	0.3580	0.3490	0.3520	0.3470	0.3680	0.3560
	Avg	**0.2430**	**0.2730**	0.2440	0.2750	0.2500	0.2760	0.2750	0.2910	0.2530	0.2760	0.2590	0.2870	0.2650	0.3170	0.2580	0.2780	0.2570	0.2800	0.2790	0.2970
ILI	24	**1.3098**	**0.7330**	1.5830	0.8020	1.9960	0.9980	2.3830	1.0040	2.3460	0.9540	2.3170	0.9340	2.3980	1.0400	2.3470	1.7310	2.3350	0.9890	2.7320	1.1000
	36	**1.5620**	**0.8000**	1.6010	0.8200	1.9060	0.9150	2.3900	0.9930	1.9980	0.9120	1.9720	0.9200	2.6460	1.0880	2.4680	0.9980	2.5610	1.0350	2.6640	1.0630
	48	**1.5030**	**0.7890**	1.7180	0.8150	1.8670	0.8680	2.3940	1.0030	1.9790	0.9120	2.2380	0.9130	2.6140	1.0860	2.4890	1.0160	2.4650	1.0220	2.6170	1.0410
	60	**1.8020**	**0.8810**	1.9200	0.9010	1.9200	0.9040	2.5620	1.0490	2.1090	0.9380	2.0270	0.9280	2.8040	1.1460	2.4710	1.0650	2.1890	0.9970	2.4780	1.0350
	Avg	**1.5440**	**0.8010**	1.7050	0.8350	1.9220	0.9210	2.4320	1.0120	2.1080	0.9290	2.1390	0.9310	2.6160	1.0900	2.4440	1.2030	2.3880	1.0110	2.6230	1.0600
Exchange	96	**0.0833**	**0.2020**	0.0850	0.2050	0.0990	0.2240	0.0870	0.2080	0.0860	0.2090	0.1070	0.2340	0.0880	0.2180	0.0860	0.2060	0.1090	0.2360	0.1480	0.2780
	192	**0.1680**	**0.2940**	0.1720	0.2960	0.1860	0.3120	0.1730	0.2990	0.1740	0.2990	0.2260	0.3440	0.1760	0.3150	0.1770	0.2990	0.2050	0.3270	0.2710	0.3800
	336	**0.3050**	**0.4000**	0.3180	0.4080	0.3640	0.4440	0.3750	0.4540	0.3190	0.4080	0.3670	0.4480	0.3130	0.4270	0.3310	0.4170	0.3560	0.4360	0.4600	0.5000
	720	**0.8280**	**0.6910**	0.8360	0.6960	0.9320	0.7350	0.8530	0.7030	0.8750	0.7010	0.9640	0.7460	0.8390	0.6950	0.8470	**0.6910**	0.8880	0.7160	1.1950	0.8410
	Avg	**0.3460**	**0.3960**	0.3530	0.4010	0.3950	0.4290	0.3720	0.4160	0.3640	0.4040	0.4160	0.4430	0.3540	0.4140	0.3600	0.4030	0.3900	0.4290	0.5190	0.5000

and its associated global residual modulation, the comparison supports the *combined cross-modal regulation mechanism*; it should not be interpreted as isolating the signed bias alone. The effect is particularly pronounced on ILI and Exchange, where removing the cross-modal regulation block produces substantially larger average errors. At individual horizons, an ablated variant occasionally matches or slightly outperforms the full model, for example on some ETTh2 and Weather settings. The advantage of the complete model is therefore better characterized by its consistent dataset-level averages across horizons than by uniform dominance at every individual horizon. Taken together, the ablations support complementary roles for the two components: SSC changes *how frequency information is represented*, whereas AdaptiveMask-based regulation changes *how aligned semantic information enters the numerical pathway*.

4.4 Mechanism Analysis

Figure 3 visualizes the learned spectral assignments, signed relation corrections, and corresponding attention changes on Weather, ILI, and Exchange.

Task-dependent spectral adaptation. The learned assignments exhibit different degrees of deviation from their initialization: Weather changes substantially, ILI shows moderate adjustment, and Exchange remains close to the initial priors. The memberships remain overlapping rather than collapsing into hard intervals. According to Equation 5, these mixtures induce frequency-dependent projection directions before the shared encoder, allowing neighboring frequencies to share structure

Table 2: Ablation study on five representative datasets. *w/o SSC* removes the soft spectral correction; *w/o AM* removes AdaptiveMask and its associated global residual modulation. H and Avg follow Table 1. Lower is better and bold denotes the best result, including ties.

Dataset	H	SFCATime		w/o SSC		w/o AM	
		MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.3660**	**0.3950**	0.3674	0.3953	0.3745	0.4018
	192	**0.4000**	**0.4170**	0.4067	0.4201	0.4098	0.4237
	336	**0.4334**	**0.4368**	0.4364	0.4399	0.4373	0.4410
	720	**0.4280**	**0.4538**	0.4398	0.4607	0.4457	0.4628
	Avg	**0.4070**	**0.4250**	0.4126	0.4290	0.4168	0.4323
ETTh2	96	**0.2774**	**0.3398**	0.2834	0.3429	0.2780	0.3405
	192	**0.3460**	0.3890	0.3617	0.3902	0.3517	**0.3843**
	336	0.3483	**0.3938**	**0.3470**	0.3939	0.3600	0.3998
	720	0.3970	0.4300	**0.3959**	**0.4265**	0.4114	0.4354
	Avg	**0.3420**	**0.3880**	0.3470	0.3884	0.3503	0.3900
Weather	96	**0.1590**	**0.2050**	0.1654	0.2180	0.1742	0.2207
	192	**0.2072**	**0.2522**	0.2073	0.2578	0.2081	0.2602
	336	0.2660	**0.2937**	**0.2650**	0.2996	0.2681	0.2959
	720	0.3407	0.3432	0.3429	0.3427	**0.3363**	**0.3420**
	Avg	**0.2430**	**0.2730**	0.2452	0.2795	0.2467	0.2797
Exchange	96	**0.0833**	**0.2020**	0.0839	0.2048	0.0837	0.2047
	192	0.1680	0.2940	**0.1679**	**0.2938**	0.1695	0.2947
	336	**0.3050**	**0.4000**	0.3321	0.4182	0.3138	0.4059
	720	**0.8280**	**0.6910**	0.8752	0.7104	0.8935	0.7210
	Avg	**0.3460**	**0.3960**	0.3648	0.4068	0.3651	0.4066
ILI	24	**1.3098**	**0.7330**	1.3751	0.7764	1.5357	0.8245
	36	**1.5620**	**0.8000**	1.5908	0.8076	1.7269	0.8448
	48	**1.5030**	**0.7890**	1.6672	0.8206	1.9235	0.8999
	60	**1.8020**	**0.8810**	1.8080	0.8814	1.8192	0.8835
	Avg	**1.5440**	**0.8010**	1.6103	0.8215	1.7513	0.8632

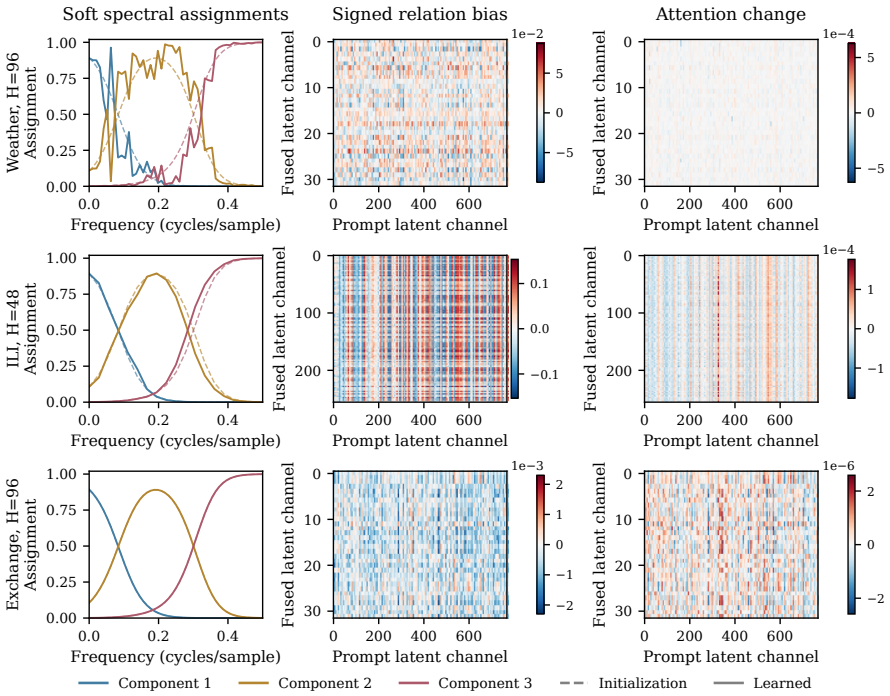

Figure 3: Learned spectral and cross-modal behavior on Weather ($H = 96$), ILI ($H = 48$), and Exchange ($H = 96$). Left: initial (dashed) and learned (solid) soft frequency assignments. Middle: learned signed relation bias. Right: attention change relative to base attention. Heatmaps average 32 test windows and the inspected alignment branches; each uses its own symmetric color scale.

while preserving the full-spectrum pathway. The assignments are learned per forecasting configuration and therefore represent task-level, rather than instance-wise, adaptation.

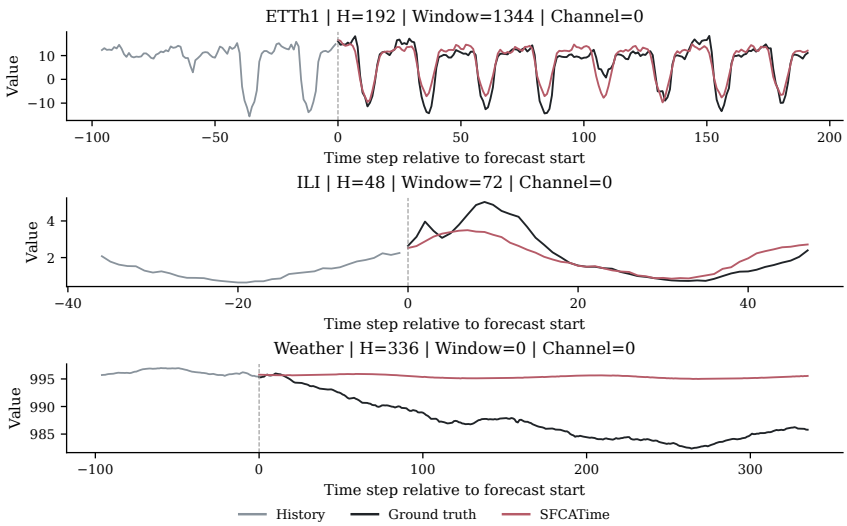

Figure 4: Forecasting examples on ETTh1 ($H = 192$, window 1344), ILI ($H = 48$, window 72), and Weather ($H = 336$, window 0). Gray, black, and red denote history, ground truth, and SFCATime prediction; dashed lines mark the forecasting boundary. Channel 0 is shown in the original scale.

