Correction-space Cross-variate Interaction for Test-time Adaptation in Time Series Forecasting

Anonymous authors
Paper under double-blind review

Abstract

Test-time adaptation (TTA) is a promising paradigm for handling distribution shift in time-series forecasting (TSF), where models adapt at inference time, often leveraging delayed observed data to refine predictions. In the multivariate setting, distribution shifts often exhibit cross-variate dependencies, yet existing TSF-TTA methods adapt each variate independently and ignore this cross-variate structure. Exploiting such structure motivates cross-variate interaction, but coupling variates through backbone predictions introduces direct pathways for mixing uncorrected errors across variates, a concern under the delayed supervision of TSF-TTA. We identify the *interaction space* as a key design choice, and show that acting on adapter corrections that refine backbone outputs, the *correction space*, rather than on the predictions themselves, avoids directly propagating backbone errors across variates. We build on this to propose CoRe (COrrection-space Interaction REfinement), realizing correction-space interaction through (i) Shared-anchor Correction Refinement (SCR), which combines each variate's correction with a shared anchor through a parameter-efficient bottleneck, and (ii) input-conditioned spectral gating, which adaptively modulates the refinement from the current input window. Across seven backbones, six datasets, and four prediction horizons, CoRe reduces MSE by 25.82% on average over backbones and 10.57% over the state-of-the-art TSF-TTA method, with stronger gains at medium-to-long horizons and modest computational overhead. Data and code are available at: `anonymous.4open.science/r/CoRe-TTA-CD98`.

1 Introduction

Time-series forecasting (TSF) supports decision-making across domains including energy, traffic, finance, and weather (Bu and Cho, 2020; Bai et al., 2020; Cheng et al., 2022; Karevan and Suykens, 2020). Although modern forecasting models achieve strong in-distribution performance (Zeng et al., 2023; Nie et al., 2023; Liu et al., 2024), their accuracy degrades after deployment as real-world time series are non-stationary and the data-generating process drifts over time (Liu et al., 2022; Fan et al., 2023; Ye et al., 2024). Existing strategies such as online learning and domain adaptation either require labeled data, assume stable domain boundaries, or incur high computational costs, limiting their applicability to continuously evolving distributions (Losing et al., 2018; Wang and Deng, 2018). Complementary approaches based on normalization aim to address distribution shift through input transformation, but operate in a static manner and do not support feedback-driven adaptation over time (Kim et al., 2021; Liu et al., 2023; 2022; Fan et al., 2023; Ye et al., 2024).

Test-time adaptation (TTA) offers a practical alternative by updating lightweight components of a frozen model using streaming test data (Liang et al., 2025). Vision TTA methods (Wang et al., 2021; Sun et al., 2020) rely on classification-specific assumptions and do not readily transfer to TSF. A key property shapes TTA design in TSF: ground-truth observations become available after the forecast horizon, enabling supervised adaptation by minimizing the forecasting loss. Recent TSF-TTA methods leverage this property to improve forecasting

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.

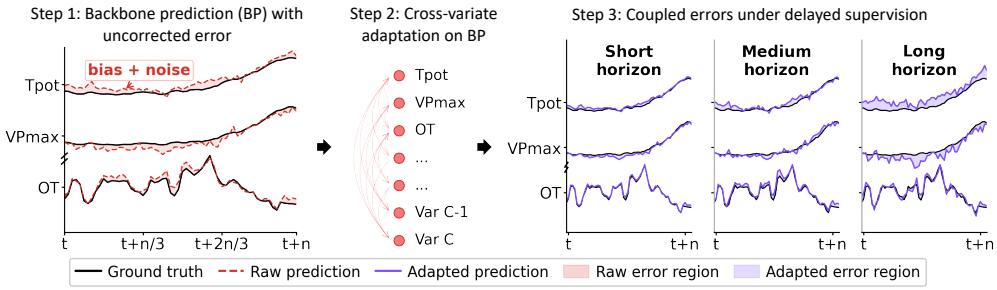

Figure 1: Direct error mixing in prediction-space cross-variate interaction. A conceptual illustration using variables from Weather (Tpot, VPmax, OT, and other variables). Cross-variate interaction applied to backbone predictions can directly mix their uncorrected errors across variates (Step 2). In TSF-TTA, ground-truth observations become available only after the prediction horizon, delaying corrective feedback (Step 3). Proposition 1 formalizes the direct error-mixing pathway, while Figure 5 empirically compares prediction-space and correction-space interaction across prediction horizons.

accuracy under distribution shift via lightweight online adaptation (Im and Kwon, 2026; Grover and Etemad, 2025; Kim et al., 2025; Wang et al., 2026).

