NOT ALL BANDS NEED ALL MODALITIES: FREQUENCY-GATED MULTIMODAL FUSION FOR TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multimodal time series forecasters augment numerical inputs with text and image views, but existing designs absorb both auxiliary views indiscriminately: regardless of what these views actually describe, one fusion decision is applied to the whole series and therefore to every frequency component alike. We argue that fusion should instead be selective in the spectral domain. For example, the trend and seasonality summarised by text describe slow, low-frequency structure. High-frequency components should therefore absorb less from the text view than low-frequency ones, since fusing it across the whole spectrum can inject noise into the forecast. To address this issue, we propose FREQGATE, a frequency-gated multimodal forecaster that assigns fusion weights to each frequency band of the input series. Soft Frequency Partitioning first learns a soft assignment of the spectrum of each input series to low-, mid- and high-frequency bands, rather than imposing a fixed partition shared by all series. Band-Gated Fusion then maps band energies to gates for the text and image encoders. Within each band, numerical tokens attend to the encoders through cross-attention weighted by the corresponding gates. As a result, each auxiliary view contributes only to the bands it can express, learned per series from its spectrum alone. Extensive experiments on eight benchmarks under full-shot, few-shot and zero-shot regimes demonstrate that FREQGATE outperforms fifteen existing forecasters.

1 INTRODUCTION

Time series forecasting informs critical decisions in energy dispatch (Jalalifar et al., 2024; Weron, 2014), traffic control (Liu et al., 2024a), clinical monitoring (Kaushik et al., 2020) and financial risk management (Cheng et al., 2022). Deep forecasters have replaced handcrafted pipelines, evolving from recurrent and convolutional encoders into architectures based on Transformers and MLPs, such as Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022), PatchTST (Nie et al., 2023), TimesNet (Wu et al., 2023), DLinear (Zeng et al., 2023) and iTransformer (Liu et al., 2024c). These methods operate solely on raw numerical series, treating forecasting as a unimodal problem.

Recent work augments numerical series with text and image views. These methods differ in where the auxiliary view enters the numerical pathway. One line injects it at the input: the series is wrapped into a textual prompt or prefixed with domain instructions, and the concatenation is processed by a shared backbone (Jin et al., 2024; Cao et al., 2024; Liu et al., 2024b). Another encodes the views separately and merges them at the representation level: Time-VLM (Zhong et al., 2025) enriches temporal, visual and textual representations with a frozen vision-language model; TimeCMA (Liu et al., 2025) retrieves the time series component of a prompt embedding through similarity across channels; M3Time (Jia et al., 2026) aligns trend and seasonal components at multiple scales with LLM semantic priors; and T3Time (Chowdhury et al., 2026) fuses time, spectral and prompt branches via adaptive alignment across modalities. In both cases the auxiliary view is absorbed at a single strength: whatever the alignment mechanism, one fusion decision is made per input and applied to the entire series. The same text and image content is therefore mixed into every frequency component alike (Figure 1b).

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.

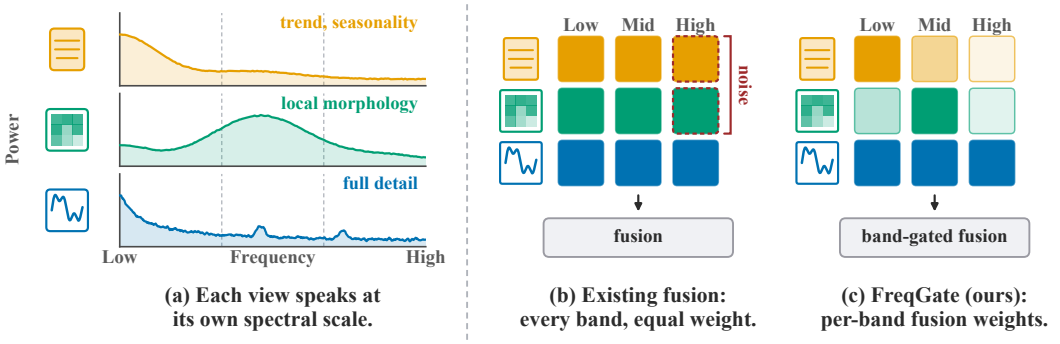


Figure 1: Motivation for FREQGATE. (a) Every view carries energy across the whole spectrum but concentrates it at its own scale: text at low frequencies, images at intermediate scales, and the numerical series extending to high frequencies. (b) Existing fusion absorbs both auxiliary views into every band alike, admitting noise where a view carries no predictive content. (c) FREQGATE assigns each band its own fusion weights; square opacity marks the absorbed weight.

We argue that fusion should instead be spectrally selective, and the reason lies in how the auxiliary views are constructed. Textual inputs verbalise summary statistics of the series: trend, seasonality and notable extrema. By construction, such summaries describe slow, low-frequency structure. Visual inputs render the series as images: they make morphology salient at intermediate scales, but quantise amplitudes into pixels, so the exact values of fast fluctuations are lost. We therefore expect the predictive content of both views to concentrate at low and mid frequencies (Figure 1a). This is an instance of a more general principle, which we call *spectral locality*: the predictive content of an auxiliary view is concentrated in the frequency range that its construction can express. Indiscriminate fusion ignores this principle and absorbs both views in full, coupling high-frequency components with content that barely describes them and risking noise injection into the forecast.

Guided by this principle, we introduce FREQGATE, a frequency-gated multimodal forecaster that assigns fusion weights per frequency band rather than per input (Figure 1c). Specifically, Soft Frequency Partitioning (SFP) splits each input spectrum into low-, mid- and high-frequency bands using a learnable band template that a lightweight offset network refines for every input. Both are trained end-to-end, so the partition adapts to the spectral semantics of each series instead of following one fixed template shared by all series. Conditioned on this partition, Band-Gated Fusion (BGF) uses a lightweight MLP to map band energies to gates for the text and image encoders. Within each band, numerical tokens attend to the encoders through cross-attention modulated by the corresponding gates, so each modality contributes only where it is informative.

Our contributions are threefold:

- We articulate a principle of spectral locality for multimodal forecasting: the predictive content of an auxiliary view is concentrated in the frequency range that its construction can express, so fusion weights should be assigned per frequency band rather than per input.
- We introduce FREQGATE, a frequency-gated multimodal forecaster with two modules: SFP learns a soft frequency partition of each input spectrum, and BGF converts band energies into gates that weight the text and image encoders band by band.
- On eight benchmarks, FREQGATE outperforms fifteen baselines under full-shot, few-shot and zero-shot regimes.

2 RELATED WORK

Unimodal time series forecasting. Forecasters for long horizons can be broadly grouped into three paradigms. Transformers capture long-range dependencies with attention that is sparse, inverted, de-stationary or coupled with decomposition, including Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022), PatchTST (Nie et al., 2023), ETSformer (Woo et al., 2022), iTransformer (Liu et al., 2024c) and Non-stationary Transformer (Liu et al., 2022). Convolu-

tional models keep computation local: TimesNet (Wu et al., 2023) folds the series along its dominant periods into a two-dimensional tensor and refines it with local kernels. MLP and linear models such as N-BEATS (Oreshkin et al., 2020), DLinear (Zeng et al., 2023) and TimeMixer (Wang et al., 2024) offer lightweight alternatives. All of these forecast from the numerical series alone.

Multimodal time series forecasting. Multimodal forecasters jointly model the numerical series and auxiliary views, which are often constructed from the series itself: Gramian Angular Fields (Wang & Oates, 2015), recurrence plots (Eckmann et al., 1987) and line charts expose visual structure, while textual prompts summarise its trend, seasonality and extrema. One group attaches a single auxiliary view, pairing the series with textual prompts, instructions or domain knowledge (Cao et al., 2024; Liu et al., 2024b; Wang et al., 2025; Wu et al., 2025). Another fuses all three views at once: Time-VLM (Zhong et al., 2025) couples chart images and text prompts with a vision-language model; M3Time (Jia et al., 2026) combines multimodal representations with temporal modelling at multiple scales; T3Time (Chowdhury et al., 2026) encodes temporal and prompt representations before aligning them with adaptive multi-head attention across modalities; and DMMV (Shen et al., 2025) constructs image and text views of the same signal and integrates them through fusion guided by decomposition. These designs differ in what they align and where they inject it, but none of them varies fusion along frequency: the auxiliary view enters at one strength for the entire spectrum.

Frequency domain modelling in time series. The spectrum has long served as an inductive bias for unimodal forecasters, whether by attending in frequency (Zhou et al., 2022), discovering periodicities through the FFT (Wu et al., 2023), interpolating the series in frequency (Xu et al., 2024) or learning Koopman predictors (Liu et al., 2023). Multimodal forecasters likewise encode spectral structure as an additional view: M3Time (Jia et al., 2026) models temporal components at multiple frequencies, and T3Time (Chowdhury et al., 2026) encodes a dedicated spectral branch alongside its time and prompt branches. Conditional computation links expert routing to frequency structure: M^{2}FMoE (Huang et al., 2026) assigns interchangeable numerical experts to disjoint Fourier and wavelet bands with learned, shared boundaries, and FM-LLM (Gu et al., 2026) reconstructs the periodic backbone with shared experts used for every input while routed ones model non-periodic residuals. Across all of these uses, however, the spectrum represents the numerical series or organises computation within it, never how much an auxiliary modality should contribute, and *spectral locality* is therefore left unexploited. To our knowledge, FREQGATE is the first forecaster to let the input spectrum decide, band by band, how much of each auxiliary modality to admit.

3 METHOD

We propose FREQGATE, a frequency-gated multimodal forecaster that absorbs the auxiliary modalities selectively, band by band (Figure 2). FREQGATE comprises two modules sharing one learned spectral coordinate system: SFP learns a soft partition of each input spectrum and decomposes its numerical representation into band views, and BGF weights each auxiliary encoder in every band with per-series gates.

Problem statement. Given a historical input $\mathbf{X} \in \mathbb{R}^{L\times C}$ of context length L with C channels, the objective is to forecast the following H steps,

$$\widehat{\mathbf{Y}}_{L+1:L+H} = f_\theta\big(\mathbf{X}, \mathbf{V}(\mathbf{X}), T(\mathbf{X})\big) \in \mathbb{R}^{H\times C}, \tag{1}$$

where $\mathbf{V}(\cdot)$ and $T(\cdot)$ are deterministic visual and textual views derived from the input alone.

3.1 FRAMEWORK OVERVIEW

We construct a visual view $\mathbf{V}$ and a textual view T from $\mathbf{X}$. The visual view comprises a Gramian Angular Field (Wang & Oates, 2015), a recurrence plot (Eckmann et al., 1987) and a line chart, while T summarises temporal trends, seasonality and notable extrema.

Text makes global statistics explicit at low frequencies and visual representations make temporal morphology salient at intermediate scales; neither retains the exact values of fast, high-frequency fluctuations, so absorbing both views in full at every band risks admitting noise. FREQGATE therefore makes fusion selective: SFP learns the frequency bands and decomposes the numerical representation into band views, and BGF determines how much of each auxiliary encoder is absorbed in every

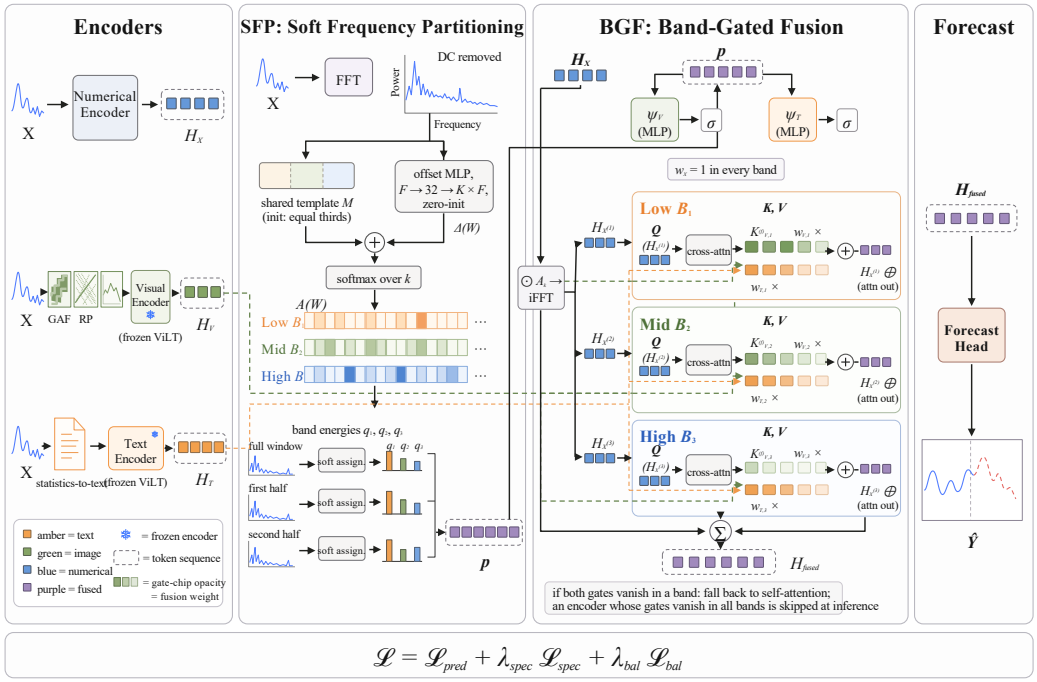


Figure 2: Overview of FREQGATE. The numerical encoder runs for every input. SFP learns a soft band decomposition of each input's spectrum, producing the multi-resolution spectrum $\mathbf{p}$ and per-series filters for the numerical band views. BGF gates the text and image encoders in every band. An encoder whose gates vanish across all bands can be skipped at inference.

band. The auxiliary views are encoded by a frozen pretrained vision-language backbone (Zhong et al., 2025); all encoder details are given in Appendix E.