Relation-level cross-modal modulation. AdaptiveMask learns both positive and negative relation corrections, while the resulting changes in attention probability remain comparatively small, especially on Exchange. This indicates that AdaptiveMask refines rather than replaces the base affinity. Equation 10 explains this behavior: attention redistribution depends on differences between relation biases within each query row, whereas a row-constant correction vanishes under softmax. The distinct patterns across datasets further suggest task-dependent modulation rather than a fixed global rescaling. These visualizations characterize how attention is modified, not the causal importance of individual attention entries (Jain & Wallace, 2019); predictive utility is established by the ablation results in Table 2.

4.5 Forecasting Cases and Failure Modes

Figure 4 illustrates both successful forecasts and a representative failure mode. On ETTh1, SFCATime captures the timing of recurrent oscillations over the 192-step horizon, although some extrema are smoothed. On ILI, it follows the broad decline and subsequent recovery, but underestimates the early peak. In contrast, on Weather the prediction remains near the recent level while the target enters a sustained downward trend. This failure suggests that SFCATime can struggle with persistent regime changes that are weakly indicated by the observed history, highlighting non-stationarity as an important direction for future work.

5 Conclusion

We presented **SFCATime**, a spectral–semantic framework for multivariate time series forecasting that introduces adaptivity both before spectral encoding and during cross-modal interaction. Soft frequency-band assignments provide frequency-dependent specialization while preserving the full spectrum, and AdaptiveMask regulates numerical–semantic relations through learned signed corrections, with residual modulation separately controlling their influence on forecasting. Across eight benchmarks, SFCATime achieves the best or tied-best average MAE on all eight datasets and the best average MSE on six. Ablation studies and representation analyses further show that adaptive spectral modeling and relation-aware cross-modal fusion provide complementary benefits. Our qualitative analysis also reveals remaining difficulty under sustained distribution changes, motivating future work on more explicit modeling of non-stationarity and regime shifts.

REFERENCES

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019. URL https://arxiv.org/abs/1907.10902.

Abdul Monaf Chowdhury, Rabeya Akter, and Safaeid Hossain Arib. T3Time: Tri-modal time series forecasting via adaptive multi-head alignment and residual fusion. *arXiv preprint arXiv:2508.04251*, 2025. URL https://arxiv.org/abs/2508.04251.

Sarthak Jain and Byron C. Wallace. Attention is not explanation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3543–3556. Association for Computational Linguistics, 2019. doi: 10.18653/v1/N19-1357. URL https://aclanthology.org/N19-1357/.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations*, 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. URL https://openreview.net/forum?id=cGDAkQo1C0p.

Jinhao Lin, Yifei Wang, Yanwu Xu, and Qi Liu. Semi-IIN: Semi-supervised intra-inter modal interaction learning network for multimodal sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 1411–1419, 2025. doi: 10.1609/aaai.v39i2.32131.

Chenxi Liu, Qianxiong Xu, Hao Miao, Sun Yang, Lingzheng Zhang, Cheng Long, Ziyue Li, and Rui Zhao. TimeCMA: Towards LLM-empowered multivariate time series forecasting via cross-modality alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18780–18788, 2025a. doi: 10.1609/aaai.v39i18.34067.

Peiyuan Liu, Hang Guo, Tao Dai, Naiqi Li, Jigang Bao, Xudong Ren, Yong Jiang, and Shu-Tao Xia. CALF: Aligning LLMs for time series forecasting via cross-modal fine-tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18915–18923, 2025b. doi: 10.1609/aaai.v39i18.34082.

Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. UniTime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM Web Conference 2024*, pp. 4095–4106, 2024a. doi: 10.1145/3589334.3645434.

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*, 2024b.

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.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. Technical report, OpenAI, 2019. URL https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

Yao-Hung Hubert Tsai, Shaojie Bai, Paul Pu Liang, J. Zico Kolter, Louis-Philippe Morency, and Ruslan Salakhutdinov. Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 6558–6569, 2019. doi: 10.18653/v1/P19-1656.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://arxiv.org/abs/1706.03762.

Hao Wang, Licheng Pan, Zhichao Chen, Degui Yang, Sen Zhang, Yifei Yang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. FreDF: Learning to forecast in the frequency domain. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2402.02399.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/a7ac8a21e5a27e7ab31a5f42a0117bdb-Abstract-Conference.html.

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*, volume 34, 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.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with 10k parameters. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/701251e1db4a2e4dd2ef23f5265d5936-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 *Advances in Neural Information Processing Systems*, volume 36, pp. 76656–76679, 2023.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023. doi: 10.1609/aaai.v37i9.26317.

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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021. doi: 10.1609/aaai.v35i12.17325.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 27268–27286, 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all: Power general time series analysis by pretrained LM. In *Advances in Neural Information Processing Systems*, volume 36, pp. 43322–43355, 2023.

A Experimental and Implementation Details

A.1 Datasets and Evaluation

Table 3 summarizes the dataset configurations. We use P for the forecast horizon, consistent with the method section; H in the main experimental tables denotes the same quantity. Each dataset–horizon pair is trained separately.

Table 3: Dataset configurations. N: number of variables; L: history length; P: forecasting horizon.

Dataset	N	Sampling interval	L	P
ETTh1	7	1 hour	96	96, 192, 336, 720
ETTh2	7	1 hour	96	96, 192, 336, 720
ETTm1	7	15 minutes	96	96, 192, 336, 720
ETTm2	7	15 minutes	96	96, 192, 336, 720
ECL	321	1 hour	96	96, 192, 336, 720
Weather	21	10 minutes	96	96, 192, 336, 720
Exchange	8	1 day	96	96, 192, 336, 720
ILI	7	1 week	36	24, 36, 48, 60

The data-loader implementation uses chronological splits. ETTh uses 8,640 training time points followed by 2,880 validation and 2,880 test time points; ETTm uses 34,560, 11,520, and 11,520, respectively. These are time-point counts, not counts of forecast windows. For the custom datasets, the training and test segments contain $\lfloor 0.7T \rfloor$ and $\lfloor 0.2T \rfloor$ time points, respectively, and the remaining points form validation, where T is the series length. Validation and test inputs include the preceding L observations as historical context. The data scaler is fitted on the training segment.

The supplied training script uses non-shuffled loaders with `drop_last=True` for training, validation, and test. The final incomplete batch is therefore omitted. This differs from the mechanism-analysis protocol, which selects windows from the complete test-window index range. These settings are reported explicitly because the batch size can affect the set of windows retained for test evaluation.

For S evaluated windows, errors are averaged over windows, variables, and forecast steps:

$$\mathrm{MSE} = \frac{1}{SPN} \sum_{s,t,n} (\widehat{y}_{s,t,n} - y_{s,t,n})^2, \qquad \mathrm{MAE} = \frac{1}{SPN} \sum_{s,t,n} |\widehat{y}_{s,t,n} - y_{s,t,n}|. \tag{14}$$

The reported Avg entries average the metric over the four horizons, with equal weight per horizon. They are distinct from pooling forecast steps across horizons. Forecast plots are converted to original units; they should not be read as plots of the numerical scale used by the standardized benchmark metrics.

A.2 Optimization and Model Selection

Forecasting modules are optimized with AdamW and MSE loss. Gradient norms are clipped at 5. The learning-rate scheduler is `CosineAnnealingLR`, stepped once per epoch, with $T_{\max} = \min(E_{\max}, 50)$ and minimum learning rate 10^{-6}, where $E_{\max}$ is the configured maximum epoch count. Cached GPT-2 embeddings have dimension 768 and remain fixed. The Transformer attention-head setting is 8; it is distinct from the four independently parameterized cross-modal alignment branches.

Selection protocol used by the supplied implementation. The tuning objective minimizes test MSE. Within a training run, improved validation loss triggers checkpoint saving during the first ten epochs. After epoch ten, validation loss below the stored threshold triggers test evaluation, and the checkpoint is updated only when test MSE improves. The validation threshold is updated with an accepted checkpoint. Thus, test performance participates in both checkpoint selection and hyperparameter selection. Reported results consequently follow this test-informed selection protocol.

Early stopping is checked after an epoch when the no-update counter reaches the configured patience and at least half of the maximum epoch count has elapsed. The tuning wrapper defaults to

Table 4: Categorical search spaces encoded in the supplied tuning script. B: batch size; C: numerical latent dimension; D_e, D_d: encoder and decoder depths. Braces indicate discrete candidates.

Dataset	P	B	C	D_e	D_d
ETTh1	96	256	$\{128, 256\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTh1	192	32	$\{128, 256\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTh1	336	16	$\{128, 256\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTh1	720	32	$\{128, 256\}$	$\{2, 3, 4, 5\}$	$\{2, 3, 4, 5\}$
ETTh2	96	256	$\{128, 256\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTh2	192	256	$\{64, 128\}$	$\{2, 3, 4\}$	$\{4, 5, 6\}$
ETTh2	336	256	$\{64, 128\}$	$\{1, 2, 3, 4\}$	$\{3, 4, 5, 6\}$
ETTh2	720	256	$\{64, 256\}$	$\{1, 2, 3, 4\}$	$\{3, 4, 5, 6\}$
ETTm1	96	64	$\{128, 256\}$	$\{1, 2\}$	$\{2, 3, 4\}$
ETTm1	192	32	$\{64, 128\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTm1	336	32	$\{64, 128\}$	$\{1, 2\}$	$\{1, 2, 3\}$
ETTm1	720	16	$\{64, 128\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTm2	96	16	$\{64, 128\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ETTm2	192	16	$\{64, 128\}$	$\{1, 2\}$	$\{1, 2\}$
ETTm2	336	16	$\{64, 128\}$	$\{1, 2\}$	$\{1, 2\}$
ETTm2	720	16	$\{16, 32\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ECL	96	8	$\{64, 128, 256\}$	$\{3, 4, 5\}$	$\{5, 6, 7\}$
ECL	192	8	$\{64, 128, 256\}$	$\{3, 4, 5\}$	$\{5, 6, 7\}$
ECL	336	$\{4, 8\}$	$\{64, 128, 256\}$	$\{3, 4, 5\}$	$\{5, 6, 7\}$
ECL	720	$\{4, 8\}$	$\{64, 128, 256\}$	$\{3, 4, 5\}$	$\{5, 6, 7\}$
Weather	96	64	$\{16, 32\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
Weather	192	64	$\{16, 32, 64\}$	$\{2, 3, 4\}$	$\{2, 3, 4\}$
Weather	336	32	$\{16, 32, 64\}$	$\{1, 2, 3\}$	$\{2, 3, 4\}$
Weather	720	$\{64, 128\}$	$\{64, 128, 256\}$	$\{1, 2\}$	$\{4, 5, 6, 7\}$
Exchange	96	$\{64, 128, 256\}$	$\{16, 32, 64\}$	$\{1, 2\}$	$\{1, 2, 3\}$
Exchange	192	$\{64, 128, 256\}$	$\{32, 64\}$	$\{1, 2\}$	$\{2, 3\}$
Exchange	336	$\{64, 128, 256\}$	$\{32, 64, 128\}$	$\{1, 2, 3\}$	$\{2, 3, 4\}$
Exchange	720	40	$\{16, 64\}$	$\{1, 2, 3\}$	$\{2, 3, 4\}$
ILI	24	16	$\{32, 64\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ILI	36	16	$\{32, 64\}$	$\{1, 2, 3\}$	$\{1, 2\}$
ILI	48	32	$\{128, 256\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$
ILI	60	32	$\{64, 128\}$	$\{1, 2, 3\}$	$\{1, 2, 3\}$

