FACT: Frequency-Aware Channel-Guided Multivariate Time Series Forecasting

Anonymous authors
Paper under double-blind review

Abstract

Forecasting Multivariate Time Series (MTS) requires capturing complex intra-channel dynamics and evolving inter-channel dependencies. However, existing methods often struggle to disentangle meaningful signals from inter-channel noise and intricate interaction patterns. To address this, we propose a novel framework that operates entirely in the frequency domain, modeling inter-channel relationships at the component level. Our approach first dynamically decomposes each time series into its constituent frequencies. A channel masking mechanism then identifies and isolates the most salient frequency components, simultaneously filtering noise and enhancing computational efficiency. This allows our model to capture time-varying inter-channel dependencies with high fidelity. Furthermore, our learning objective effectively balances accuracy against regularization constraints for both computational efficiency and interpretability. Extensive experiments on diverse, real-world datasets demonstrate that our method achieves competitive performance. Code is available at this repository: https://anonymous.4open.science/r/FACT.

1 Introduction

Multivariate time series (MTS) forecasting supports power scheduling, weather prediction and industrial control, where accuracy, robustness and interpretability are equally critical (Zhou et al., 2021; Haixu Wu & Long, 2021; Zhou et al., 2022). Existing research largely falls into two paradigms. Channel-dependent (CD) models explicitly mix variables but easily introduce spurious correlations and face scalability issues in high dimensions (Zhang & Yan, 2023; Liu et al., 2023; Xue Wang & Jin, 2023); channel-independent (CI) models improve robustness by per-channel processing, but sacrifice genuine couplings and physical interpretability (Nie et al., 2023; Han et al., 2024). This tension indicates a need for fine-grained, controllable interaction modelling.

Our key observation is physical: different spectral components of each channel carry distinct semantics—amplitude corresponds to intensity/energy, and phase encodes temporal alignment or spatial shift. Daily/weekly/yearly patterns (high/mid/low frequencies) in electricity demand are not "just noise"; they interact across channels (coordination or antagonism), and can trigger cross-frequency effects (e.g., a sudden cold snap induces sustained changes in low-frequency heating demand). Hence, interactions should be modeled at the channel–frequency cell level with both magnitude and phase.

However, most spectral approaches focus on single-sequence decomposition or global reweighting (Wu et al., 2023; Yi et al., 2023b; Xu et al., 2024; Yi et al., 2024; Zhang et al., 2024b;a; Yang et al., 2023; Yi et al., 2023a), overlooking dynamic, cell-level dependencies and the indispensable role of phase. This leaves room for methods that are physically grounded yet computationally efficient.

We introduce FACT (Frequency-Adaptive Complex Transformer), which lifts interaction modelling from raw channels to frequency components. FACT (i) performs dynamic frequency-band decomposition (DynFBD) with a frequency selector to obtain sparse, confidence-weighted tokens; (ii) estimates magnitude coherence Γ and phase offsets Φ via

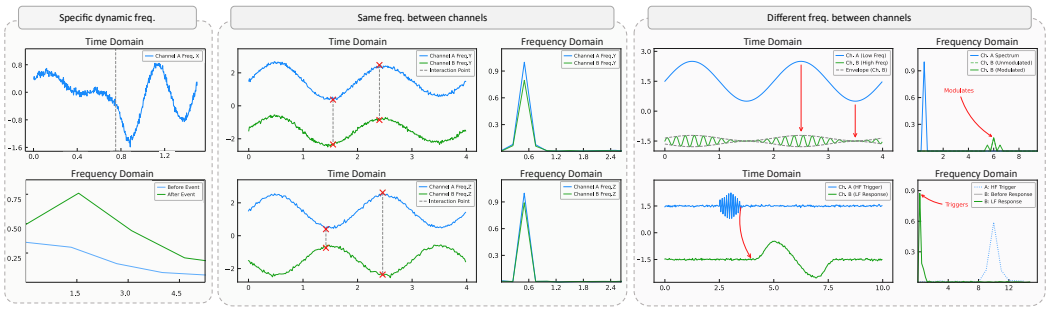

Figure 1: Representative channel–frequency interactions: dynamic drift within a channel (left), same-frequency coordination/antagonism (middle), and cross-frequency modulation/triggering (right).

ChannelPriorMixer to form channel priors and guided gating; and (iii) fuses tokens with the original spectrum through complex cross-attention and encodes them with a complex Transformer. During training, coherence/phase regularizers align priors with the learned attention, and cached signals provide diagnosis.

Contributions.

- Frequency-level interaction paradigm. We treat the channel–frequency cell as the basic unit and design a sparse token pipeline (DynFBD + selector) to suppress noisy bands while preserving physically meaningful signals.
- Magnitude/phase-aware priors and fusion. We introduce ChannelPriorMixer and adaptive fusion to encode Γ/Φ-aware gating and attention, and integrate coherence/phase regularizers for training-time guidance and post-hoc diagnostics.
- Model-agnostic plug-in with consistent gains. By lifting interaction modelling to frequency components, FACT more effectively separates informative signals from noise than raw-channel mixing. The design plugs into diverse backbones (Transformer/MLP/Linear) and yields consistent improvements across datasets.

We validate these claims through comprehensive experiments: ablations on each component, regularization sweeps, and interpretability visualizations. Results demonstrate positive correlation between our interpretability metrics and accuracy, and consistent gains across backbones. Details are provided in Section 5.

2 Related Work

2.1 Channel Interaction Modelling

Early multivariate forecasting adopted RNN/CNN backbones with local dependencies (Hochreiter & Schmidhuber, 1997; Shaojie Bai & Koltun, 2018), later extended by graph and multi-task formulations that encode handcrafted adjacencies (Zonghan Wu & Zhang, 2020; Xinle Wu & Jensen, 2021; Yue Cui & Zhou, 2021). Transformers broaden the receptive field (Vaswani et al., 2017; Zhou et al., 2021; Haixu Wu & Long, 2021; Zhou et al., 2022), but how to model variable interactions remains contentious. Channel-independent (CI) designs (e.g., PatchTST, iTransformer) favor per-channel tokenization for robustness to noise/drift (Nie et al., 2023; Liu et al., 2023); some even argue high-amplitude frequencies dominate prediction (Dai et al., 2024; Xu et al., 2024). Channel-dependent (CD) methods (Crossformer, CARD, SOFTS, TimePro, DUET) reintroduce interactions via cross-dimension routes, alignment-aware attention, global cores or routing/clustering (Zhang & Yan, 2023; Xue Wang & Jin, 2023; Han et al., 2024; Ma et al., 2025; Qiu et al., 2025). CI may discard genuine couplings; CD often mixes signals coarsely and is sensitive to noise—motivating frequency-aware, fine-grained priors as a middle ground.

2.2 Time–Frequency Methods and Physical Priors

Spectral approaches provide efficiency but typically treat amplitude as the sole carrier of information, whereas phase determines temporal alignment/lag and spatial shift. TimeMixer/TimeMixer++ mix frequency bands for long contexts yet collapse phase cues into shared representations (Wang et al.; 2025). FredFormer and TSMixer refine spectra via normalization or MLP mixing, but channel fusion remains entangled and phase alignment implicit (Piao et al., 2024; Ekambaram et al., 2023). FreTS/FITS recalibrate responses (Kun Yi & Niu, 2023; Xu et al., 2024), yet they average across channels and cannot reveal which variable drives a specific band or how cross-frequency triggering unfolds. A complementary line emphasizes that spectral components should not be treated uniformly: FreDF shows frequency utility is scenario-dependent and benefits from dynamic fusion (Zhang et al., 2024a); periodicity decoupling highlights the role of high-frequency harmonics beyond mere noise (Dai et al., 2024). These observations motivate modelling interactions at the channel–frequency cell with explicit magnitude/phase priors and channel-specific reweighting—precisely what FACT operationalizes. Beyond accuracy, recent work values robustness and interpretability. CI strategies offer stability but little diagnosis (Lu Han & Zhan, 2023); CD designs (SOFTS/CARD) balance the two via global cores or alignment penalties (Han et al., 2024; Xue Wang & Jin, 2023). FACT inherits spectral efficiency and contributes a physically grounded, fine-grained interaction paradigm that plugs into diverse backbones.

3 Preliminaries

We briefly state the basic forecasting setup and notations; full details are deferred to Appendix H.