3.2 SOFT FREQUENCY PARTITIONING (SFP)

To read which spectral content an input carries, SFP decomposes its power spectrum into frequency bands that are learned from data and adapted to each series, rather than fixed a priori. For a temporal window $\mathbf{W} \in \mathbb{R}^{S\times C}$, we first apply mean and variance normalisation per channel and denote the result by $\widetilde{\mathbf{W}}$. Its average non-DC spectral power is

$$E_f(\mathbf{W}) = \frac{1}{C}\sum_{c=1}^{C}\left|\mathrm{DFT}(\widetilde{\mathbf{W}}_{:,c})_f\right|^2, \qquad f = 1,\dots,\lfloor S/2\rfloor. \tag{2}$$

Excluding the DC bin makes the spectrum sensitive to temporal variation rather than absolute level or scale. Since windows of different lengths yield different numbers of bins, every spectrum is then linearly resampled onto a canonical grid of $F = \lfloor L/2\rfloor$ bins, the resolution of the full context, and E_f everywhere below refers to the resampled spectrum. A single template and offset network are thus shared by the full window and its halves, and across datasets with the same context length.

Learnable soft band decomposition. Instead of partitioning the frequency axis into fixed bands of equal width, SFP learns the partition itself and adapts it to every input. A matrix of band logits $\mathbf{M} \in \mathbb{R}^{K\times F}$, initialised to a template of equal thirds, provides a shared partition, which a lightweight offset network then refines for each window. Concretely, the normalised spectrum $\bar{E}_f(\mathbf{W}) = E_f(\mathbf{W})/\sum_{f'} E_{f'}(\mathbf{W})$ is mapped by a two-layer MLP to an offset $\Delta(\mathbf{W}) \in \mathbb{R}^{K\times F}$, and the softmax of their sum over the K bands yields a soft assignment of every frequency bin that is specific to the window:

$$A_{k,f}(\mathbf{W}) = \frac{\exp\big(M_{k,f} + \Delta_{k,f}(\mathbf{W})\big)}{\sum_{k'=1}^{K}\exp\big(M_{k',f} + \Delta_{k',f}(\mathbf{W})\big)}, \qquad q_k(\mathbf{W}) = \frac{\sum_f A_{k,f}(\mathbf{W})\,E_f(\mathbf{W})}{\sum_{f,k'} A_{k',f}(\mathbf{W})\,E_f(\mathbf{W})}. \tag{3}$$

The offset network's output layer is initialised to zero, so training starts from the shared template and the partition gradually adapts to each series. We use $K = 3$ in all experiments. Because the assignments are learned end-to-end, the bands adapt to the data's spectral semantics and may become non-contiguous, for instance grouping a fundamental frequency with its harmonics. When a window has zero non-DC energy, we set $\bar{\mathbf{E}}(\mathbf{W}) = \mathbf{0}$ and $q_k(\mathbf{W}) = 1/K$.

Multi-resolution spectrum. Concatenating the band spectra of the full window and its two halves yields the multi-resolution spectrum,

$$\mathbf{p} = \left[\, q(\mathbf{X}) \,\|\, q(\mathbf{X}_{1:\lfloor L/2 \rfloor}) \,\|\, q(\mathbf{X}_{\lfloor L/2 \rfloor + 1:L}) \,\right] \in \mathbb{R}^{3K}. \tag{4}$$

The spectra of the two halves expose how the spectral composition drifts over time within the context, so the gates can react to where in time each band's energy concentrates. The spectrum is computed before constructing the auxiliary modalities, enabling sparse inference. The same masks also provide the band filters that BGF applies to the numerical representation (§3.3), so the coordinate system that reads the series is the one the fusion operates in.

3.3 Band-Gated Fusion (BGF)

To absorb each modality only where it is informative, BGF fuses the auxiliary encoders band by band under per-series band gates. The numerical encoder produces $\mathbf{H}_X \in \mathbb{R}^{C \times N \times d}$ with one latent token per input step, $N = L$, so the token axis of $\mathbf{H}_X$ is in one-to-one correspondence with the time axis of $\mathbf{X}$ and the two share a discrete frequency axis; this is what allows a band mask read off the input spectrum to act directly on $\mathbf{H}_X$ in Eq. (5). The visual and textual encoders produce global window representations $\mathbf{H}_V, \mathbf{H}_T \in \mathbb{R}^{N \times d}$, broadcast across channels; their canonical text token or visual patch order is resampled into N ordered latent tokens, so that all three token streams share one length. Because the spectrum in Eq. (2) is averaged across channels, the band assignments $\mathbf{A}(\mathbf{X})$ and the gates $w_{m,k}$ are shared by all C channels: they are per-window, not per-channel. Downstream, the band decomposition of $\mathbf{H}_X$, the cross-attention and the forecasting head are applied to each channel independently with weights shared across channels. We suppress the channel index below.

Numerical band views. Only the numerical representation is decomposed into bands. Applying the soft assignments produced by SFP on the DFT of $\mathbf{H}_X$ along the token axis yields one view per band:

$$\mathbf{H}_X^{(k)} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{H}_X) \odot \mathbf{A}_k(\mathbf{X})), \qquad k = 1, \ldots, K, \tag{5}$$

where $\mathbf{A}_k(\mathbf{X}) \in \mathbb{R}^F$ is the k-th soft band mask of the *full* input window, $F = \lfloor L/2 \rfloor$, extended with DC assigned to the low band and conjugate-symmetrised to the $\lfloor N/2 \rfloor + 1$ non-negative token frequencies; the spectra of the two halves in Eq. (4) enter only through the gates. Since the numerical tokens retain temporal order, their band decomposition carries genuine frequency semantics; the auxiliary encoders are kept whole.

Band gates. For each auxiliary modality $m \in \{T, V\}$, a lightweight MLP maps the spectrum $\mathbf{p}$ to one gate per band,

$$w_{m,k} = \sigma\big(\psi_m(\mathbf{p})_k\big), \qquad k = 1, \ldots, K, \tag{6}$$

a per-series gate that scales how much of encoder m is absorbed in band k. The gates are trained end-to-end jointly with all other trainable parameters.

Gated band fusion. In band k, the band's numerical tokens query the unsplit auxiliary encoders under their gates:

$$\mathbf{H}_{\text{fused}}^{(k)} = \mathbf{H}_X^{(k)} + \sum_{m \in \{T,V\}} w_{m,k}\ \mathrm{Attn}\big(\mathbf{H}_X^{(k)}, \mathbf{H}_m, \mathbf{H}_m\big). \tag{7}$$

The attention operator uses query, key and value projections shared across bands. When the gates of both auxiliary encoders vanish in a band, we instead apply self-attention to the numerical tokens, preserving a purely numerical path. The fused band views are summed and passed to a linear head shared across channels to produce $\widehat{\mathbf{Y}}$.

Gate specialisation and inference sparsity. To discourage the two auxiliary encoders from crowding the same bands, we penalise the overlap of their gate distributions averaged over the batch:

$$\mathcal{L}_{\text{spec}} = \sum_{k=1}^{K} \bar{w}_T(k)\, \bar{w}_V(k), \qquad \bar{w}_m(k) = \frac{1}{|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} w_{i,m,k}. \tag{8}$$

The usage floor term $\mathcal{L}_{\text{bal}} = \sum_{m \in \{T,V\}} \big[\max\big(0,\, u_0 - \bar{w}_m\big)\big]^2$, with $\bar{w}_m$ the mean gate of encoder m over batch and bands and u_0 a small floor, prevents an encoder's gates from collapsing to zero globally (Appendix E). At inference, an encoder whose gates fall below a small threshold in every band can be skipped entirely: its view is not constructed and its encoder is not run, so the realised cost depends on the input. Gates are strictly positive at training time, so all reported results use the full forward pass; thresholded skipping is an optional optimisation at deployment (Appendix D).

3.4 Training Objective and Complexity

The training objective is

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda_{\text{spec}} \mathcal{L}_{\text{spec}} + \lambda_{\text{bal}} \mathcal{L}_{\text{bal}}, \qquad \mathcal{L}_{\text{pred}} = (HC)^{-1} \|\widehat{\mathbf{Y}} - \mathbf{Y}\|_F^2. \tag{9}$$

We select $\lambda_{\text{spec}} = 0.1$ and $\lambda_{\text{bal}} = 0.01$ on the ETTh1 validation split and reuse them on all remaining benchmarks.

SFP costs $\mathcal{O}(CL \log L)$ and adds only the $K \times F$ template logits and the lightweight offset network as parameters. BGF requires one FFT and K inverse FFTs on the numerical tokens, plus K cross-attention operations of cost $\mathcal{O}(N^2 d)$ per active auxiliary encoder.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on eight widely used multivariate forecasting benchmarks that together span diverse sampling rates, domains and spectral characteristics: ETTh1, ETTh2, ETTm1 and ETTm2 (Wu et al., 2021), Weather (Wu et al., 2021), Electricity (ECL), Traffic and Exchange (Lai et al., 2018).

Baselines. We compare against fifteen baselines organised into four families: (i) recent multimodal forecasters Time-VLM (Zhong et al., 2025), DMMV (Shen et al., 2025), Aurora (Wu et al., 2025) and VisionTS (Chen et al., 2025); (ii) LLM-augmented forecasters Time-LLM (Jin et al., 2024) and GPT4TS (Zhou et al., 2023); (iii) strong unimodal forecasters iTransformer (Liu et al., 2024c), PatchTST (Nie et al., 2023), TimesNet (Wu et al., 2023), DLinear (Zeng et al., 2023), Non-stationary Transformer (Liu et al., 2022) and ETSformer (Woo et al., 2022); and (iv) classical Transformer forecasters FEDformer (Zhou et al., 2022), Autoformer (Wu et al., 2021) and Informer (Zhou et al., 2021).

Implementation Details. Following Liu et al. (2024c), we evaluate all models under a unified protocol with input length L=96 and report MSE and MAE over four horizons $H \in \{96, 192, 336, 720\}$; the main tables average both metrics over horizons, and complete per-horizon results are provided in Appendix A. The numerical encoder of FreqGate is trained from scratch, while the visual and textual encoders are the frozen pretrained vision and text encoders of the backbone used by Time-VLM (Zhong et al., 2025) (details in Appendix E). All models are trained with Adam (learning rate 10^{-4}, batch size 256), with early stopping on validation MSE. Each result is the mean over three random seeds.

4.2 Long-term Forecasting

Setting. We first evaluate full-shot long-term forecasting, where every model is trained on the complete training split of each benchmark (Table 1).

Results. FreqGate attains the best average MSE and MAE and ranks first on six of the eight benchmarks, improving on the strongest baseline VisionTS by 2.5% in average MSE and 1.8% in

Table 1: Long-term forecasting results averaged over horizons $H \in \{96, 192, 336, 720\}$. All fifteen baselines are shown; per-horizon results appear in Table 7 (Appendix A). Red: best, blue underline: second best.

	ETTh1		ETTh2		ETTm1		ETTm2		Weather		ECL		Traffic		Exchange		**Avg**	
Model	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
FREQGATE (Ours)	0.400	0.416	0.358	0.393	0.339	0.365	0.255	0.317	0.218	0.253	0.152	0.244	0.377	0.247	0.385	0.423	0.311	0.332
VisionTS	0.409	0.417	0.353	0.387	0.349	0.377	0.275	0.335	0.227	0.259	0.159	0.249	0.387	0.254	0.393	0.422	0.319	0.338
Time-VLM	0.431	0.442	0.359	0.397	0.349	0.381	0.264	0.322	0.232	0.268	0.178	0.278	0.440	0.310	0.388	0.419	0.330	0.352
DMMV	0.526	0.477	0.603	0.526	0.361	0.381	0.297	0.357	0.231	0.295	0.179	0.281	0.496	0.341	3.679	1.353	0.797	0.501
Aurora	2.132	1.116	0.532	0.535	1.695	0.994	0.350	0.429	0.731	0.605	0.833	0.545	1.545	0.667	1.688	1.063	1.188	0.744
Time-LLM	0.432	0.440	0.359	0.399	0.368	0.390	0.266	0.325	0.246	0.281	0.177	0.276	0.439	0.310	0.411	0.437	0.337	0.357
GPT4TS	0.421	0.433	0.371	0.411	0.352	0.384	0.274	0.331	0.234	0.270	0.167	0.262	0.421	0.297	0.447	0.450	0.336	0.355
iTransformer	0.451	0.455	0.383	0.413	0.366	0.395	0.281	0.334	0.238	0.273	0.177	0.273	0.450	0.324	0.410	0.441	0.345	0.364
PatchTST	0.487	0.470	0.374	0.413	0.353	0.385	0.270	0.327	0.228	0.264	0.174	0.273	0.429	0.301	0.394	0.424	0.339	0.357
TimesNet	0.594	0.530	0.410	0.439	0.460	0.443	0.316	0.355	0.269	0.298	0.203	0.304	0.622	0.338	0.935	0.627	0.476	0.417
DLinear	0.453	0.462	0.531	0.503	0.363	0.384	0.302	0.363	0.246	0.299	0.221	0.330	0.627	0.420	0.376	0.470	0.390	0.404
Non-stat. Trans.	0.695	0.596	0.414	0.441	0.545	0.500	0.390	0.401	0.274	0.308	0.258	0.350	0.720	0.420	0.490	0.495	0.473	0.439
ETSformer	0.597	0.557	0.452	0.491	0.701	0.602	0.341	0.403	0.284	0.356	0.376	0.458	1.088	0.647	0.509	0.517	0.544	0.504
FEDformer	0.498	0.502	0.425	0.458	0.454	0.462	0.323	0.372	0.337	0.385	0.286	0.383	0.747	0.457	0.743	0.631	0.477	0.456
Autoformer	0.542	0.511	0.418	0.455	0.613	0.528	0.330	0.379	0.354	0.392	0.260	0.360	0.721	0.445	0.764	0.652	0.500	0.465
Informer	1.145	0.806	3.240	1.516	0.878	0.701	2.070	1.069	0.830	0.647	0.877	0.745	1.377	0.759	0.971	0.824	1.424	0.883

Table 2: Few-shot long-term forecasting on 5% training data. Dataset scores are averaged over the available horizons in $H \in \{96, 192, 336, 720\}$; horizons for which the 5% prefix is too short are excluded. Per-horizon results appear in Table 8 (Appendix A). Red: best, blue underline: second best.