a patience of 50 and passes it to the training script, whose standalone default is 25. The search wrapper defaults to 30 additional trials per invocation and resumes an existing SQLite study when present. Consequently, 30 is not necessarily the total number of trials in a completed study.

A.3 Hyperparameter Search Spaces

Tables 4 and 5 specify the spaces encoded in the supplied tuning script. Batch size, numerical latent dimension, encoder depth, and decoder depth are categorical choices. Learning rate and weight decay are sampled logarithmically, while dropout is sampled linearly. A singleton candidate set is fixed. These tables describe the search configuration; they are not lists of selected optimum values.

The additional categorical parameter `mask_strength_init` uses $\{0, 0.5, 0.7443967, 1\}$. In the archived model implementation, this initializes the global residual parameter ρ, giving $m_0 = \sigma(-\rho_0)$. It is not the initialization of the signed relation scale β_h, which starts at zero. This distinction separates global residual modulation from relation-level attention correction.

A.4 Mechanism-Analysis Checkpoints

Table 6 identifies the 20 models recorded in the mechanism-analysis archive. These recorded checkpoints connect the visualized spectral assignments and relation corrections to their training configurations. The accompanying configuration file additionally records maximum epochs, seeds, stopping patience, and checkpoint hashes for reproducibility.

Table 5: Continuous hyperparameter search spaces. Learning rate and weight decay use logarithmic sampling; dropout uses linear sampling.

Dataset	P	Learning rate	Dropout	Weight decay
ETTh1	96	$[5 \times 10^{-5}, 4 \times 10^{-4}]$	$[0.25, 0.55]$	$[1 \times 10^{-4}, 3 \times 10^{-3}]$
ETTh1	192	$[1 \times 10^{-4}, 3 \times 10^{-4}]$	$[0.4, 0.8]$	$[5 \times 10^{-5}, 1 \times 10^{-3}]$
ETTh1	336	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.5, 0.8]$	$[7 \times 10^{-5}, 1 \times 10^{-3}]$
ETTh1	720	$[5 \times 10^{-5}, 5 \times 10^{-4}]$	$[0.35, 0.7]$	$[8 \times 10^{-5}, 1 \times 10^{-3}]$
ETTh2	96	$[8 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.1, 0.4]$	$[5 \times 10^{-5}, 5 \times 10^{-4}]$
ETTh2	192	$[6 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.1, 0.4]$	$[1 \times 10^{-4}, 1 \times 10^{-3}]$
ETTh2	336	$[8 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.25, 0.55]$	$[8 \times 10^{-5}, 8 \times 10^{-4}]$
ETTh2	720	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.25, 0.6]$	$[1 \times 10^{-4}, 5 \times 10^{-3}]$
ETTm1	96	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.35, 0.65]$	$[8 \times 10^{-5}, 1 \times 10^{-3}]$
ETTm1	192	$[7 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.45, 0.75]$	$[1 \times 10^{-4}, 3 \times 10^{-3}]$
ETTm1	336	$[5 \times 10^{-5}, 2 \times 10^{-4}]$	$[0.45, 0.8]$	$[8 \times 10^{-5}, 1 \times 10^{-3}]$
ETTm1	720	$[3 \times 10^{-5}, 2 \times 10^{-4}]$	$[0.5, 0.8]$	$[1 \times 10^{-4}, 2 \times 10^{-3}]$
ETTm2	96	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.4, 0.75]$	$[1 \times 10^{-4}, 1 \times 10^{-3}]$
ETTm2	192	$[5 \times 10^{-5}, 2 \times 10^{-4}]$	$[0.3, 0.7]$	$[1 \times 10^{-4}, 3 \times 10^{-3}]$
ETTm2	336	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.5, 0.8]$	$[1 \times 10^{-4}, 3 \times 10^{-3}]$
ETTm2	720	$[5 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.35, 0.7]$	$[1 \times 10^{-4}, 4 \times 10^{-3}]$
ECL	96	$[5 \times 10^{-4}, 1.5 \times 10^{-3}]$	$[0.1, 0.4]$	$[1 \times 10^{-6}, 2 \times 10^{-3}]$
ECL	192	$[5 \times 10^{-4}, 1.5 \times 10^{-3}]$	$[0.1, 0.4]$	$[1 \times 10^{-6}, 2 \times 10^{-3}]$
ECL	336	$[1 \times 10^{-4}, 2 \times 10^{-3}]$	$[0.1, 0.4]$	$[1 \times 10^{-6}, 2 \times 10^{-3}]$
ECL	720	$[1 \times 10^{-4}, 2 \times 10^{-3}]$	$[0.1, 0.4]$	$[1 \times 10^{-6}, 2 \times 10^{-3}]$
Weather	96	$[4 \times 10^{-4}, 2 \times 10^{-3}]$	$[0.05, 0.3]$	$[2 \times 10^{-4}, 1 \times 10^{-3}]$
Weather	192	$[2 \times 10^{-4}, 2 \times 10^{-3}]$	$[0.15, 0.4]$	$[1 \times 10^{-5}, 3 \times 10^{-4}]$
Weather	336	$[6 \times 10^{-5}, 6 \times 10^{-4}]$	$[0.1, 0.3]$	$[1 \times 10^{-5}, 1 \times 10^{-4}]$
Weather	720	$[6 \times 10^{-5}, 6 \times 10^{-4}]$	$[0.1, 0.5]$	$[9 \times 10^{-6}, 4 \times 10^{-4}]$
Exchange	96	$[1 \times 10^{-4}, 1 \times 10^{-3}]$	$[0.05, 0.35]$	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
Exchange	192	$[5 \times 10^{-5}, 8 \times 10^{-4}]$	$[0.05, 0.35]$	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
Exchange	336	$[5 \times 10^{-5}, 5 \times 10^{-4}]$	$[0.05, 0.35]$	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
Exchange	720	$[3 \times 10^{-5}, 8 \times 10^{-5}]$	$[0.05, 0.3]$	$[3 \times 10^{-5}, 1 \times 10^{-4}]$
ILI	24	$[3 \times 10^{-5}, 2 \times 10^{-4}]$	$[0.02, 0.3]$	$[1 \times 10^{-5}, 3 \times 10^{-4}]$
ILI	36	$[8 \times 10^{-5}, 3 \times 10^{-4}]$	$[0.05, 0.3]$	$[8 \times 10^{-5}, 3 \times 10^{-4}]$
ILI	48	$[5 \times 10^{-4}, 4 \times 10^{-3}]$	$[0.3, 0.65]$	$[1 \times 10^{-5}, 1 \times 10^{-3}]$
ILI	60	$[3 \times 10^{-5}, 4 \times 10^{-4}]$	$[0.05, 0.3]$	$[1 \times 10^{-5}, 5 \times 10^{-4}]$

B Additional Architectural Details

B.1 Spectral Representation and Initialization

The spectral input contains $L_f = \lfloor L/2 \rfloor + 1$ magnitudes per variable: 49 bins for $L = 96$ and 19 bins for $L = 36$. Retaining the full spectrum refers to magnitude coverage, not Fourier phase. The $R = 3$ assignment components are shared across samples and variables within each trained model. Input-specific variation enters through $v_{n\ell r} = F_{n\ell} s_{\ell r}$.

The initialization used for the spectral priors is

$$A_{\ell r} = -\left(\frac{u_\ell - \mu_r}{\tau_r}\right)^2, \quad u_\ell = \frac{\ell - 1}{L_f - 1}, \quad \boldsymbol{\mu} = (0, 0.35, 1), \quad \boldsymbol{\tau} = (0.22, 0.24, 0.38). \tag{15}$$

The component projection starts at zero, so the initial tokens coincide with the full-spectrum projection. Learned assignment logits need not retain the smoothness or ordering of the initial priors. Assignment determines how a frequency bin is represented before encoding, whereas attention pooling weights the encoded bins afterward. The two operations are therefore not interchangeable.

Table 6: Configurations of the 20 models used for mechanism analysis. B: batch size; C: numerical latent dimension; D_e, D_d: encoder and decoder depths; WD: weight decay; ρ_0: global residual-parameter initialization.