Task. Given $\mathbf{X} \in \mathbb{R}^{L\times C}$, predict T future steps $\mathbf{Y} \in \mathbb{R}^{T\times C}$ with a squared loss. We process sequences in the frequency domain via rFFT, obtaining complex coefficients $\mathbf{X}_{\text{fft}}(f, c) = A(f, c)e^{\mathrm{i}\theta(f,c)}$.

Key operators. DynFBD applies learnable Gaussian masks to form B_f soft bands and compresses each into K-dimensional tokens; a frequency selector produces low-dimensional mask/weight summaries used as priors. Channel priors include amplitude coherence Γ and phase offsets Φ, which summarize cross-channel coupling. Complex Linear/LayerNorm are used to preserve amplitude/phase; a guided gating normalizes weighted amplitudes to $[0, 1]$ for stabilizing learning.

Please refer to Appendix H for the complete formulas and derivations.

4 Method

FACT addresses the CI–CD dilemma by modelling interactions at the channel–frequency level with explicit magnitude/phase priors. We first outline the pipeline (Fig. 2), then introduce the key modules and the training-time regularizers. Basic notation and operators are given in Section 3.

4.1 Architecture and Complexity Overview

Figure 2 overviews the pipeline: (i) RevIN normalization and rFFT transformation; (ii) Adaptive Band Decomposition using Gaussian filters to generate frequency bands; (iii) Complex Linear Projection to create multi-scale tokens and extract mask/weight information; (iv) Feature Alignment through cross-attention and gated networks; (v) Complex encoder with coherence (L_{coh}) and phase (L_{phase}) regularization losses.

Complexity overview. A concise summary of the per-module complexity is deferred to Appendix (Table 3).

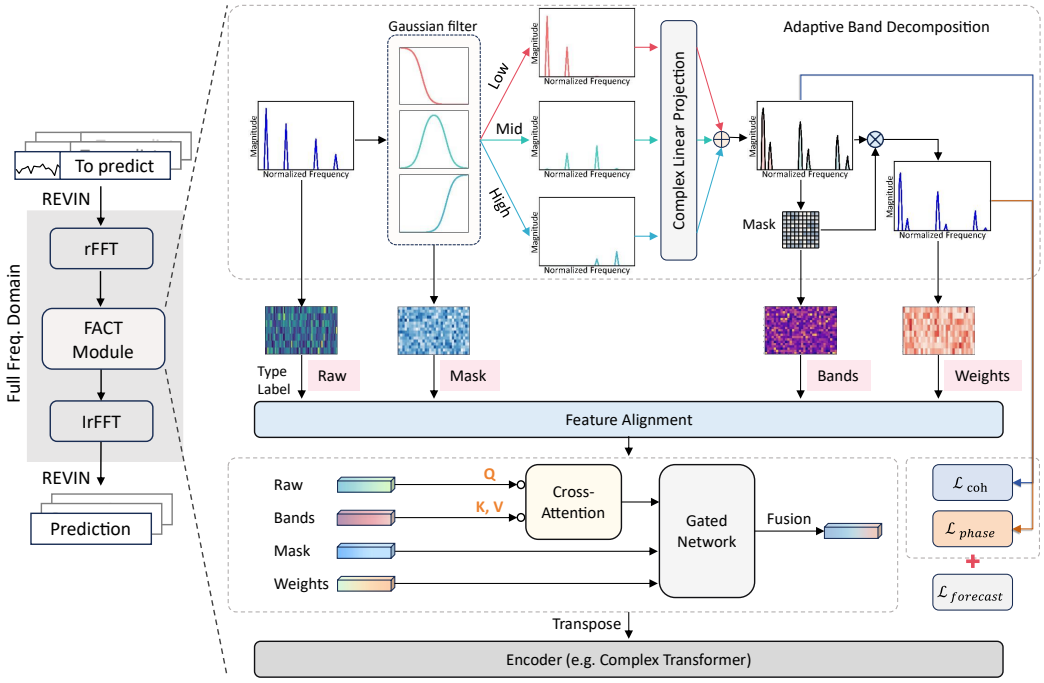

Figure 2: Overall FACT pipeline: input sequences undergo RevIN normalization and rFFT transformation to frequency domain. Gaussian filters perform adaptive band decomposition generating low/mid/high frequency bands, mask, and weight information. Complex linear projection creates multi-scale tokens, followed by Feature Alignment using cross-attention with gated networks. The encoder processes aligned features with coherence and phase regularization losses, finally recovering time-domain predictions through inverse operations.

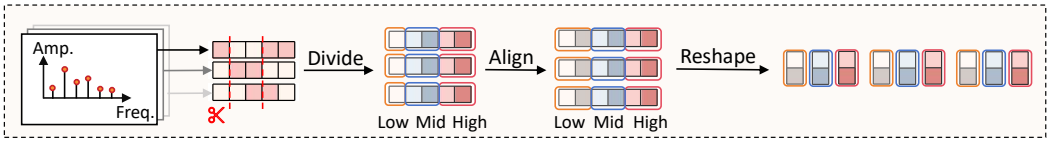

Figure 3: Fixed frequency band division illustration: the frequency axis is divided into low/medium/high three segments according to preset thresholds, each segment is compressed through independent complex linear branches and then concatenated into unified token representation.

4.2 Adaptive Band Decomposition and Frequency Selection

Motivation. Multi-scale frequencies correspond to seasonalities and lags. Adaptive Band Decomposition uses learnable Gaussian filters to softly separate low/mid/high components, avoiding aliasing and supplying priors for cross-channel modelling.

Apply learnable Gaussian filters to each channel to obtain B_f soft frequency bands, then generate tokens through ComplexLinear:

$$\mathbf{Z}_i = \text{ComplexLinear}(\omega_i \odot \mathbf{X}_{\text{fft}}) \in \mathbb{C}^{K \times C}, \quad i = 1, \ldots, B_f. \tag{1}$$

Concatenating gives $\mathbf{Z} \in \mathbb{C}^{K \times CB_f}$. A frequency selector produces two summaries $\mathbf{P}_{\text{mask}}, \mathbf{P}_{\text{weight}}$ that act as priors for channel mixing and attention bias.

Starting from fixed thresholds. An intuitive approach is to use fixed thresholds to divide the frequency axis into low/medium/high three segments and construct independent complex linear branches for each segment. This approach can quickly obtain frequency domain

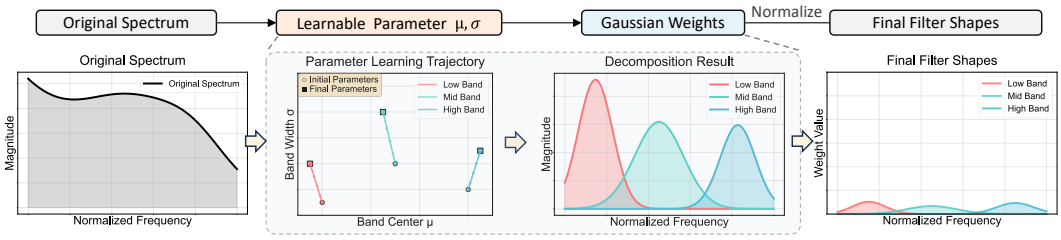

Figure 4: DynFBD's learnable Gaussian filters: raw spectrum, (μ, σ) trajectories, soft-band decomposition, and normalized filter shapes.

tokens of consistent length and finally concatenate them into $(B, K, 3C)$ channel-frequency representation; however, it only outputs a set of amplitude weights and no additional mask signals, making it difficult to guide downstream modules to highlight key frequency bands.

Gaussian adaptive division. To break free from manual threshold dependence, we further introduce learnable Gaussian kernels to perform soft division of frequency points for each channel. The softplus-constrained (μ, σ) parameters are normalized within each frequency band to obtain (B, C, bands, F) soft masks, which are point-wise multiplied with the original spectrum and then projected back to $(B, K, 3C)$ through shared complex linear layers. Meanwhile, the resulting masks and weights are compressed into low-dimensional summaries $\mathbf{P}^{\text{proj}}_{\text{mask}} \in \mathbb{R}^{B\times F\times d_m}$ and $\mathbf{P}^{\text{proj}}_{\text{weight}} \in \mathbb{R}^{B\times K\times d_w}$, providing interpretable attention bias and gating priors for adaptive fusion and channel mixing. Fig. 4 visualizes how (μ, σ) drift during training and yield normalized filter shapes that persist as frequency-domain priors.