	ETTh1		ETTh2		ETTm1		ETTm2		Weather		ECL		Traffic		Exchange		**Avg**	
Model	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
FREQGATE (Ours)	0.424	0.413	0.354	0.373	0.397	0.382	0.284	0.324	0.254	0.290	0.175	0.261	0.387	0.252	0.217	0.318	0.312	0.327
VisionTS	0.419	0.408	0.361	0.379	0.406	0.396	0.316	0.346	0.276	0.305	0.180	0.268	0.407	0.266	0.214	0.316	0.322	0.336
Time-VLM	0.749	0.567	0.472	0.470	0.498	0.471	0.296	0.339	0.265	0.303	0.324	0.406	0.799	0.536	1.514	0.844	0.615	0.492
DMMV	1.219	0.826	3.298	1.385	1.257	0.811	2.343	1.164	0.404	0.421	0.436	0.443	0.885	0.538	5.809	2.092	1.956	0.960
Aurora	1.361	0.859	0.528	0.531	1.587	0.903	0.394	0.448	0.740	0.617	0.831	0.565	1.301	0.611	0.561	0.614	0.913	0.644
Time-LLM	1.180	0.742	0.430	0.443	0.722	0.552	0.387	0.402	0.283	0.310	0.221	0.317	0.430	0.299	0.233	0.357	0.486	0.428
GPT4TS	0.744	0.563	0.427	0.441	0.502	0.467	0.310	0.352	0.270	0.306	0.182	0.279	0.445	0.316	0.407	0.475	0.411	0.400
iTransformer	0.746	0.587	0.424	0.444	0.551	0.502	0.312	0.359	0.290	0.326	0.337	0.409	0.765	0.513	0.360	0.445	0.473	0.448
PatchTST	0.774	0.588	0.470	0.471	0.502	0.471	0.325	0.361	0.279	0.315	0.240	0.336	0.492	0.371	0.371	0.453	0.432	0.421
TimesNet	0.967	0.654	0.602	0.529	0.822	0.621	0.405	0.418	0.345	0.364	0.494	0.505	1.035	0.604	0.645	0.589	0.664	0.536
DLinear	0.687	0.572	0.405	0.452	0.602	0.521	0.329	0.381	0.295	0.352	0.616	0.634	1.210	0.714	0.429	0.500	0.572	0.516
Non-stat. Trans.	0.911	0.638	0.504	0.491	0.832	0.599	0.382	0.410	0.328	0.357	0.921	0.773	1.604	0.874	0.669	0.598	0.769	0.593
ETSformer	1.573	0.938	0.772	0.661	1.417	0.870	0.486	0.514	0.505	0.514	1.510	0.969	2.496	1.135	0.486	0.536	1.156	0.767
FEDformer	0.944	0.697	0.647	0.612	0.778	0.608	0.424	0.457	0.407	0.448	1.021	0.806	1.690	0.919	0.733	0.675	0.831	0.653
Autoformer	0.817	0.648	0.484	0.515	0.739	0.586	0.389	0.432	0.351	0.390	0.598	0.600	1.325	0.763	0.616	0.616	0.665	0.569
Informer	1.481	0.934	3.066	1.367	1.673	0.929	3.784	1.544	0.553	0.528	1.274	0.920	1.683	0.883	3.383	1.497	2.112	1.075

average MAE. The two exceptions are informative. ETTh2 is dominated by low-frequency structure, where fusing the auxiliary views injects little noise, so indiscriminate fusion already helps; VisionTS wins by only 1.4%. Exchange lacks stable periodic structure, so the auxiliary views carry little predictive content and the purely numerical DLinear attains the best MSE; with its gates almost closed (Table 5), FREQGATE falls back to a near-numerical forecaster and is a close second on MSE. The benefit of auxiliary modalities thus correlates with the input's spectral characteristics rather than being uniform across datasets. Classical forecasters and the ungated multimodal baselines DMMV and Aurora trail the leading group by a wide margin, suggesting that multimodal augmentation alone is insufficient without selective gating.

4.3 FEW-SHOT FORECASTING

Setting. We evaluate few-shot long-term forecasting with only 5% or 10% of the training data. The few-shot split is a *contiguous chronological prefix* of the training split: windows sampled uniformly at random would leak the full temporal extent through adjacent, overlapping windows. Validation and test splits are unchanged, all baselines are re-run under this identical protocol, and averages are taken over the available horizons.

Table 3: Zero-shot transfer results averaged over horizons $H \in \{96, 192, 336, 720\}$ per source→target pair. h1, h2, m1 and m2 denote ETTh1, ETTh2, ETTm1 and ETTm2. All baselines supporting this protocol are shown; per-horizon results appear in Table 10 (Appendix A). Red: best, blue underline: second best.

	h1→h2		h2→h1		h1→m2		m1→m2		m2→m1		**Avg**	
Model	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
FREQGATE (Ours)	0.333	0.377	0.410	0.400	0.291	0.329	0.258	0.313	0.422	0.400	0.343	0.364
VisionTS	0.372	0.393	0.427	0.420	0.314	0.338	0.314	0.338	0.415	0.393	0.368	0.376
Time-VLM	0.347	0.394	0.599	0.548	0.306	0.359	0.267	0.322	0.540	0.490	0.412	0.423
Aurora	0.439	0.508	1.286	0.873	0.316	0.404	0.316	0.404	1.345	0.890	0.740	0.616
Time-LLM	0.350	0.395	0.535	0.505	0.303	0.359	0.268	0.323	0.449	0.434	0.381	0.403
GPT4TS	0.375	0.411	0.551	0.519	0.332	0.378	0.269	0.326	0.488	0.454	0.403	0.418
iTransformer	0.363	0.403	0.711	0.597	0.316	0.369	0.278	0.332	0.582	0.516	0.450	0.443
PatchTST	0.360	0.401	0.510	0.493	0.310	0.363	0.267	0.323	0.462	0.444	0.382	0.405
TimesNet	0.456	0.464	1.275	0.713	0.447	0.432	0.354	0.382	1.132	0.666	0.733	0.531
DLinear	0.466	0.467	0.477	0.484	0.409	0.442	0.276	0.341	0.411	0.425	0.408	0.432
Non-stat. Trans.	0.519	0.502	0.929	0.662	0.464	0.454	0.397	0.414	1.780	0.866	0.818	0.580
ETSformer	0.495	0.515	1.068	0.745	0.417	0.472	0.389	0.453	1.012	0.697	0.676	0.576
FEDformer	0.463	0.496	0.779	0.627	0.379	0.431	0.434	0.465	0.758	0.593	0.563	0.522
Autoformer	0.609	0.573	0.829	0.651	0.444	0.484	0.449	0.476	1.187	0.672	0.704	0.571
Informer	2.329	1.191	1.077	0.768	1.550	0.936	2.061	1.196	0.839	0.636	1.571	0.945

Table 4: Ablation study: MSE averaged over four horizons and three seeds; Δ is relative to the full model (first row).

Variant	**ETTh1**	**ETTh2**	**ETTm1**	**ETTm2**	**Weather**	**ECL**	**Traffic**	**Exchange**	**Avg**	Δ
Full FREQGATE	0.400	0.358	0.339	0.255	0.218	0.152	0.377	0.385	0.311	—
w/o SFP	0.411	0.369	0.354	0.268	0.229	0.169	0.445	0.409	0.332	+6.8%
w/o $\mathcal{L}_{\text{spec}}$	0.417	0.374	0.356	0.269	0.231	0.164	0.412	0.406	0.329	+5.8%
w/o BGF	0.432	0.384	0.361	0.273	0.235	0.159	0.398	0.400	0.330	+6.1%
w/o Multi-res. spectrum	0.408	0.367	0.359	0.271	0.236	0.159	0.385	0.399	0.323	+3.9%
Fixed gates	0.423	0.380	0.365	0.276	0.240	0.166	0.415	0.418	0.335	+7.7%

Results. As shown in Table 2 (5%) and Table 6 (10%, Appendix A), FREQGATE attains the best average MSE and MAE at both budgets, ranking first on six of the eight datasets in each. Its advantage is most pronounced under data scarcity: from full-shot to 5%, its average MSE changes by only 0.001, from 0.311 to 0.312, whereas PatchTST, iTransformer and Time-VLM degrade by 27.4%, 37.1% and 86.4%; at 10% it still matches its full-shot average and improves on VisionTS by 3.4% in MSE. These gains stem from spectral gating rather than from pretrained knowledge: Time-VLM, using the same frozen backbone, degrades by 86.4% under the same protocol.

4.4 ZERO-SHOT FORECASTING

Setting. We evaluate zero-shot transfer: the model is trained on a source dataset and evaluated on an unseen target without any target-domain gradient step. Following Jin et al. (2024), we use five source → target pairs: ETTh1 → ETTh2, ETTh2 → ETTh1, ETTh1 → ETTm2, ETTm1 → ETTm2 and ETTm2 → ETTm1. No prediction head is adapted on target-domain data. DMMV does not admit cross-dataset transfer and is therefore excluded from this protocol.

Results. Table 3 summarises the results for each transfer pair. FREQGATE ranks first on four of the five pairs by both MSE and MAE, and its average MSE improves on VisionTS by 6.8% and on Time-LLM by 10.0%. The only pair FREQGATE does not win is ETTm2 → ETTm1, where DLinear attains the best MSE and VisionTS the best MAE, with FREQGATE within 2.7% of the best on both metrics.

4.5 ABLATION STUDIES

We ablate five variants of FREQGATE on all eight datasets, each removing or replacing a single component (Table 4); two observations emerge.

Table 5: Per-band gates $w_{m,k}$ of the two auxiliary encoders, averaged over the test series of each dataset; $\bar{w}_m$ averages over the $K{=}3$ bands.

	Text encoder				Image encoder			
Dataset	$w_{T,\text{low}}$	$w_{T,\text{mid}}$	$w_{T,\text{high}}$	$\bar{w}_T$	$w_{V,\text{low}}$	$w_{V,\text{mid}}$	$w_{V,\text{high}}$	$\bar{w}_V$
ETTh1 (low)	0.92	0.88	0.72	0.84	0.70	0.82	0.61	0.71
ETTh2 (low)	0.89	0.85	0.69	0.81	0.66	0.78	0.60	0.68
ETTm1 (low-mid)	0.78	0.62	0.46	0.62	0.62	0.76	0.60	0.66
ETTm2 (low-mid)	0.74	0.58	0.42	0.58	0.58	0.74	0.57	0.63
Weather (mid)	0.48	0.28	0.17	0.31	0.68	0.84	0.64	0.72
ECL (mid-high)	0.38	0.19	0.09	0.22	0.34	0.48	0.23	0.35
Traffic (high)	0.07	0.03	0.02	0.04	0.15	0.21	0.09	0.15
Exchange (weak per.)	0.16	0.10	0.07	0.11	0.14	0.16	0.09	0.13

Gating and specialisation. Removing SFP degrades performance on every dataset, with the increase reaching 18.0% on Traffic and 11.2% on ECL. Replacing the learned gates with a fixed frequency-to-modality rule is worse still, with a 7.7% increase in average MSE, so the learned, series-dependent gates matter. Removing $\mathcal{L}_{\text{spec}}$ increases average MSE by 5.8%, supporting distinct spectral responses from the two auxiliary encoders.

Band-conditioned fusion. Removing BGF increases average MSE by 6.1%, with the increase reaching 8.0% on ETTh1 and 7.8% on Weather; this ablation removes the full band-conditioned mechanism rather than cross-modal attention alone. The multi-resolution spectrum also matters: replacing it with a single-resolution FFT increases average MSE by 3.9% (K is swept in Appendix C).