Dataset	P	B	C	D_e	D_d	Learning rate	WD	Dropout	ρ_0
ETTh1	96	256	128	1	3	3.6204×10^{-4}	7.6962×10^{-4}	0.543	0.7443967
ETTh1	192	32	256	2	3	2.0761×10^{-4}	7.318×10^{-5}	0.566	0.5
ETTh1	336	16	128	3	3	5.161×10^{-5}	2.846×10^{-4}	0.619	0.0
ETTh1	720	32	256	4	4	5.753×10^{-5}	3.7234×10^{-4}	0.609	0.7443967
ETTh2	96	256	128	1	1	1.1971×10^{-4}	1.2146×10^{-4}	0.239	0.7443967
ETTh2	192	256	64	4	5	2.3587×10^{-4}	5.9309×10^{-4}	0.236	0.0
ETTh2	336	256	64	1	6	8.043×10^{-5}	5.101×10^{-4}	0.384	1.0
ETTh2	720	256	64	2	3	1.0177×10^{-4}	2.9388×10^{-4}	0.334	1.0
Weather	96	64	32	3	2	1.05344×10^{-3}	2.6501×10^{-4}	0.134	0.0
Weather	192	64	16	3	4	8.5294×10^{-4}	1.3478×10^{-4}	0.328	0.0
Weather	336	32	16	3	4	2.2346×10^{-4}	2.916×10^{-5}	0.105	1.0
Weather	720	128	128	2	6	1.6167×10^{-4}	2.9243×10^{-4}	0.111	1.0
Exchange	96	128	32	1	2	1.1067×10^{-4}	1.7965×10^{-4}	0.304	0.0
Exchange	192	256	32	1	2	1.8907×10^{-4}	7.3754×10^{-4}	0.273	0.0
Exchange	336	128	32	1	3	6.654×10^{-5}	1.091×10^{-5}	0.348	0.7443967
Exchange	720	40	16	2	3	4.765×10^{-5}	3.304×10^{-5}	0.146	0.7443967
ILI	24	16	32	2	2	1.3123×10^{-4}	4.289×10^{-5}	0.035	1.0
ILI	36	16	32	3	1	1.4695×10^{-4}	9.343×10^{-5}	0.054	0.5
ILI	48	32	256	3	2	2.61789×10^{-3}	5.993×10^{-5}	0.416	0.0
ILI	60	32	64	3	1	3.254×10^{-5}	2.033×10^{-5}	0.066	0.5

B.2 Conditioning, Alignment, and Residual Control

The horizon input $P/1000$ is fixed within a trained configuration. The temporal–spectral gate can still vary across samples through its temporal summary. This is configuration-level horizon conditioning, not training one model on changing horizons.

There are $M = 4$ independent alignment branches, rather than four internal heads of one conventional multi-head attention layer. Numerical and prompt latent channels provide C queries and E keys, with their profiles across the N variables supplying the feature dimension $d = N$. Attention and bias matrices have shape $C \times E$. Queries and keys determine the bias; values enter only the attention aggregation. The branch weights $\pi_{n,h}$ are normalized over branches for each variable.

The residual coefficient is $m\boldsymbol{\alpha}$, with

$$m = \exp[-\mathrm{softplus}(\rho)] = \frac{1}{1 + \exp(\rho)}. \tag{16}$$

Although $m \in (0, 1)$, the channel coefficients $\boldsymbol{\alpha}$ are unconstrained, so this is not necessarily a convex mixture. AM changes relative attention before value aggregation; the global factor scales the channel-wise residual correction after branch aggregation.

C Detailed Theoretical Derivations

C.1 Permutation Invariance of the Position-Agnostic Pathway

For one variable, let $\mathbf{f} \in \mathbb{R}^{L_f}$ be the magnitude spectrum, Φ the shared scalar projection, $\mathcal{T}$ the spectral Transformer, and a a shared scalar pooling scorer. Write

$$\mathcal{G}_0(\mathbf{f}) = \mathbf{H}^\top \mathbf{p}, \quad \mathbf{H} = \mathcal{T}(\Phi(\mathbf{f})), \quad \mathbf{p} = \mathrm{softmax}(a(\mathbf{H})). \tag{17}$$

For a permutation matrix $\mathbf{\Pi}$, the shared projection satisfies $\Phi(\mathbf{\Pi f}) = \mathbf{\Pi}\Phi(\mathbf{f})$. Permuting the tokens permutes queries, keys, and values in the same way. Thus,

$$\begin{aligned}\mathrm{Attn}(\mathbf{\Pi U}) &= \mathrm{softmax}_{\mathrm{row}}\left(\mathbf{\Pi}\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\mathbf{\Pi}^\top\right)\mathbf{\Pi V} \\ &= \mathbf{\Pi}\,\mathrm{softmax}_{\mathrm{row}}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right)\mathbf{V}.\end{aligned} \tag{18}$$

Shared token-wise feed-forward layers, layer normalization, and residual connections preserve equivariance. With positional information and position-dependent masks absent, and dropout disabled, the Transformer is therefore permutation equivariant. The scorer and its softmax weights permute with the tokens, giving

$$\mathcal{G}_0(\mathbf{\Pi f}) = (\mathbf{\Pi H})^\top(\mathbf{\Pi p}) = \mathbf{H}^\top\mathbf{p} = \mathcal{G}_0(\mathbf{f}). \tag{19}$$

This concerns the specified spectral pathway, not the complete model: the temporal pathway independently retains the ordered history.

C.2 Frequency-Dependent Effective Projection

Writing $\phi_{\mathrm{full}}(f) = f\mathbf{w}_{\mathrm{full}} + \mathbf{b}_{\mathrm{full}}$ yields

$$\mathbf{U}_{n\ell} = F_{n\ell}\mathbf{w}_\ell^{\mathrm{eff}} + \mathbf{b}_{\mathrm{full}}, \qquad \mathbf{w}_\ell^{\mathrm{eff}} = \mathbf{w}_{\mathrm{full}} + \lambda_f \sum_{r=1}^{R} s_{\ell r}\mathbf{W}_{b,r:}. \tag{20}$$

For bins ℓ and k,

$$\mathbf{w}_\ell^{\mathrm{eff}} - \mathbf{w}_k^{\mathrm{eff}} = \lambda_f \sum_r (s_{\ell r} - s_{kr})\mathbf{W}_{b,r:}. \tag{21}$$

Let $(\mathbf{\Pi f})_\ell = f_{\pi(\ell)}$, while memberships remain attached to their original frequency indices. The token discrepancy is

$$\mathbf{U}_\ell(\mathbf{\Pi f}) - (\mathbf{\Pi U}(\mathbf{f}))_\ell = f_{\pi(\ell)}(\mathbf{w}_\ell^{\mathrm{eff}} - \mathbf{w}_{\pi(\ell)}^{\mathrm{eff}}). \tag{22}$$

It need not vanish. SSC can therefore break the token-level symmetry through frequency-dependent effective projections. This capacity depends jointly on memberships and projection weights, explaining how SSC supplies frequency-indexed structure to the shared encoder.

Because $\sum_r s_{\ell r} = 1$, weighted magnitudes satisfy $\sum_r v_{n\ell r} = F_{n\ell}$. The identity preserves magnitude allocation before projection, not an orthogonal energy decomposition or invertibility of the final representation.

C.3 Attention Reweighting and Relative Preference

Omitting the branch index, let $p_{ij}^0 = \mathrm{softmax}_j(S_{ij})$ and $p_{ij} = \mathrm{softmax}_j(S_{ij} + B_{ij})$. Then

$$p_{ij} = \frac{p_{ij}^0 \exp(B_{ij})}{\sum_k p_{ik}^0 \exp(B_{ik})}. \tag{23}$$

Taking ratios cancels the normalizers:

$$\log\frac{p_{ij}}{p_{ik}} - \log\frac{p_{ij}^0}{p_{ik}^0} = B_{ij} - B_{ik}. \tag{24}$$

Only within-row differences alter relative preferences. A row-constant bias leaves the distribution unchanged. Since $B_{ij} = \beta\tanh(T_{ij})$,

$$|B_{ij} - B_{ik}| \le 2|\beta|. \tag{25}$$

This bounds log-odds changes for a fixed learned scale, not forecasting error or the trainable scale itself. An individual probability increases precisely when

$$B_{ij} > \log\left(\sum_k p_{ik}^0 \exp(B_{ik})\right). \tag{26}$$

Positive bias therefore need not imply increased probability. All these identities concern probabilities before attention dropout. Finite biases do not explicitly remove connections.

D Ablation Definitions

Without SSC. The component-based correction is removed, leaving $\mathbf{U}_{n\ell} = \phi_{\text{full}}(F_{n\ell})$. The full-spectrum encoder and cross-modal pathway remain. This compares the added frequency-dependent structure against the retained spectral pathway, rather than against a model without frequency information.

Without AM. The signed relation bias is removed and the associated global residual factor is fixed to one. Base attention, adaptive branch aggregation, and channel-wise residual coefficients remain:

$$\mathbf{P}_h = \mathrm{softmax}_{\mathrm{key}}(\mathbf{S}_h), \qquad \mathbf{\Theta} = \mathbf{H}_g + \boldsymbol{\alpha} \odot (\mathbf{\Lambda} - \mathbf{H}_g). \tag{27}$$

This removal evaluates signed relation modulation and its associated residual control jointly. The per-horizon results in the main ablation table test the complementary roles of structured spectral inputs and regulated semantic alignment.

E Additional Mechanism Visualizations

E.1 Computation and Interpretation

The following five figures cover ETTh1, ETTh2, Weather, Exchange, and ILI at every investigated horizon. They are reconstructed from the stored arrays without modifying numerical values. Solid and dashed curves show learned and initialized memberships. These are parameters shared across samples and variables, not instance-wise assignments.

The analysis script uses 32 test-window indices distributed across the complete test-window range. The saved arrays contain window-averaged quantities for four inspected branches. For window s and branch h, define the base and adjusted probabilities from the same trained branch as

$$\mathbf{P}^{0,(s,h)} = \mathrm{softmax}_{\mathrm{key}}(\mathbf{S}^{(s,h)}), \qquad \mathbf{P}^{(s,h)} = \mathrm{softmax}_{\mathrm{key}}(\mathbf{S}^{(s,h)} + \mathbf{B}^{(s,h)}). \tag{28}$$

The heatmaps display

$$\overline{\mathbf{B}} = \frac{1}{32M} \sum_{s=1}^{32} \sum_{h=1}^{M} \mathbf{B}^{(s,h)}, \qquad \overline{\Delta\mathbf{P}} = \frac{1}{32M} \sum_{s=1}^{32} \sum_{h=1}^{M} (\mathbf{P}^{(s,h)} - \mathbf{P}^{0,(s,h)}), \tag{29}$$

with $M = 4$. Base attention is not taken from a separately retrained ablation. The difference is computed before averaging, not by applying softmax to an averaged bias.

Rows and columns index numerical and prompt latent channels. Each panel uses its own symmetric color scale. The same color in different panels need not represent the same magnitude. Opposing signed changes may cancel under averaging. Together, the two heatmaps separate the learned relation correction from its effect after normalization, directly illustrating the relative-preference analysis in Appendix C.

