ANCHORSCOPE: LEARNING TEXTUAL FORECAST REVISIONS AT THE RIGHT TEMPORAL SCALE

Anonymous authors
https://anonymous.4open.science/status/Anchor-0103

ABSTRACT

Forecasting systems often begin with a prediction from numerical history and then revise it when contemporaneous reports reveal conditions that the observed trajectory has not yet fully expressed. Text is useful in this role when it is correctly paired in time, relevant to the current numerical dynamics, and prevented from overwhelming the numerical evidence. We formulate text-aware forecasting as forecast revision: numerical history establishes a reference from which the horizon is predicted, and paired text proposes a limited update to that reference. AnchorScope distils document histories into semantic candidates, routes them to ordered adaptive frequency bands, and calibrates each revision with bounded amplitude, phase, and residual components. Under a row-level causal availability rule for Climate, aligned text outperforms no text, shuffled text, and random text. Controlled retraining holds encoders, text modulators, budgets, and seeds fixed: anchored adaptive bands outperform fixed, anchor-free, and global allocation rules on four datasets, and configured bounds outperform a threefold relaxation. Broad comparisons on nine Time-MMD datasets show that these revisions retain competitive forecasting performance. AnchorScope learns where a textual revision belongs in the numerical spectrum and how large that revision should be.

1 INTRODUCTION

Time series forecasting supports decisions in agriculture, energy, public health, transportation, and other settings in which numerical history is accompanied by contextual records. Agricultural reports describe supply conditions, policy narratives accompany economic indicators, and incident reports provide context for traffic or security measurements. Such records can reveal changes before their aggregate effect appears in the observed trajectory. Time-MMD formalises this setting through paired numerical and textual streams from nine domains Liu et al. (2024a). Context Is Key further studies settings in which text provides essential forecasting information Ashok et al. (2024). These settings call for a forecast revision: numerical history supplies the initial prediction, while contextual records can update it with information about conditions surrounding the observation window.

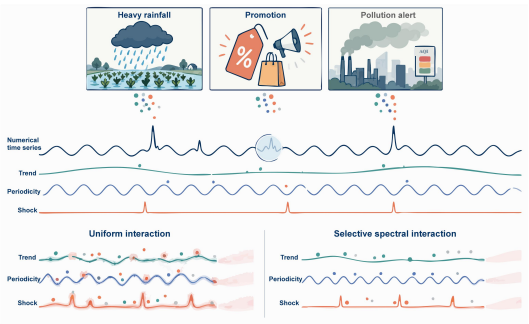


Figure 1: Semantic evidence can support forecast revisions at different temporal scales. AnchorScope locates the revision in an adaptive spectral range and controls its size.

The revision is difficult because contextual evidence has uneven relevance. A crop report may describe a persistent supply change, a promotion may accompany recurring demand, and a disruption may coincide with a brief deviation. Their numerical signatures depend on the target variable, observation interval, and forecast horizon. A record that informs a slowly evolving component may offer little guidance about rapid variation in the same series. Repetition and weak association with the target further complicate the use of text. Temporal pairing provides context, but the forecaster must still determine which numerical structures the text should revise.

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.

This motivates revision at the level of temporal structure. Figure 1 illustrates the issue: one semantic update can alter several numerical components even when its useful contribution is concentrated in a smaller range. Frequency provides an ordered coordinate system for locating the revision. Lower frequencies represent slower variation, while higher frequencies represent rapid changes. The forecasting problem therefore has two coupled questions: where should paired text revise the numerical state, and how large should that revision be?

Numerical forecasters make temporal structure explicit through efficient attention in Informer Zhou et al. (2021), decomposition in Autoformer Wu et al. (2021), and frequency modeling in FEDformer Zhou et al. (2022b). Other architectures reorganize temporal information through representations in two dimensions in TimesNet Wu et al. (2023), patch tokens in PatchTST Nie et al. (2023), variable tokens in iTransformer Liu et al. (2024b), or spectral interpolation in FITS Xu et al. (2024). Multimodal forecasters introduce language representations through GPT4TS Zhou et al. (2023), reprogramming through Time-LLM Jin et al. (2024), and multimodal prompting through GPT4MTS Jia et al. (2024). TimeCMA Liu et al. (2025a) and CALF Liu et al. (2025b) study alignment between modalities, while TaTS Li et al. (2026) organizes paired documents over time. These approaches motivate the use of contextual signals. AnchorScope asks where text should revise numerical dynamics and how large that revision should be.

AnchorScope treats the numerical spectrum as the reference for a forecast revision. The numerical branch encodes the observed spectrum, while the text branch distils the document history into a compact set of semantic candidates. Ordered adaptive soft bands locate the spectral range in which a candidate can revise the reference. A gate and bounded amplitude, phase, and residual coefficients control the size of that revision before decoding the requested horizon. This gives the revision a location and a magnitude that can be tested separately. Figure 2 presents the architecture.

The evaluation tests each part of this account. Broad comparisons on nine domains measure forecasting performance. The Climate study reconstructs each historical text field from documents available by its observation end date, then tests aligned, removed, shuffled, and random text. Controlled retraining holds encoders, modulators, budgets, and seeds fixed while changing the allocation rule or the correction bounds. The experiments test whether text can support a revision, where the revision should enter the spectrum, and how strongly it should change the numerical reference.

Our contributions are threefold:

- We formulate text-aware forecasting as forecast revision. Numerical history provides a reference for the prediction, and paired text supplies an update with an explicit spectral location and size.
- We develop AnchorScope, which locates text-derived revisions with ordered adaptive bands and controls their amplitude, phase, and residual components with bounded calibration.
- We evaluate the full mechanism through broad forecasting comparisons, a causal Climate text study, and controlled retraining experiments for revision location and size.

2 Related Work

Numerical time-series forecasting. Forecasting architectures encode historical dynamics through several complementary structures. Informer uses sparse attention for long contexts Zhou et al. (2021); Autoformer separates trend and seasonal components Wu et al. (2021); and FEDformer models temporal structure in frequency coordinates Zhou et al. (2022b). TimesNet represents periods in two dimensions Wu et al. (2023), PatchTST forms local temporal tokens Nie et al. (2023), and iTransformer treats variables as tokens Liu et al. (2024b). Crossformer captures cross-variable dependencies Zhang & Yan (2023), while ModernTCN uses temporal convolutions Luo & Wang (2024). Other approaches use hierarchical interpolation Challu et al. (2023), sample interaction Liu et al. (2022), all-MLP mixing Chen et al. (2023), exponential smoothing Woo et al. (2022), simple linear extrapolation Zeng et al. (2023), multi-scale mixing Wang et al. (2024a), adaptive paths Chen et al. (2024a), or spectral interpolation Xu et al. (2024). These models provide strong numerical forecasts from observed history. AnchorScope uses a numerical spectrum from this setting as the reference that contextual evidence may revise.

Foundation models and external context. Time-series foundation models learn transferable temporal representations from large and diverse corpora, including Chronos Ansari et al. (2024), TimesFM Das et al. (2024), and Moirai Woo et al. (2024). Their central challenge is broad transfer across numerical series. A related line incorporates external variables into a forecaster, as in TimeXer Wang et al. (2024b). Paired textual records create a different input: they describe selected conditions surrounding an observation and can vary in temporal relevance. The present work addresses how such records should modify a forecast reference already formed from the numerical history.

Text-aware and multimodal forecasting. Time-MMD provides paired numerical and textual streams across real-world domains Liu et al. (2024a), and Context Is Key studies settings where language supplies essential forecasting information Ashok et al. (2024). GPT4TS and Time-LLM adapt language-model representations to time-series prediction Zhou et al. (2023); Jin et al. (2024); GPT4MTS introduces multimodal prompting Jia et al. (2024). TimeCMA and CALF study cross-modal alignment Liu et al. (2025a;b), while Time-VLM, TaTS, and Aurora organize richer multimodal context over time Zhong et al. (2025); Li et al. (2026); Wu et al. (2026). These studies establish the value of semantic context. AnchorScope focuses on the numerical effect of that context by treating it as a forecast revision with an explicit spectral location and bounded size.

Spectral revisions from text. Frequency provides an ordered coordinate system for relating document evidence to slow and rapid numerical variation. Existing forecasters use frequency-aware representations to preserve long-range structure Zhou et al. (2022a) or jointly model temporal and frequency-domain features Chen et al. (2024b). AnchorScope addresses a distinct question: how paired text should revise a numerical spectral reference. It learns soft bands whose frequency locations adapt to the current numerical spectrum, then bounds the amplitude, phase, and residual terms that revise each band. The controlled studies in Section 4 evaluate the location and size of these revisions directly.

3 ANCHORSCOPE: SPECTRAL FORECAST REVISION

AnchorScope treats text as a revision to a forecast built from numerical history. Given a numerical history $\mathbf{X}_{1:T}$ and documents $\mathcal{D}_{1:T}$ available at those observation times, it predicts $\widehat{\mathbf{Y}}_{T+1:T+H}$. The numerical path establishes a spectral reference for the forecast. The text path identifies the spectral range that a document history can revise and controls the size of the resulting update. Ordered anchors locate the range, and bounded calibration limits the update. The architecture has four stages: it constructs a numerical reference, summarizes temporally aligned documents, assigns semantic evidence to a learned spectral range, and applies the revision before decoding the horizon. Figure 2 gives an overview.

3.1 FORECAST REVISION SETUP

The input is a C-variate history $\mathbf{X}_{1:T} \in \mathbb{R}^{T\times C}$ and the accompanying document sequence $\mathcal{D}_{1:T} = \{d_1, \ldots, d_T\}$. The forecast horizon is H. Each d_t is assembled only from documents released by the end of observation interval t, so the text branch has the same information boundary as the numerical branch. This pairing matters because a report that becomes available after the observation interval cannot explain a decision made at that interval.

This construction gives the textual stream a precise role. At time t, the model receives a record of information that could have informed a forecast after observing $\mathbf{X}_{1:t}$; it never receives documents associated with target steps $t+1$ through $t+H$. A window can contain repeated reporting, slowly changing background information, and isolated developments. The text encoder therefore retains their temporal order and distils the full document history into a small set of candidate summaries. The candidates are not manually labeled event types. They are learned representations that allow the model to preserve several potentially relevant pieces of context before deciding which one belongs with a numerical structure.