Why turn to soft division. Fixed thresholds gradually expose two bottlenecks on multi-source data: first, hard segmentation when processing low-amplitude high-frequency signals causes some frequency bands to almost lose energy, leading to projection outputs close to zero and gradient sparsity; second, when seasonality or sampling rates change, manual readjustment of thresholds is needed, making it difficult to handle cross-dataset drift. Learnable Gaussian masks not only bring smooth gradients but also form a "frequency band $\rightarrow$ mask $\rightarrow$ attention bias" closed loop with Feature Alignment, enabling key frequency bands to be emphasized by gating in early training. Experiments show that on benchmarks like ETTh1, ETTm2, ECL, the Gaussian version reduces sMAPE by approximately $1.3\% \sim 2.1\%$ on average compared to the fixed version, and also provides direct mask signals for interpretability visualization in subsequent sections.

4.3 Channel Prior Mixer

Motivation. Direct attention on high-dimensional channels is costly and prone to introducing noise. Channel Prior Mixer adopts a centralized aggregation-distribution strategy: first aggregating into shared priors, then distributing to individual channels.

Based on the computed amplitude coherence γ and phase offsets ϕ, we obtain the mixing matrix using learnable scalars α, β and temperature τ:

$$\mathbf{M}_{\text{mix}} = \text{softmax}\Big(\frac{\alpha\gamma + \beta\phi}{\tau}\Big) + \delta\,\mathbf{I}. \tag{2}$$

where $\mathbf{M}_{\text{mix}} \in \mathbb{R}^{C\times C}$, and γ and ϕ are computed from channel priors. The mixed spectrum is interpolated with strength 0.1; guided gating compresses amplitudes to $[0, 1]$.

4.4 Encoder Pluggability

The frequency frontend outputs unified complex tokens, so the encoder can be chosen by budget: a Complex Transformer (best when channels are large), a Complex MLP (linear cost in $BLd_{\text{model}}d_{\text{ff}}$), or a single-layer Complex Linear (lightest). Full comparisons are given in Appendix.

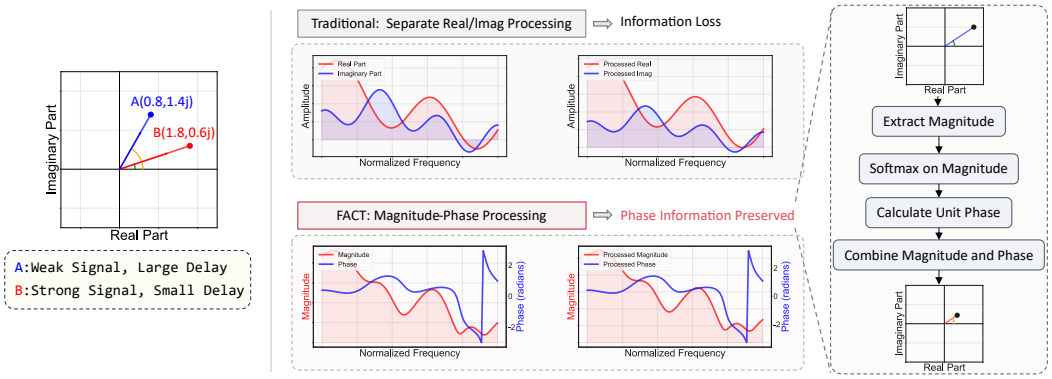

Figure 5: Complex feature handling: traditional real/imaginary split (top) vs. FACT's magnitude–phase processing (bottom). Right: magnitude-softmax and unit-phase reconstruction for complex attention values.

4.5 Feature Alignment

Motivation. Tokens and the raw spectrum are misaligned in length and channels. Simple concatenation causes leakage and ignores priors. We therefore adopt complex cross-attention where the raw spectrum queries the tokens, while prior-driven gating and bias highlight key bands and suppress noise.

This magnitude–phase pipeline (Fig. 5) allows Feature Alignment to gate strong/weak responses based on amplitude while retaining phase delays, which are essential for identifying cross-channel lead–lag relations.

Structure Analysis. The module contains three sub-pathways: (i) query/key projection, which splits complex inputs into real and imaginary parts and then generates multi-head vectors through linear layers; (ii) value projection employs complex linear mappings to preserve phase information; (iii) a gating generator consumes the projected mask and weight summaries, learning injection strength and attention bias for each head through lightweight affine layers. The specific form is:

$$\mathbf{Q} = \mathbf{W}_Q[\Re(\mathbf{X}_{\text{fft}}); \Im(\mathbf{X}_{\text{fft}})], \quad \mathbf{K} = \mathbf{W}_K[\Re(\mathbf{Z}); \Im(\mathbf{Z})], \quad \mathbf{V} = \text{ComplexLinear}(\mathbf{Z}). \tag{3}$$

Prior gating and bias are written as

$$\mathbf{G} = \sigma\big(\mathcal{A}_m(\mathbf{M})\big) \odot \sigma\big(\mathcal{A}_w(\mathbf{W})\big), \quad \mathbf{B} = \mathcal{B}(\mathbf{M}, \mathbf{W}), \tag{4}$$

where $\mathbf{M}$ and $\mathbf{W}$ correspond to the projected mask and weight summaries respectively, and $\mathcal{A}_m$, $\mathcal{A}_w$, $\mathcal{B}$ are lazily initialized linear mappings as needed. The attention output is

$$\mathbf{H}_{\text{fused}} = \text{Softmax}\Big(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}} + \mathbf{B}\Big)\big(\mathbf{V} \odot \mathbf{G}\big). \tag{5}$$

The final result is residually interpolated with the original spectrum according to $\alpha = 0.7$ and normalized by ComplexLayerNorm; complex dropout ensures amplitude consistency. The module caches attention weights, gating strength, and entropy metrics to provide data for subsequent interpretability visualization. The overall complexity remains $\mathcal{O}(n_{\text{heads}} K d^2)$, but thanks to prior gating, it can focus on key frequency bands early in training.

4.6 Complex Transformer Encoder

Motivation. After completing alignment in the frequency domain, it is still necessary to model dependencies over long time spans. Complex Transformer Encoder preserves amplitude-phase information and can simultaneously handle interactions between time and channel dimensions.

The encoder consists of two ComplexFullAttentionLayer layers, with input being the fused features from Section 4.5:

$$\mathbf{H}_{\ell+1} = \text{ComplexLayerNorm}\big(\mathbf{H}_\ell + \text{ComplexMultiHeadAttn}(\mathbf{H}_\ell, \mathbf{H}_\ell, \mathbf{H}_\ell)\big), \tag{6}$$

$$\mathbf{H}_{\ell+1} = \text{ComplexLayerNorm}\big(\mathbf{H}_{\ell+1} + \text{ComplexConv1d}(\mathbf{H}_{\ell+1})\big). \tag{7}$$

where ComplexMultiHeadAttn reuses the weights from Equation equation 3 and adds the same prior bias, and ComplexConv1d performs depthwise separable convolution in the frequency dimension to capture local smoothness. The output is mapped back to $\mathbb{C}^{F\times C}$ through ComplexProjection, and then recovered to time-domain predictions through irFFT and the inverse operations of RevIN and the auxiliary normalization layer.

4.7 Interpretability Regularization

Motivation. During the training phase, directly imposing constraints on cached attention, gating, and priors can avoid the fragmentation of ẗrain first, interpret laterändd enable the model to naturally align with physical mechanisms in the optimization objective.

The specific approach is to cache fusion representations $\hat{\mathbf{H}}$, gating vectors $\mathbf{g}$, mixing matrices $\mathbf{M}_{\text{mix}}$, and frequency-domain phases within each mini-batch. After averaging these cached tensors over the frequency dimension, we estimate amplitude correlations and phase differences: the final-channel embeddings are modulated and normalized to obtain $\hat{\gamma}$, while the mean phase of complex attention yields $\widehat{\Delta\theta}$. These statistics drive the coherence and phase regularizers and are logged for post-hoc interpretability analysis.

$$L_{\text{coh}} = \|\hat{\gamma} - \gamma\|_2^2, \quad \hat{\gamma} = \text{corr}\left(|\hat{\mathbf{H}}|\right), \tag{8}$$

$$L_{\text{phase}} = 1 - \cos\left(\widehat{\Delta\theta} - \phi\right), \tag{9}$$

where γ and ϕ come from the amplitude/phase priors. The total loss is

$$\mathcal{L} = \mathcal{L}_{\text{forecast}} + \lambda_{\text{coh}} L_{\text{coh}} + \lambda_{\text{phase}} L_{\text{phase}} \tag{10}$$

Default weights and sensitivity analyses are reported in Section 5 and Appendix.