The complete horizon coverage exposes configuration-dependent memberships and relation corrections; ablations assess their predictive contributions. Weather at $P = 96$ exhibits a pronounced change from the initial memberships, whereas several Exchange configurations retain a shape close to the prior. This illustrates the flexibility of learned spectral assignments: optimization can reshape frequency specialization or preserve an already useful initial structure. The signed-bias and attention-change panels further show how relation corrections translate into normalized redistribution across prompt latent channels. The main case figure shows channel 0 of ETTh1 ($P = 192$, window 1344), ILI ($P = 48$, window 72), and Weather ($P = 336$, window 0) over the full forecast interval in original units. The complete archived set is shown in Appendix F.

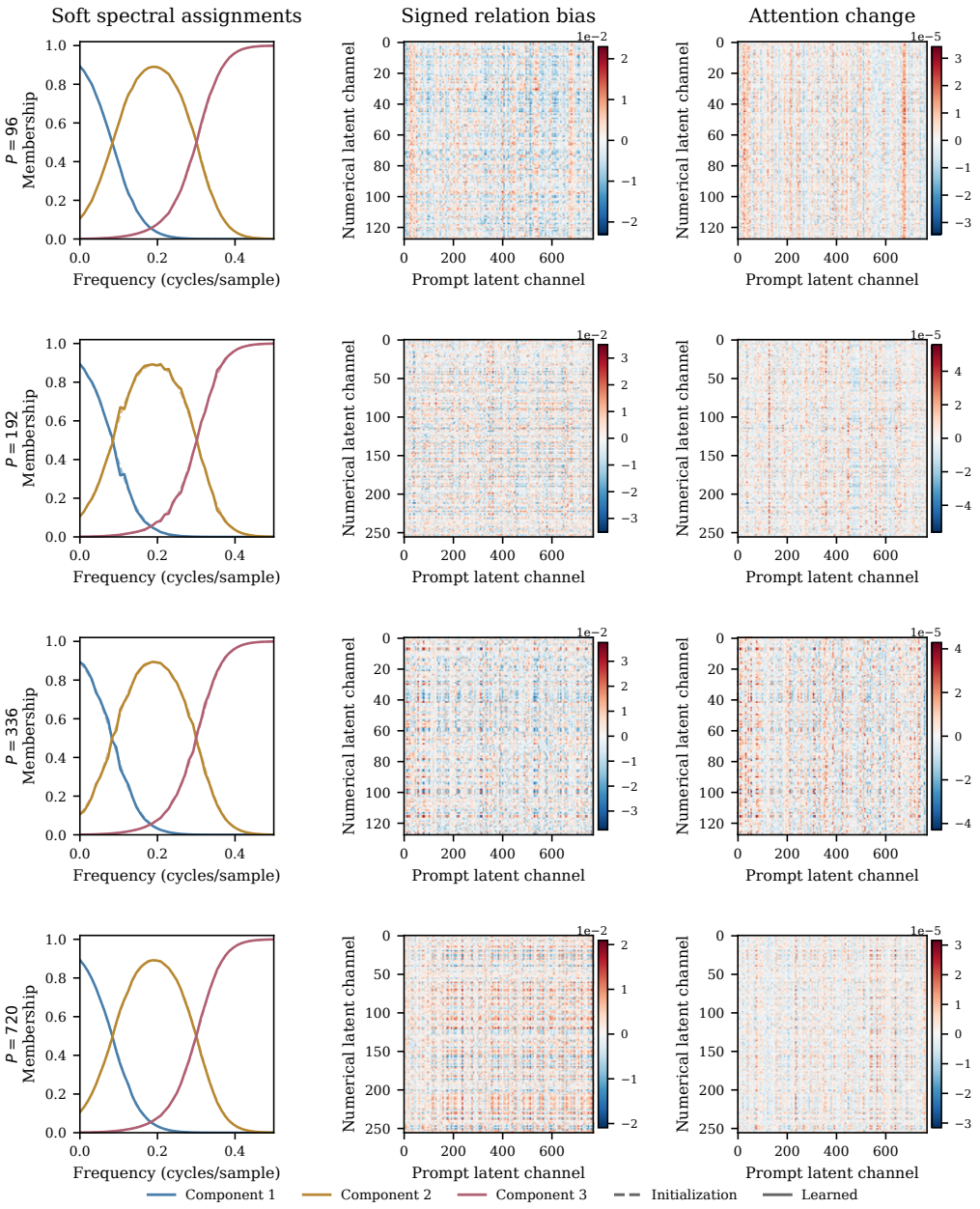

Figure 5: Additional mechanism analysis on ETTh1. Rows correspond to forecast horizons 96, 192, 336, and 720 from top to bottom. Left: initialized (dashed) and learned (solid) spectral memberships. Middle: signed relation bias. Right: adjusted minus base attention probabilities. Heatmaps average 32 windows and four inspected alignment branches, using separate symmetric color scales.

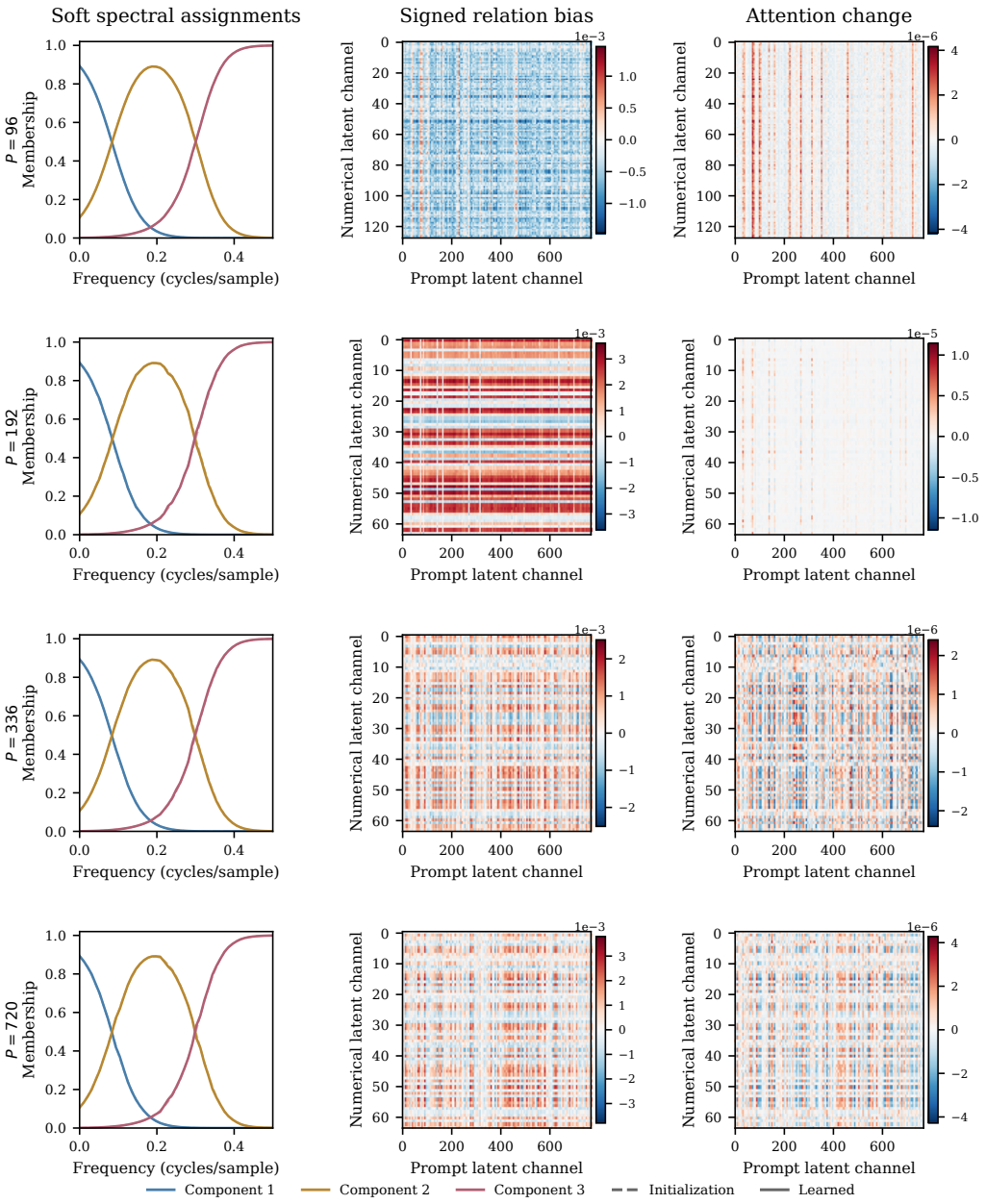

Figure 6: Additional mechanism analysis on ETTh2. Rows correspond to forecast horizons 96, 192, 336, and 720 from top to bottom. Left: initialized (dashed) and learned (solid) spectral memberships. Middle: signed relation bias. Right: adjusted minus base attention probabilities. Heatmaps average 32 windows and four inspected alignment branches, using separate symmetric color scales.