Existing TSF-TTA methods predominantly adapt each variate independently, yet distribution shifts in multivariate time series often affect variates jointly (Lin et al., 2025) (Figure 2). Such shifts may induce shared structure in the adaptation corrections required across variates, making cross-variate correction sharing potentially beneficial. Our analysis shows that the leading singular mode captures approximately 73% of correction energy on average (Appendix E.1), further supporting this hypothesis. Such shared structure motivates cross-variate interaction. Prior work has explored *how* to model this interaction. For example, Im and Kwon (2026) investigated cross-variable attention and shared specific components on backbone predictions, yet found inconsistent benefits across datasets. We identify a more fundamental design choice that has been overlooked: *where* should cross-variate interaction take place?

This choice of interaction space is particularly important in TSF-TTA under delayed supervision. Backbone predictions made under distribution shift may contain substantial uncorrected errors. Applying cross-variate interaction directly to these predictions introduces pathways through which errors from one variate can influence the refinement of others (Figure 1). Since ground-truth observations become available only after the prediction horizon, such error propagation cannot be immediately corrected through supervised feedback. We therefore investigate an alternative: performing cross-variate interaction on adapter-produced corrections (the *correction space*) rather than on backbone predictions. This design avoids directly mixing uncorrected backbone errors during cross-variate interaction.

In this work, we formalize the distinction between prediction- and correction-space interaction and characterize their different direct dependencies on frozen-backbone errors (Proposition 1). Building on this analysis, we propose CoRe (COrrection-space Interaction REfinement), a framework that performs structured cross-variate interaction in the correction space. CoRe introduces Shared-anchor Correction Refinement (SCR), which combines each variate's correction with a shared anchor through a parameter-efficient bottleneck. Input-conditioned spectral gating further modulates the refinement using information from the current input window. Controlled experiments demonstrate the importance of interaction space, particularly at medium-to-long prediction horizons. We summarize our contributions as follows:

1. We identify the *interaction space* as a key design choice for cross-variate TSF-TTA, showing that correction-space interaction enables information sharing across variates without directly mixing uncorrected backbone errors.
2. We propose CoRe, a framework to perform structured cross-variate adaptation in the *correction space*, through Shared-anchor Correction Refinement (SCR), com-

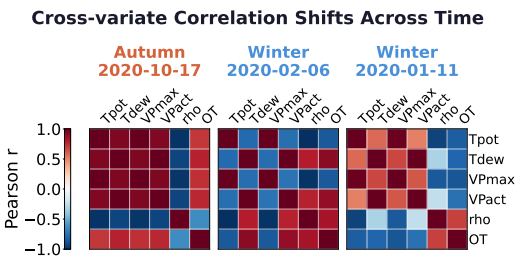

Figure 2: Cross-variate dependencies in the Weather dataset.

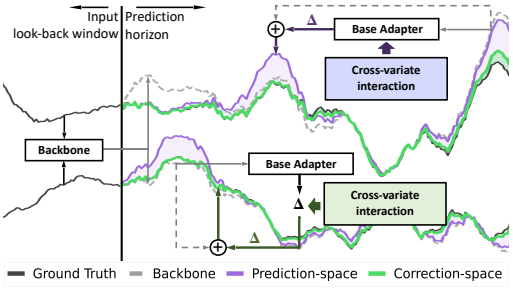

Figure 3: Prediction-space vs. correction-space cross-variate interaction.

plemented by a lightweight spectral gate that modulates the refinement using the current input window.

3. Extensive experiments across seven backbones, six datasets, and four horizons demonstrate consistent improvements over existing TSF-TTA methods. Controlled comparisons show an advantage of correction-space over prediction-space interaction, and the gains are not explained by additional capacity alone.

2 Related Work

Multivariate Time-Series Forecasting from a Channel Perspective. Multivariate time-series forecasting models temporal patterns in sequential data with multiple observed variables, whose interactions may vary across time and environments (Qiu et al., 2025a). There have been several approaches to modeling cross-variate relationships. From a channel perspective, multivariate forecasters are often contrasted as channel-independent (CI) (Nie et al., 2023; Zeng et al., 2023; Toner and Darlow, 2024) versus channel-dependent (CD) (Zhou et al., 2021; Wang et al., 2023; Liu et al., 2024), depending on whether they couple variables. However, prior findings show that cross-variate modeling can be harmful when correlations are weak, noisy, or non-stationary, and may degrade robustness under distribution shift (Chen et al., 2024; He et al., 2025). These findings suggest that the question is not whether to model cross-variate dependencies, but how to introduce them without amplifying noise under distribution shift.

Test-Time Adaptation for Time-Series Forecasting. Test-time adaptation is an emerging paradigm for addressing non-stationarity and distribution shifts in time-series forecasting. In TSF-TTA, representative baselines include TAFAS (Kim et al., 2025), PETSA (Medeiros et al., 2025), DynaTTA (Grover and Etemad, 2025), and COSA (Im and Kwon, 2026). Recent work further introduces a matured-ground-truth protocol for TSF-TTA (Wang et al., 2026). Despite recent progress, existing methods typically adapt each variate independently, overlooking cross-variate structure. Although cross-variate interaction on backbone predictions has been explored through attention and shared structures (Im and Kwon, 2026), its benefits are inconsistent across datasets, and operating on backbone predictions introduces a direct pathway for mixing their uncorrected errors across variates. CoRe instead performs cross-variate interaction in the *correction space*, removing this direct backbone-error mixing pathway. See Appendix A for further related work.