Summary. FACT composes Adaptive Band Decomposition with Gaussian filters, channel priors and complex Feature Alignment/encoding, with coherence/phase regularizers for training-time guidance and diagnosis.

5 Experiments

5.1 Datasets

We follow the public SOFTS benchmarks (Han et al., 2024): ETT (4 subsets), Traffic, Electricity, Weather, Solar-Energy, and PEMS (4 subsets). These cover electricity, transportation and energy scenarios with heterogeneous channels and sampling rates. Full statistics (channels, horizons, splits, sampling) are provided in Appendix G (Table 4).

5.2 Training and Implementation Settings

Key hyperparameters (optimizer, depth, hidden size, subset protocol) are summarized in Appendix (Section C).

5.3 Main Results and Ablation

Baselines. We compare against representative linear/MLP (DLinear, TSMixer, TiDE), Transformer (FEDformer, Stationary, PatchTST, Crossformer, iTransformer), and CNN (SCINet, TimesNet) models.

Protocol. We follow standard long-sequence settings (Zhou et al., 2021; Liu et al., 2022): historical window $L = 96$, horizons as in prior work, and MSE/MAE as metrics. Full implementation details are in Appendix C.

Results. Table 5 presents the comprehensive comparison between FACT and state-of-the-art baselines across 12 datasets, showing average performance metrics (MSE/MAE). FACT demonstrates competitive performance on Solar/Weather datasets while leaving headroom for improvement on high-channel regimes (Traffic/PEMS). Notably, FACT achieves 15 first-place results in MSE and 22 first-place results in MAE metrics across all dataset-horizon combinations. The results validate FACT's effectiveness in multivariate long-term forecasting while maintaining interpretability through frequency-domain analysis. Ablations and sensitivity analysis (Appendix J, K) show that removing DynFBD or channel priors consistently degrades accuracy and Γ/Φ, and moderate regularization improves both interpretability metrics and prediction.

Table 1: Multivariate forecasting results with prediction lengths $H \in \{12, 24, 48, 96\}$ for PEMS and $H \in \{96, 192, 336, 720\}$ for others and fixed lookback window length $L = 96$. The results are taken from SOFTS (Han et al., 2024) and iTransformer (Liu et al., 2023).

Models		FACT (ours)		SOFTS		iTransformer		PatchTST		TSMixer		Crossformer		TiDE		TimesNet		DLinear		SCINet		FEDformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	Avg	0.407	0.409	0.393	0.403	0.407	0.410	0.396	0.406	0.398	0.407	0.513	0.496	0.419	0.419	0.400	0.406	0.403	0.407	0.485	0.481	0.448	0.452
ETTm2	Avg	0.298	0.340	0.287	0.330	0.288	0.332	0.287	0.330	0.289	0.333	0.757	0.610	0.358	0.404	0.291	0.333	0.350	0.401	0.571	0.537	0.305	0.349
ETTh1	Avg	0.451	0.446	0.449	0.442	0.454	0.447	0.453	0.446	0.463	0.452	0.529	0.522	0.541	0.507	0.458	0.450	0.456	0.452	0.747	0.647	0.440	0.460
ETTh2	Avg	0.383	0.407	0.373	0.400	0.383	0.407	0.385	0.410	0.401	0.417	0.942	0.684	0.611	0.550	0.414	0.427	0.559	0.515	0.954	0.723	0.437	0.449
ECL	Avg	0.179	0.272	0.174	0.264	0.178	0.270	0.189	0.276	0.186	0.287	0.244	0.334	0.251	0.344	0.192	0.295	0.212	0.300	0.268	0.365	0.214	0.327
Traffic	Avg	0.453	0.290	0.409	0.267	0.428	0.282	0.454	0.286	0.522	0.357	0.550	0.304	0.760	0.473	0.620	0.336	0.625	0.383	0.804	0.509	0.610	0.376
Weather	Avg	0.251	0.279	0.255	0.278	0.258	0.278	0.256	0.279	0.256	0.279	0.259	0.315	0.271	0.320	0.259	0.287	0.265	0.317	0.292	0.363	0.309	0.360
Solar-Energy	Avg	0.229	0.265	0.229	0.256	0.233	0.262	0.236	0.266	0.260	0.297	0.641	0.639	0.347	0.417	0.301	0.319	0.330	0.401	0.282	0.375	0.291	0.381
PEMS03	Avg	0.116	0.222	0.104	0.210	0.113	0.221	0.137	0.240	0.119	0.233	0.169	0.281	0.326	0.419	0.147	0.248	0.278	0.375	0.114	0.224	0.213	0.327
PEMS04	Avg	0.111	0.223	0.102	0.208	0.111	0.221	0.145	0.249	0.103	0.215	0.209	0.314	0.353	0.437	0.129	0.241	0.295	0.388	0.092	0.202	0.231	0.337
PEMS07	Avg	0.090	0.185	0.087	0.184	0.101	0.204	0.144	0.233	0.112	0.217	0.235	0.315	0.380	0.440	0.124	0.225	0.329	0.395	0.119	0.234	0.165	0.283
PEMS08	Avg	0.147	0.230	0.138	0.219	0.150	0.226	0.200	0.275	0.165	0.261	0.268	0.307	0.441	0.464	0.193	0.271	0.379	0.416	0.158	0.244	0.286	0.358

The results in Table 5 demonstrate several key findings: (1) FACT achieves strong performance across diverse datasets, particularly excelling on Solar-Energy and Weather forecasting tasks; (2) The frequency-domain approach proves effective for capturing temporal dependencies while maintaining computational efficiency; (3) FACT's interpretable design does not compromise prediction accuracy, establishing a favorable trade-off between performance and explainability in multivariate time series forecasting.

5.4 Interpretability Visualization

We visualize attention, channel priors and gating trajectories on the interpretability subset; see Appendix L (Fig. 6, Fig. 7).

5.5 Regularization Impact

Sweeping λ_{coh} and λ_{phase} yields a positive correlation between interpretability metrics and accuracy; see Appendix K (tables) and Appendix L (plots).

6 Conclusion

This paper addresses the core challenge of modeling channel interactions in multivariate time series. We identify that existing methods mostly process correlations at the original channel dimension level, often struggling to balance noise suppression with preserving effective information, especially in high-dimensional or long-sequence tasks where they either lose fine-grained mechanisms or introduce high computational complexity. To address this, FACT elevates channel interactions to the frequency-domain component level, more effectively distinguishing signals from noise through dynamic frequency band decomposition and prior-guided complex modeling, while directly integrating fine-grained interpretability constraints into the training process, thereby improving prediction stability while maintaining efficiency.

Beyond serving as an independent predictor, FACT's frequency domain channel modeling ideas can also be integrated into mainstream sequence models in a modular fashion. Experimental results show that using FACT as a model-agnostic plugin can bring consistent performance gains across multiple backbone networks. We believe this direction provides a new perspective for building efficient and interpretable time series systems in the future, and look forward to further validating its potential on larger-scale data and richer tasks.

References

Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Jigang Bao, Yong Jiang, and Shu-Tao Xia. Periodicity decoupling framework for long-term series forecasting. In International Conference on Learning Representations (ICLR), 2024. URL https://openreview.net/forum?id=OpW8q3K45D.

Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 459–469, 2023.

Jianmin Wang Haixu Wu, Jiehui Xu and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, 2021. In NeurIPS, pages 101–112, 2021a.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. Softs: Efficient multivariate time series forecasting with series-core fusion. arXiv preprint arXiv:2404.14197, 2024.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory, 1997. Neural computation, 9 (8): 1735–1780, 1997.

Wei Fan Shoujin Wang Pengyang Wang Hui He Ning An Defu Lian Longbing Cao Kun Yi, Qi Zhang and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting, 2023. In NeurIPS, 2023.

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 Advances in Neural Information Processing Systems, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625, 2023.

Han-Jia Ye Lu Han and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting, 2023. CoRR, abs/2304.05206, 2023.

Xiaowen Ma, Zhen-Liang Ni, Shuai Xiao, and Xinghao Chen. Timepro: Efficient multivariate long-term time series forecasting with variable-and time-aware hyper-state. In Forty-second International Conference on Machine Learning, 2025.

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

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, pp. 2400–2410, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1, pp. 1185–1196, 2025.

J. Zico Kolter Shaojie Bai and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018. CoRR, abs/1803.01271, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

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

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 ICLR, 2025.

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.