AnchorScope starts from the premise that numerical history provides the base prediction. Text supplies evidence about changes that the history alone may represent incompletely, such as a localized event, policy change, or contextual shift. The model first locates the part of the current spectrum that

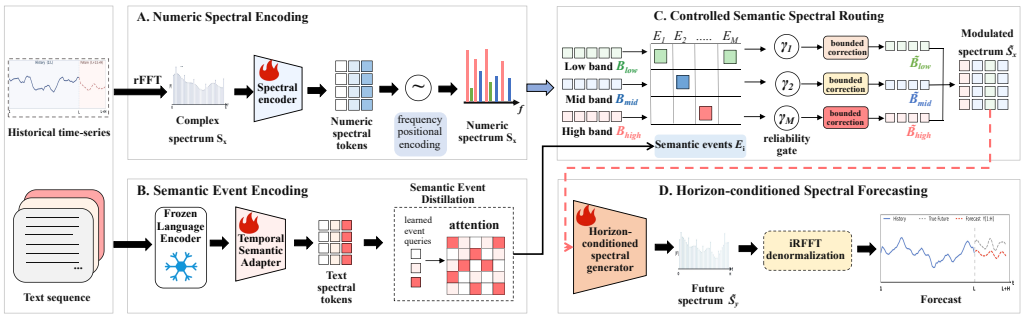


Figure 2: Overview of AnchorScope. The numerical encoder forms the spectral reference $\mathbf{S}_x$ from the raw spectrum $\mathbf{S}_x^0$. The text encoder distils M semantic candidates $\{\mathbf{E}_m\}_{m=1}^M$. Ordered soft bands pool the reference into K tokens $\{\mathbf{B}_k\}_{k=1}^K$ and determine where each candidate can revise it. The gate γ_k and bounded modulation set the size of the revision, producing $\widetilde{\mathbf{B}}_k$. The revised bands are distributed over the original grid to obtain $\widetilde{\mathbf{S}}_x$, which is decoded into the future spectrum $\widetilde{\mathbf{S}}_y$ and forecast. The three illustrated bands are examples; K and M are separate hyperparameters.

is relevant to a semantic signal, then determines the size of the revision. These decisions correspond to spectral allocation and bounded calibration.

3.2 Numerical Reference and Textual Evidence

The numerical branch normalizes each channel using statistics from the observed window, applies a real Fourier transform, and uses a learned complex projection to obtain the reference spectrum:

$$\mathbf{S}_x = E_{\text{num}}(\mathcal{F}_r(\text{Norm}(\mathbf{X}_{1:T}))) \in \mathbb{C}^{F\times d}, \tag{1}$$

where F is the number of nonredundant real-FFT bins and d is the latent width. $\mathbf{S}_x$ retains the numerical structure that the decoder will extrapolate. It is the reference that text may revise.

For the text branch, a frozen language encoder maps each causally available document collection to a document representation. A trainable temporal adapter processes their order and distils the history into M semantic candidates:

$$\{\mathbf{E}_m\}_{m=1}^M = \Phi_{\text{text}}(\mathcal{D}_{1:T}), \qquad \mathbf{E}_m \in \mathbb{R}^d. \tag{2}$$

Using a small candidate set summarizes the document history before it interacts with the spectrum. The adapter and all fusion layers are trained with the forecasting task; the language encoder remains fixed. The resulting semantic representation depends on temporal context while the numerical reference remains available to the decoder.

The two branches meet only after each has formed a compact representation of the observed window. This separation is useful when text is weakly related to the target: the numerical encoder can still form a reference from the series itself, and the text branch can be assigned little influence by the subsequent routing and gate. When a document history contains complementary evidence, its candidates provide an additional source of information without replacing the numerical path. The causal text experiments in Section 4 test this distinction by preserving the model and changing only the relationship between a window and its text.

3.3 Locating Textual Revisions in the Spectrum

The model next turns the reference spectrum into K soft band tokens. Every frequency bin f has normalized position r_f and descriptor $\mathbf{u}_f$, formed from the real part, imaginary part, and magnitude of $\mathbf{S}_{x,f}$. Ordered anchors $c_1 < \cdots < c_K$ give each band a distinct spectral neighborhood. The locality weight ρ and temperature τ set the strength and softness of this prior, while a learned score shifts the allocation when the observed numerical state supports it:

$$W_{k,f} = \text{softmax}_k\left(\frac{g_{\text{band}}(\mathbf{u}_f)_k - \rho(r_f - c_k)^2}{\tau}\right). \tag{3}$$

The soft assignments pool $\mathbf{S}_x$ into $\{\mathbf{B}_k\}_{k=1}^K$. Each band token queries the semantic candidates and receives a context vector for that spectral region. The same text history can therefore revise different temporal scales in different ways. The anchors encode an ordered frequency prior, while the learned logits adapt the revision range to the observed spectrum.

Every bin is assigned across all bands, so the soft decomposition retains coverage of the full numerical spectrum while allowing neighboring bands to overlap. The anchor term gives the bands stable low-to-high identities across windows. The learned score can broaden, narrow, or shift their effective support when the spectrum calls for it. This is the intended division of labor: anchors provide a common coordinate system for interpreting a revision, and the data-dependent scores decide how that coordinate system should be used for the current forecasting state. The allocation ablation compares this mechanism with fixed equal-width bands, anchor-free bands, and one global correction.

3.4 Controlling Revision Size and Decoding the Forecast

For each band, a modulator receives the numerical token and its semantic context. It proposes an amplitude update, a phase update, and a residual vector, while a learned gate scales their contribution. The proposals are explicitly bounded:

$$|\Delta a_k| \le a_{\max}, \qquad |\Delta \varphi_k| \le \varphi_{\max}, \qquad 0 \le \beta_k \le \beta_{\max}. \tag{4}$$

Amplitude and phase modify the existing spectral coefficient; the residual gives the model limited capacity to add content that is not expressed as a simple rotation or rescaling. The bounds limit the size of a textual revision, so the numerical reference remains dominant when semantic evidence is weak or poorly matched to the current state.

More specifically, the band token attends over the semantic candidates to obtain a routed context $\mathbf{R}_k$, and a gate $\gamma_k \in [0, 1]$ is computed from $[\mathrm{Re}(\mathbf{B}_k), \mathrm{Im}(\mathbf{B}_k), \mathbf{R}_k]$. The gate scales each proposed amplitude, phase, and residual update before it is applied to $\mathbf{B}_k$. The amplitude term rescales the existing coefficient, the phase term rotates it in the complex plane, and the residual supplies a bounded additive component. Thus the model can make a small revision in a selected band without treating text as an unrestricted replacement for the history-derived spectrum. The fixed-gate sweep and bound study evaluate these two controls in Section 4.

The corrected band tokens are distributed back to the original frequency grid:

$$\widetilde{\mathbf{S}}_{x,f} = \mathbf{S}_{x,f} + \sum_{k=1}^{K} W_{k,f} \left(\widetilde{\mathbf{B}}_k - \mathbf{B}_k \right). \tag{5}$$

The horizon-conditioned decoder maps $\widetilde{\mathbf{S}}_x$ to a future spectrum, applies the inverse real Fourier transform, and restores the window statistics to produce $\widehat{\mathbf{Y}}_{T+1:T+H}$. We minimize mean squared error over all forecast steps and channels. Equation 3 determines the location of a revision, and Equation 4 controls its size. Section 4 tests both choices directly. Appendix A specifies the tensor operations, decoder, training objective, and complete inference procedure.

The decoder receives the revised history representation rather than a text-only representation. It maps the F bins induced by the observation window to the F_H bins required by the requested horizon, then applies an inverse real transform with explicit output length H. This makes the target horizon part of the generation problem: a revision that is useful for a short forecast need not be propagated in the same way for a longer forecast. The same location and calibration mechanism is used at every horizon, while the decoder learns the appropriate extrapolation from the revised numerical state. Detailed tensor shapes, complex projections, and the horizon-conditioned generator are given in Appendix A.

4 Experiments

4.1 Experimental Setup

Datasets and forecasting protocol. We evaluate AnchorScope on the nine Time-MMD datasets with paired text: Agriculture, Climate, Economy, Energy, Environment, Health, Security, SocialGood,

and Traffic. They cover agricultural production, climate, economics, energy, environmental monitoring, health, public security, social welfare, and traffic. Each dataset pairs textual context with numerical histories at a different sampling frequency. A forecasting window consumes only the `fact` fields attached to its historical rows; text from target rows is not provided to the model. Agriculture, Climate, Economy, Security, SocialGood, and Traffic use an input length of 12 and prediction horizons $\{6, 8, 10, 12\}$. Energy and Health use an input length of 24 and horizons $\{12, 24, 36, 48\}$, whereas Environment uses an input length of 96 and horizons $\{48, 96, 192, 336\}$. All methods use the same window configuration within each dataset. MSE and MAE are averaged across the four horizons, and complete results for every horizon are reported in the appendix.

Causal text availability for Climate. For the Climate text-condition study, we reconstruct each `fact` field from raw documents for its observation row. A document d is eligible for row i only if $\texttt{available_date}(d) \leq \texttt{end_date}_i$; the field concatenates its eligible documents. Hence, a length-12 history from rows $t-11$ through t contains twelve text fields that were each available by the end of their corresponding observation interval. Forecasting begins after $\texttt{end_date}_t$, and neither future observation rows nor their documents are visible. This rule excludes a later-published report even when it describes an earlier climate interval. Document reassignment in the shuffled condition also follows the eligibility rule for its recipient row.

The nine-domain comparisons use the official Time-MMD forecasting protocol. The separately reconstructed Climate study tests real-time document availability at the row level. Each dataset is evaluated at four prediction horizons; we report their arithmetic mean within the original error scale and provide individual results in the appendix.

Table 1: Forecasting performance across prediction horizons. Results are averaged across horizons for each dataset. Lower MSE and MAE indicate better performance. Best results are in bold and second best results are underlined.

Dataset	AnchorScope (ours)		Aurora		CALF		TimeCMA		GPT4TS		GPT4MTS		Time-LLM		Time-VLM	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	<u>0.230</u>	<u>0.307</u>	0.384	0.391	0.256	0.322	**0.114**	**0.230**	0.242	0.310	0.256	0.318	0.397	0.405	0.232	0.308
Climate	**0.935**	**0.778**	1.153	0.880	1.148	0.871	1.237	0.903	1.097	0.856	1.082	0.850	1.147	0.874	<u>1.078</u>	<u>0.849</u>
Economy	**0.010**	**0.077**	0.032	0.139	<u>0.015</u>	<u>0.098</u>	0.031	0.144	0.020	0.114	0.023	0.123	0.030	0.140	0.018	0.107
Energy	**0.242**	**0.349**	0.321	0.426	0.259	0.371	0.351	0.459	0.264	0.377	0.268	0.385	0.328	0.442	<u>0.243</u>	<u>0.352</u>
Environment	**0.408**	**0.479**	0.604	0.567	0.540	0.516	0.579	0.547	<u>0.506</u>	0.513	0.511	0.530	0.559	0.534	0.513	<u>0.512</u>
Health	**1.360**	**0.787**	2.516	1.154	<u>1.571</u>	<u>0.841</u>	2.584	1.141	1.816	0.916	1.737	0.901	2.602	1.162	1.653	0.868
Security	<u>70.364</u>	**3.757**	82.083	4.444	73.150	3.995	77.150	4.261	72.587	4.128	73.907	4.281	**64.969**	4.018	71.303	<u>3.983</u>
SocialGood	**0.919**	**0.414**	1.027	0.539	1.019	<u>0.433</u>	1.134	0.591	<u>0.921</u>	0.461	1.044	0.445	1.078	0.585	0.967	0.467
Traffic	**0.179**	**0.233**	0.278	0.418	0.211	<u>0.258</u>	0.277	0.389	0.198	0.262	0.206	0.266	0.263	0.390	<u>0.196</u>	0.279

Comparison methods and metrics. Table 1 compares AnchorScope with Aurora Wu et al. (2026), CALF Liu et al. (2025b), TimeCMA Liu et al. (2025a), GPT4TS Zhou et al. (2023), GPT4MTS Jia et al. (2024), Time-LLM Jin et al. (2024), and Time-VLM Zhong et al. (2025). Appendix Table 2 reports common-protocol comparisons across numerical backbones and text-fusion plugins, including TaTS Li et al. (2026) and MM-TSFlib implementations released with Time-MMD Liu et al. (2024a). We report average MSE and MAE over four horizons; lower values are better. The language encoder is frozen and the remaining AnchorScope modules are trained jointly.