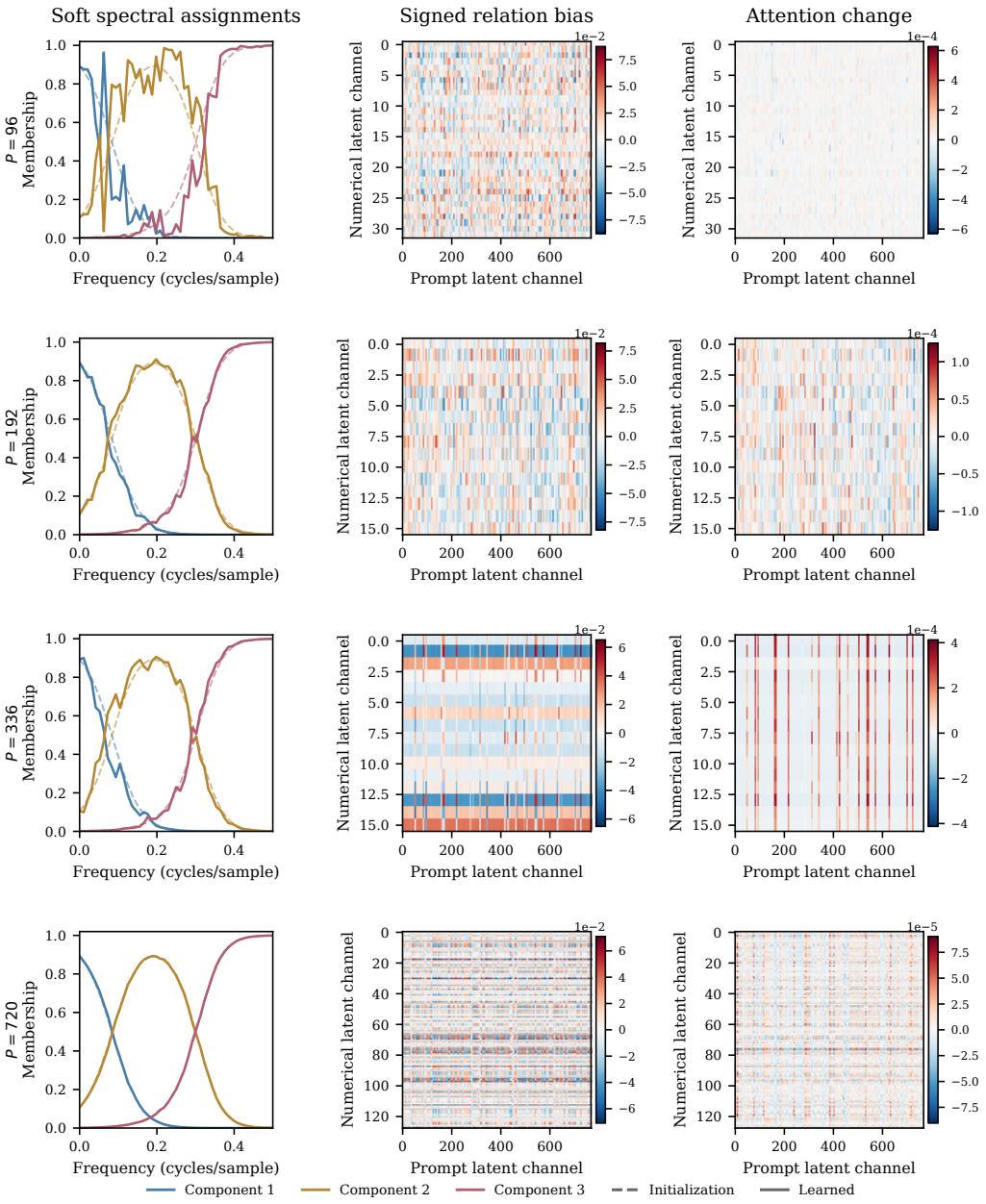

Figure 7: Additional mechanism analysis on Weather. Rows correspond to forecast horizons 96, 192, 336, and 720 from top to bottom. Left: initialized (dashed) and learned (solid) spectral memberships. Middle: signed relation bias. Right: adjusted minus base attention probabilities. Heatmaps average 32 windows and four inspected alignment branches, using separate symmetric color scales.

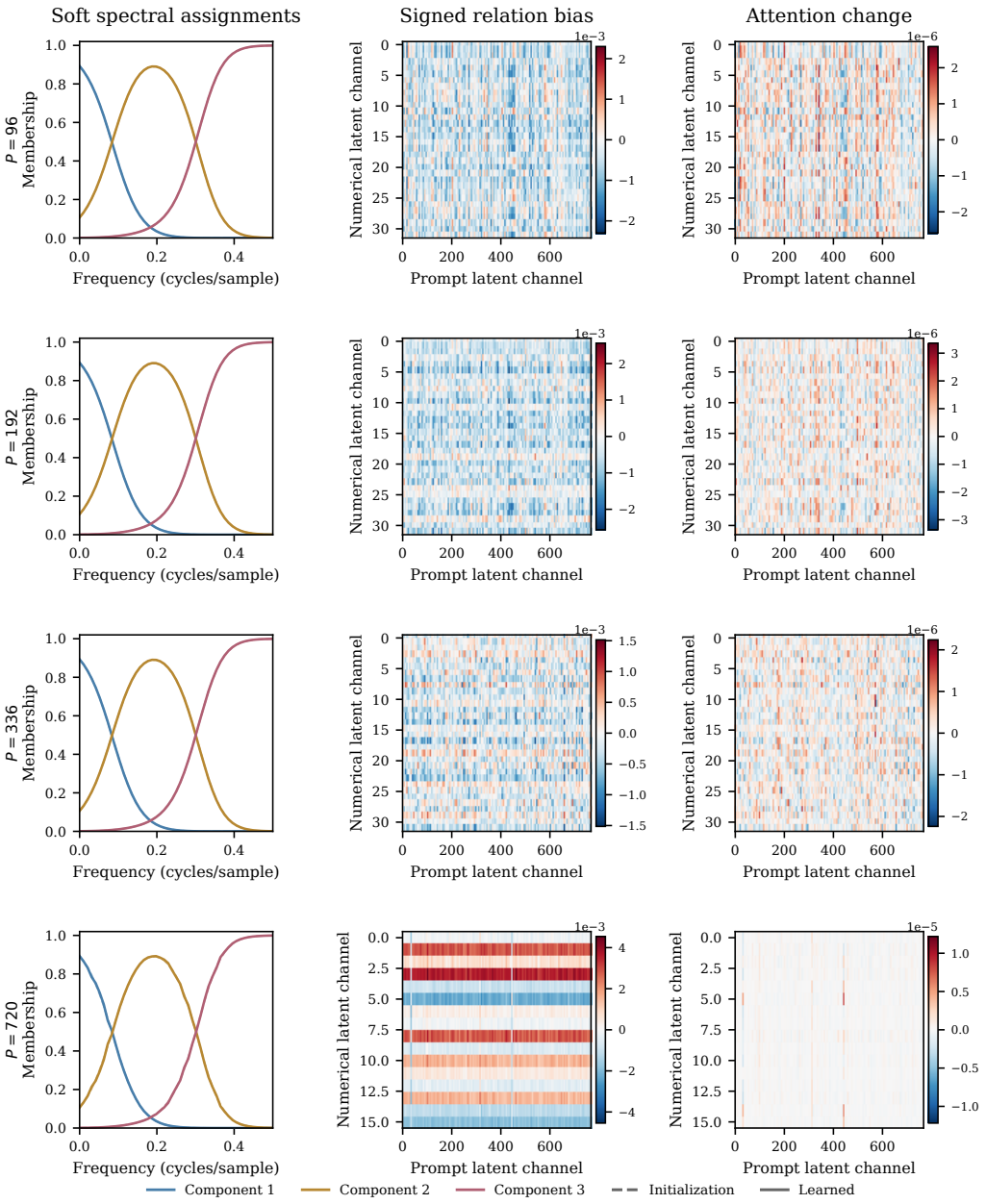

Figure 8: Additional mechanism analysis on Exchange. Rows correspond to forecast horizons 96, 192, 336, and 720 from top to bottom. Left: initialized (dashed) and learned (solid) spectral memberships. Middle: signed relation bias. Right: adjusted minus base attention probabilities. Heatmaps average 32 windows and four inspected alignment branches, using separate symmetric color scales.

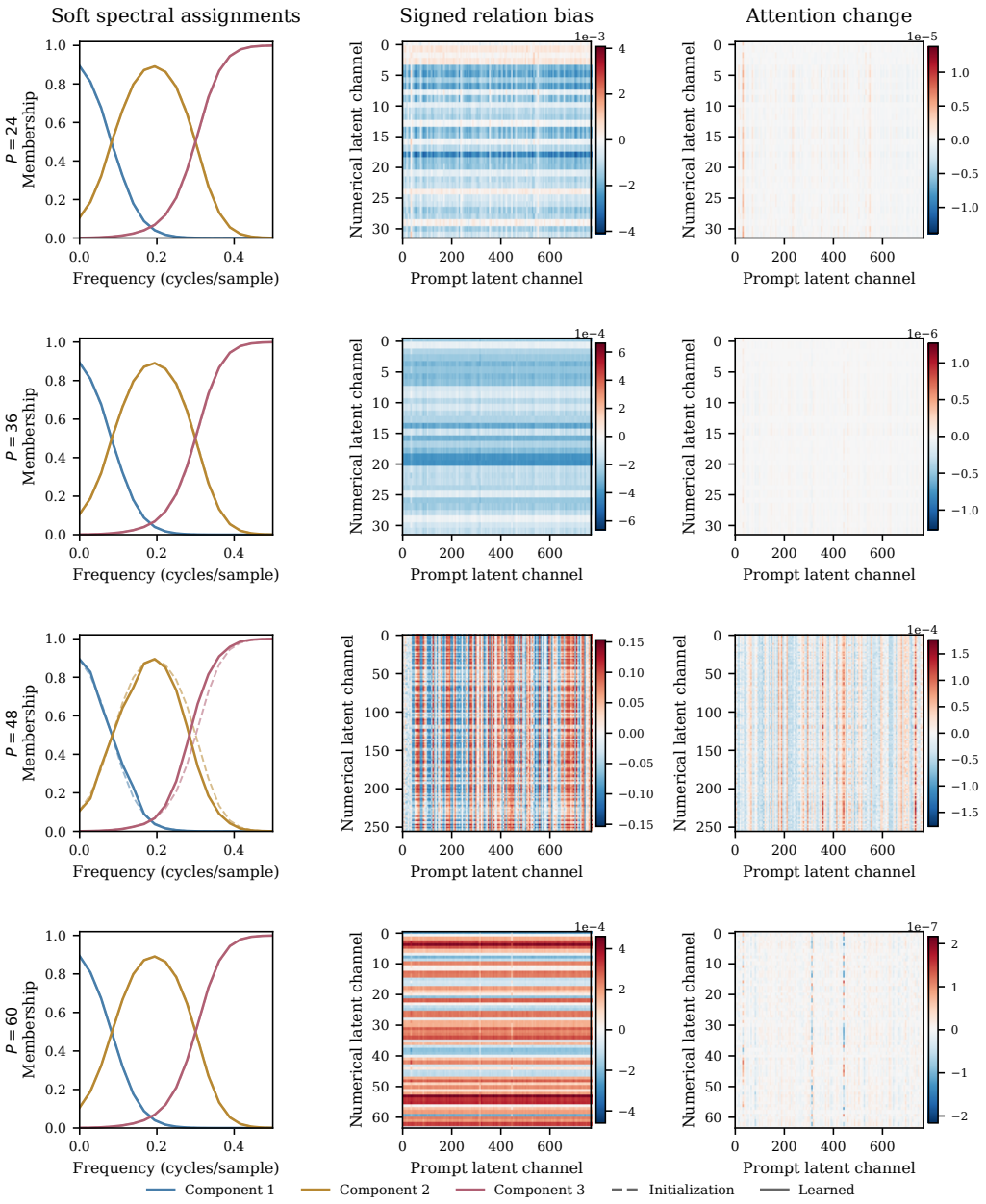

Figure 9: Additional mechanism analysis on ILI. Rows correspond to forecast horizons 24, 36, 48, and 60 from top to bottom. Left: initialized (dashed) and learned (solid) spectral memberships. Middle: signed relation bias. Right: adjusted minus base attention probabilities. Heatmaps average 32 windows and four inspected alignment branches, using separate symmetric color scales.