Chenjuan Guo Chaoyang He Bin Yang Xinle Wu, Dalin Zhang and Christian S. Jensen. Autocts: Automated correlated time series forecasting, 2021. Proc. VLDB Endow., 15 (4): 971–983, 2021b.

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

Qingsong Wen Jinyang Gao Bolin Ding Xue Wang, Tian Zhou and Rong Jin. Make transformer great again for time series forecasting: Channel aligned robust dual transformer, 2023. CoRR, abs/2305.12095, 2023a.

Fuhao Yang, Xin Li, MIng Wang, Hongyu Zang, Wei Pang, and Mingzhong Wang. Waveform: Graph enhanced wavelet learning for long sequence forecasting of multivariate time series. In Proceedings of the 37th AAAI Conference on Artificial Intelligence, 2023.

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 Thirty-seventh Conference on Neural Information Processing Systems, 2023a.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series forecasting. In Thirty-seventh Conference on Neural Information Processing Systems, 2023b.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. In Advances in Neural Information Processing Systems 37 (NeurIPS 2024), 2024.

Dingshan Cui Jiandong Xie Liwei Deng Feiteng Huang Yue Cui, Kai Zheng and Xiaofang Zhou. Metro: A generic graph neural network framework for multivariate time series forecasting, 2021. Proc. VLDB Endow., 15 (2): 224–236, 2021.

Xingyu Zhang, Siyu Zhao, Zeen Song, Huijie Guo, Jianqi Zhang, Changwen Zheng, and Wenwen Qiang. Not all frequencies are created equal: Towards a dynamic fusion of frequencies in time-series forecasting. In Proceedings of the 32nd ACM International Conference on Multimedia, pp. 4729–4737, 2024a.

Xinyu Zhang, Shanshan Feng, Jianghong Ma, Huiwei Lin, Xutao Li, Yunming Ye, Fan Li, and Yew Soon Ong. Frnet: Frequency-based rotation network for long-term time series forecasting. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3586–3597, 2024b.

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

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International conference on machine learning, pp. 27268–27286. PMLR, 2022.

Guodong Long Jing Jiang Xiaojun Chang Zonghan Wu, Shirui Pan and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks, 2020. In SIGKDD, pages 753–763, 2020.

A Symbol Extensions and Inference Pseudocode

To facilitate reproduction, we supplement the key steps of FACT inference based on the symbols in the main text. The pseudocode mirrors the repository implementation, but we present it here using conceptual module names for clarity:

1. Input tensor $X \in \mathbb{R}^{B\times L\times C}$. If RevIN is enabled, execute $X \leftarrow \mathrm{RevIN}(X)$ to obtain normalized representation; if reversible normalization is enabled, additionally cache mean and variance.
2. Compute $X_{\mathrm{fft}} = \mathcal{F}_{\mathrm{rfft}}(X)$, and pass it through the dynamic frequency-band preprocessor to obtain sparse frequency-domain tokens $\mathbf{Z}$, mask priors $\mathbf{M}$, and frequency-band weights $\boldsymbol{\omega}$.
3. Apply the frequency selector to smooth these weights, producing low-dimensional mask and weight summaries that will act as priors in later stages.
4. When channel mixing is enabled, estimate amplitude coherence γ and phase priors ϕ, construct mixing matrices and guided gating, and cache the resulting channel priors for regularization use.
5. Activate Adaptive Feature Fusion to re-weight frequency-domain representations through complex cross-attention informed by the aforementioned priors; otherwise, directly reuse the mixed spectrum X_{fft}.
6. Transform features back to the time domain and feed them into the chosen complex encoder (Transformer/MLP/Linear), obtaining prediction hidden states through the complex projection layer.
7. If reversible normalization or RevIN reverse process is enabled, restore original scale at output and extract the last T step results.

B Dataset and Preprocessing Details

This paper follows the divisions published in SOFTS (Han et al., 2024), with related statistics in Table ?? in the main text. Due to size limitations, the anonymous code package only includes Solar-137 examples. The loader implementation in the supplementary code package follows the considerations below:

- Data format: By default reads comma-separated floating-point text; for CSV files, skips the header row.

Table 2: FACT default hyperparameters (consistent with open-source implementation).

Module	Key Parameters	Default Values / Notes
RevIN	use_revin, use_complex_revin, ε	true, false, 1×10^{-5}
Frequency Embedding	d_{model}, per-channel scale/bias	128, learnable
BandPreprocessor	B_f, K, mask_proj_dim, weights_proj_dim	3, 128, 16, 8
Channel Prior Mixer	mixing_topk, τ, mixing_strength, diag_bias, α, β	16, 1.0, 0.1, 0.2, learnable
Guided Gating	gate_bias, gate_scale	0.5, 0.5
Adaptive Feature Fusion	n_{heads}, dropout, α	8, 0.1, 0.7
Complex Encoder	e_{layers}, d_{ff}	2 (main exp.) / 1 (interpretability subset), 512

- Split strategy: Splits training/validation/test in chronological order according to 70/10/20, and fits the normalizer on the training set to prevent information leakage.
- Window parameters: the default window configuration $[96, 48, 96]$ is maintained as in the main experiments; the optional subsampling limit is set to 2000 rows for quick validation and can be disabled to load complete files.
- Temporal features: The anonymous release only supports the multivariate setting with standard time-encoding flags, consistent with Solar examples.

C Training and Implementation Configuration

Training uses the public entry point, with key hyperparameter default values as follows:

- Optimizer uses AdamW with learning rate 5×10^{-4}, combined with cosine annealing and linear warmup.
- Batch size 32, training epochs 10, early stopping patience 3. Interpretability subset scripts reduce the number of training epochs to three to shorten visualization generation time.
- Regularization coefficients λ_{coh} and λ_{phase} default to 0.01, and are skipped automatically when channel priors are unavailable.
- Complex attention defaults to two layers, hidden dimension 128, feedforward dimension 512; the token length produced by DynFBD is 128.
- Hardware configuration needs to be specified via the environment variable CUDA_VISIBLE_DEVICES before execution; the reference launch script in the supplementary package defaults to single-card operation.

D Complexity and Parameter Count Supplement

A concise per-module complexity summary is reported in Table 3. Additional notes: DynFBD uses only three Gaussian kernels with token length $K = 128$; the channel mixer computes correlations on the top-k frequency bands with complexity $\mathcal{O}(BCk)$; adaptive fusion inherits the multi-head attention structure with cost $\mathcal{O}(n_{\text{heads}} K d^2)$.

E Additional Experimental Results

Detailed interpretability metrics and regularization sensitivity statistics for Solar and Weather datasets are provided with accompanying CSV files, with values consistent with the main text analysis and can be directly accessed in the accompanying CSV tables.

Table 3: Time complexity overview of main modules (default $B_f = 3$, $K = 128$, top-k=16).

Module	Main Complexity	Description
rFFT	$\mathcal{O}(LC \log L)$	One rFFT per channel
DynFBD	$\mathcal{O}(B_f KC)$	Complex linear mapping, band projection
Channel Prior Mixer	$\mathcal{O}(Ck)$	Aggregation after top-k selection
Guided Gating	$\mathcal{O}(CF)$	Weighted amplitude normalization
Adaptive Fusion	$\mathcal{O}(n_{\text{heads}} K d^2)$	Complex cross-attention
Complex Encoder	$\mathcal{O}(n_{\text{layers}} d^2 K)$	Two ComplexFullAttentionLayer layers

Table 4: Dataset statistics (channels, horizons, splits, sampling rates).