4.2 Do Textual Revisions Improve Forecasting?

Table 1 evaluates the revised forecasts across nine domains. AnchorScope has the lowest average MSE on seven datasets and the lowest MAE on eight. Its MSE is 1.360 on Health, 0.408 on Environment, and 0.179 on Traffic, compared with the next best values of 1.571, 0.506, and 0.196. The result varies by domain: TimeCMA leads on Agriculture, and Time-LLM has lower Security MSE while AnchorScope has lower Security MAE. Table 2 adds common-protocol comparisons with numerical backbones and text-fusion plugins. AnchorScope leads on both metrics for Climate, Health, Security, SocialGood, and Traffic.

4.3 Why Does a Textual Revision Need Structure?

Table 3 compares numerical forecasting, global text fusion, and the components used by AnchorScope. Relative to numerical forecasting, AnchorScope reduces MSE from 0.2795 to 0.2308

Table 2: Comparison of AnchorScope with numerical backbones and text fusion plugins under a common Time-MMD protocol. All methods use the same dataset version, chronological split, history window, prediction horizons, and evaluation pipeline, while retaining their respective backbone and fusion design. Results are averaged across four prediction horizons for each dataset. Lower MSE and MAE indicate better performance. Best results are in bold and second best results are underlined.

Dataset	AnchorScope (ours)		Numeric-iTransformer		Numeric-PatchTST		TaTS-iTransformer		TaTS-PatchTST		MM-TSFlib-iTransformer		MM-TSFlib-PatchTST	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	0.230	0.307	**0.198**	**0.301**	0.223	0.309	0.293	0.369	0.282	0.356	0.199	0.302	0.210	0.316
Climate	**0.935**	**0.778**	1.130	0.866	1.164	0.879	0.981	0.800	0.995	0.809	1.068	0.845	1.136	0.868
Economy	0.010	**0.077**	0.013	0.090	0.015	0.098	**0.008**	0.078	0.008	0.077	0.015	0.097	0.019	0.112
Energy	0.242	**0.349**	0.237	0.351	0.260	0.370	0.297	0.416	0.273	0.396	**0.236**	0.357	0.248	0.368
Environment	0.408	0.479	0.432	0.491	0.539	0.536	**0.263**	**0.375**	0.267	0.378	0.416	0.478	0.475	0.506
Health	**1.360**	**0.787**	2.343	0.909	1.748	0.874	1.407	0.814	1.395	0.800	1.971	0.850	1.669	0.854
Security	**70.364**	**3.757**	78.477	4.558	78.037	4.510	75.295	4.353	76.288	4.410	80.155	4.634	77.509	4.496
SocialGood	**0.919**	**0.414**	1.304	0.505	1.234	0.486	0.961	0.430	0.973	0.448	1.176	0.473	1.218	0.503
Traffic	**0.179**	**0.233**	0.203	0.239	0.187	0.235	0.191	0.235	0.181	0.235	0.197	0.236	0.191	0.237

Table 3: Ablation of components used to form textual forecast revisions. Results are reported as MSE and MAE on six representative datasets. Lower values indicate better performance. Best results are in bold and second best results are underlined.

Variant	Agriculture		Climate		Energy		Environment		Health		Security	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Numerical only	0.2795	0.3358	0.9440	0.7777	0.2533	0.3639	0.4219	0.4823	1.4219	0.8032	70.5405	3.7567
Global text fusion	0.3133	0.3663	0.9494	0.7840	0.2670	0.3794	0.4193	0.4819	1.8669	0.9277	77.9474	4.3900
w/o Temporal Semantic Adapter	0.2641	0.3379	0.9395	0.7773	0.2441	0.3547	0.4116	0.4803	1.7275	0.8518	70.5277	3.7592
w/o Semantic Frequency Modulation	0.2823	0.3374	0.9362	0.7778	0.2533	0.3638	0.4249	0.4841	1.4227	0.8036	70.5119	3.7576
AnchorScope	**0.2308**	**0.3074**	**0.9350**	**0.7769**	**0.2423**	**0.3493**	**0.4078**	**0.4791**	**1.3601**	**0.7867**	**70.3638**	**3.7566**

on Agriculture and from 1.4219 to 1.3601 on Health. Global text fusion increases Health MSE to 1.8669 and Security MSE to 77.9474. These results motivate representing text as a structured revision of the numerical reference. Appendix Table 6 reports record-level diagnostics that contextualize these domain differences.

4.4 Does Causally Available Text Support a Revision?

Table 4 tests whether the Climate revision is supported by the paired text under the row-level availability rule. Removing text increases MSE from 0.935 to 0.962. Shuffling the causally available text fields increases it to 0.978, and replacing them with random text increases it to 0.991. This ordered degradation attributes the gain to temporal pairing and textual content.

4.5 Where and How Large Should a Revision Be?

Table 5 tests the location and size of a revision through controlled retraining. With the same encoders, text modulator, training budget, and random seed, the anchored adaptive-band variant has the lowest MSE and MAE on Agriculture, Health, Security, and Traffic when compared with fixed equal-width bands, anchor-free soft bands, and a matched global revision. The variants retain the same encoders and text-modulation capacity; they differ only in how a textual revision is assigned to the numerical spectrum. The bounded configuration is also lower than the threefold relaxation on all four datasets. The table therefore supports both parts of the revision: it should enter an appropriate frequency range and remain limited in size.

4.6 Sensitivity of Learned Revisions

Figure 3(a) evaluates the learned gate by replacing it with fixed interaction strengths on four representative datasets. The curves have different slopes and preferred strengths across Agriculture, Health, Energy, and Security. A single global interaction strength therefore cannot represent the contribution of paired text in every forecasting setting; the learned gate adapts the size of a textual revision to the current numerical and semantic context.

Table 4: Causal Climate text study. Every document assigned to historical row i, including under shuffling, satisfies $\texttt{available_date} \leq \texttt{end_date}_i$. All conditions use the same causal chronological split, architecture, training budget, and random seed; the historical text condition is the only change.

Text condition	MSE	MAE	ΔMSE vs. aligned
Aligned text	**0.935**	**0.778**	**0.0%**
No text	0.962	0.790	$+2.9\%$
Shuffled text	0.978	0.799	$+4.6\%$
Random text	0.991	0.807	$+6.0\%$

Table 5: Controlled retraining for the location and size of textual revisions. All variants use identical encoders, text modulators, training budgets, random seeds, and Time-MMD evaluation; the named component is the only change. Values are MSE/MAE. The final AnchorScope row is the reference for both panels and uses $(a_{\max}, \varphi_{\max}, \beta_{\max}) = (0.25, 0.20, 0.12)$; loose bounds use three times these values.

Variant	Agriculture	Health	Security	Traffic
(a) Location of textual revision				
Fixed equal-width bands	0.241 / 0.315	1.397 / 0.801	71.024 / 3.793	0.182 / 0.235
Anchor-free soft bands	0.247 / 0.319	1.412 / 0.807	71.286 / 3.811	0.184 / 0.237
Matched global revision	0.267 / 0.337	1.523 / 0.837	72.104 / 3.900	0.189 / 0.243
(b) Size of textual revision				
Loose revision bounds ($3\times$)	0.241 / 0.315	1.415 / 0.803	73.741 / 3.844	0.187 / 0.239
AnchorScope: anchored, bounded revision	**0.230 / 0.307**	**1.360 / 0.787**	**70.364 / 3.757**	**0.179 / 0.233**

The sweep provides a direct view of the same mechanism tested by the controlled retraining studies. Appendix Tables 7 and 8 extend the analysis across all 36 dataset and horizon settings. Masking a high-band revision has the largest mean absolute relative MSE change (4.25%), and relaxing the revision bounds by $3\times$ raises average MSE by 4.56%. Together, these results show that the relevant revision strength depends on both the temporal scale and the forecasting setting.

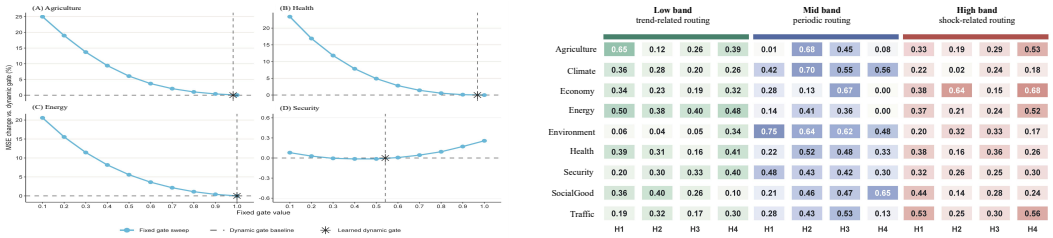
The intervention fixes only the strength of semantic interaction after the model has formed its numerical reference, semantic candidates, and band assignments. It therefore isolates the magnitude of a revision from the question of where that revision enters the spectrum. The heterogeneous responses in Figure 3(a) complement the controlled bounds comparison: the useful amount of contextual revision depends on the forecasting state rather than on a dataset-independent constant.

Figure 3(b) shows the fitted allocation preferences behind this variation. Agriculture gives its largest preference to the low band at the first horizon (0.65), the middle band at the second (0.68), and the high band at the fourth (0.53). Traffic favors the high band at the first and fourth horizons (0.53 and 0.56), whereas Environment gives its largest preference to the middle band at every horizon. The pattern links the controlled allocation study to the learned behavior of the final model: paired text is directed to different temporal scales as the dataset and forecast horizon change.

4.7 Case Study: Revisions Across Temporal Scales

Figure 4 illustrates the revision mechanism in four paired Agriculture windows. Panel (A) locates the records on the target trajectory. Panel (C) shows excerpts about broiler production, consumption, and export conditions. The windows describe related conditions at distinct points in the trajectory, so they need not support the same forecast revision.

Panel (B) compares the numerical amplitude spectrum with trajectories derived from the paired and shuffled text windows. The paired text has local peaks near marked numerical frequencies, and the shuffled sequence changes this pattern. The figure shows how records at different points in the trajectory can support revisions at different temporal scales. Table 5 tests the functional contribution of locating and bounding those revisions.



(a) Fixed-gate sensitivity **(b)** Learned band preferences

Figure 3: Revision magnitude and location. (a) Relative MSE change when semantic interaction strength is fixed on Agriculture, Health, Energy, and Security. The responses differ across datasets, motivating a learned revision size. (b) Learned preferences for low, middle, and high bands across datasets and horizons. H1 to H4 denote ascending prediction lengths; scale annotations provide illustrative temporal interpretations.