4.6 Band-Gating Analysis

Frequency-dependent modality utility. Section 1 hypothesised that the predictive content of the auxiliary views concentrates at low and mid frequencies. A controlled branch-removal ablation on a single tri-modal architecture (Time-VLM (Zhong et al., 2025)) supports this premise: auxiliary branches help or are neutral on the low- and mid-frequency datasets, whereas the numerical-only variant is best on ECL, Traffic and Exchange (Table 11, Appendix B), consistent with Table 1 and the w/o SFP ablation in Section 4.5.

Gate behaviour. Table 5 reports the per-band gates $w_{m,k}$ of each auxiliary encoder. Within every dataset, the text gate decays monotonically from the low to the high band—e.g., $0.92 \to 0.88 \to 0.72$ on ETTh1 and $0.07 \to 0.03 \to 0.02$ on Traffic—while the image gate peaks at the mid band, precisely the spectral ranges that each view's construction can express. The overall level scales with the dataset's spectral character: $\bar{w}_T$ falls from 0.84 on low-frequency ETTh1 to 0.04 on high-frequency Traffic, and $\bar{w}_V$ from 0.63–0.72 on the low- and mid-frequency datasets to 0.13–0.15 on Traffic and Exchange. Since the gates are conditioned only on the spectrum $\mathbf{p}$, this modality–band relationship is learned rather than hand-coded: the gates recover the spectral-locality principle of Section 1 from data. Gates are soft weights, not binary activation rates.

5 Conclusion

In this paper, we articulate a principle of *spectral locality* for multimodal forecasting: the predictive content of an auxiliary view concentrates in the frequency range that its construction can express. Guided by this principle, we present FreqGate, a frequency-gated multimodal forecaster that absorbs the auxiliary modalities selectively, band by band, rather than fusing both auxiliary views indiscriminately: SFP learns a soft, per-series partition of the input spectrum, and BGF weights each auxiliary encoder in every band. Across eight benchmarks and fifteen baselines, FreqGate achieves the best average performance under full-shot, few-shot and zero-shot regimes, and with only 5% of the training data its average MSE is virtually unchanged from full-shot. The learned band gates conform to this principle without hand-coded rules.

Limitations and future work. Our evaluation is limited to auxiliary views derived deterministically from the input series itself; whether the spectrum is the right gating criterion for exogenous signals remains open. Future work will validate the approach on exogenous multimodal signals to verify its generalisation.

AI USE STATEMENT

In this work, we used generative AI tools for language editing of author-written prose to improve grammar, concision and readability, and for minor LaTeX formatting assistance such as table and figure layout. We have not used generative AI tools for research ideation, for the design of FREQGATE, SFP or BGF, for the design or execution of any experiment, or for the generation, selection or interpretation of any reported result; every number in this paper comes from runs executed by the authors. Generative AI was not used to retrieve or summarise related work, and all cited references were read and verified by the authors against their original sources. We have reviewed all AI-assisted work: every AI-edited passage was checked line by line against the authors' intended meaning, and the authors take full responsibility for the final content of this paper, including its text, claims, figures and artefacts.

ETHICS STATEMENT

This work studies long-horizon forecasting on eight public multivariate benchmarks (ETT, Weather, Electricity, Traffic and Exchange) that are distributed under standard research licenses and are already widely used in the forecasting literature. We collect no new data, and none of the benchmarks contains personally identifiable information or human-subject data; the Traffic and Electricity series are aggregate sensor and client-level load measurements. Both auxiliary modalities are derived deterministically from the numerical input itself, namely a statistics-to-text template and fixed visual representations, so no external corpus, annotator or scraped content enters the pipeline, and the frozen pretrained vision-language encoder is never fine-tuned on third-party data. We foresee no direct path to harm beyond the risks generic to forecasting: as with any forecaster, deploying FREQGATE in energy dispatch, traffic control or financial settings without calibrated uncertainty could support over-confident decisions, and we therefore report accuracy on public benchmarks rather than claim deployment readiness. The authors declare no conflicts of interest or sponsorship that could bias the reported results.

REPRODUCIBILITY STATEMENT

We have taken the following steps to make this work reproducible. The method is specified completely in Section 3: the spectrum in Eqs. (2)–(4), the band gates in Eq. (6), the band-gated fusion in Eqs. (5)–(7), and the training objective in Eqs. (8) and (9). Section 4.1 states the evaluation protocol (input length, prediction horizons, metrics, the three training regimes, the optimiser settings and the number of seeds), and Appendix E lists every hyper-parameter, including the values tuned on the ETTh1 validation split and reused unchanged on the remaining seven datasets. Appendix A reports the complete per-horizon results for every model and regime, and Appendix D specifies the protocol for future conditional-compute measurements. All datasets are publicly available and we use the canonical train/validation/test splits of prior work (Appendix E). Our implementation, including the data-preprocessing scripts and the configuration files for every reported run, is included as anonymised source code in the supplementary material.

REFERENCES

Defu Cao, Furong Jia, Sercan Ö. Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. TEMPO: Prompt-based generative pre-trained transformer for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024.

Mouxiang Chen, Lefei Shen, Zhuo Li, Xiaoyun Joy Wang, Jianling Sun, and Chenghao Liu. VisionTS: Visual masked autoencoders are free-lunch zero-shot time series forecasters. In *International Conference on Machine Learning (ICML)*, 2025.

Dawei Cheng, Fangzhou Yang, Sheng Xiang, and Jin Liu. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition*, 121:108218, 2022.

Abdul Monaf Chowdhury, Rabeya Akter, and Safaeid Hossain Arib. T3Time: Tri-modal time series forecasting via adaptive multi-head alignment and residual fusion. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.

J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle. Recurrence plots of dynamical systems. *Europhysics Letters*, 4(9):973–977, 1987.

Rentao Gu, Yihang Ding, Junjie Li, Yi Ding, Weijing Sang, Xiaoli Huo, Xin Qin, and Yuefeng Ji. FM-LLM: A frequency-enhanced mixture-of-experts framework for adapting LLMs to time series forecasting. *Knowledge-Based Systems*, 341:115776, 2026.

Yaohui Huang, Runmin Zou, Yun Wang, Laeeq Aslam, and Ruipeng Dong. M^2FMoE: Multi-resolution multi-view frequency mixture-of-experts for extreme-adaptive time series forecasting. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.

R. Jalalifar, M. R. Delavar, and S. F. Ghaderi. SAC-ConvLSTM: A novel spatio-temporal deep learning-based approach for a short term power load forecasting. *Expert Systems with Applications*, 237:121487, 2024.

Shuning Jia, Baijun Song, Canming Ye, and Chun Yuan. M3Time: LLM-enhanced multi-modal, multi-scale, and multi-frequency multivariate time series forecasting. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2026.

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 (ICLR)*, 2024.

Shruti Kaushik, Abhinav Choudhury, Pankaj Kumar Sheron, Nataraj Dasgupta, Sayee Natarajan, Larry A. Pickett, and Varun Dutt. AI in healthcare: Time-series forecasting using statistical, neural, and ensemble architectures. *Frontiers in Big Data*, 3:4, 2020.

Wonjae Kim, Bokyung Son, and Ildoo Kim. ViLT: Vision-and-language transformer without convolution or region supervision. In *International Conference on Machine Learning (ICML)*, 2021.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pp. 95–104, 2018.

Chenxi Liu, Sun Yang, Qianxiong Xu, Z. Li, Cheng Long, Ziyue Li, and Rui Zhao. Spatial-temporal large language model for traffic prediction. In *IEEE International Conference on Mobile Data Management (MDM)*, pp. 31–40, 2024a.

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 *AAAI Conference on Artificial Intelligence (AAAI)*, 2025.

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 *The Web Conference (WWW)*, pp. 4095–4106, 2024b.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

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

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

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

ChengAo Shen, Wenchao Yu, Ziming Zhao, Dongjin Song, Wei Cheng, Haifeng Chen, and Jingchao Ni. Multi-modal view enhanced large vision models for long-term time series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Chengsen Wang, Qi Qi, Jingyu Wang, Haifeng Sun, Zirui Zhuang, Jinming Wu, Lei Zhang, and Jianxin Liao. ChatTime: A unified multimodal time series foundation model bridging numerical and textual data. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 12694–12702, 2025.

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

Zhiguang Wang and Tim Oates. Imaging time-series to improve classification and imputation. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3939–3945, 2015.

Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.

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

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 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 (ICLR)*, 2023.

Xingjian Wu, Jianxin Jin, Wanghui Qiu, Peng Chen, Yang Shu, Bin Yang, and Chenjuan Guo. Aurora: Towards universal generative multimodal time series forecasting. *arXiv preprint arXiv:2509.22295*, 2025.

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

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

Siru Zhong, Weilin Ruan, Ming Jin, Huan Li, Qingsong Wen, and Yuxuan Liang. Time-VLM: Exploring multimodal vision-language models for augmented time series forecasting. In *International Conference on Machine Learning (ICML)*, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *AAAI Conference on Artificial Intelligence (AAAI)*, 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 (ICML)*, 2022.

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 (NeurIPS)*, 2023.

Table 6: Few-shot long-term forecasting on 10% training data. Results are averaged over the available forecasting horizons in $H \in \{96, 192, 336, 720\}$ ("-" means that 10% of the time series is not sufficient to constitute a training set for that horizon). A lower value indicates better performance. Full per-horizon results are provided in Table 9. Red: best, blue underline: second best.

	ETTh1		ETTh2		ETTm1		ETTm2		Weather		ECL		Traffic		Exchange		**Avg**	
Model	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**MAE**
FREQGATE (Ours)	0.435	0.435	0.353	0.385	0.377	0.378	0.269	0.320	0.232	0.269	0.174	0.263	0.421	0.283	0.230	0.339	0.311	0.334
VisionTS	0.423	0.419	0.369	0.393	0.386	0.387	0.304	0.337	0.269	0.290	0.179	0.265	0.426	0.284	0.222	0.331	0.322	0.338
Time-VLM	0.669	0.538	0.389	0.422	0.451	0.443	0.278	0.328	0.237	0.275	0.244	0.339	0.649	0.441	0.471	0.491	0.424	0.410
DMMV	1.148	0.801	3.124	1.355	0.932	0.703	1.327	0.879	0.328	0.361	0.253	0.336	0.706	0.461	6.371	2.197	1.774	0.887
Aurora	1.401	0.925	0.474	0.521	1.620	0.937	0.361	0.432	0.773	0.631	0.790	0.534	1.371	0.620	0.859	0.788	0.956	0.674
Time-LLM	0.896	0.654	0.459	0.462	0.678	0.519	0.353	0.374	0.245	0.283	0.198	0.293	0.439	0.303	0.319	0.408	0.448	0.412
GPT4TS	0.615	0.518	0.387	0.423	0.556	0.470	0.291	0.337	0.244	0.282	0.184	0.279	0.459	0.331	0.309	0.411	0.381	0.381
iTransformer	0.700	0.576	0.422	0.442	0.528	0.491	0.292	0.346	0.268	0.304	0.238	0.333	0.662	0.458	0.347	0.438	0.432	0.424
PatchTST	0.677	0.563	0.417	0.438	0.444	0.436	0.299	0.341	0.247	0.285	0.203	0.302	0.458	0.339	0.350	0.440	0.387	0.393
TimesNet	1.183	0.713	0.534	0.503	0.876	0.637	0.403	0.416	0.334	0.343	0.337	0.406	0.894	0.520	0.898	0.642	0.682	0.523
DLinear	0.637	0.555	0.402	0.443	0.530	0.483	0.303	0.358	0.264	0.316	0.419	0.505	1.044	0.630	0.472	0.529	0.509	0.477
Non-stat. Trans.	0.908	0.640	0.454	0.471	0.848	0.607	0.368	0.400	0.303	0.333	0.664	0.629	1.451	0.819	0.676	0.595	0.709	0.562
ETSformer	1.523	0.922	0.698	0.626	1.112	0.754	0.420	0.472	0.389	0.437	1.185	0.857	2.216	1.051	0.520	0.556	1.008	0.709
FEDformer	0.872	0.665	0.547	0.553	0.719	0.572	0.361	0.405	0.347	0.396	0.762	0.678	1.539	0.856	0.827	0.706	0.747	0.604
Autoformer	0.757	0.622	0.470	0.504	0.768	0.603	0.378	0.420	0.352	0.378	0.433	0.498	1.130	0.667	0.639	0.618	0.616	0.539
Informer	1.501	0.913	3.297	1.419	1.695	0.955	3.787	1.534	0.496	0.496	1.195	0.890	1.608	0.848	3.444	1.493	2.128	1.069

Table 7: Full long-term forecasting results with forecasting horizons $H \in \{96, 192, 336, 720\}$. We use the same protocol as in Table 1. A lower value indicates better performance. Red: best, blue underline: second best.