Dataset	Channels	Prediction Horizon H	Data Split (Train, Val, Test)	Sampling Rate	Domain
ETTh1, ETTh2	7	$\{96, 192, 336, 720\}$	(8545, 2881, 2881)	Hourly	Electricity
ETTm1, ETTm2	7	$\{96, 192, 336, 720\}$	(34465, 11521, 11521)	15min	Electricity
Weather	21	$\{96, 192, 336, 720\}$	(36792, 5271, 10540)	10min	Weather
ECL	321	$\{96, 192, 336, 720\}$	(18317, 2633, 5261)	Hourly	Electricity
Traffic	862	$\{96, 192, 336, 720\}$	(12185, 1757, 3509)	Hourly	Traffic
Solar-Energy	137	$\{96, 192, 336, 720\}$	(36601, 5161, 10417)	10min	Energy
PEMS03	358	$\{12, 24, 48, 96\}$	(15617, 5135, 5135)	5min	Traffic
PEMS04	307	$\{12, 24, 48, 96\}$	(10172, 3375, 3375)	5min	Traffic
PEMS07	883	$\{12, 24, 48, 96\}$	(16911, 5622, 5622)	5min	Traffic
PEMS08	170	$\{12, 24, 48, 96\}$	(10690, 3548, 3548)	5min	Traffic

F Encoder Pluggability Experiments

To echo the discussion in Section 4.4 of the main text, we evaluate three backends: Complex Transformer, Complex MLP, and Complex Linear under the unified configuration of seq_len=96 and pred_len=96. Core data as follows:

- Electricity: Linear/MLP/Transformer MSE are 0.1547/0.1527/0.1454, MAE are 0.2541/0.2516/0.2428; single epoch times are 43.14s/45.72s/99.37s respectively, with lightweight backends compressing training time by half while error increases by less than 5%.
- Solar-137: Linear/MLP/Transformer MSE are 0.2109/0.1982/0.1921, MAE are 0.2642/0.2489/0.2356; single epoch times are 39.84s/43.39s/74.59s, similarly demonstrating the trend of "lightweight backends significantly saving computational resources".

Therefore, under the condition of keeping the frequency-domain frontend and interpretability regularization unchanged, Linear and MLP can provide more cost-effective options for computationally constrained deployment scenarios.

G Dataset Statistics

Full statistics of the reused benchmarks are reported in Table 4.

H Preliminaries (Full)

H.1 Multivariate Long-term Forecasting Setup

Let the input sequence be $\mathbf{X} \in \mathbb{R}^{B \times L \times C}$. The target is to predict $\mathbf{Y} \in \mathbb{R}^{B \times T \times C}$ with loss $\mathcal{L}_{\text{forecast}} = \frac{1}{BCT} \sum_{b,t,c} (Y_{b,t,c} - \hat{Y}_{b,t,c})^2$.

H.2 Real Fast Fourier Transform and Complex Representation

Stack the time series as $\mathbf{X} \in \mathbb{R}^{L\times C}$, rFFT yields $\mathbf{X}_{\text{fft}} = \mathcal{F}_{\text{rfft}}(\mathbf{X}) \in \mathbb{C}^{F\times C}$ with $F = L/2+1$. For frequency f and channel c, $\mathbf{X}_{\text{fft}}(f,c) = A(f,c)e^{\mathrm{i}\theta(f,c)}$.

H.3 Dynamic Frequency-Band Decomposition

For band i, the Gaussian weight is

$$\omega_i(f) = \frac{\exp\big(-(f-\mu_i)^2/(2\sigma_i^2)\big)}{\sum_{j=1}^{B_f}\exp\big(-(f-\mu_j)^2/(2\sigma_j^2)\big)}, \tag{11}$$

where μ_i, σ_i are learnable and $B_f = 3$ by default. Each band is compressed into K-dimensional tokens via complex linear projection.

H.4 Frequency Selection and Projection

Given $\mathbf{Z} \in \mathbb{C}^{K\times CB_f}$, the selector computes

$$\boldsymbol{\alpha} = \text{softmax}\Big(\text{Mean}_b(\sigma(|\mathbf{W}_1\mathbf{Z}|))\Big), \tag{12}$$

and projects it into mask/weight summaries $\mathbf{P}_{\text{mask}} \in \mathbb{R}^{F\times d_m}$ and $\mathbf{P}_{\text{weight}} \in \mathbb{R}^{K\times d_w}$ for subsequent priors and attention bias.

H.5 Channel Correlation and Phase Priors

Weighted amplitudes $\mathbf{A}_{c,f} = w_{\text{eff}}(f)(A(f,c) - \text{Mean}_f A(f,c))$ lead to

$$\gamma = \mathbf{A}\mathbf{D}^{-1}\mathbf{A}^\top, \tag{13}$$

where $\mathbf{D}$ normalizes $\gamma \in [-1,1]^{C\times C}$. Phase offsets summarize lead/lag:

$$\phi = \frac{\sin\boldsymbol{\theta}\ \cos\boldsymbol{\theta}^\top - \cos\boldsymbol{\theta}\ \sin\boldsymbol{\theta}^\top}{\max|\sin\boldsymbol{\theta}\ \cos\boldsymbol{\theta}^\top - \cos\boldsymbol{\theta}\ \sin\boldsymbol{\theta}^\top|}, \tag{14}$$

where $\sin\boldsymbol{\theta}, \cos\boldsymbol{\theta} \in \mathbb{R}^C$ are weighted by frequency.

H.6 Complex Operators and Guided Gating

For $\mathbf{z} = \mathbf{z}_r + \mathrm{i}\,\mathbf{z}_i$, a complex linear layer is

$$\text{ComplexLinear}(\mathbf{z}) = (\mathbf{W}_r\mathbf{z}_r - \mathbf{W}_i\mathbf{z}_i) + \mathrm{i}(\mathbf{W}_i\mathbf{z}_r + \mathbf{W}_r\mathbf{z}_i). \tag{15}$$

Guided gating compresses weighted amplitudes to $[0,1]$ via

$$\mathbf{s} = \text{Norm}_c(\text{Mean}_f\, w_{\text{eff}}(f)|\mathbf{X}_{\text{fft}}(f,\cdot)|), \quad \mathbf{g} = \text{gate\_bias} + \text{gate\_scale}\cdot\text{clip}(\mathbf{s},0,1), \tag{16}$$

which stabilizes optimization and supports interpretability regularization.

I Full Results

The complete cross-dataset comparison Table ?? is provided below.

J Ablations and Regularization Tables

We report full ablation and regularization sweeps.

Table 5: Multivariate forecasting results with prediction lengths $H \in \{12, 24, 48, 96\}$ for PEMS and $H \in \{96, 192, 336, 720\}$ for others and fixed lookback window length $L = 96$. The results are taken from SOFTS and iTransformer (?).