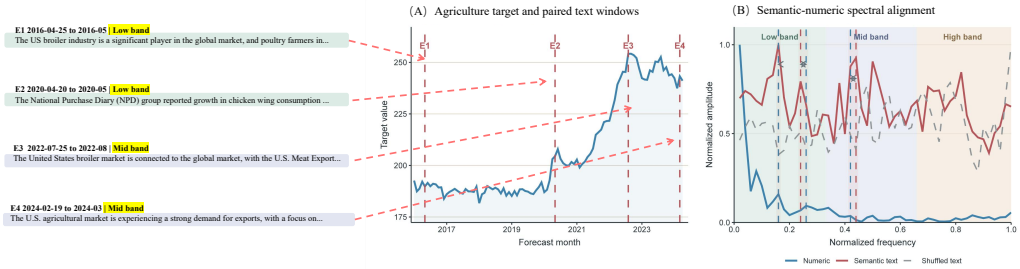


Figure 4: Paired textual context and spectral structure in Agriculture. (A) Four selected text windows on the numerical target trajectory. (B) Numerical and semantic text spectra, with shuffled text as a comparison and low, middle, and high frequency ranges indicated in the background. (C) Text excerpts associated with the selected windows. The panels illustrate the records and spectral correspondence used to analyze forecast revisions across frequency ranges.

The selected windows make the allocation result concrete. The first two records are associated with the low-band portion of the spectrum, whereas the later records are associated with the middle band. Their excerpts describe different market conditions, and their positions on the trajectory show why a common text update would discard useful temporal structure. The paired spectrum links each collection of records to the numerical context in which it was observed; shuffling breaks that correspondence. This qualitative view connects the causal text study with the allocation ablation: temporal pairing supplies meaningful evidence, and anchored bands determine the scale at which that evidence can revise the forecast.

5 Conclusion

AnchorScope frames text-aware forecasting as forecast revision. Numerical history establishes a reference for the horizon, and paired text revises that reference when it contains useful contextual evidence. Ordered anchors locate the revision in the spectrum, while bounded calibration controls its size.

The experiments evaluate this account from three directions. In the causal Climate study, aligned text outperforms removed, shuffled, and random text. In controlled retraining, anchored allocation and bounded calibration outperform their tested alternatives under fixed encoders, modulators, budgets, and seeds. Broad results on nine Time-MMD domains show competitive forecasting performance. Together, the results support a simple view of text-aware forecasting: text is most useful when it revises the numerical forecast at an appropriate temporal scale and with an appropriate magnitude.

AI USE STATEMENT

Generative AI tools, including GPT-5.6, assisted with language editing and manuscript organization. The authors critically reviewed and revised all AI-assisted material and remain fully responsible for the manuscript's content, claims, and accompanying artifacts.

REFERENCES

A. F. Ansari, L. Stella, C. Turkmen, X. Zhang, P. Mercado, H. Shen, O. Shchur, S. S. Rangapuram, S. P. Arango, and S. Kapoor. Chronos: Learning the language of time series, 2024. arXiv:2403.07815, 2024.

A. Ashok, A. R. Williams, E. Marcotte, V. Zantedeschi, J. Subramanian, R. Riachi, J. Requeima, A. Lacoste, I. Rish, and N. Chapados. Context is key: A benchmark for forecasting with essential textual information, 2024. in: *NeurIPS Workshop on Time Series in the Age of Large Models*, 2024.

C. Challu, K. G. Olivares, B. N. Oreshkin, F. G. Ramirez, M. M. Canseco, and A. Dubrawski. N-hits: Neural hierarchical interpolation for time series forecasting, 2023. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (2023) 6989–6997.

P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang, and C. Guo. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting, 2024a. in: *International Conference on Learning Representations*, 2024.

S.-A. Chen, C.-L. Li, N. Yoder, S. O. Arik, and T. Pfister. Tsmixer: An all-mlp architecture for time series forecasting, 2023. *Transactions on Machine Learning Research* (2023).

Y. Chen, S. Liu, J. Yang, H. Jing, W. Zhao, and G. Yang. A joint time-frequency domain transformer for multivariate time series forecasting, 2024b. *Neural Networks*, 176 (2024) 106334.

A. Das, W. Kong, R. Sen, and Y. Zhou. A decoder-only foundation model for time-series forecasting, 2024. in: *Proceedings of the 41st International Conference on Machine Learning*, 2024.

F. Jia, K. Wang, Y. Zheng, D. Cao, and Y. Liu. Gpt4mts: Prompt-based large language model for multimodal time-series forecasting, 2024. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 38 (2024) 23343–23351.

M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen. Time-llm: Time series forecasting by reprogramming large language models, 2024. in: *International Conference on Learning Representations*, 2024.

Z. Li, X. Lin, Z. Liu, J. Zou, Z. Wu, L. Zheng, D. Fu, Y. Zhu, H. Hamann, H. Tong, and J. He. Language in the flow of time: Time-series-paired texts weaved into a unified temporal narrative. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=f8068e220f73f110fe07c11de5a1d975ce6ecb86.

C. Liu, Q. Xu, H. Miao, S. Yang, L. Zhang, C. Long, Z. Li, and R. Zhao. Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment, 2025a. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 39 (2025) 18780–18788.

H. Liu, S. Xu, Z. Zhao, L. Kong, H. Kamarthi, A. B. Sasanur, M. Sharma, J. Cui, Q. Wen, C. Zhang, and B. A. Prakash. Time-mmd: Multi-domain multimodal dataset for time series analysis. In *Advances in Neural Information Processing Systems*, volume 37, pp. 77888–77933, 2024a. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/8e7768122f3eeec6d77cd2b424b72413-Abstract-Datasets_and_Benchmarks_Track.html.

M. Liu, A. Zeng, M. Chen, Z. Xu, Q. Lai, L. Ma, and Q. Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction, 2022. in: *Advances in Neural Information Processing Systems*, 35 (2022) 5816–5828.

P. Liu, H. Guo, T. Dai, N. Li, J. Bao, X. Ren, Y. Jiang, and S.-T. Xia. Calf: Aligning llms for time series forecasting via cross-modal fine-tuning, 2025b. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 39 (2025) 18915–18923.

Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long. itransformer: Inverted transformers are effective for time series forecasting, 2024b. in: *International Conference on Learning Representations*, 2024.

D. Luo and X. Wang. Moderntcn: A modern pure convolution structure for general time series analysis, 2024. in: *International Conference on Learning Representations*, 2024.

Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers, 2023. in: *International Conference on Learning Representations*, 2023.

S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Zhang, and J. Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting, 2024a. in: *International Conference on Learning Representations*, 2024.

Y. Wang, H. Wu, J. Dong, G. Qin, H. Zhang, Y. Liu, Y. Qiu, J. Wang, and M. Long. Timexer: Empowering transformers for time series forecasting with exogenous variables, 2024b. in: *Advances in Neural Information Processing Systems*, 37 (2024) 469–498.

G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi. Etsformer: Exponential smoothing transformers for time-series forecasting, 2022. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 36 (2022) 11105–11112.

G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo. Unified training of universal time series forecasting transformers, 2024. in: *Proceedings of the 41st International Conference on Machine Learning*, 2024.

H. Wu, J. Xu, J. Wang, and M. Long. Autoformer: Decomposition transformers with autocorrelation for long-term series forecasting, 2021. in: *Advances in Neural Information Processing Systems*, 34 (2021) 22419–22430.

H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long. Timesnet: Temporal 2d-variation modeling for general time series analysis, 2023. in: *International Conference on Learning Representations*, 2023.

X. Wu, J. Jin, W. Qiu, P. Chen, Y. Shu, B. Yang, and C. Guo. Aurora: Towards universal generative multimodal time series forecasting, 2026. in: *International Conference on Learning Representations*, 2026.

Z. Xu, A. Zeng, and Q. Xu. Fits: Modeling time series with 10k parameters, 2024. in: *International Conference on Learning Representations*, 2024.

A. Zeng, M. Chen, L. Zhang, and Q. Xu. Are transformers effective for time series forecasting?, 2023. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 37 (2023) 11121–11128.

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

S. Zhong, W. Ruan, M. Jin, H. Li, Q. Wen, and Y. Liang. Time-vlm: Exploring multimodal vision-language models for augmented time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 78478–78497, 2025. URL https://proceedings.mlr.press/v267/zhong25a.html.

H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021. in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 35 (2021) 11106–11115.

T. Zhou, Z. Ma, Q. Wen, L. Sun, T. Yao, W. Yin, and R. Jin. Film: Frequency improved legendre memory model for long-term time series forecasting, 2022a. in: *Advances in Neural Information Processing Systems*, 35 (2022) 12677–12690.

T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting, 2022b. in: *Proceedings of the 39th International Conference on Machine Learning*, 2022, pp. 27268–27286.

T. Zhou, P. Niu, L. Sun, and R. Jin. One fits all: Power general time series analysis by pretrained lm, 2023. in: *Advances in Neural Information Processing Systems*, 36 (2023) 43322–43355.

Algorithm 1 AnchorScope forecasting with anchored spectral revision

Require: numerical history $\mathbf{X}_{1:T}$, causally available documents $\mathcal{D}_{1:T}$, horizon H, anchors $\{c_k\}_{k=1}^{K}$
Ensure: forecast $\widehat{\mathbf{Y}}_{T+1:T+H}$
$\mathbf{S}_x \leftarrow E_{\mathrm{num}}(\mathcal{F}_r(\mathrm{Norm}(\mathbf{X}_{1:T})))$
$\{\mathbf{E}_m\}_{m=1}^{M} \leftarrow \Phi_{\mathrm{text}}(\mathcal{D}_{1:T})$
for each frequency bin f **do**
    compute descriptor $\mathbf{u}_f$ and assignments $\{W_{k,f}\}_{k=1}^{K}$
end for
for each anchored band k **do**
    pool $\mathbf{B}_k$ from $\mathbf{S}_x$ using $\{W_{k,f}\}_{f=1}^{F}$
    retrieve semantic context from $\{\mathbf{E}_m\}_{m=1}^{M}$ using $\mathbf{B}_k$
    predict gated, bounded amplitude, phase, and residual updates
    construct the corrected band token $\widetilde{\mathbf{B}}_k$
end for
distribute $\{\widetilde{\mathbf{B}}_k - \mathbf{B}_k\}_{k=1}^{K}$ to obtain $\widetilde{\mathbf{S}}_x$
generate the horizon spectrum, apply inverse real FFT, and restore window statistics
return $\widehat{\mathbf{Y}}_{T+1:T+H}$

APPENDIX

A DETAILED ANCHORSCOPE SPECIFICATION

We consider a numerical history $\mathbf{X}_{1:T} \in \mathbb{R}^{T\times C}$ and its paired document sequence $\mathcal{D}_{1:T} = \{d_1, \ldots, d_T\}$, where d_t denotes the text associated with observation step t. The objective is to predict $\mathbf{Y}_{T+1:T+H} \in \mathbb{R}^{H\times C}$. AnchorScope constructs an encoded numerical reference from the observed window and learns textual revisions to selected spectral components.

Figure 2 presents four stages. Numerical encoding produces a spectral reference, and semantic encoding summarizes the document history. Controlled routing assigns semantic candidates to adaptive bands and regulates their revisions. Finally, a generator maps the revised representation to the future spectrum. Batch indices are suppressed until the training objective. The equations retain the channel dimension in the input and output projections and use width d for the latent representation.