F Additional Forecasting Examples

Figures 10–14 show all 60 archived cases: five datasets, four horizons, and three windows per horizon. The script selects integer indices with `linspace(0, S-1, 3)`, where S is the number of test windows. Columns show the first, middle, and last selected windows. Channel 0 is plotted after inverse scaling, including the complete history and forecast interval.

These examples complement MSE and MAE by exposing period, phase, level, and amplitude behavior. ETTh1 illustrates recurring fluctuations across horizons, while ILI displays broader rises and declines. Combined with the mechanism visualizations and component ablations, the cases connect adaptive spectral representation and cross-modal alignment to observable forecast trajectories.

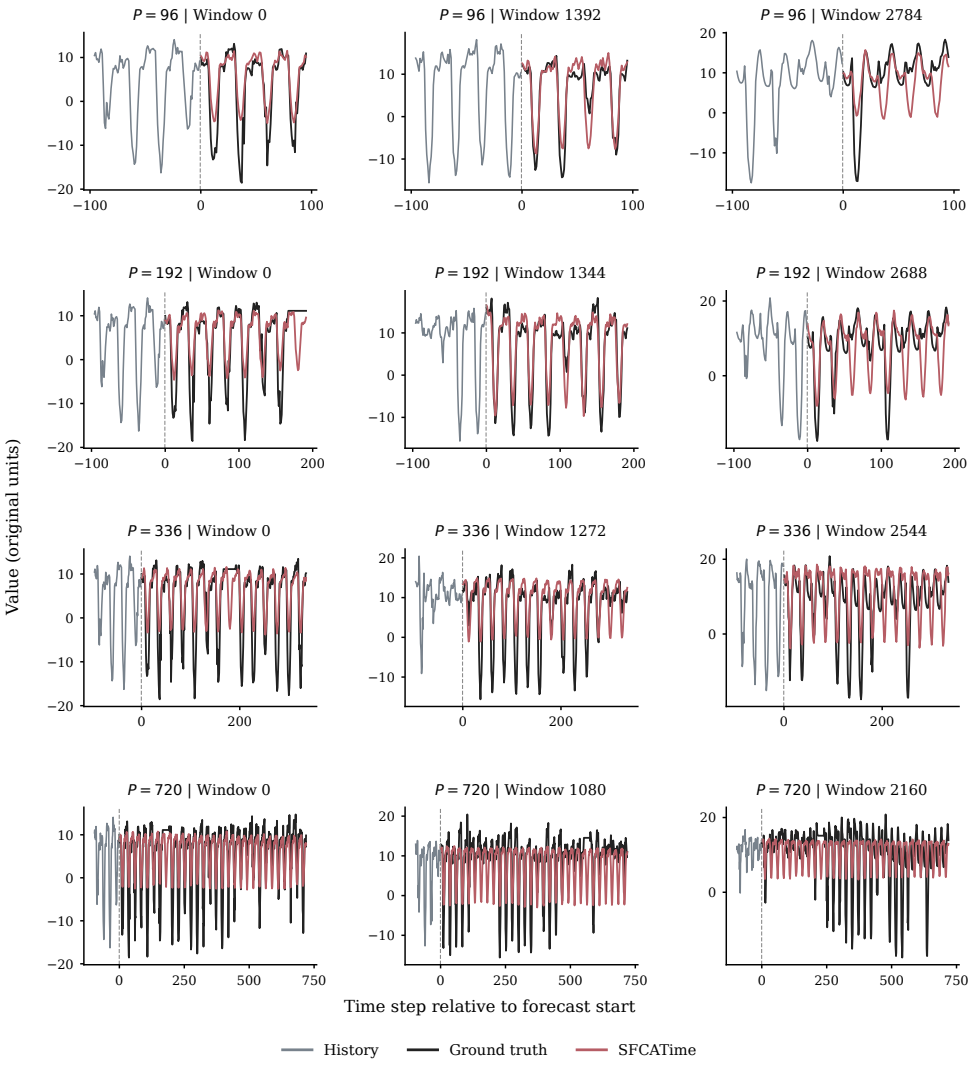

Figure 10: Additional forecasting examples on ETTh1. Rows show horizons 96, 192, 336, and 720; columns show the first, middle, and last sampled test windows, with indices printed above each panel. Channel 0 is shown in original units over the complete history and forecast interval. Gray: history; black: ground truth; red: SFCATime. Dashed lines mark the forecast boundary.

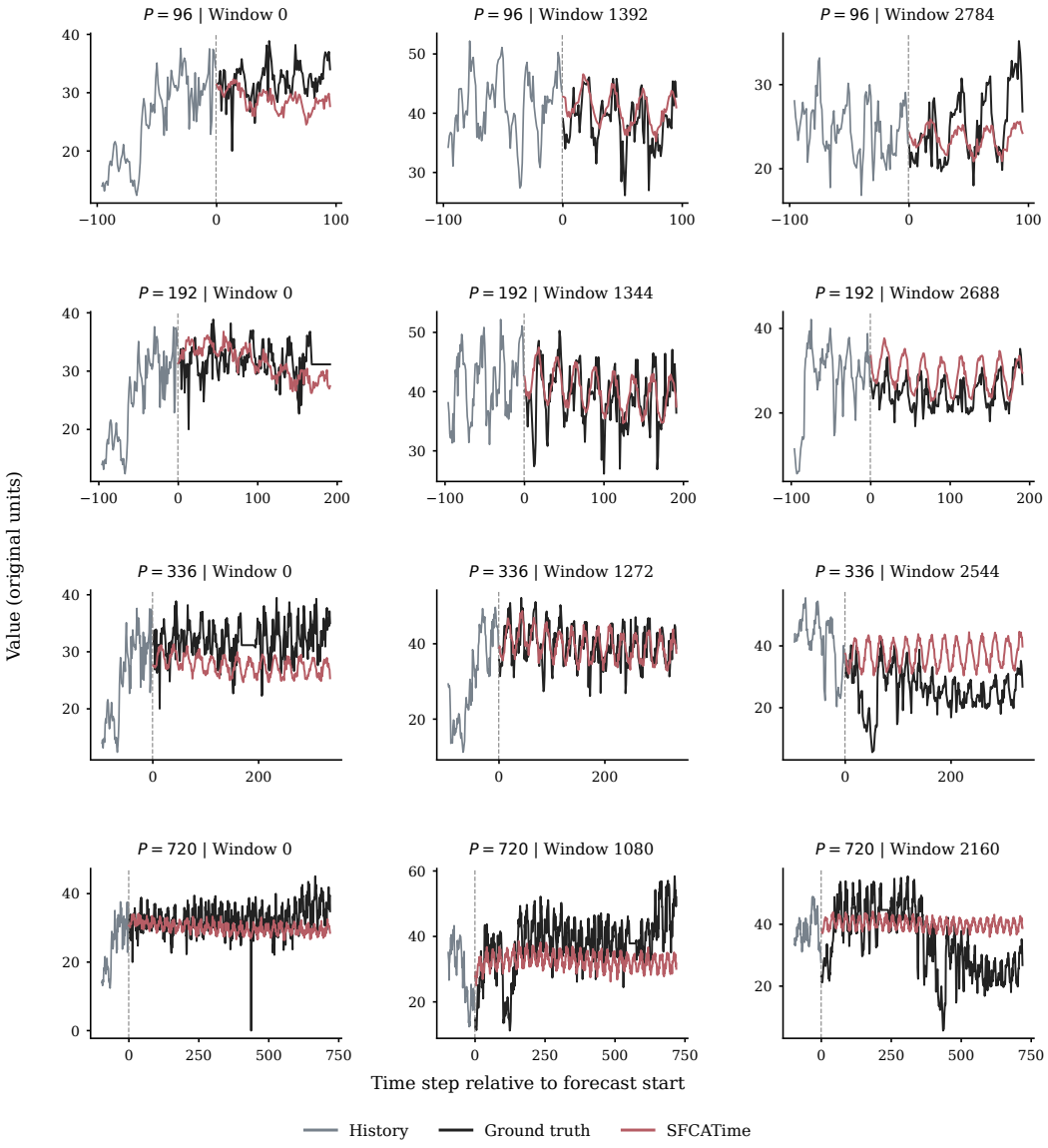

Figure 11: Additional forecasting examples on ETTh2. Rows show horizons 96, 192, 336, and 720; columns show the first, middle, and last sampled test windows, with indices printed above each panel. Channel 0 is shown in original units over the complete history and forecast interval. Gray: history; black: ground truth; red: SFCATime. Dashed lines mark the forecast boundary.