Models		FACT (ours)		SOFTS		iTransformer		PatchTST		TSMixer		Crossformer		TiDE		TimesNet		DLinear		SCINet		FEDformer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	0.327	0.361	0.325	0.361	0.334	0.368	0.329	0.365	0.323	0.363	0.404	0.426	0.364	0.387	0.338	0.375	0.345	0.372	0.418	0.438	0.379	0.419
	192	0.376	0.392	0.375	0.389	0.377	0.391	0.380	0.394	0.376	0.392	0.450	0.451	0.398	0.404	0.374	0.387	0.380	0.389	0.439	0.450	0.426	0.441
	336	0.422	0.418	0.405	0.412	0.426	0.420	0.400	0.410	0.407	0.413	0.532	0.515	0.428	0.425	0.410	0.411	0.413	0.413	0.490	0.485	0.445	0.459
	720	0.502	0.463	0.466	0.447	0.491	0.459	0.475	0.453	0.485	0.459	0.666	0.589	0.487	0.461	0.478	0.450	0.474	0.453	0.595	0.550	0.543	0.490
	Avg	0.407	0.409	0.393	0.403	0.407	0.410	0.396	0.406	0.398	0.407	0.513	0.496	0.419	0.419	0.400	0.406	0.403	0.407	0.485	0.481	0.448	0.452
ETTm2	96	0.193	0.275	0.180	0.261	0.180	0.264	0.184	0.264	0.182	0.266	0.287	0.366	0.207	0.305	0.187	0.267	0.193	0.292	0.286	0.377	0.203	0.287
	192	0.271	0.329	0.246	0.306	0.250	0.309	0.246	0.306	0.249	0.309	0.414	0.492	0.290	0.364	0.249	0.309	0.284	0.362	0.399	0.445	0.269	0.328
	336	0.312	0.349	0.319	0.352	0.311	0.348	0.308	0.346	0.309	0.347	0.597	0.542	0.377	0.422	0.321	0.351	0.369	0.427	0.637	0.591	0.325	0.366
	720	0.417	0.408	0.405	0.401	0.412	0.407	0.409	0.402	0.416	0.408	1.730	1.042	0.558	0.524	0.408	0.403	0.554	0.522	0.960	0.735	0.421	0.415
	Avg	0.298	0.340	0.287	0.330	0.288	0.332	0.287	0.330	0.289	0.333	0.757	0.610	0.358	0.404	0.291	0.333	0.350	0.401	0.571	0.537	0.305	0.349
ETTh1	96	0.384	0.404	0.381	0.399	0.386	0.405	0.394	0.406	0.401	0.412	0.423	0.448	0.479	0.464	0.384	0.402	0.386	0.400	0.654	0.599	0.376	0.419
	192	0.436	0.436	0.435	0.431	0.441	0.436	0.440	0.435	0.452	0.442	0.471	0.474	0.525	0.492	0.436	0.429	0.437	0.432	0.719	0.631	0.420	0.448
	336	0.480	0.458	0.480	0.452	0.487	0.458	0.491	0.462	0.492	0.463	0.570	0.546	0.565	0.515	0.491	0.469	0.481	0.459	0.778	0.659	0.459	0.465
	720	0.504	0.486	0.499	0.488	0.503	0.491	0.487	0.479	0.507	0.490	0.653	0.621	0.594	0.558	0.521	0.500	0.519	0.516	0.836	0.699	0.506	0.507
	Avg	0.451	0.446	0.449	0.442	0.454	0.447	0.453	0.446	0.463	0.452	0.529	0.522	0.541	0.507	0.458	0.450	0.456	0.452	0.747	0.647	0.440	0.460
ETTh2	96	0.307	0.356	0.297	0.347	0.297	0.349	0.288	0.340	0.319	0.361	0.745	0.584	0.400	0.440	0.340	0.374	0.333	0.387	0.707	0.621	0.358	0.397
	192	0.383	0.400	0.373	0.394	0.380	0.400	0.376	0.395	0.402	0.410	0.877	0.656	0.528	0.509	0.402	0.414	0.477	0.476	0.860	0.689	0.429	0.439
	336	0.422	0.430	0.410	0.426	0.428	0.432	0.440	0.451	0.444	0.446	1.043	0.731	0.643	0.571	0.452	0.452	0.594	0.541	1.000	0.744	0.496	0.487
	720	0.422	0.442	0.411	0.433	0.427	0.445	0.436	0.453	0.441	0.450	1.104	0.763	0.874	0.679	0.462	0.468	0.831	0.657	1.249	0.838	0.463	0.474
	Avg	0.383	0.407	0.373	0.400	0.383	0.407	0.385	0.410	0.401	0.417	0.942	0.684	0.611	0.550	0.414	0.427	0.559	0.515	0.954	0.723	0.437	0.449
ECL	96	0.146	0.241	0.143	0.233	0.148	0.240	0.164	0.251	0.157	0.260	0.219	0.314	0.237	0.329	0.168	0.272	0.197	0.282	0.247	0.345	0.193	0.308
	192	0.178	0.268	0.158	0.248	0.162	0.253	0.173	0.262	0.173	0.274	0.231	0.322	0.236	0.330	0.184	0.289	0.196	0.285	0.257	0.355	0.201	0.315
	336	0.187	0.280	0.178	0.269	0.178	0.269	0.190	0.279	0.192	0.295	0.246	0.337	0.249	0.344	0.198	0.300	0.209	0.301	0.269	0.369	0.214	0.329
	720	0.206	0.300	0.218	0.305	0.225	0.317	0.230	0.313	0.223	0.318	0.280	0.363	0.284	0.373	0.220	0.320	0.245	0.333	0.299	0.390	0.246	0.355
	Avg	0.179	0.272	0.174	0.264	0.178	0.270	0.189	0.276	0.186	0.287	0.244	0.334	0.251	0.344	0.192	0.295	0.212	0.300	0.268	0.365	0.214	0.327
Traffic	96	0.409	0.273	0.376	0.251	0.395	0.268	0.427	0.272	0.493	0.336	0.522	0.290	0.805	0.493	0.593	0.321	0.650	0.396	0.788	0.499	0.587	0.366
	192	0.427	0.279	0.398	0.261	0.417	0.276	0.454	0.289	0.497	0.351	0.530	0.293	0.756	0.474	0.617	0.336	0.598	0.370	0.789	0.505	0.604	0.373
	336	0.465	0.294	0.415	0.269	0.433	0.283	0.450	0.282	0.528	0.361	0.558	0.305	0.762	0.477	0.629	0.336	0.605	0.373	0.797	0.508	0.621	0.383
	720	0.512	0.315	0.447	0.287	0.467	0.302	0.484	0.301	0.569	0.380	0.589	0.328	0.719	0.449	0.640	0.350	0.645	0.394	0.841	0.523	0.626	0.382
	Avg	0.453	0.290	0.409	0.267	0.428	0.282	0.454	0.286	0.522	0.357	0.550	0.304	0.760	0.473	0.620	0.336	0.625	0.383	0.804	0.509	0.610	0.376
Weather	96	0.167	0.213	0.166	0.208	0.174	0.214	0.176	0.217	0.166	0.210	0.158	0.230	0.202	0.261	0.172	0.220	0.196	0.255	0.221	0.306	0.217	0.296
	192	0.214	0.255	0.217	0.253	0.221	0.254	0.221	0.256	0.215	0.256	0.206	0.277	0.242	0.298	0.219	0.261	0.237	0.296	0.261	0.340	0.276	0.336
	336	0.273	0.299	0.282	0.300	0.278	0.296	0.275	0.296	0.287	0.300	0.272	0.335	0.287	0.335	0.280	0.306	0.283	0.335	0.309	0.378	0.339	0.380
	720	0.350	0.349	0.356	0.351	0.358	0.347	0.352	0.346	0.355	0.348	0.398	0.418	0.351	0.386	0.365	0.359	0.345	0.381	0.377	0.427	0.403	0.428
	Avg	0.251	0.279	0.255	0.278	0.258	0.278	0.256	0.279	0.256	0.279	0.259	0.315	0.271	0.320	0.259	0.287	0.265	0.317	0.292	0.363	0.309	0.360
Solar-Energy	96	0.192	0.236	0.200	0.230	0.203	0.237	0.205	0.246	0.221	0.275	0.310	0.331	0.312	0.399	0.250	0.292	0.290	0.378	0.237	0.344	0.242	0.342
	192	0.233	0.269	0.229	0.253	0.233	0.261	0.237	0.267	0.268	0.306	0.734	0.725	0.339	0.416	0.296	0.318	0.320	0.398	0.280	0.380	0.285	0.380
	336	0.240	0.275	0.243	0.269	0.248	0.273	0.250	0.276	0.272	0.294	0.750	0.735	0.368	0.430	0.319	0.330	0.353	0.415	0.304	0.389	0.282	0.376
	720	0.251	0.280	0.245	0.272	0.249	0.275	0.252	0.275	0.281	0.313	0.769	0.765	0.370	0.425	0.338	0.337	0.356	0.413	0.308	0.388	0.357	0.427
	Avg	0.229	0.265	0.229	0.256	0.233	0.262	0.236	0.266	0.260	0.297	0.641	0.639	0.347	0.417	0.301	0.319	0.330	0.401	0.282	0.375	0.291	0.381
PEMS03	12	0.063	0.166	0.064	0.165	0.071	0.174	0.073	0.178	0.075	0.186	0.090	0.203	0.178	0.305	0.085	0.192	0.122	0.243	0.066	0.172	0.126	0.251
	24	0.084	0.191	0.083	0.188	0.093	0.201	0.105	0.212	0.095	0.210	0.121	0.240	0.257	0.371	0.118	0.223	0.201	0.317	0.085	0.198	0.149	0.275
	48	0.127	0.234	0.114	0.223	0.125	0.236	0.159	0.264	0.121	0.240	0.202	0.317	0.379	0.463	0.155	0.260	0.333	0.425	0.127	0.238	0.227	0.348
	96	0.191	0.296	0.156	0.264	0.164	0.275	0.210	0.305	0.184	0.295	0.262	0.367	0.490	0.539	0.228	0.317	0.457	0.515	0.178	0.287	0.348	0.434
	Avg	0.116	0.222	0.104	0.210	0.113	0.221	0.137	0.240	0.119	0.233	0.169	0.281	0.326	0.419	0.147	0.248	0.278	0.375	0.114	0.224	0.213	0.327
PEMS04	12	0.075	0.179	0.074	0.176	0.078	0.183	0.085	0.189	0.079	0.188	0.098	0.218	0.219	0.340	0.087	0.195	0.148	0.272	0.073	0.177	0.138	0.262
	24	0.091	0.200	0.088	0.194	0.095	0.205	0.115	0.222	0.089	0.201	0.131	0.256	0.292	0.398	0.103	0.215	0.224	0.340	0.084	0.193	0.177	0.293
	48	0.118	0.233	0.110	0.219	0.120	0.233	0.167	0.273	0.111	0.222	0.205	0.326	0.409	0.478	0.136	0.250	0.355	0.437	0.099	0.211	0.270	0.368
	96	0.162	0.280	0.135	0.244	0.150	0.262	0.211	0.310	0.133	0.247	0.402	0.457	0.492	0.532	0.190	0.303	0.452	0.504	0.114	0.227	0.341	0.427
	Avg	0.111	0.223	0.102	0.208	0.111	0.221	0.145	0.249	0.103	0.215	0.209	0.314	0.353	0.437	0.129	0.241	0.295	0.388	0.092	0.202	0.231	0.337
PEMS07	12	0.056	0.150	0.057	0.152	0.067	0.165	0.068	0.163	0.073	0.181	0.094	0.200	0.173	0.304	0.082	0.181	0.115	0.242	0.068	0.171	0.109	0.225
	24	0.072	0.168	0.073	0.173	0.088	0.190	0.102	0.201	0.090	0.199	0.139	0.247	0.271	0.383	0.101	0.204	0.210	0.329	0.119	0.225	0.125	0.244
	48	0.098	0.196	0.096	0.195	0.110	0.215	0.170	0.261	0.124	0.231	0.311	0.369	0.446	0.495	0.134	0.238	0.398	0.458	0.149	0.237	0.165	0.288
	96	0.133	0.227	0.120	0.218	0.139	0.245	0.236	0.308	0.163	0.255	0.396	0.442	0.628	0.577	0.181	0.279	0.594	0.553	0.141	0.234	0.262	0.376
	Avg	0.090	0.185	0.087	0.184	0.101	0.204	0.144	0.233	0.112	0.217	0.235	0.315	0.380	0.440	0.124	0.225	0.329	0.395	0.119	0.234	0.165	0.283
PEMS08	12	0.074	0.173	0.074	0.171	0.079	0.182	0.098	0.205	0.083	0.189	0.165	0.214	0.227	0.343	0.112	0.212	0.154	0.276	0.087	0.184	0.173	0.273
	24	0.098	0.198	0.104	0.201	0.115	0.219	0.162	0.266	0.117	0.226	0.215	0.260	0.318	0.409	0.141	0.238	0.248	0.353	0.122	0.221	0.210	0.301
	48	0.149	0.241	0.164	0.253	0.186	0.235	0.238	0.311	0.196	0.299	0.315	0.355	0.497	0.510	0.198	0.283	0.440	0.470	0.189	0.270	0.320	0.394
	96	0.265	0.307	0.211	0.253	0.221	0.267	0.303	0.318	0.266	0.331	0.377	0.397	0.721	0.592	0.320	0.351	0.674	0.565	0.236	0.300	0.442	0.465
	Avg	0.147	0.230	0.138	0.219	0.150	0.226	0.200	0.275	0.165	0.261	0.268	0.307	0.441	0.464	0.193	0.271	0.379	0.416	0.158	0.244	0.286	0.358