A.1 PROBLEM SETUP AND NOTATION

Let B, C, T, and H denote batch size, the number of target channels, lookback length, and prediction horizon, respectively, and let d be the latent width. For $T \geq 2$, the historical real Fourier transform has $F = \lfloor T/2 \rfloor + 1$ nonredundant bins; a future sequence has $F_H = \lfloor H/2 \rfloor + 1$ bins. We write $\mathcal{F}_r$ for the real FFT and $\mathcal{F}_{r,H}^{-1}$ for its inverse with output length H. Real and imaginary parts are denoted by Re and Im, i is the imaginary unit, and $|\cdot|$ denotes elementwise complex magnitude. Square brackets denote feature concatenation unless an explicit stacking operation is specified.

The raw numerical spectrum is $\mathbf{S}_x^0 \in \mathbb{C}^{F\times C}$ and the encoded reference is $\mathbf{S}_x \in \mathbb{C}^{F\times d}$. The semantic spectrum $\mathbf{S}_t \in \mathbb{C}^{F\times d}$ uses the same number of bins; the subscript t in this symbol identifies the textual branch. AnchorScope pools $\mathbf{S}_x$ into K band tokens $\mathbf{B}_k \in \mathbb{C}^d$ and distils $\mathbf{S}_t$ into M semantic candidates $\mathbf{E}_m \in \mathbb{R}^d$. A tilde on a band or spectrum denotes its state after semantic correction, and $\widehat{\mathbf{Y}}$ denotes the forecast. The normalized history $\widetilde{\mathbf{X}}$ is defined separately below. The band count K and candidate count M are independent hyperparameters; the low, middle, and high rows in the overview illustrate their roles.

A.2 NUMERIC ENCODING IN THE FREQUENCY DOMAIN

Stage A converts the observed history into a representation in which temporal scales are explicit. AnchorScope first normalizes every channel using statistics computed only from the observed win-

dow:

$$\begin{aligned}\boldsymbol{\mu} &= \frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t,\\ \boldsymbol{\sigma} &= \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\mathbf{X}_t-\boldsymbol{\mu})^2},\\ \widetilde{\mathbf{X}}_{1:T} &= \frac{\mathbf{X}_{1:T}-\boldsymbol{\mu}}{\boldsymbol{\sigma}+\epsilon}.\end{aligned} \tag{6}$$

Here and below, $\epsilon > 0$ is a fixed numerical stabilizer. The normalized sequence is transformed by

$$\mathbf{S}_x^0 = \mathcal{F}_r(\widetilde{\mathbf{X}}_{1:T}) \in \mathbb{C}^{F\times C}. \tag{7}$$

The encoder projects the C input channels into width d. With learned matrices $\mathbf{W}_r, \mathbf{W}_i \in \mathbb{R}^{C\times d}$, complex multiplication gives

$$\begin{aligned}\mathbf{Z}_x = {}& \mathrm{Re}(\mathbf{S}_x^0)\mathbf{W}_r - \mathrm{Im}(\mathbf{S}_x^0)\mathbf{W}_i\\ &+ i\big(\mathrm{Re}(\mathbf{S}_x^0)\mathbf{W}_i + \mathrm{Im}(\mathbf{S}_x^0)\mathbf{W}_r\big).\end{aligned} \tag{8}$$

A learned frequency positional embedding $\mathbf{P} \in \mathbb{R}^{F\times d}$, whose fth row is $\mathbf{P}_f$, preserves the identity of each bin. The complex threshold operator Shrink_η, with threshold $\eta \geq 0$, is applied componentwise to suppress weak coefficients:

$$\mathbf{S}_x = \mathrm{Shrink}_\eta\left(\mathbf{Z}_x + \mathbf{P}\right). \tag{9}$$

The resulting $\mathbf{S}_x \in \mathbb{C}^{F\times d}$ is the numerical reference for the remaining stages. The text path does not replace this representation; Stage C decides whether, where, and how much of it should be revised.

A.3 Semantic Event Encoding

Stage B represents the document history at the resolution needed for routing. A frozen language encoder $\mathrm{Enc}_{\mathrm{lang}}$ processes each document. Pooling its token representations gives $\mathbf{E}_t^0 \in \mathbb{R}^{d_{\mathrm{lang}}}$, where d_{lang} is the language representation width:

$$\mathbf{E}_t^0 = \mathrm{Pool}\big(\mathrm{Enc}_{\mathrm{lang}}(d_t)\big). \tag{10}$$

Let $\mathrm{Proj} : \mathbb{R}^{d_{\mathrm{lang}}} \to \mathbb{R}^d$ be a trainable projection. The temporal semantic adapter A_{text} then models document order and maps the projected sequence directly to text spectral tokens:

$$\mathbf{S}_t' = A_{\mathrm{text}}\big(\mathrm{Proj}(\mathbf{E}_{1:T}^0)\big) \in \mathbb{C}^{F_t\times d}. \tag{11}$$

The adapter contains the trainable temporal and spectral projection layers shown in Stage B. It maps variation in the document sequence to complex semantic tokens. Its output has F_t bins, which need not equal the numerical count F. The operator Align_F pads or truncates the frequency axis:

$$\mathbf{S}_t = \mathrm{Shrink}_\eta(\mathrm{Align}_F(\mathbf{S}_t')) \in \mathbb{C}^{F\times d}. \tag{12}$$

Alignment gives the two branches compatible tensor dimensions for fusion. Correspondence between their content is learned in the subsequent routing stage, where semantic candidates are conditioned on numerical band queries.

A.4 Controlled Semantic Spectral Routing

Stage C is the information fusion core of AnchorScope. It first constructs numerical band tokens that define where semantic intervention may occur. It then compresses the text representation into latent semantic candidates and aggregates them for each band. Finally, it parameterizes a bounded spectral correction. This sequence separates the location of an interaction from the magnitude of the numerical change that follows.

Adaptive semantic scope through spectral decomposition. The numerical reference is partitioned into K soft, ordered bands. For each frequency bin $f \in \{1, \dots, F\}$, AnchorScope forms a descriptor from its real part, imaginary part, and magnitude:

$$\mathbf{u}_f = [\mathrm{Re}(\mathbf{S}_{x,f}), \mathrm{Im}(\mathbf{S}_{x,f}), |\mathbf{S}_{x,f}|] \in \mathbb{R}^{3d}. \tag{13}$$

For bin f, let $r_f = (f-1)/(F-1) \in [0,1]$ be its normalized coordinate, and let $0 \le c_1 < \cdots < c_K \le 1$ denote ordered band anchors. The multilayer perceptron $g_{\mathrm{band}} : \mathbb{R}^{3d} \to \mathbb{R}^K$ produces a logit for each band. A quadratic preference for proximity to the anchor is added before normalization across bands:

$$\ell_{k,f} = g_{\mathrm{band}}(\mathbf{u}_f)_k - \rho\,(r_f - c_k)^2, \qquad \mathbf{W}_{k,f} = \mathrm{softmax}_k\left(\frac{\ell_{k,f}}{\tau}\right). \tag{14}$$

Here $\rho \ge 0$ weights the locality preference and $\tau > 0$ is the softmax temperature. The assignment matrix $\mathbf{W} \in \mathbb{R}^{K \times F}$ satisfies $\mathbf{W}_{k,f} \ge 0$ and $\sum_{k=1}^{K} \mathbf{W}_{k,f} = 1$. Ordered anchors encourage locality while the learned logits adapt the assignments to the input. Band pooling uses a separate normalization over frequency bins, followed by a trainable complex projection $\Pi_{\mathbb{C}} : \mathbb{C}^d \to \mathbb{C}^d$:

$$\mathbf{B}_k = \Pi_{\mathbb{C}}\left(\frac{\sum_{f=1}^{F} \mathbf{W}_{k,f}\mathbf{S}_{x,f}}{\sum_{f=1}^{F} \mathbf{W}_{k,f} + \epsilon}\right) \in \mathbb{C}^d. \tag{15}$$

The resulting $\{\mathbf{B}_k\}_{k=1}^{K}$ represent numerical dynamics at different spectral scales. The weights $\mathbf{W}_{k,f}$ are retained to project a correction from each band back to its frequency bins. They define the scope of a correction without imposing a fixed hard partition on all windows.

Semantic candidate distillation. One lookback window may contain several events, repeated reports, and generic context. AnchorScope summarizes these inputs with M semantic candidate tokens. The real and imaginary components of $\mathbf{S}_t$ are concatenated and projected to a real feature space:

$$\mathbf{T}_f = \phi_{\mathrm{text}}\left([\mathrm{Re}(\mathbf{S}_{t,f}), \mathrm{Im}(\mathbf{S}_{t,f})]\right) \in \mathbb{R}^d. \tag{16}$$

The trainable map $\phi_{\mathrm{text}} : \mathbb{R}^{2d} \to \mathbb{R}^d$ places these tokens in a real feature space. Learned queries $\mathbf{Q}_e \in \mathbb{R}^{M \times d}$ attend to $\mathbf{T} = [\mathbf{T}_1, \dots, \mathbf{T}_F]^\top$. Here Attn denotes scaled dot product attention with query, key, and value projections. Its output is refined by a feedforward map $\mathrm{FFN} : \mathbb{R}^d \to \mathbb{R}^d$ applied to each row and layer normalization LN:

$$\mathbf{E}_{1:M} = \mathrm{LN}(\mathbf{A}_e + \mathrm{FFN}(\mathbf{A}_e)), \quad \mathbf{A}_e = \mathrm{Attn}(\mathbf{Q}_e, \mathbf{T}, \mathbf{T}) \in \mathbb{R}^{M \times d}. \tag{17}$$

Each candidate $\mathbf{E}_m \in \mathbb{R}^d$ is a learned latent summary of temporally adapted textual evidence. The term semantic event refers to this representation and does not require a discrete event annotation. The value of M controls the capacity of the summary independently of the band count K.