Dataset	H	FREQGATE (Ours)		VisionTS		Time-VLM		GPT4TS		Time-LLM		PatchTST		iTransformer		DLinear		Non-stat. Trans.		TimesNet		FEDformer		Autoformer		ETSformer		DMMV		Aurora		Informer	
	96	0.353	0.376	0.358	0.378	0.374	0.400	0.375	0.398	0.394	0.413	0.568	0.496	0.404	0.419	0.417	0.431	0.637	0.561	0.466	0.467	0.450	0.474	0.580	0.519	0.565	0.533	0.669	0.522	2.134	1.121	0.925	0.698
	192	0.413	0.414	0.397	0.405	0.408	0.421	0.420	0.424	0.429	0.432	0.429	0.433	0.448	0.447	0.444	0.448	0.627	0.564	0.535	0.507	0.494	0.498	0.505	0.490	0.600	0.552	0.440	0.432	2.320	1.149	1.045	0.758
ETTh1	336	0.411	0.412	0.421	0.423	0.434	0.440	0.448	0.446	0.447	0.445	0.457	0.454	0.458	0.457	0.460	0.462	0.682	0.571	0.541	0.517	0.500	0.499	0.505	0.491	0.616	0.565	0.478	0.456	1.985	1.054	1.180	0.822
	720	0.421	0.460	0.460	0.462	0.508	0.505	0.441	0.465	0.459	0.472	0.496	0.495	0.494	0.499	0.490	0.507	0.835	0.690	0.835	0.627	0.548	0.536	0.576	0.543	0.606	0.578	0.516	0.500	2.090	1.141	1.432	0.948
	Avg	0.400	0.416	0.409	0.417	0.431	0.442	0.421	0.433	0.432	0.440	0.487	0.470	0.451	0.455	0.453	0.462	0.695	0.596	0.594	0.530	0.498	0.502	0.542	0.511	0.597	0.557	0.526	0.477	2.132	1.116	1.145	0.806
	96	0.284	0.344	0.289	0.337	0.287	0.346	0.306	0.362	0.293	0.351	0.305	0.363	0.308	0.363	0.349	0.403	0.373	0.409	0.359	0.404	0.384	0.429	0.391	0.436	0.365	0.436	0.362	0.403	0.469	0.495	2.399	1.265
	192	0.344	0.379	0.349	0.381	0.360	0.389	0.364	0.401	0.354	0.390	0.366	0.403	0.367	0.400	0.436	0.456	0.433	0.446	0.405	0.436	0.411	0.449	0.409	0.446	0.432	0.481	0.479	0.474	0.524	0.526	3.528	1.585
ETTh2	336	0.385	0.415	0.371	0.400	0.380	0.409	0.382	0.420	0.378	0.413	0.396	0.428	0.424	0.437	0.527	0.511	0.413	0.444	0.424	0.448	0.418	0.458	0.404	0.446	0.483	0.513	0.557	0.521	0.556	0.550	3.369	1.541
	720	0.418	0.435	0.403	0.431	0.409	0.443	0.430	0.460	0.411	0.442	0.428	0.458	0.431	0.452	0.814	0.643	0.438	0.465	0.453	0.470	0.486	0.497	0.470	0.492	0.527	0.535	1.012	0.707	0.578	0.567	3.663	1.672
	Avg	0.358	0.393	0.353	0.387	0.359	0.397	0.371	0.411	0.359	0.399	0.374	0.413	0.383	0.413	0.531	0.503	0.414	0.441	0.410	0.439	0.425	0.458	0.418	0.455	0.452	0.491	0.603	0.526	0.532	0.535	3.240	1.516
	96	0.287	0.334	0.297	0.345	0.288	0.343	0.295	0.350	0.317	0.362	0.291	0.345	0.305	0.358	0.310	0.353	0.476	0.465	0.340	0.381	0.407	0.439	0.636	0.534	0.582	0.537	0.296	0.342	1.606	0.987	0.750	0.629
	192	0.323	0.343	0.327	0.364	0.329	0.368	0.331	0.371	0.346	0.377	0.329	0.371	0.342	0.380	0.342	0.371	0.539	0.500	0.423	0.420	0.453	0.459	0.666	0.538	0.649	0.573	0.336	0.368	1.652	0.989	0.877	0.701
ETTm1	336	0.343	0.377	0.358	0.384	0.360	0.391	0.366	0.392	0.380	0.396	0.366	0.393	0.378	0.403	0.374	0.390	0.600	0.522	0.435	0.435	0.460	0.466	0.527	0.507	0.725	0.627	0.375	0.390	1.764	1.014	0.941	0.735
	720	0.401	0.407	0.412	0.416	0.418	0.425	0.418	0.423	0.430	0.424	0.425	0.432	0.438	0.437	0.427	0.423	0.563	0.516	0.643	0.536	0.498	0.486	0.622	0.533	0.849	0.671	0.435	0.424	1.757	0.987	0.945	0.740
	Avg	0.339	0.365	0.349	0.377	0.349	0.381	0.352	0.384	0.368	0.390	0.353	0.385	0.366	0.395	0.363	0.384	0.545	0.500	0.460	0.443	0.454	0.462	0.613	0.528	0.701	0.602	0.361	0.381	1.695	0.994	0.878	0.701
	96	0.171	0.253	0.178	0.269	0.173	0.260	0.173	0.261	0.179	0.268	0.173	0.262	0.179	0.268	0.180	0.278	0.223	0.310	0.201	0.283	0.257	0.332	0.269	0.343	0.210	0.315	0.213	0.308	0.262	0.379	0.666	0.642
	192	0.225	0.301	0.244	0.313	0.235	0.303	0.240	0.309	0.231	0.303	0.236	0.306	0.252	0.315	0.246	0.329	0.312	0.366	0.266	0.325	0.290	0.352	0.302	0.365	0.284	0.368	0.281	0.353	0.315	0.410	1.100	0.848
ETTm2	336	0.263	0.334	0.292	0.347	0.282	0.336	0.290	0.344	0.280	0.336	0.298	0.346	0.306	0.349	0.319	0.380	0.438	0.428	0.332	0.369	0.333	0.379	0.331	0.379	0.364	0.422	0.304	0.358	0.367	0.438	1.650	1.019
	720	0.359	0.380	0.388	0.413	0.367	0.388	0.394	0.408	0.373	0.392	0.371	0.394	0.387	0.401	0.462	0.467	0.587	0.500	0.467	0.442	0.413	0.425	0.418	0.431	0.508	0.506	0.390	0.408	0.456	0.490	4.863	1.767
	Avg	0.255	0.317	0.275	0.335	0.264	0.322	0.274	0.331	0.266	0.325	0.270	0.327	0.281	0.334	0.302	0.363	0.390	0.401	0.316	0.355	0.323	0.372	0.330	0.379	0.341	0.403	0.297	0.357	0.350	0.429	2.070	1.069
	96	0.149	0.194	0.146	0.191	0.156	0.206	0.154	0.203	0.176	0.228	0.150	0.199	0.163	0.213	0.179	0.244	0.172	0.234	0.178	0.230	0.279	0.346	0.302	0.365	0.196	0.285	0.156	0.230	0.264	0.376	0.531	0.512
	192	0.184	0.227	0.191	0.236	0.199	0.246	0.197	0.245	0.214	0.260	0.194	0.241	0.205	0.251	0.218	0.277	0.234	0.287	0.242	0.283	0.314	0.372	0.339	0.386	0.249	0.332	0.196	0.269	0.520	0.527	0.454	0.478
Weather	336	0.239	0.270	0.247	0.279	0.249	0.285	0.252	0.287	0.264	0.295	0.246	0.282	0.255	0.290	0.263	0.313	0.312	0.337	0.301	0.320	0.358	0.403	0.355	0.390	0.311	0.380	0.250	0.313	0.848	0.664	0.836	0.676
	720	0.301	0.319	0.324	0.330	0.322	0.336	0.334	0.344	0.331	0.341	0.321	0.335	0.330	0.340	0.325	0.364	0.376	0.376	0.357	0.357	0.398	0.419	0.418	0.426	0.380	0.426	0.324	0.369	1.291	0.854	1.498	0.921
	Avg	0.218	0.253	0.227	0.259	0.232	0.268	0.234	0.270	0.246	0.281	0.228	0.264	0.238	0.273	0.246	0.299	0.274	0.308	0.269	0.298	0.337	0.385	0.354	0.392	0.284	0.356	0.231	0.295	0.731	0.605	0.830	0.647
	96	0.122	0.214	0.127	0.218	0.149	0.253	0.138	0.236	0.147	0.249	0.144	0.247	0.142	0.240	0.197	0.309	0.238	0.335	0.178	0.284	0.267	0.369	0.228	0.337	0.339	0.433	0.152	0.251	0.502	0.402	0.735	0.647
	192	0.140	0.235	0.148	0.238	0.164	0.266	0.153	0.249	0.162	0.262	0.159	0.261	0.161	0.259	0.210	0.320	0.252	0.345	0.191	0.294	0.278	0.377	0.262	0.361	0.363	0.450	0.162	0.263	0.607	0.471	0.915	0.773
ECL	336	0.157	0.249	0.163	0.253	0.181	0.282	0.169	0.266	0.180	0.281	0.177	0.277	0.181	0.279	0.223	0.332	0.264	0.355	0.217	0.315	0.285	0.383	0.248	0.351	0.385	0.464	0.185	0.287	0.764	0.486	0.922	0.775
	720	0.188	0.279	0.199	0.287	0.218	0.311	0.206	0.298	0.219	0.313	0.215	0.308	0.224	0.315	0.255	0.359	0.279	0.366	0.227	0.323	0.315	0.404	0.303	0.393	0.417	0.483	0.218	0.321	1.461	0.819	0.936	0.785
	Avg	0.152	0.244	0.159	0.249	0.178	0.278	0.167	0.262	0.177	0.276	0.174	0.273	0.177	0.273	0.221	0.330	0.258	0.350	0.203	0.304	0.286	0.383	0.260	0.360	0.376	0.458	0.179	0.281	0.833	0.545	0.877	0.745
	96	0.347	0.227	0.351	0.233	0.417	0.300	0.393	0.285	0.414	0.297	0.404	0.289	0.411	0.298	0.601	0.410	0.690	0.414	0.597	0.328	0.728	0.452	0.726	0.452	1.046	0.630	0.477	0.334	1.532	0.670	1.226	0.680
	192	0.367	0.243	0.377	0.247	0.431	0.306	0.414	0.292	0.427	0.303	0.419	0.295	0.434	0.314	0.617	0.417	0.726	0.425	0.611	0.334	0.739	0.455	0.701	0.433	1.077	0.644	0.480	0.330	1.518	0.658	1.387	0.753
Traffic	336	0.363	0.248	0.389	0.253	0.445	0.312	0.425	0.297	0.441	0.310	0.433	0.302	0.455	0.328	0.630	0.421	0.713	0.412	0.627	0.342	0.742	0.453	0.701	0.433	1.099	0.651	0.505	0.342	1.528	0.663	1.422	0.792
	720	0.430	0.268	0.432	0.280	0.468	0.323	0.451	0.312	0.475	0.330	0.460	0.317	0.499	0.356	0.659	0.434	0.751	0.429	0.655	0.347	0.781	0.466	0.755	0.462	1.129	0.662	0.523	0.359	1.600	0.675	1.473	0.810
	Avg	0.377	0.247	0.387	0.254	0.440	0.310	0.421	0.297	0.439	0.310	0.429	0.301	0.450	0.324	0.627	0.420	0.720	0.420	0.622	0.338	0.747	0.457	0.721	0.445	1.088	0.647	0.496	0.341	1.545	0.667	1.377	0.759
	96	0.106	0.212	0.090	0.210	0.096	0.217	0.105	0.232	0.103	0.232	0.091	0.215	0.099	0.227	0.157	0.312	0.153	0.287	0.191	0.319	0.359	0.447	0.381	0.465	0.150	0.294	0.177	0.307	0.683	0.657	0.731	0.711
	192	0.218	0.333	0.196	0.317	0.190	0.311	0.226	0.344	0.205	0.327	0.194	0.317	0.216	0.338	0.256	0.397	0.275	0.390	0.400	0.470	0.476	0.515	0.519	0.548	0.311	0.427	0.320	0.417	1.506	1.047	0.973	0.828
Exchange	336	0.414	0.452	0.390	0.456	0.357	0.434	0.411	0.472	0.375	0.449	0.384	0.452	0.400	0.470	0.424	0.508	0.461	0.511	0.665	0.608	0.693	0.635	0.718	0.659	0.520	0.553	5.639	2.093	2.343	1.395	1.076	0.879
	720	0.802	0.695	0.898	0.703	0.910	0.714	1.046	0.750	0.960	0.739	0.907	0.712	0.923	0.727	0.669	0.664	1.070	0.794	2.483	1.112	1.444	0.927	1.438	0.936	1.055	0.794	8.580	2.594	2.220	1.153	1.105	0.876
	Avg	0.385	0.423	0.393	0.422	0.388	0.419	0.447	0.450	0.411	0.437	0.394	0.424	0.410	0.441	0.376	0.470	0.490	0.495	0.935	0.627	0.743	0.631	0.764	0.652	0.509	0.517	3.679	1.353	1.688	1.063	0.971	0.824

A COMPLETE FORECASTING RESULTS

Every MSE and MAE value reported in the main text is an unweighted mean over the available forecasting horizons. This section provides the complete per-horizon results for all evaluated models: full-shot forecasting (Table 7), few-shot forecasting with 5% and 10% training data (Tables 8 and 9) and zero-shot transfer (Table 10).

The horizon-averaged summary under 10% training data is given in Table 6.

Table 8: Full few-shot learning results on 5% training data with forecasting horizons $H \in \{96, 192, 336, 720\}$. A lower value indicates better performance. "-" means that 5% of the time series is not sufficient to constitute a training set for that horizon; dataset averages are taken over the available horizons. Red: best, blue underline: second best.