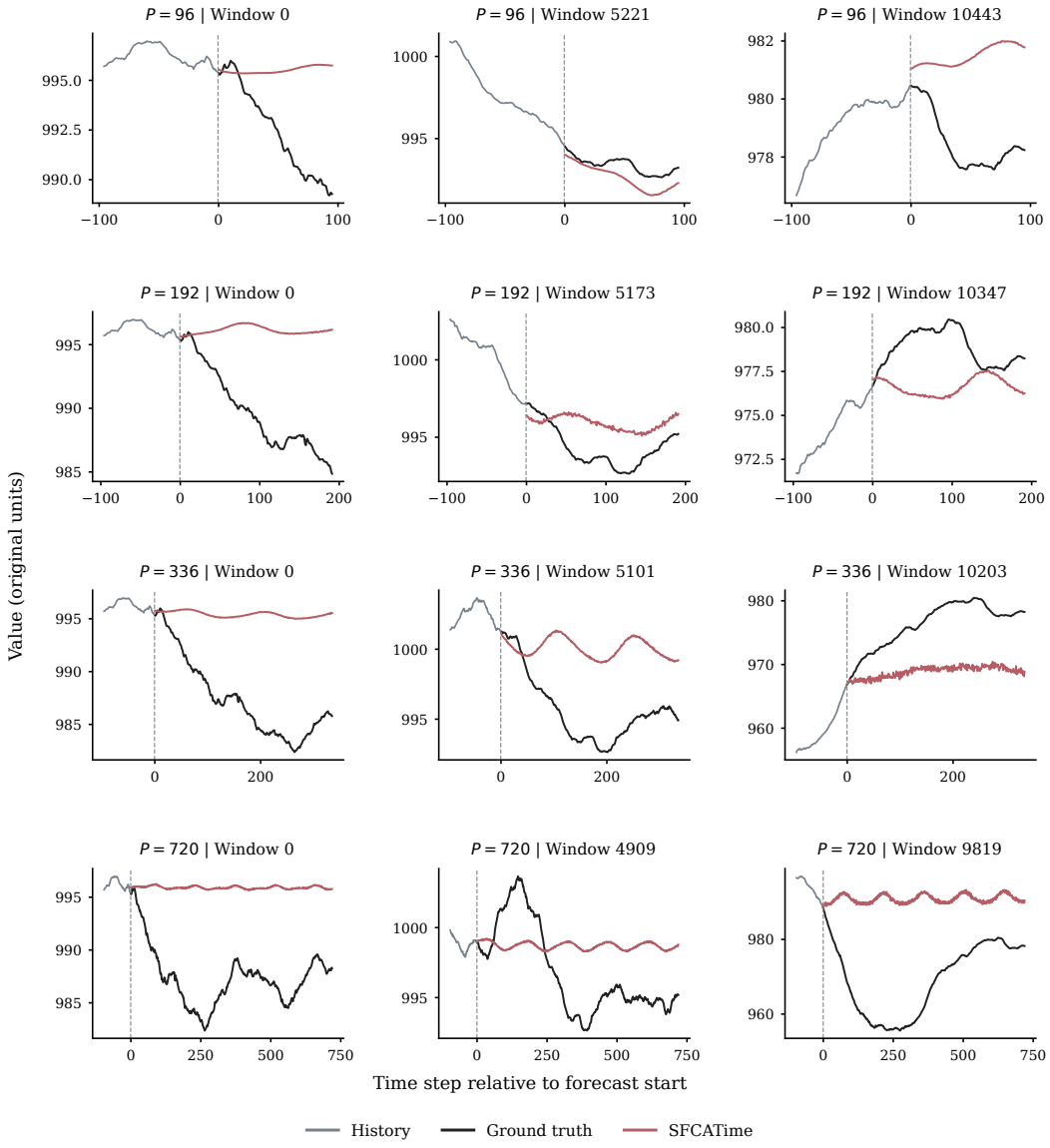

Figure 12: Additional forecasting examples on Weather. Rows show horizons 96, 192, 336, and 720; columns show the first, middle, and last sampled test windows, with indices printed above each panel. Channel 0 is shown in original units over the complete history and forecast interval. Gray: history; black: ground truth; red: SFCATime. Dashed lines mark the forecast boundary.

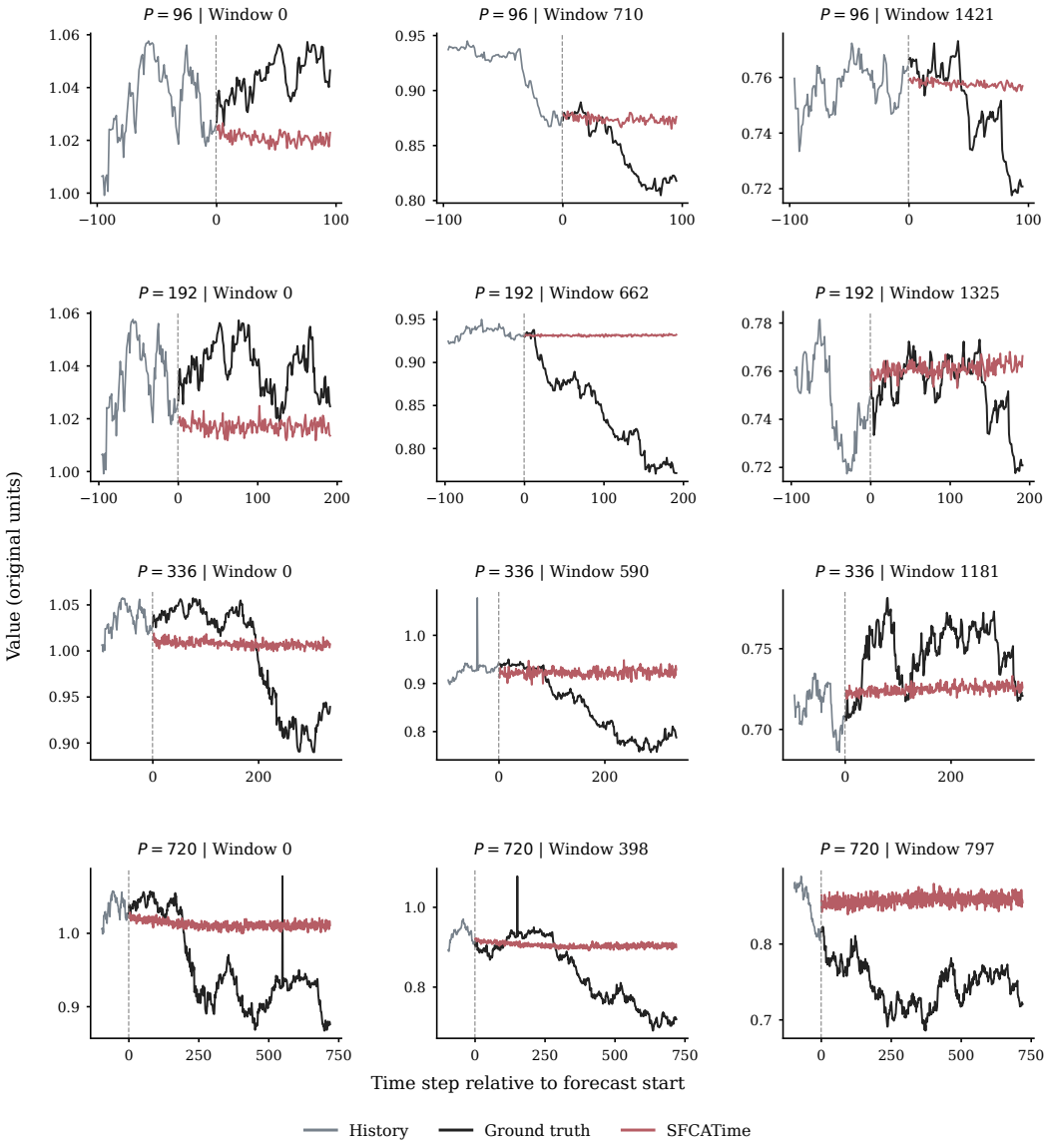

Figure 13: Additional forecasting examples on Exchange. Rows show horizons 96, 192, 336, and 720; columns show the first, middle, and last sampled test windows, with indices printed above each panel. Channel 0 is shown in original units over the complete history and forecast interval. Gray: history; black: ground truth; red: SFCATime. Dashed lines mark the forecast boundary.

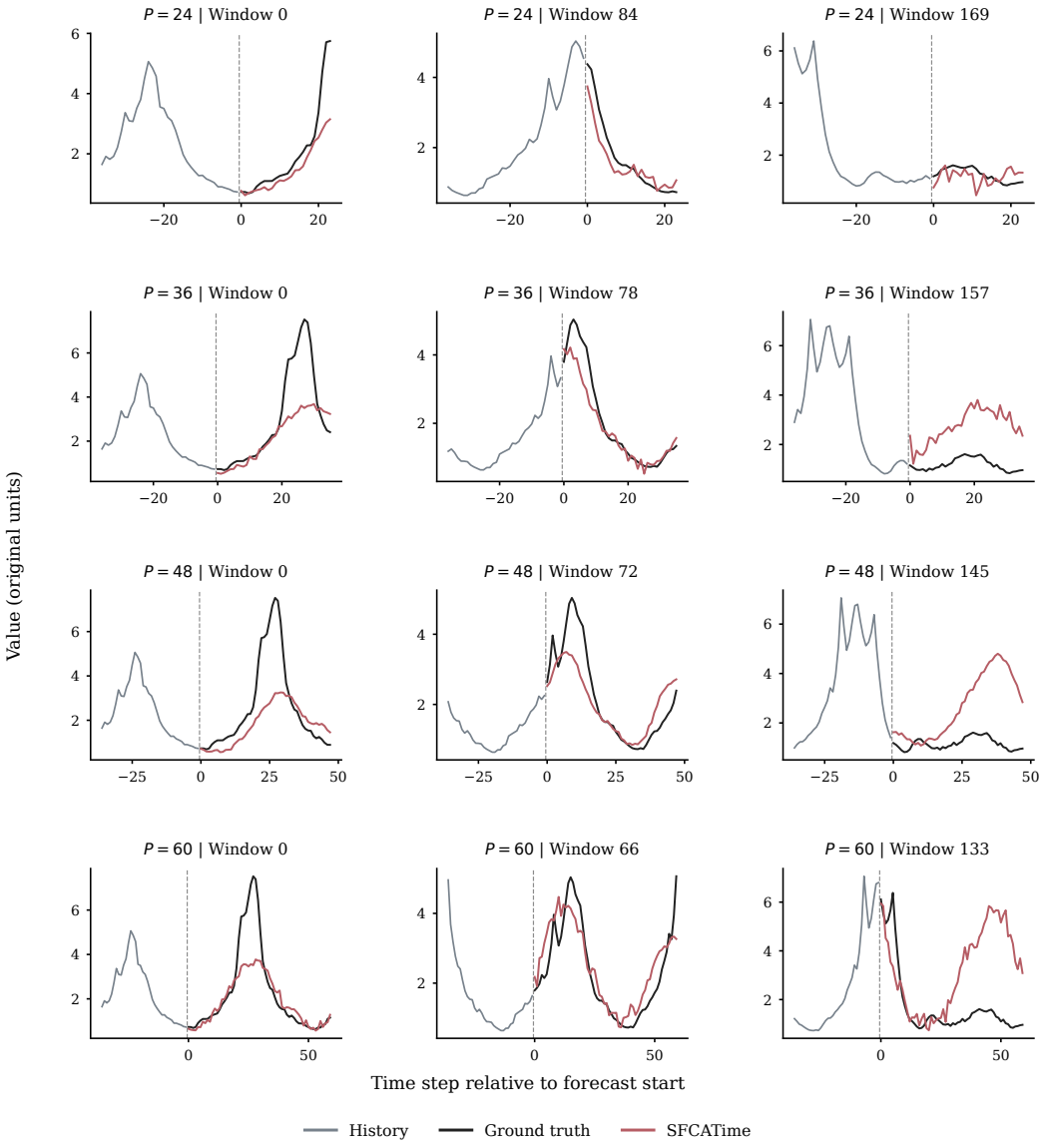

Figure 14: Additional forecasting examples on ILI. Rows show horizons 24, 36, 48, and 60; columns show the first, middle, and last sampled test windows, with indices printed above each panel. Channel 0 is shown in original units over the complete history and forecast interval. Gray: history; black: ground truth; red: SFCATime. Dashed lines mark the forecast boundary.