Routing and correction calibration. Each numerical band queries the semantic candidates. Let $\mathbf{b}_k = [\mathrm{Re}(\mathbf{B}_k), \mathrm{Im}(\mathbf{B}_k)] \in \mathbb{R}^{2d}$. The query map $\mathbf{q} : \mathbb{R}^{2d} \to \mathbb{R}^d$ and key and value maps $\mathbf{k}, \mathbf{v} : \mathbb{R}^d \to \mathbb{R}^d$ define the attention weights and routed summary:

$$\alpha_{k,m} = \mathrm{softmax}_m\left(\frac{\mathbf{q}(\mathbf{b}_k)^\top \mathbf{k}(\mathbf{E}_m)}{\sqrt{d}}\right), \qquad \mathbf{R}_k = \sum_{m=1}^{M} \alpha_{k,m}\mathbf{v}(\mathbf{E}_m). \tag{18}$$

The routed summary $\mathbf{R}_k \in \mathbb{R}^d$ is concatenated with $\mathbf{b}_k$. A trainable modulator $h_{\mathrm{mod}} : \mathbb{R}^{3d} \to \mathbb{R}^{5d}$ produces five proposal vectors in $\mathbb{R}^d$ for amplitude, phase, residual weight, and the real and imaginary residual components:

$$(\mathbf{u}_k^a, \mathbf{u}_k^\varphi, \mathbf{u}_k^\beta, \mathbf{u}_k^r, \mathbf{u}_k^i) = h_{\mathrm{mod}}\left([\mathbf{b}_k, \mathbf{R}_k]\right). \tag{19}$$

A separate trainable map $g_\gamma : \mathbb{R}^{3d} \to \mathbb{R}$ produces one scalar gate for each band:

$$\gamma_k = \sigma\big(g_\gamma\big([\mathbf{b}_k, \mathbf{R}_k]\big)\big) \in [0,1]. \tag{20}$$

Here σ denotes the logistic sigmoid. The gate is optimized through the forecasting loss and supplies a learned modulation weight for the routed interaction. Let $s \ge 0$ denote the global modulation

scale. The constants $a_{\max} > 0$, $\varphi_{\max} > 0$, and $\beta_{\max} > 0$ bound the direct amplitude, phase, and residual injection coefficients:

$$\begin{aligned} \Delta a_k &= a_{\max} \tanh(\mathbf{u}_k^a)\gamma_k s, \\ \Delta\varphi_k &= \varphi_{\max} \tanh(\mathbf{u}_k^\varphi)\gamma_k s, \\ \beta_k &= \beta_{\max}\sigma(\mathbf{u}_k^\beta)\gamma_k s. \end{aligned} \tag{21}$$

All proposal transforms and the following products are elementwise, with γ_k broadcast over d features. Thus Δa_k, $\Delta\varphi_k$, and β_k denote vectors despite their compact notation. The real and imaginary residual proposals are injected before separate feature normalizations:

$$\begin{aligned} \mathbf{z}_k^r &= (1+\Delta a_k)\left[\mathrm{Re}(\mathbf{B}_k)\cos(\Delta\varphi_k) - \mathrm{Im}(\mathbf{B}_k)\sin(\Delta\varphi_k)\right] + \beta_k \mathbf{u}_k^r, \\ \mathbf{z}_k^i &= (1+\Delta a_k)\left[\mathrm{Re}(\mathbf{B}_k)\sin(\Delta\varphi_k) + \mathrm{Im}(\mathbf{B}_k)\cos(\Delta\varphi_k)\right] + \beta_k \mathbf{u}_k^i, \\ \widetilde{\mathbf{B}}_k &= \mathrm{LN}_r(\mathbf{z}_k^r) + i\,\mathrm{LN}_i(\mathbf{z}_k^i). \end{aligned} \tag{22}$$

The two layer normalizations stabilize the feature scale after modulation. The parameter bounds therefore control the direct amplitude, phase, and residual injection coefficients; they do not impose a pointwise bound on the final difference between normalized band tokens. Finally, the correction from each band is returned to the full frequency grid through the same soft assignments:

$$\widetilde{\mathbf{S}}_{x,f} = \mathbf{S}_{x,f} + \sum_{k=1}^{K} \mathbf{W}_{k,f}\left(\widetilde{\mathbf{B}}_k - \mathbf{B}_k\right). \tag{23}$$

The update retains an additive path from $\mathbf{S}_x$ to $\widetilde{\mathbf{S}}_x$. Setting γ_k to zero removes the direct amplitude, phase, and residual injection terms for band k before normalization after the update. With nonzero gates, $\mathbf{W}$ distributes the band updates across frequencies, while routing determines their content. Band masking in Section 4 suppresses one of these contributions; gate and bound interventions change the scale of the direct modulation coefficients.

A.5 Future Spectrum Generation for Each Horizon

Stage D converts the controlled historical spectrum into the requested future horizon. The input has F bins and the output has F_H bins, so a single identity mapping cannot serve every forecast length. AnchorScope first applies a complex base projection $U_{F\to F_H}$, implemented by separate real and imaginary projections along the frequency dimension:

$$\mathbf{S}_y^{\mathrm{base}} = U_{F\to F_H}(\widetilde{\mathbf{S}}_x) \in \mathbb{C}^{F_H\times d}. \tag{24}$$

In parallel, a spectral generator G conditioned on the forecast horizon receives the corrected spectrum and a learned embedding of the requested prediction length. Let H_{ref} be a fixed reference horizon used only to normalize the scalar input to the embedding network ω:

$$\mathbf{h}_H = \omega(H/H_{\mathrm{ref}}), \qquad \mathbf{S}_y^{\mathrm{hsg}} = G(\widetilde{\mathbf{S}}_x, \mathbf{h}_H) \in \mathbb{C}^{F_H\times d}. \tag{25}$$

The base projection and the spectrum generated for the requested horizon are combined with the configured blend coefficient $\lambda \in [0,1]$:

$$\widetilde{\mathbf{S}}_y = (1-\lambda)\mathbf{S}_y^{\mathrm{base}} + \lambda \mathbf{S}_y^{\mathrm{hsg}}. \tag{26}$$

The complex output map $\mathrm{Dec}_{\mathbb{C}} : \mathbb{C}^{F_H\times d} \to \mathbb{C}^{F_H\times C}$ restores the target channels. The inverse real FFT uses the explicit output length H, including when H is odd. The result is rescaled using the historical standard deviation and mean defined in Equation (6):

$$\widehat{\mathbf{Y}}_{T+1:T+H} = \mathcal{F}_{r,H}^{-1}\left(\mathrm{Dec}_{\mathbb{C}}(\widetilde{\mathbf{S}}_y)\right)\boldsymbol{\sigma} + \boldsymbol{\mu}. \tag{27}$$

The prediction horizon is therefore explicit when the future spectrum is generated. By decoding after semantic correction, AnchorScope keeps fusion tied to the observed numerical state while allowing the final generator to adapt its spectral resolution to the requested forecast length.

A.6 Training Objective and Ablation Interfaces

AnchorScope is trained end to end by minimizing mean squared error over the prediction horizon:

$$\mathcal{L}_{\text{forecast}} = \frac{1}{BHC} \sum_{b=1}^{B} \sum_{h=1}^{H} \sum_{c=1}^{C} \left(\widehat{Y}_{b,h,c} - Y_{b,h,c} \right)^2 . \tag{28}$$

The language encoder remains frozen. The trainable components are Proj and A_{text}, the numerical spectral encoder, g_{band} and $\Pi_{\mathbb{C}}$, ϕ_{text} and the candidate queries $\mathbf{Q}_e$, the routing maps $\mathbf{q}$, $\mathbf{k}$, and $\mathbf{v}$, the modulator h_{mod} and gate g_γ, and the decoder conditioned on the forecast horizon. This keeps the language representation stable while learning allocation and calibration specific to the forecasting task.

The ablations remove the textual branch, replace the structured interaction with global text fusion, or remove the Temporal Semantic Adapter or Semantic Frequency Modulation. Separate matched structural ablations retrain variants with identical encoders, modulators, training budgets, and random seeds while changing only the frequency allocation rule or the direct modulation bounds. These comparisons assess the contribution of each component after training. They are complemented by interventions on the trained model, which isolate changes to a band correction or its control while retaining the numerical reference.

B Supplementary Results

This appendix reports complete results for each horizon and additional analyses of band controls referenced in the main paper.

B.1 Extended Comparisons and Component Ablations

Table 6: Text quality and temporal alignment diagnostics for individual paired records on representative datasets. Explicit dates in a supplied `fact` field are compared with the `end_date` of its paired observation row. A mismatched row contains at least one later date marker, and a positive median offset means that the latest marker predates the end of the observation interval. Values are MSE/MAE.

Dataset	Unique text ↑	Repeat ↓	Mismatched rows ↓	Median text date offset	Alignment score ↑	Numerical	Global fusion	AnchorScope (Table 1)
Health	99.6%	0.4%	0.0%	23.5 days	0.735	1.4219/0.8032	1.8669/0.9277	1.360/0.787
Agriculture	100.0%	0.0%	0.0%	49.0 days	0.773	0.2795/0.3358	0.3133/0.3663	0.230/0.307
Environment	14.3%	85.7%	6.6%	19.0 days	0.069	0.4219/0.4823	0.4193/0.4819	0.408/0.479
Security	100.0%	0.0%	0.0%	49.0 days	0.996	70.5405/3.7567	77.9474/4.3900	70.364/3.757

B.2 Post-training Revision Interventions

The following analyses alter trained checkpoints. They characterize sensitivity to learned revisions and complement the controlled retraining results in Table 5.

Table 7: Effect of masking textual revisions in each band on the same trained checkpoints. There are 36 combinations of dataset and horizon. Sensitivity is the mean absolute relative MSE change from AnchorScope. Counts separate settings in which masking worsens or improves MSE. Error reductions summarize the subset in which masking improves MSE.

Masked band	Sensitivity $\lvert\Delta\text{MSE}\rvert$	Masking outcome (cases)		Reduction when masking helps	
		MSE worsens	MSE improves	MSE	MAE
Low	2.27%	15/36	21/36	2.66%	0.32%
Mid	1.42%	16/36	20/36	1.81%	0.46%
High	4.25%	23/36	13/36	2.87%	1.50%

Across the 36 dataset and horizon settings, high-band masking has the largest average sensitivity and worsens MSE in 23 settings. The direction of this intervention varies across settings, which is consistent with a revision whose useful range depends on the forecast state.

Table 8: Effect of changing revision size on the same trained checkpoints. Relative changes are measured against AnchorScope; positive values indicate larger errors. The second group reports the subset in which band masking improves MSE. Mean gate is the average weight applied to modulation and is fixed at 1 when the gate is removed. Semantic delta reports the magnitude of the semantic update.

Variant	All cases		Masking improves MSE		Mean gate	Semantic delta
	ΔMSE	ΔMAE	ΔMSE	ΔMAE		
AnchorScope	+0.00%	+0.00%	+0.00%	+0.00%	0.938	1.007
No gate	+0.02%	+0.03%	+0.03%	+0.04%	1.000	1.029
Loose bounds (3×)	+4.56%	+2.27%	+5.53%	+2.75%	0.938	1.557
No gate + loose bounds	+4.71%	+2.36%	+5.71%	+2.87%	1.000	1.603

Relaxing bounds increases the semantic update magnitude from 1.007 to 1.557 and raises average MSE by 4.56%. Together, these interventions characterize learned revisions and complement the controlled retraining results.

B.3 Supplementary Diagnostic Figures

Figure 6 compares numerical amplitude spectra with dominant frequencies extracted from TF-IDF/SVD text trajectories. Agriculture, Energy, and Health concentrate strongly at lower numerical frequencies, while the text markers occupy different ranges across domains. Climate has no stable semantic peak under this diagnostic. The figure describes correspondence in the supplied signals.