Dataset	H	FreqGate (Ours)		VisionTS		GPT4TS		PatchTST		iTransformer		Time-LLM		DLinear		Time-VLM		TimesNet		Autoformer		Non-stat. Trans.		FEDformer		Aurora		ETSformer		DMMV		Informer	
ETTh1	96	0.391	0.381	0.385	0.381	0.625	0.516	0.638	0.532	0.635	0.542	0.877	0.645	0.661	0.560	0.666	0.526	0.868	0.611	0.763	0.617	0.905	0.612	0.835	0.628	1.363	0.833	1.537	0.922	1.277	0.847	1.520	0.948
	192	0.417	0.414	0.420	0.412	0.766	0.567	0.821	0.596	0.756	0.587	1.446	0.826	0.703	0.576	0.755	0.562	0.933	0.643	0.845	0.668	0.924	0.646	0.984	0.719	1.301	0.841	1.559	0.933	1.242	0.840	1.465	0.932
	336	0.464	0.445	0.451	0.432	0.841	0.608	0.864	0.636	0.846	0.632	1.216	0.755	0.698	0.581	0.827	0.613	1.101	0.708	0.844	0.658	0.904	0.656	1.014	0.745	1.418	0.903	1.623	0.958	1.138	0.791	1.459	0.923
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.424	0.413	0.419	0.408	0.744	0.563	0.774	0.588	0.746	0.587	1.180	0.742	0.687	0.572	0.749	0.567	0.967	0.654	0.817	0.648	0.911	0.638	0.944	0.697	1.361	0.859	1.573	0.938	1.219	0.826	1.481	0.934
ETTh2	96	0.306	0.337	0.311	0.342	0.348	0.396	0.468	0.466	0.373	0.411	0.378	0.418	0.398	0.447	0.464	0.465	0.500	0.483	0.476	0.509	0.423	0.452	0.544	0.551	0.412	0.472	0.659	0.612	3.280	1.374	2.672	1.269
	192	0.360	0.377	0.373	0.386	0.435	0.442	0.473	0.473	0.445	0.453	0.406	0.423	0.406	0.451	0.480	0.472	0.648	0.547	0.478	0.514	0.546	0.509	0.678	0.630	0.501	0.520	0.795	0.673	3.279	1.381	3.166	1.396
	336	0.395	0.405	0.398	0.410	0.497	0.485	0.467	0.474	0.455	0.468	0.507	0.488	0.411	0.457	0.474	0.472	0.658	0.556	0.498	0.521	0.542	0.511	0.718	0.654	0.671	0.601	0.863	0.699	3.334	1.400	3.358	1.437
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.354	0.373	0.361	0.379	0.427	0.441	0.470	0.471	0.424	0.444	0.430	0.443	0.405	0.452	0.472	0.470	0.602	0.529	0.484	0.515	0.504	0.491	0.647	0.612	0.528	0.531	0.772	0.661	3.298	1.385	3.066	1.367
ETTm1	96	0.335	0.333	0.347	0.353	0.416	0.414	0.395	0.413	0.420	0.432	0.494	0.454	0.530	0.479	0.374	0.400	0.789	0.598	0.727	0.578	0.799	0.573	0.700	0.556	1.499	0.873	1.208	0.798	0.858	0.686	1.459	0.878
	192	0.371	0.361	0.373	0.371	0.456	0.437	0.440	0.439	0.472	0.464	0.603	0.513	0.561	0.496	0.420	0.431	0.817	0.609	0.732	0.578	0.804	0.566	0.723	0.572	1.622	0.910	1.336	0.844	1.408	0.852	1.582	0.909
	336	0.399	0.384	0.404	0.399	0.514	0.472	0.487	0.464	0.532	0.494	0.698	0.551	0.608	0.521	0.484	0.468	0.768	0.588	0.745	0.590	0.794	0.574	0.760	0.598	1.722	0.927	1.433	0.873	1.369	0.849	1.702	0.944
	720	0.481	0.449	0.500	0.459	0.624	0.547	0.685	0.568	0.780	0.620	1.094	0.689	0.711	0.587	0.713	0.588	0.913	0.687	0.753	0.598	0.932	0.682	0.930	0.708	1.504	0.903	1.692	0.963	1.394	0.856	1.948	0.986
	Avg	0.397	0.382	0.406	0.396	0.502	0.467	0.502	0.471	0.551	0.502	0.722	0.552	0.602	0.521	0.498	0.471	0.822	0.621	0.739	0.586	0.832	0.599	0.778	0.608	1.587	0.903	1.417	0.870	1.257	0.811	1.673	0.929
ETTm2	96	0.195	0.274	0.237	0.294	0.217	0.296	0.228	0.302	0.216	0.301	0.265	0.333	0.258	0.336	0.200	0.280	0.330	0.378	0.340	0.410	0.296	0.365	0.328	0.397	0.281	0.382	0.356	0.445	1.641	0.987	2.902	1.356
	192	0.247	0.308	0.290	0.329	0.271	0.330	0.285	0.337	0.272	0.335	0.325	0.370	0.293	0.357	0.254	0.314	0.359	0.392	0.364	0.419	0.357	0.394	0.374	0.426	0.364	0.430	0.436	0.489	2.183	1.145	3.597	1.520
	336	0.300	0.335	0.336	0.362	0.324	0.362	0.340	0.371	0.323	0.368	0.405	0.416	0.330	0.382	0.307	0.348	0.440	0.436	0.395	0.434	0.399	0.419	0.425	0.456	0.434	0.477	0.512	0.529	2.378	1.195	4.101	1.620
	720	0.393	0.380	0.402	0.400	0.427	0.421	0.449	0.432	0.438	0.431	0.554	0.490	0.436	0.449	0.421	0.416	0.492	0.466	0.456	0.464	0.477	0.462	0.568	0.549	0.497	0.504	0.643	0.593	3.172	1.331	4.535	1.681
	Avg	0.284	0.324	0.316	0.346	0.310	0.352	0.325	0.361	0.312	0.359	0.387	0.402	0.329	0.381	0.296	0.339	0.405	0.418	0.389	0.432	0.382	0.410	0.424	0.457	0.394	0.448	0.486	0.514	2.343	1.164	3.784	1.544
Weather	96	0.170	0.232	0.183	0.233	0.175	0.233	0.196	0.254	0.202	0.257	0.180	0.234	0.235	0.308	0.177	0.234	0.274	0.309	0.310	0.369	0.246	0.301	0.317	0.388	0.268	0.385	0.362	0.432	0.312	0.361	0.510	0.506
	192	0.226	0.257	0.239	0.276	0.234	0.284	0.246	0.292	0.257	0.303	0.238	0.283	0.270	0.333	0.229	0.279	0.337	0.359	0.332	0.381	0.303	0.341	0.357	0.415	0.359	0.549	0.447	0.485	0.371	0.399	0.585	0.541
	336	0.284	0.304	0.312	0.334	0.296	0.327	0.297	0.330	0.313	0.347	0.309	0.331	0.311	0.362	0.288	0.322	0.364	0.379	0.360	0.394	0.339	0.368	0.421	0.460	0.844	0.672	0.551	0.542	0.428	0.437	0.573	0.539
	720	0.336	0.368	0.371	0.378	0.376	0.382	0.377	0.386	0.388	0.397	0.405	0.392	0.362	0.404	0.367	0.376	0.405	0.406	0.401	0.415	0.425	0.419	0.532	0.530	1.289	0.864	0.661	0.597	0.503	0.487	0.543	0.526
	Avg	0.254	0.290	0.276	0.305	0.270	0.306	0.279	0.315	0.290	0.326	0.283	0.310	0.295	0.352	0.265	0.303	0.345	0.364	0.351	0.390	0.328	0.357	0.407	0.448	0.740	0.617	0.505	0.514	0.404	0.421	0.553	0.528
ECL	96	0.137	0.230	0.142	0.232	0.149	0.251	0.189	0.293	0.217	0.318	0.181	0.281	0.507	0.572	0.251	0.349	0.314	0.394	0.489	0.537	0.824	0.737	0.844	0.731	0.647	0.440	1.298	0.900	0.238	0.327	1.284	0.922
	192	0.158	0.242	0.160	0.250	0.164	0.262	0.206	0.308	0.257	0.351	0.196	0.297	0.581	0.618	0.263	0.358	0.333	0.407	0.556	0.581	0.880	0.760	0.979	0.791	0.565	0.465	1.418	0.940	0.294	0.371	1.278	0.922
	336	0.176	0.263	0.178	0.270	0.184	0.281	0.229	0.328	0.296	0.382	0.215	0.318	0.591	0.622	0.276	0.368	0.359	0.432	0.544	0.572	0.904	0.768	1.017	0.805	0.872	0.607	1.461	0.955	0.417	0.442	1.267	0.919
	720	0.229	0.310	0.239	0.320	0.232	0.322	0.337	0.415	0.580	0.584	0.290	0.372	0.784	0.725	0.505	0.549	0.972	0.786	0.802	0.710	1.076	0.827	1.245	0.896	1.238	0.746	1.862	1.080	0.793	0.631	1.268	0.917
	Avg	0.175	0.261	0.180	0.268	0.182	0.279	0.240	0.336	0.337	0.409	0.221	0.317	0.616	0.634	0.324	0.406	0.494	0.505	0.598	0.600	0.921	0.773	1.021	0.806	0.831	0.565	1.510	0.969	0.436	0.443	1.274	0.920
Traffic	96	0.366	0.240	0.388	0.256	0.433	0.310	0.458	0.347	0.640	0.461	0.418	0.291	1.162	0.695	0.739	0.505	0.916	0.543	1.262	0.739	1.524	0.851	1.617	0.895	1.277	0.614	2.365	1.099	0.803	0.500	1.660	0.876
	192	0.390	0.253	0.409	0.266	0.445	0.314	0.484	0.365	0.745	0.506	0.428	0.298	1.175	0.698	0.786	0.530	0.993	0.584	1.278	0.741	1.573	0.863	1.669	0.913	1.305	0.595	2.413	1.113	0.872	0.534	1.661	0.874
	336	0.405	0.262	0.425	0.276	0.457	0.322	0.535	0.400	0.909	0.573	0.443	0.307	1.291	0.748	0.872	0.572	1.197	0.684	1.434	0.809	1.715	0.908	1.784	0.951	1.322	0.623	2.711	1.194	0.981	0.581	1.729	0.898
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.387	0.252	0.407	0.266	0.445	0.316	0.492	0.371	0.765	0.513	0.430	0.299	1.210	0.714	0.799	0.536	1.035	0.604	1.325	0.763	1.604	0.874	1.690	0.919	1.301	0.611	2.496	1.135	0.885	0.538	1.683	0.883
Exchange	96	0.134	0.247	0.123	0.242	0.269	0.387	0.305	0.412	0.243	0.369	0.197	0.330	0.382	0.471	0.322	0.425	0.529	0.533	0.570	0.591	0.564	0.555	0.640	0.632	0.449	0.544	0.436	0.504	5.267	1.986	3.374	1.499
	192	0.299	0.388	0.305	0.390	0.546	0.564	0.437	0.495	0.477	0.522	0.270	0.385	0.476	0.529	2.706	1.264	0.760	0.644	0.662	0.641	0.775	0.642	0.826	0.718	0.673	0.684	0.535	0.567	6.350	2.197	3.392	1.494
	336	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.217	0.318	0.214	0.316	0.407	0.475	0.371	0.453	0.360	0.445	0.233	0.357	0.429	0.500	1.514	0.844	0.645	0.589	0.616	0.616	0.669	0.598	0.733	0.675	0.561	0.614	0.486	0.536	5.809	2.092	3.383	1.497

Table 9: Full few-shot learning results on 10% training data with forecasting horizons $H \in \{96, 192, 336, 720\}$. A lower value indicates better performance. "-" means that 10% of the time series is not sufficient to constitute a training set for that horizon; dataset averages are taken over the available horizons. Red: best, blue underline: second best.