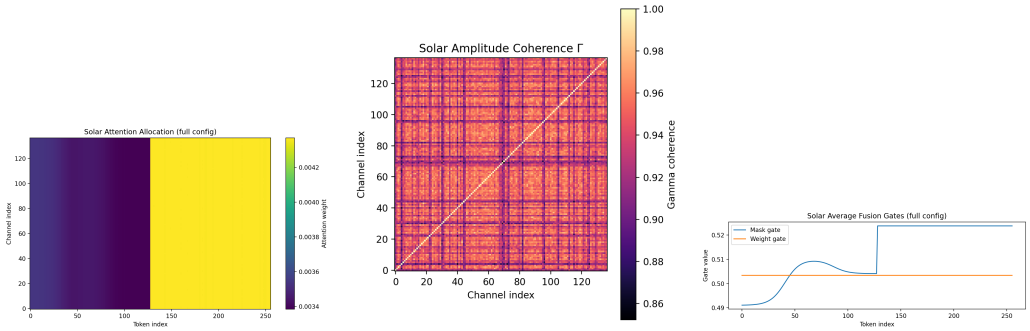

Figure 6: Attention, Γ heatmaps and gating trajectories for Solar-137 interpretability subset.

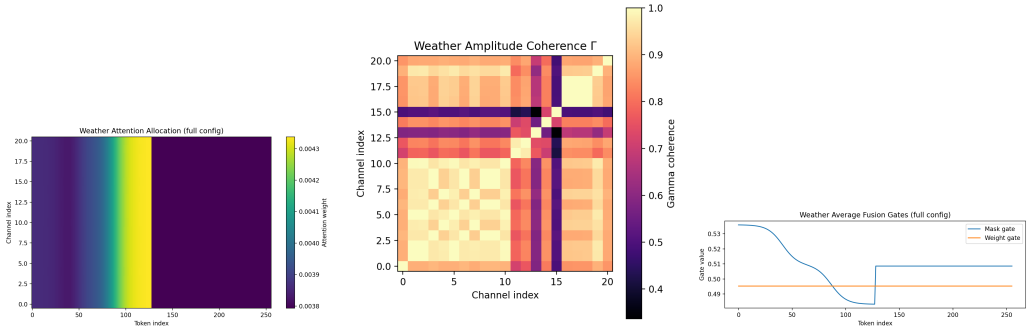

Figure 7: Attention, Γ heatmaps and gating trajectories for Weather interpretability subset.

K Regularization Sweeps

L Interpretability Visualizations (Full)

Additional attention heatmaps, channel priors, and gating trajectories for Solar and Weather. Figures would be included here when images are available

(Figures of fixed-threshold baseline and learnable Gaussian filters are presented in the main text; we omit duplicates here.) Figures would be included here when images are available

M Reproduction Workflow Summary

All figures and tables can be automatically generated through the auxiliary scripts shipped with the supplementary package. We keep the outline below at a high level and redact internal file names. [Placeholder: provide the finalized reproducibility checklist for the camera-ready submission.]

- Main results: run the standard FACT training recipe on Solar with DynFBD, channel mixing, and adaptive fusion enabled.
- Interpretability subset: execute the lightweight configuration on curated Solar/Weather subsets (4,096 samples, $e_{\text{layers}} = 1$, 3 epochs).
- Attention heatmaps: post-process cached interpretability tensors to render attention and gating visualizations for Solar.
- Physical alignment: consolidate interpretability caches to compute Γ/Φ alignment statistics against meteorological variables.
- Regularization analysis: sweep coherence/phase regularization coefficients and export the summarized metrics.

The README in the supplementary scripts directory provides dataset-specific parameter examples that extend to domains such as Traffic and ECL.

N Reproducibility Checklist

High-level command reference for reproducing the main results and analyses:

- Main results: run the standard FACT training recipe with DynFBD, channel mixing, and adaptive fusion enabled.
- Interpretability subset: execute the lightweight configuration on Solar/Weather (4,096 samples, one encoder layer, three epochs).
- Heatmaps: post-process cached tensors to render attention and gating visualizations.
- Physical alignment: compute alignment between Γ/Φ and meteorological variables.
- Regularization: sweep $\lambda_{\text{coh}}/\lambda_{\text{phase}}$ and export summary tables.

O Ethics Statement

This research complies with the ICLR Code of Ethics. All experiments are based on public benchmarks.

The release and use of publicly available datasets respect their respective licenses and intended purposes. The proposed methodology is developed for scientific research and carries minimal risk of harmful applications. We acknowledge the broader concerns of fairness and bias in machine learning models, and we have taken steps to evaluate model robustness and to mitigate unintended discrimination.

No sensitive personal attributes were included in training or evaluation. This work does not involve conflicts of interest, unauthorized sponsorship, or activities that may compromise privacy, security, or research integrity.

P Reproducibility Statement

To facilitate the verification and extension of our work, we provide the following resources:

- Code Availability: The complete implementation is available at: https://anonymous.4open.science/r/FACT
- Datasets: All experiments are based on public benchmarks (ETT, Traffic, Electricity, Weather, Solar-Energy).
- Key Components: The core innovations include:
  - Dynamic Frequency-Band Decomposition (DynFBD)
  - ChannelPriorMixer for amplitude-phase priors
  - Complex cross-attention fusion
- Training Setup: We employ standard hyperparameters (learning rate=5e-4, batch size=32) alongside coherence and phase regularization.

We confirm that all reported results can be reproduced with minimal error using the provided resources and configuration.

Q LLM Usage

Large Language Models (LLMs) were used exclusively for polishing the language and writing of this manuscript. The LLM contributed neither to the research conception nor to the core intellectual content. We bear full responsibility for the work presented herein.