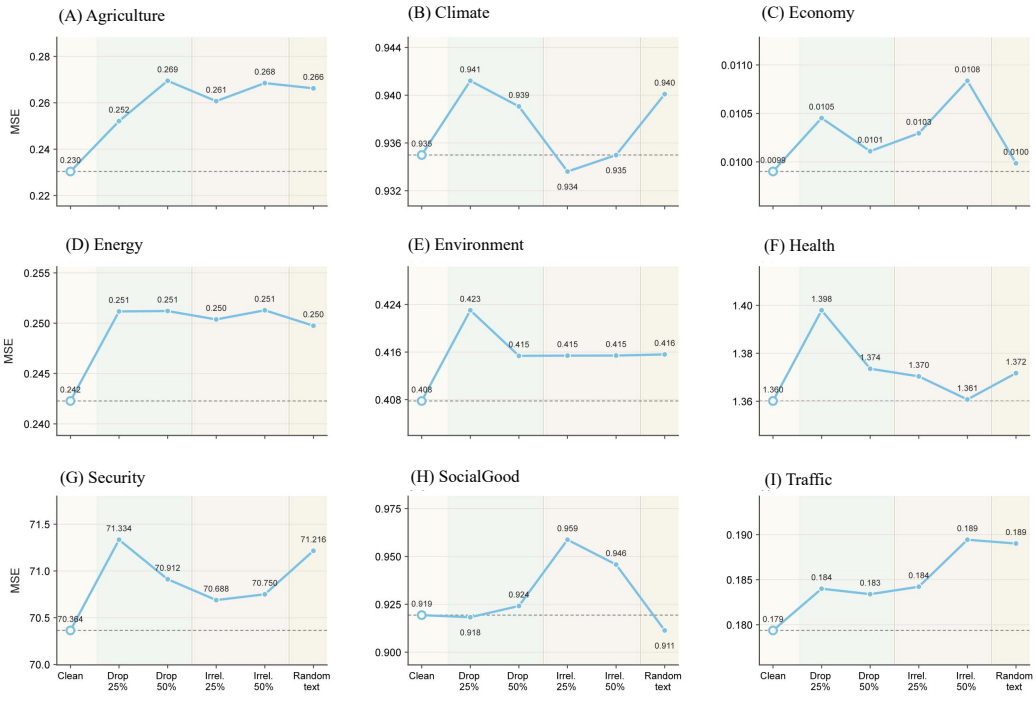


Figure 5: Robustness to imperfect textual evidence across datasets. Each panel reports MSE with clean text, partial text removal, irrelevant text injection, and random text. Values are averaged across prediction horizons. Clean values are taken from Table 1. Values under altered text are obtained by retraining under each condition with the same data split and horizons.

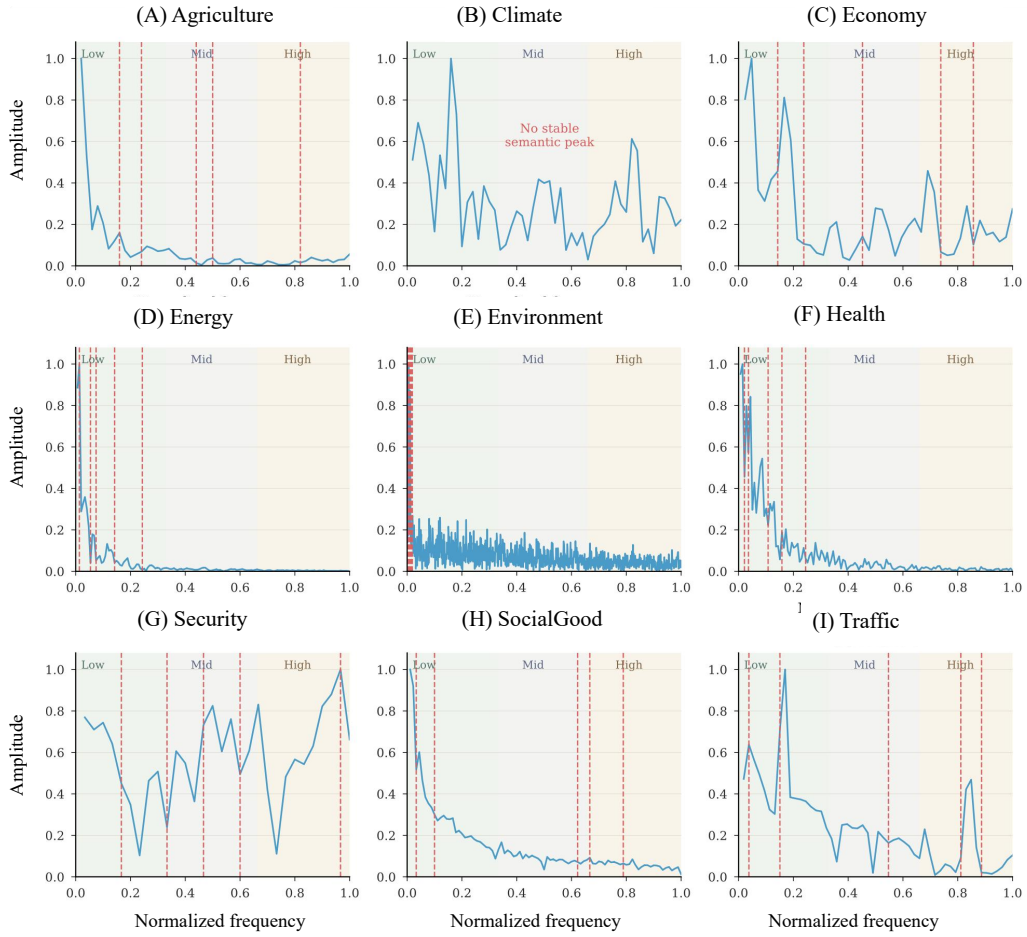


Figure 6: Numerical spectra and dominant textual frequencies across nine domains. Blue curves show normalized numerical amplitude spectra. Red dashed lines mark dominant frequencies extracted from TF-IDF/SVD text trajectories independently of AnchorScope. Shading indicates low, middle, and high frequency ranges. Climate has no stable semantic peak under this diagnostic. The figure describes signal correspondence; Table 7 evaluates the effects of masking textual revisions.

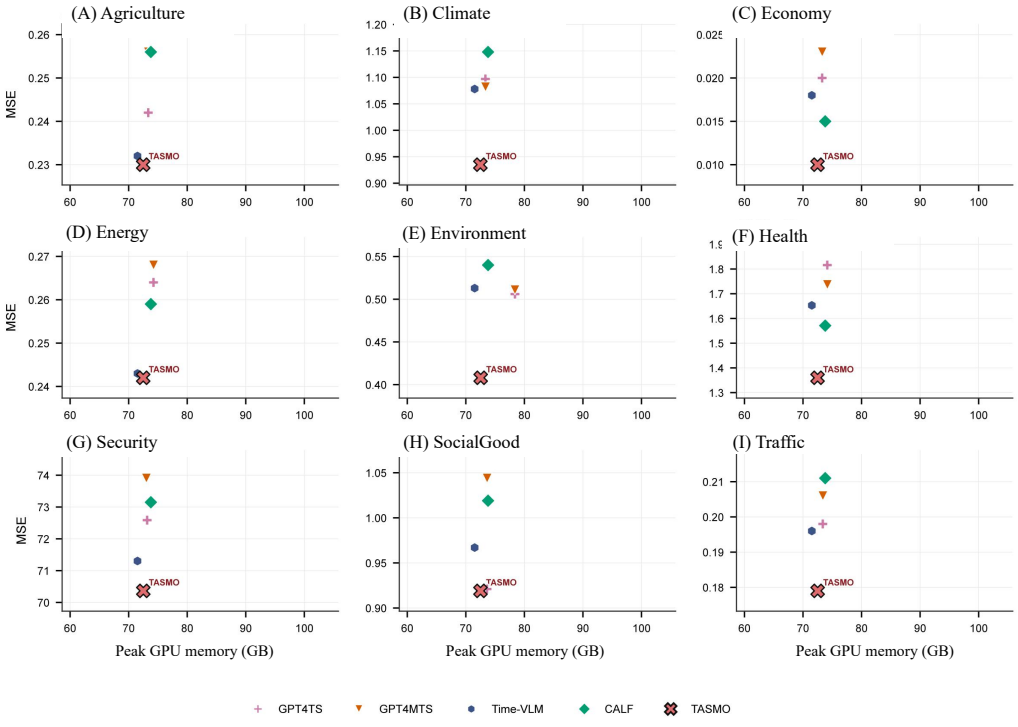


Figure 7: Memory and accuracy across datasets. MSE values are taken from Table 1. Points are shown only for methods with comparable peak GPU memory profiles. AnchorScope achieves low error on several datasets without using the largest memory footprint.

Table 9: Detailed forecasting performance at each prediction horizon. Lower MSE and MAE indicate better performance. Best results are in bold and second best results are underlined.