Dataset	H	FreqGate (Ours)		VisionTS		GPT4TS		PatchTST		Time-VLM		iTransformer		Time-LLM		DLinear		Autoformer		TimesNet		Non-stat. Trans.		FEDformer		Aurora		ETSformer		DMMV		Informer	
ETTh1	96	0.387	0.405	0.382	0.380	0.485	0.455	0.527	0.494	0.576	0.489	0.567	0.523	0.561	0.508	0.600	0.527	0.731	0.608	0.881	0.620	0.831	0.596	0.783	0.609	1.294	0.904	1.394	0.870	1.207	0.818	1.471	0.903
	192	0.436	0.417	0.418	0.411	0.560	0.490	0.588	0.524	0.619	0.507	0.654	0.554	0.707	0.579	0.611	0.537	0.780	0.628	0.996	0.667	0.908	0.621	0.828	0.634	1.313	0.907	1.555	0.922	1.188	0.820	1.610	0.937
	336	0.458	0.440	0.445	0.432	0.627	0.531	0.731	0.586	0.660	0.529	0.721	0.578	1.038	0.710	0.635	0.557	0.723	0.606	1.439	0.781	0.990	0.677	0.880	0.673	1.399	0.952	1.600	0.954	1.138	0.791	1.503	0.912
	720	0.458	0.479	0.445	0.453	0.788	0.597	0.862	0.648	0.820	0.628	0.856	0.651	1.278	0.818	0.700	0.600	0.794	0.647	1.416	0.783	0.903	0.668	0.997	0.743	1.600	0.937	1.543	0.941	1.059	0.776	1.420	0.900
	Avg	0.435	0.435	0.423	0.419	0.615	0.518	0.677	0.563	0.669	0.538	0.700	0.576	0.896	0.654	0.637	0.555	0.757	0.622	1.183	0.713	0.908	0.640	0.872	0.665	1.401	0.925	1.523	0.922	1.148	0.801	1.501	0.913
ETTh2	96	0.299	0.340	0.307	0.342	0.321	0.375	0.343	0.385	0.303	0.358	0.343	0.391	0.391	0.422	0.360	0.411	0.458	0.494	0.492	0.484	0.400	0.437	0.427	0.476	0.416	0.485	0.572	0.566	2.839	1.283	2.819	1.301
	192	0.355	0.367	0.368	0.385	0.376	0.413	0.423	0.439	0.377	0.412	0.424	0.437	0.485	0.474	0.396	0.437	0.458	0.495	0.472	0.479	0.433	0.459	0.462	0.501	0.457	0.508	0.655	0.604	2.857	1.296	3.275	1.414
	336	0.369	0.408	0.395	0.413	0.391	0.429	0.417	0.442	0.392	0.428	0.439	0.454	0.494	0.480	0.419	0.458	0.449	0.491	0.450	0.468	0.452	0.477	0.558	0.567	0.495	0.545	0.770	0.658	3.334	1.400	3.487	1.462
	720	0.390	0.425	0.407	0.432	0.461	0.472	0.484	0.489	0.486	0.489	0.482	0.488	0.467	0.473	0.433	0.465	0.515	0.535	0.721	0.581	0.529	0.513	0.740	0.667	0.529	0.545	0.795	0.678	3.467	1.442	3.607	1.499
	Avg	0.353	0.385	0.369	0.393	0.387	0.423	0.417	0.438	0.389	0.422	0.422	0.442	0.459	0.462	0.402	0.443	0.470	0.504	0.534	0.503	0.454	0.471	0.547	0.553	0.474	0.521	0.698	0.626	3.124	1.355	3.297	1.419
ETTm1	96	0.329	0.345	0.337	0.355	0.422	0.409	0.390	0.407	0.386	0.403	0.420	0.431	0.477	0.436	0.481	0.455	0.745	0.587	0.697	0.569	0.760	0.551	0.688	0.552	1.584	0.944	1.008	0.705	0.793	0.655	1.462	0.901
	192	0.352	0.356	0.364	0.368	0.469	0.439	0.419	0.419	0.420	0.424	0.487	0.470	0.603	0.491	0.503	0.468	0.772	0.603	0.775	0.587	0.809	0.584	0.701	0.560	1.556	0.915	1.079	0.737	0.843	0.664	1.610	0.939
	336	0.379	0.384	0.383	0.386	0.567	0.478	0.450	0.434	0.456	0.447	0.567	0.512	0.787	0.552	0.540	0.487	0.782	0.612	0.846	0.625	0.846	0.606	0.726	0.578	1.634	0.933	1.128	0.763	1.060	0.760	1.711	0.966
	720	0.447	0.426	0.463	0.438	0.764	0.553	0.518	0.483	0.542	0.497	0.637	0.550	0.844	0.596	0.595	0.520	0.775	0.608	1.186	0.767	0.976	0.687	0.762	0.598	1.707	0.958	1.236	0.811	1.032	0.731	1.998	1.015
	Avg	0.377	0.378	0.386	0.387	0.556	0.470	0.444	0.436	0.451	0.443	0.528	0.491	0.678	0.519	0.530	0.483	0.768	0.603	0.876	0.637	0.848	0.607	0.719	0.572	1.620	0.937	1.112	0.754	0.932	0.703	1.695	0.955
ETTm2	96	0.187	0.268	0.227	0.286	0.198	0.276	0.207	0.284	0.189	0.270	0.198	0.285	0.234	0.304	0.232	0.315	0.312	0.376	0.319	0.365	0.290	0.357	0.291	0.363	0.283	0.390	0.310	0.412	1.055	0.798	2.987	1.365
	192	0.232	0.300	0.269	0.313	0.253	0.313	0.264	0.320	0.247	0.308	0.254	0.324	0.286	0.339	0.272	0.339	0.344	0.395	0.363	0.397	0.328	0.377	0.322	0.381	0.323	0.408	0.374	0.449	1.125	0.814	3.619	1.509
	336	0.277	0.325	0.307	0.340	0.297	0.342	0.312	0.349	0.292	0.338	0.308	0.358	0.377	0.393	0.314	0.366	0.384	0.423	0.423	0.429	0.383	0.406	0.366	0.409	0.387	0.444	0.442	0.485	1.603	0.974	4.035	1.588
	720	0.380	0.386	0.413	0.407	0.415	0.417	0.413	0.411	0.385	0.395	0.408	0.416	0.516	0.461	0.395	0.414	0.471	0.484	0.505	0.471	0.473	0.460	0.465	0.469	0.450	0.485	0.554	0.542	1.526	0.931	4.505	1.674
	Avg	0.269	0.320	0.304	0.337	0.291	0.337	0.299	0.341	0.278	0.328	0.292	0.346	0.353	0.374	0.303	0.358	0.378	0.420	0.403	0.416	0.368	0.400	0.361	0.405	0.361	0.432	0.420	0.472	1.327	0.879	3.787	1.534
Weather	96	0.155	0.205	0.164	0.210	0.161	0.214	0.175	0.229	0.159	0.211	0.193	0.243	0.163	0.216	0.205	0.273	0.307	0.356	0.249	0.278	0.228	0.280	0.302	0.370	0.264	0.378	0.278	0.365	0.243	0.301	0.414	0.452
	192	0.213	0.255	0.221	0.260	0.210	0.259	0.219	0.262	0.206	0.254	0.241	0.284	0.215	0.264	0.240	0.300	0.337	0.371	0.304	0.324	0.293	0.325	0.325	0.384	0.530	0.530	0.349	0.414	0.297	0.339	0.500	0.506
	336	0.256	0.288	0.310	0.321	0.270	0.305	0.265	0.300	0.262	0.298	0.288	0.322	0.268	0.304	0.278	0.327	0.356	0.380	0.368	0.368	0.319	0.348	0.352	0.398	0.913	0.709	0.426	0.463	0.352	0.377	0.507	0.505
	720	0.302	0.326	0.380	0.368	0.334	0.350	0.329	0.345	0.320	0.338	0.349	0.365	0.335	0.350	0.332	0.364	0.409	0.405	0.416	0.401	0.372	0.381	0.408	0.430	1.386	0.906	0.504	0.505	0.421	0.425	0.563	0.521
	Avg	0.232	0.269	0.269	0.290	0.244	0.282	0.247	0.285	0.237	0.275	0.268	0.304	0.245	0.283	0.264	0.316	0.352	0.378	0.334	0.343	0.303	0.333	0.347	0.396	0.773	0.631	0.389	0.437	0.328	0.361	0.496	0.496
ECL	96	0.144	0.234	0.138	0.229	0.146	0.247	0.167	0.272	0.214	0.317	0.179	0.285	0.163	0.264	0.365	0.467	0.408	0.481	0.300	0.380	0.552	0.565	0.617	0.607	0.467	0.381	1.065	0.813	0.187	0.281	1.180	0.880
	192	0.151	0.244	0.155	0.246	0.164	0.264	0.182	0.285	0.223	0.324	0.209	0.311	0.179	0.278	0.375	0.473	0.408	0.481	0.315	0.391	0.571	0.576	0.686	0.642	0.549	0.424	1.104	0.827	0.205	0.304	1.190	0.887
	336	0.174	0.263	0.176	0.265	0.182	0.280	0.202	0.302	0.239	0.336	0.243	0.340	0.196	0.295	0.427	0.513	0.440	0.504	0.331	0.403	0.668	0.635	0.797	0.698	0.878	0.570	1.190	0.859	0.237	0.329	1.192	0.890
	720	0.226	0.311	0.245	0.320	0.242	0.325	0.262	0.347	0.301	0.379	0.321	0.396	0.253	0.334	0.511	0.569	0.476	0.527	0.400	0.450	0.865	0.739	0.948	0.766	1.266	0.762	1.380	0.928	0.433	0.429	1.219	0.905
	Avg	0.174	0.263	0.179	0.265	0.184	0.279	0.203	0.302	0.244	0.339	0.238	0.333	0.198	0.293	0.419	0.505	0.433	0.498	0.337	0.406	0.664	0.629	0.762	0.678	0.790	0.534	1.185	0.857	0.253	0.336	1.195	0.890
Traffic	96	0.386	0.260	0.395	0.268	0.433	0.322	0.427	0.320	0.599	0.414	0.534	0.398	0.416	0.288	0.937	0.582	1.028	0.624	0.800	0.464	1.349	0.780	1.359	0.785	1.325	0.652	1.996	0.987	0.600	0.409	1.563	0.832
	192	0.440	0.287	0.411	0.277	0.446	0.324	0.440	0.327	0.613	0.420	0.590	0.428	0.425	0.295	1.008	0.616	1.082	0.648	0.844	0.484	1.410	0.807	1.506	0.847	1.324	0.616	2.134	1.030	0.622	0.420	1.593	0.844
	336	0.389	0.272	0.415	0.278	0.453	0.327	0.451	0.335	0.624	0.425	0.660	0.461	0.437	0.303	1.022	0.619	1.081	0.644	0.897	0.512	1.426	0.811	1.552	0.862	1.382	0.601	2.193	1.046	0.673	0.448	1.600	0.843
	720	0.470	0.311	0.482	0.315	0.502	0.350	0.515	0.376	0.762	0.505	0.865	0.547	0.477	0.325	1.209	0.705	1.327	0.753	1.035	0.619	1.621	0.878	1.737	0.931	1.452	0.610	2.542	1.142	0.929	0.566	1.676	0.872
	Avg	0.421	0.283	0.426	0.284	0.459	0.331	0.458	0.339	0.649	0.441	0.662	0.458	0.439	0.303	1.044	0.630	1.130	0.667	0.894	0.520	1.451	0.819	1.539	0.856	1.371	0.620	2.216	1.051	0.706	0.461	1.608	0.848
Exchange	96	0.103	0.232	0.097	0.222	0.197	0.331	0.204	0.342	0.186	0.319	0.202	0.334	0.162	0.297	0.348	0.458	0.484	0.535	0.653	0.550	0.539	0.516	0.613	0.607	0.714	0.698	0.398	0.483	5.262	1.928	3.428	1.498
	192	0.199	0.333	0.195	0.324	0.288	0.403	0.361	0.450	0.337	0.429	0.306	0.419	0.314	0.415	0.428	0.508	0.580	0.587	0.714	0.595	0.658	0.582	0.801	0.695	0.711	0.707	0.479	0.538	6.350	2.197	3.507	1.500
	336	0.389	0.451	0.375	0.447	0.443	0.500	0.484	0.528	0.891	0.726	0.535	0.560	0.482	0.511	0.639	0.621	0.852	0.732	1.327	0.780	0.831	0.687	1.067	0.817	1.151	0.960	0.682	0.646	7.495	2.407	3.397	1.483
	720	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	0.230	0.339	0.222	0.331	0.309	0.411	0.350	0.440	0.471	0.491	0.347	0.438	0.319	0.408	0.472	0.529	0.639	0.618	0.898	0.642	0.676	0.595	0.827	0.706	0.859	0.788	0.520	0.556	6.371	2.197	3.444	1.493

B Controlled Modality-Utility Ablation

This section reports the controlled ablation referenced in Section 4.6: within a single tri

Table 10: Full zero-shot learning results with forecasting horizons $H \in \{96, 192, 336, 720\}$. We use the same protocol as in Table 3. A lower value indicates better performance. Red: best, blue underline: second best.