Dataset	Horizon	AnchorScope (ours)		Aurora		CALF		TimeCMA		GPT4TS		GPT4MTS		Time-LLM		Time-VLM	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	6	0.145	0.247	0.276	0.328	0.161	0.262	**0.080**	**0.193**	0.147	0.248	0.161	0.263	0.291	0.346	0.142	0.248
Agriculture	8	0.199	0.300	0.344	0.369	0.197	0.291	**0.099**	**0.216**	0.220	0.303	0.236	0.303	0.383	0.397	0.202	0.295
Agriculture	10	0.256	0.328	0.415	0.411	0.310	0.353	**0.123**	**0.242**	0.278	0.331	0.287	0.335	0.419	0.418	0.241	0.321
Agriculture	12	0.321	0.355	0.503	0.455	0.356	0.380	**0.152**	**0.270**	0.323	0.359	0.339	0.370	0.497	0.458	0.343	0.367
Climate	6	**0.944**	**0.782**	1.136	0.873	1.122	0.861	1.194	0.888	1.069	0.847	1.073	0.843	1.139	0.869	1.054	0.840
Climate	8	**0.931**	**0.776**	1.149	0.881	1.071	0.843	1.233	0.901	1.099	0.857	1.077	0.847	1.142	0.875	1.074	0.846
Climate	10	**0.922**	**0.772**	1.166	0.884	1.322	0.937	1.257	0.909	1.101	0.857	1.083	0.852	1.155	0.876	1.089	0.852
Climate	12	**0.944**	**0.782**	1.162	0.884	1.077	0.842	1.262	0.913	1.118	0.863	1.095	0.857	1.152	0.875	1.094	0.857
Economy	6	**0.009**	**0.078**	0.029	0.135	0.013	0.093	0.029	0.141	0.017	0.104	0.026	0.133	0.027	0.135	0.017	0.103
Economy	8	**0.010**	**0.081**	0.031	0.137	0.014	0.096	0.030	0.143	0.022	0.120	0.021	0.118	0.029	0.137	0.018	0.105
Economy	10	**0.010**	**0.083**	0.032	0.140	0.017	0.106	0.032	0.144	0.019	0.111	0.025	0.128	0.031	0.141	0.019	0.108
Economy	12	**0.010**	**0.084**	0.034	0.145	0.016	0.098	0.032	0.145	0.023	0.123	0.020	0.113	0.033	0.145	0.020	0.113
Energy	12	**0.096**	**0.211**	0.179	0.321	0.114	0.245	0.231	0.370	0.144	0.282	0.134	0.273	0.218	0.364	0.099	0.219
Energy	24	0.209	**0.323**	0.275	0.398	0.215	0.342	0.309	0.429	0.221	0.349	0.232	0.372	0.297	0.423	**0.201**	0.326
Energy	36	**0.291**	**0.401**	0.379	0.466	0.308	0.416	0.376	0.477	0.307	0.409	0.305	0.408	0.360	0.465	0.298	0.402
Energy	48	**0.372**	**0.462**	0.453	0.517	0.400	0.481	0.490	0.558	0.383	0.470	0.402	0.488	0.439	0.517	0.376	0.463
Environment	48	**0.393**	**0.467**	0.551	0.532	0.484	0.491	0.573	0.547	0.465	0.487	0.475	0.497	0.513	0.513	0.461	0.484
Environment	96	**0.409**	**0.480**	0.619	0.572	0.546	0.521	0.593	0.552	0.505	0.508	0.517	0.517	0.572	0.539	0.511	0.512
Environment	192	**0.409**	**0.487**	0.638	0.590	0.585	0.541	0.590	0.546	0.540	0.537	0.580	0.581	0.587	0.541	0.564	0.536
Environment	336	**0.419**	**0.482**	0.607	0.574	0.545	0.512	0.561	0.541	0.512	0.521	0.474	0.526	0.563	0.544	0.515	0.516
Health	12	**0.988**	**0.683**	2.173	1.071	1.109	0.702	2.489	1.094	1.462	0.817	1.437	0.830	2.427	1.103	1.246	0.734
Health	24	**1.361**	**0.802**	2.760	1.225	1.586	0.842	2.641	1.138	1.832	0.934	1.900	0.946	2.801	1.206	1.750	0.886
Health	36	**1.528**	**0.836**	2.602	1.171	1.791	0.902	2.637	1.169	1.989	0.945	1.800	0.910	2.703	1.198	1.806	0.924
Health	48	**1.563**	**0.825**	2.530	1.149	1.797	0.918	2.569	1.164	1.981	0.967	1.810	0.919	2.479	1.139	1.811	0.928
Security	6	64.575	**3.524**	75.660	4.219	67.108	3.772	73.961	4.020	66.966	3.939	72.012	4.470	**59.604**	3.854	65.247	3.718
Security	8	69.030	**3.714**	79.554	4.354	72.344	4.039	76.873	4.252	70.653	4.044	72.512	4.292	**58.960**	3.741	69.537	3.914
Security	10	72.283	**3.855**	84.436	4.517	74.779	4.009	78.116	4.342	75.922	4.348	73.024	4.046	**61.930**	4.028	73.464	4.113
Security	12	**75.568**	**3.934**	88.682	4.686	78.370	4.159	79.653	4.430	76.809	4.179	78.080	4.316	79.384	4.450	76.963	4.186
SocialGood	6	0.785	**0.366**	0.893	0.467	0.997	0.406	1.019	0.528	0.839	0.405	0.893	0.388	0.993	0.549	**0.767**	0.377
SocialGood	8	0.849	**0.396**	0.982	0.516	0.935	0.403	1.106	0.573	0.853	0.427	0.916	0.418	1.043	0.562	**0.847**	0.408
SocialGood	10	**0.989**	0.444	1.074	0.563	1.000	**0.444**	1.169	0.614	0.997	0.534	1.156	0.484	1.110	0.601	1.030	0.504
SocialGood	12	1.054	**0.448**	1.159	0.608	1.145	0.477	1.243	0.651	**0.998**	0.476	1.212	0.488	1.165	0.629	1.224	0.579
Traffic	6	**0.168**	**0.228**	0.274	0.415	0.200	0.240	0.261	0.379	0.183	0.261	0.184	0.254	0.250	0.373	0.187	0.273
Traffic	8	**0.176**	**0.230**	0.278	0.418	0.214	0.260	0.266	0.391	0.194	0.252	0.210	0.266	0.255	0.391	0.195	0.280
Traffic	10	**0.184**	**0.238**	0.279	0.419	0.211	0.264	0.271	0.390	0.212	0.266	0.217	0.273	0.265	0.393	0.199	0.280
Traffic	12	**0.189**	**0.235**	0.280	0.418	0.217	0.266	0.310	0.397	0.203	0.269	0.215	0.269	0.281	0.402	0.202	0.284

Table 10: Comparison between AnchorScope and plugin designs with different backbones at each horizon. Lower MSE and MAE indicate better performance. Best results are in bold and second best results are underlined.

Dataset	Horizon	AnchorScope (ours)		Numeric-iTransformer		Numeric-PatchTST		TaTS-iTransformer		TaTS-PatchTST		MM-TSFlib-iTransformer		MM-TSFlib-PatchTST	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Agriculture	6	0.145	**0.247**	0.130	0.251	0.143	0.255	0.423	0.283	0.175	0.291	**0.120**	0.251	0.139	0.268
Agriculture	8	0.199	0.300	**0.170**	0.293	0.197	0.300	0.433	0.351	0.247	0.333	0.172	**0.286**	0.187	0.308
Agriculture	10	0.256	0.328	**0.219**	**0.314**	0.232	0.318	0.520	0.414	0.321	0.383	0.228	0.320	0.222	0.322
Agriculture	12	0.321	0.355	**0.274**	**0.346**	0.322	0.363	0.599	0.430	0.383	0.419	0.274	0.352	0.292	0.364
Climate	6	**0.944**	**0.782**	1.114	0.872	1.163	0.875	1.070	0.805	0.992	0.803	1.051	0.838	1.102	0.855
Climate	8	**0.931**	**0.776**	1.116	0.857	1.163	0.879	1.092	0.797	1.004	0.829	1.057	0.842	1.158	0.881
Climate	10	**0.922**	**0.772**	1.143	0.870	1.159	0.875	1.104	0.800	0.998	0.803	1.072	0.846	1.173	0.874
Climate	12	**0.944**	**0.782**	1.146	0.865	1.172	0.888	1.104	0.798	0.987	0.800	1.090	0.855	1.113	0.862
Economy	6	0.009	**0.072**	0.013	0.090	0.014	0.096	0.023	0.074	0.009	0.077	0.014	0.094	0.019	0.111
Economy	8	0.010	**0.076**	0.013	0.093	0.015	0.098	0.020	0.076	0.008	0.076	0.014	0.097	0.019	0.114
Economy	10	0.010	0.079	0.012	0.088	0.015	0.098	0.021	0.078	0.008	**0.077**	0.015	0.096	0.020	0.114
Economy	12	0.010	0.081	0.012	0.088	0.016	0.100	0.023	0.084	0.008	**0.080**	0.016	0.102	0.018	0.107
Energy	12	0.096	**0.211**	**0.095**	0.213	0.102	0.225	0.109	0.283	0.133	0.276	0.106	0.238	0.105	0.237
Energy	24	0.209	0.323	0.187	0.323	0.211	0.341	0.207	0.402	0.242	0.381	**0.177**	**0.319**	0.209	0.342
Energy	36	0.291	0.401	0.284	**0.399**	0.343	0.440	0.298	0.471	0.331	0.440	**0.281**	0.401	0.299	0.416
Energy	48	**0.372**	**0.462**	0.381	0.469	0.383	0.474	0.410	0.506	0.387	0.487	0.379	0.472	0.380	0.478
Environment	48	0.393	0.467	0.415	0.479	0.475	0.495	0.297	**0.375**	**0.268**	0.378	0.408	0.474	0.437	0.485
Environment	96	0.409	0.480	0.439	0.497	0.521	0.537	0.322	**0.383**	0.279	0.383	0.413	0.476	0.465	0.497
Environment	192	0.409	0.487	0.447	0.506	0.632	0.593	0.298	**0.373**	0.264	0.375	0.423	0.484	0.504	0.525
Environment	336	0.419	0.482	0.427	0.484	0.525	0.520	0.271	**0.368**	0.257	0.374	0.420	0.478	0.494	0.517
Health	12	**0.988**	0.683	2.072	0.806	1.219	0.725	1.240	0.687	1.195	0.696	1.640	0.703	1.094	**0.681**
Health	24	1.361	0.802	2.165	0.878	1.799	0.894	1.524	0.825	**1.312**	**0.762**	1.727	0.822	1.815	0.909
Health	36	1.528	**0.836**	2.530	0.954	1.908	0.913	1.588	0.851	**1.481**	0.853	2.003	0.885	1.810	0.907
Health	48	**1.563**	**0.825**	2.605	1.001	2.066	0.966	1.694	0.891	1.591	0.890	2.513	0.988	1.958	0.921
Security	6	**64.575**	**3.524**	74.340	4.553	72.026	4.326	74.014	3.902	69.107	4.133	77.451	4.568	71.650	4.223
Security	8	**69.030**	**3.714**	74.965	4.433	77.405	4.474	75.371	4.449	75.321	4.376	78.237	4.709	77.397	4.559
Security	10	**72.283**	**3.855**	81.359	4.580	80.338	4.611	77.122	4.519	77.566	4.443	81.387	4.633	78.927	4.569
Security	12	**75.568**	**3.934**	83.243	4.664	82.381	4.630	77.651	4.544	83.157	4.687	83.544	4.626	82.063	4.634
SocialGood	6	**0.785**	**0.366**	1.154	0.452	1.149	0.450	1.027	0.383	0.876	0.400	0.961	0.395	0.997	0.419
SocialGood	8	**0.849**	**0.396**	1.212	0.486	1.052	0.452	1.153	0.411	0.935	0.426	1.118	0.462	1.173	0.497
SocialGood	10	**0.989**	**0.444**	1.337	0.514	1.297	0.498	1.130	0.457	1.004	0.455	1.281	0.489	1.322	0.521
SocialGood	12	1.054	**0.448**	1.512	0.569	1.439	0.542	1.199	0.468	1.078	0.510	1.345	0.545	1.382	0.574
Traffic	6	**0.168**	0.228	0.196	0.234	0.174	0.218	0.233	0.215	0.170	0.229	0.183	**0.215**	0.192	0.246
Traffic	8	0.176	0.230	0.200	0.233	0.184	0.236	0.215	**0.222**	0.177	0.237	0.190	0.237	0.184	0.229
Traffic	10	**0.184**	0.238	0.205	0.241	0.191	0.242	0.229	0.243	0.185	0.238	0.205	0.241	0.189	**0.234**
Traffic	12	**0.189**	**0.235**	0.210	0.246	0.200	0.244	0.272	0.260	0.192	0.236	0.210	0.251	0.197	0.237

Table 11: Control intervention grouped by the semantic band for which masking improves performance during evaluation. ΔMSE and ΔMAE are relative to AnchorScope under the same dataset and horizon. Larger positive values indicate that removing or relaxing a control amplifies excessive semantic correction.

Band where masking helps	Variant	ΔMSE ↓	ΔMAE ↓	Semantic delta	Interpretation
High	AnchorScope	+0.00%	+0.00%	0.896	Controlled semantic correction at high frequency
High	Loose bounds (3×)	+10.55%	+4.83%	1.494	Relaxed bounds strongly amplify excessive correction at high frequency
High	No gate + loose bounds	+10.68%	+4.87%	1.543	An open gate further increases excessive correction at high frequency
Low	AnchorScope	+0.00%	+0.00%	1.122	Controlled semantic correction at low frequency
Low	Loose bounds (3×)	+3.88%	+1.56%	1.671	Relaxed bounds amplify excessive correction of trends
Low	No gate + loose bounds	+3.83%	+1.57%	1.691	Open gate provides little additional benefit when bounds are relaxed
Mid	AnchorScope	+0.00%	+0.00%	0.900	Controlled semantic correction at middle frequency
Mid	Loose bounds (3×)	+4.19%	+2.99%	1.342	Relaxed bounds amplify excessive periodic correction
Mid	No gate + loose bounds	+4.76%	+3.35%	1.450	An open gate and loose bounds increase excessive correction at middle frequency