Dataset	H	FreqGate (Ours)		VisionTS		Time-LLM		PatchTST		GPT4TS		DLinear		Time-VLM		iTransformer		FEDformer		ETSformer		Autoformer		TimesNet		Aurora		Non-stat. Trans.		Informer	
ETTh1 → ETTh2	96	0.274	0.327	0.311	0.343	0.279	0.344	0.284	0.346	0.282	0.349	0.322	0.378	0.278	0.341	0.287	0.349	0.421	0.467	0.397	0.460	0.525	0.532	0.376	0.416	0.361	0.453	0.432	0.452	1.820	1.042
	192	0.327	0.372	0.372	0.386	0.339	0.381	0.354	0.393	0.351	0.392	0.399	0.428	0.340	0.384	0.356	0.393	0.436	0.478	0.488	0.509	0.554	0.551	0.453	0.457	0.429	0.504	0.552	0.521	2.041	1.113
	336	0.344	0.391	0.397	0.409	0.363	0.405	0.383	0.417	0.378	0.418	0.466	0.474	0.372	0.411	0.388	0.419	0.446	0.487	0.536	0.537	0.686	0.607	0.451	0.465	0.472	0.534	0.523	0.507	2.622	1.291
	720	0.385	0.419	0.408	0.432	0.418	0.449	0.419	0.449	0.487	0.484	0.678	0.588	0.399	0.438	0.419	0.451	0.551	0.552	0.559	0.555	0.670	0.601	0.543	0.518	0.492	0.541	0.568	0.528	2.834	1.319
	Avg	0.333	0.377	0.372	0.393	0.350	0.395	0.360	0.401	0.375	0.411	0.466	0.467	0.347	0.394	0.363	0.403	0.463	0.496	0.495	0.515	0.609	0.573	0.456	0.464	0.439	0.508	0.519	0.502	2.329	1.191
ETTh2 → ETTh1	96	0.367	0.366	0.388	0.383	0.489	0.470	0.441	0.442	0.488	0.473	0.432	0.446	0.517	0.496	0.553	0.516	0.750	0.606	1.013	0.709	0.782	0.616	0.824	0.606	1.171	0.817	0.928	0.663	0.914	0.678
	192	0.415	0.395	0.423	0.412	0.539	0.503	0.499	0.480	0.553	0.519	0.464	0.469	0.609	0.551	0.717	0.599	0.777	0.622	1.073	0.743	0.848	0.653	0.878	0.634	1.256	0.860	0.909	0.651	0.992	0.735
	336	0.427	0.412	0.451	0.432	0.538	0.508	0.548	0.518	0.520	0.502	0.476	0.485	0.591	0.546	0.743	0.611	0.755	0.618	1.120	0.772	0.804	0.639	0.970	0.668	1.317	0.887	0.951	0.666	1.051	0.780
	720	0.432	0.426	0.446	0.454	0.576	0.539	0.553	0.532	0.641	0.583	0.535	0.538	0.679	0.601	0.830	0.661	0.833	0.663	1.066	0.758	0.880	0.694	2.429	0.944	1.401	0.926	0.929	0.669	1.349	0.878
	Avg	0.410	0.400	0.427	0.420	0.535	0.505	0.510	0.493	0.551	0.519	0.477	0.484	0.599	0.548	0.711	0.597	0.779	0.627	1.068	0.745	0.829	0.651	1.275	0.713	1.286	0.873	0.929	0.662	1.077	0.768
ETTh1 → ETTm2	96	0.212	0.292	0.247	0.291	0.214	0.305	0.220	0.311	0.222	0.313	0.244	0.342	0.207	0.301	0.224	0.316	0.313	0.388	0.307	0.408	0.383	0.448	0.293	0.359	0.228	0.346	0.335	0.389	1.428	0.900
	192	0.256	0.304	0.287	0.319	0.265	0.337	0.279	0.346	0.289	0.354	0.315	0.391	0.267	0.338	0.280	0.349	0.344	0.407	0.379	0.449	0.415	0.462	0.369	0.394	0.283	0.383	0.459	0.457	1.476	0.912
	336	0.311	0.337	0.324	0.347	0.317	0.368	0.328	0.374	0.346	0.389	0.407	0.451	0.325	0.371	0.332	0.380	0.387	0.433	0.449	0.487	0.464	0.495	0.523	0.462	0.337	0.418	0.486	0.464	1.585	0.946
	720	0.385	0.383	0.398	0.396	0.414	0.426	0.413	0.422	0.470	0.456	0.671	0.583	0.423	0.425	0.429	0.433	0.470	0.494	0.531	0.544	0.512	0.531	0.604	0.514	0.415	0.467	0.576	0.505	1.712	0.987
	Avg	0.291	0.329	0.314	0.338	0.303	0.359	0.310	0.363	0.332	0.378	0.409	0.442	0.306	0.359	0.316	0.369	0.379	0.431	0.417	0.472	0.444	0.484	0.447	0.432	0.316	0.404	0.464	0.454	1.550	0.936
ETTm1 → ETTm2	96	0.169	0.258	0.247	0.291	0.182	0.267	0.182	0.267	0.183	0.268	0.185	0.278	0.179	0.264	0.188	0.275	0.375	0.435	0.287	0.389	0.374	0.436	0.227	0.309	0.228	0.346	0.278	0.346	1.838	1.111
	192	0.226	0.285	0.287	0.319	0.234	0.300	0.236	0.303	0.237	0.304	0.236	0.315	0.231	0.299	0.238	0.308	0.409	0.453	0.352	0.431	0.395	0.443	0.282	0.342	0.283	0.383	0.352	0.391	2.257	1.270
	336	0.279	0.329	0.324	0.347	0.286	0.336	0.287	0.336	0.285	0.338	0.292	0.353	0.286	0.336	0.296	0.344	0.438	0.465	0.429	0.484	0.470	0.493	0.380	0.402	0.337	0.418	0.420	0.429	2.429	1.328
	720	0.357	0.380	0.398	0.396	0.368	0.387	0.364	0.385	0.373	0.393	0.391	0.418	0.371	0.389	0.390	0.401	0.514	0.507	0.488	0.510	0.557	0.533	0.527	0.475	0.415	0.467	0.537	0.488	1.718	1.074
	Avg	0.258	0.313	0.314	0.338	0.268	0.323	0.267	0.323	0.269	0.326	0.276	0.341	0.267	0.322	0.278	0.332	0.434	0.465	0.389	0.453	0.449	0.476	0.354	0.382	0.316	0.404	0.397	0.414	2.061	1.196
ETTm2 → ETTm1	96	0.396	0.362	0.386	0.367	0.402	0.407	0.413	0.416	0.428	0.414	0.372	0.400	0.445	0.433	0.440	0.436	0.746	0.589	0.946	0.660	2.501	0.920	0.687	0.534	1.290	0.837	1.332	0.725	0.786	0.596
	192	0.397	0.376	0.394	0.377	0.446	0.432	0.453	0.440	0.470	0.447	0.396	0.414	0.516	0.478	0.538	0.495	0.753	0.591	1.013	0.691	0.739	0.583	0.911	0.621	1.330	0.884	2.244	0.966	0.815	0.615
	336	0.434	0.411	0.414	0.396	0.455	0.438	0.475	0.453	0.491	0.462	0.417	0.429	0.558	0.503	0.613	0.536	0.758	0.591	1.086	0.724	0.738	0.585	1.551	0.769	1.371	0.913	2.046	0.939	0.821	0.633
	720	0.462	0.452	0.467	0.432	0.492	0.460	0.508	0.466	0.561	0.493	0.459	0.455	0.640	0.547	0.735	0.597	0.776	0.603	1.003	0.713	0.768	0.600	1.379	0.742	1.388	0.925	1.499	0.835	0.934	0.700
	Avg	0.422	0.400	0.415	0.393	0.449	0.434	0.462	0.444	0.488	0.454	0.411	0.425	0.540	0.490	0.582	0.516	0.758	0.593	1.012	0.697	1.187	0.672	1.132	0.666	1.345	0.890	1.780	0.866	0.839	0.636

Table 11: Controlled modality ablation within a single tri-modal architecture. Time-VLM is re-run under our unified protocol with its text and image branches removed in all combinations. MSE is averaged over four horizons; Δ = Tri-modal − Num-only (negative: auxiliary modalities help). Red marks the best variant per dataset.

Dataset	Spec.	Num-only	+Text	+Image	Tri-modal	Δ
ETTh2 (avg)	Low	0.362	0.364	0.356	0.359	−0.003
ETTm2 (avg)	Low-Mid	0.274	0.263	0.270	0.264	−0.010
Weather (avg)	Mid	0.238	0.240	0.231	0.232	−0.006
ECL (avg)	Mid-High	0.170	0.176	0.174	0.178	+0.008
Traffic (avg)	High	0.396	0.427	0.415	0.440	+0.044
Exchange (avg)	Weak per.	0.381	0.386	0.384	0.388	+0.007

C Sensitivity to the Band Count

We vary $K \in \{2, 3, 4, 5, 8\}$ on representative low-, mid- and high-frequency datasets. K=3 yields the lowest average MSE and the lowest converged $\mathcal{L}_{\text{spec}}$ value. Both coarser and finer partitions increase error and encoder-response overlap, supporting three bands as a robust default for the evaluated benchmarks rather than a universal optimum.

D Conditional-Compute Measurement Protocol

An auxiliary encoder whose band gates vanish in every band is skipped at inference. To quantify this conditional-compute property, future measurements should report, for each benchmark, the per-band gate statistics of each auxiliary encoder, model parameters, forward FLOPs under a fixed input shape and batch-size-one wall-clock latency on a stated device. Per-band gates in Table 5 are not substitutes for these measurements.

E Reproducibility

Hyper-parameters. FreqGate uses K=3 frequency bands, hidden dimension d=512, harmonised token count N=96, band-logit template $\mathbf{M}$ initialised to equal thirds, a per-series offset network $\Delta(\mathbf{W})$ implemented as a two-layer MLP $\mathbb{R}^F \to \mathbb{R}^{32} \to \mathbb{R}^{K\times F}$ with GELU whose output layer is zero-initialised, anti-collapse floor u_0=0.02, an inference-time gate threshold of 0.02, and loss weights λ_{spec}=0.1 and λ_{bal}=0.01. All hyper-parameters are tuned on the ETTh1 validation split and reused across every other dataset without re-tuning.

Table 12: Sensitivity to the band count K. MSE is averaged over four horizons on three representative datasets (low / mid / high frequency); "Avg" is their mean. The $\mathcal{L}_{\text{spec}}$ convergence value is the spectral overlap between the auxiliary encoders' normalised responses (Eq. (8)), so lower indicates cleaner separation.

K	ETTh1 (low)	Weather (mid)	Traffic (high)	Avg	$\mathcal{L}_{\text{spec}}$ conv.
2	0.410	0.227	0.382	0.340	0.18
3 (default)	**0.400**	**0.218**	**0.377**	**0.332**	**0.14**
4	0.402	0.220	0.378	0.333	0.16
5	0.406	0.225	0.381	0.337	0.19
8	0.412	0.229	0.386	0.342	0.22

Anti-collapse floor. For a mini-batch $\mathcal{Q}$, let

$$\bar{w}_m = \frac{1}{K\,|\mathcal{Q}|} \sum_{i \in \mathcal{Q}} \sum_{k=1}^{K} w_{i,m,k}, \qquad \mathcal{L}_{\text{bal}} = \sum_{m \in \{T,V\}} \big[\max(0, u_0 - \bar{w}_m)\big]^2. \tag{10}$$

This one-sided floor penalises only an almost unused auxiliary encoder and does not impose a target gate level on individual input series.

Encoder architectures. All three encoders map into $d{=}512$ and emit $N{=}96$ tokens; only the numerical encoder is trained from scratch, while the auxiliary encoders are frozen pretrained ViLT (Kim et al., 2021) encoders, following Zhong et al. (2025). *Numerical encoder* Enc_X: instance normalisation, then a linear embedding of each scalar time step into $\mathbb{R}^d$, then 2 pre-norm Transformer layers (4 heads, head dimension 128, feed-forward width $2d{=}1024$, GELU, dropout 0.1) with learned additive positional embeddings over the $L{=}96$ steps. The encoder is *channel-independent*: one set of weights is applied to each of the C channels as an independent univariate sequence, so no weight tensor has an extent that depends on C and the $\mathbb{R}^{N\times d} \to \mathbb{R}^H$ forecasting head is likewise shared across channels. This is what makes the zero-shot pairs below well posed. *Visual encoder* Enc_V: the three image representations (Gramian angular field, recurrence plot, line chart) are stacked into a three-channel image and encoded by the frozen ViLT vision encoder (`vilt-b32-finetuned-coco`), following Zhong et al. (2025); the patch embeddings are projected to $\mathbb{R}^d$ and resampled to $N{=}96$ tokens by a learned linear map acting on the token axis. *Textual encoder* Enc_T: the template string is tokenised by the ViLT tokenizer and encoded by the frozen ViLT text encoder; the token embeddings are projected to $\mathbb{R}^d$ and resampled to $N{=}96$ tokens by a learned linear map acting on the token axis. The two token-axis maps harmonise the auxiliary representations to the shared latent-token length N used by BGF; they do not impose a one-to-one correspondence between auxiliary tokens and physical timestamps. Each gate network ψ_m is a two-layer MLP $\mathbb{R}^{3K} \to \mathbb{R}^{32} \to \mathbb{R}^K$ with GELU. BGF uses 4 attention heads per band and shares its projection weights across the $K{=}3$ bands.

Optimisation. Adam optimiser, initial learning rate 10^{-4}, batch size 256, early stopping on validation MSE with patience 10, three seeds per configuration. All experiments are conducted on a single NVIDIA H800 GPU (80GB).

Splits and metrics. We follow the standardised protocol of Liu et al. (2024c): input length $L{=}96$, prediction horizons $H \in \{96, 192, 336, 720\}$, the canonical train/validation/test splits of prior work (Wu et al., 2021; 2023; Nie et al., 2023) and MSE/MAE as evaluation metrics. The 5% and 10% few-shot settings take the earliest 5% or 10% of the training split as a contiguous chronological prefix and extract sliding windows from it, leaving validation and test untouched; the zero-shot setting trains on a source dataset and evaluates on the target with no target-domain gradient step, over the five pairs ETTh1 $\to$ ETTh2, ETTh2 $\to$ ETTh1, ETTh1 $\to$ ETTm2, ETTm1 $\to$ ETTm2 and ETTm2 $\to$ ETTm1. Both protocols are stated in full in Section 4.1 and are applied identically to every baseline.

Code and data. All datasets used are publicly available under standard research licenses. Anonymised source code is included in the supplementary material.