3 Methodology

3.1 Problem Formulation

We study test-time adaptation for multivariate time-series forecasting under delayed supervision. At each time step t, a frozen backbone f_θ maps a past input window $\mathbf{X}_t \in \mathbb{R}^{L\times C}$, with lookback length L and C variates, to an H-horizon forecast $\hat{\mathbf{Y}}_t^{\text{base}} \in \mathbb{R}^{H\times C}$. A lightweight adapter $\mathcal{A}_\phi$ produces per-variate corrections $\mathbf{\Delta}_t = \mathcal{A}_\phi(\hat{\mathbf{Y}}_t^{\text{base}}, \mathbf{I}_t) \in \mathbb{R}^{H\times C}$, where $\mathbf{\Delta}_t^{(c)} \in \mathbb{R}^H$

denotes the correction for variate c, and $\mathbf{I}_t$ denotes optional auxiliary information available to the base adapter. The adapted forecast is

$$\hat{\mathbf{Y}}_t = \hat{\mathbf{Y}}_t^{\text{base}} + \boldsymbol{\Delta}_t. \tag{1}$$

During deployment, ground-truth observations become available with a delay of H steps, and the adapter is updated only when the corresponding supervision becomes available. Thus, each prediction depends only on past observations, ensuring a leakage-free setting. Adapter-specific instantiations and update protocols are detailed in Appendix B.3 and C.2.

3.2 Cross-Variate Interaction in TSF-TTA

Current TSF-TTA methods adapt each variate independently. Yet distribution shifts may induce shared structure in the corrections required across variates. The decomposition in Equation 1 offers two distinct places to model this cross-variate structure: the frozen-backbone prediction $\hat{\mathbf{Y}}_t^{\text{base}}$ and the adapter-produced correction $\boldsymbol{\Delta}_t$. Applying cross-variate interaction to each defines the *prediction space* and the *correction space* (Figure 3). We analyze how backbone error propagates in each, and show that prediction-space interaction directly spreads this error across variates while correction-space interaction avoids such direct propagation.

Let $\mathcal{M}_{\text{pred}}$ and $\mathcal{M}_{\text{corr}}$ denote cross-variate refinement operators applied in the prediction and correction spaces, respectively. The two interaction forms are

$$\begin{aligned} \hat{\mathbf{Y}}_t^{\text{pred}} &= \hat{\mathbf{Y}}_t^{\text{base}} + \boldsymbol{\Delta}_t + \mathcal{M}_{\text{pred}}(\hat{\mathbf{Y}}_t^{\text{base}}), \\ \hat{\mathbf{Y}}_t^{\text{corr}} &= \hat{\mathbf{Y}}_t^{\text{base}} + \boldsymbol{\Delta}_t + \mathcal{M}_{\text{corr}}(\boldsymbol{\Delta}_t). \end{aligned} \tag{2}$$

Let $\mathbf{e}_t^{\text{base}} = \hat{\mathbf{Y}}_t^{\text{base}} - \mathbf{Y}_t$ denote the frozen-backbone error. Prediction-space interaction therefore operates on $\mathbf{Y}_t + \mathbf{e}_t^{\text{base}}$, whereas correction-space interaction operates on $\boldsymbol{\Delta}_t$.

Proposition 1 (Structural Separation)**.** *Define the backbone-error-induced prediction-space refinement as*

$$\boldsymbol{\Gamma}_t^{\text{pred}} := \mathcal{M}_{\text{pred}}(\mathbf{Y}_t + \mathbf{e}_t^{\text{base}}) - \mathcal{M}_{\text{pred}}(\mathbf{Y}_t).$$

For linear cross-variate mixing $\mathcal{M}_{\text{pred}}(\cdot) = (\cdot)\mathbf{W}_{\text{pred}}^\top$,

$$\boldsymbol{\Gamma}_{t,:,c}^{\text{pred}} = \sum_{j=1}^{C} W_{\text{pred},cj}\, \mathbf{e}_{t,:,j}^{\text{base}}.$$

Thus, each nonzero off-diagonal $W_{\text{pred},cj}$ provides a direct path from the backbone error of variate j to the refinement of variate c. The same structural distinction extends locally to nonlinear operators via their Jacobians (Appendix H.2). In correction space, the backbone error remains outside the cross-variate operator.

This proposition establishes a *structural* distinction in how the two interaction spaces directly depend on backbone errors. It does not imply that adapter corrections are independent of backbone errors, since the base adapter may itself depend on backbone predictions. Nor does it establish an accuracy ordering between the two interaction spaces. We investigate their empirical consequences in Section 4.3.

3.3 The CoRe Framework

Building on the above design principle, we propose CoRe as an instantiation of cross-variate adaptation in the correction space. The framework is illustrated in Figure 4. Given the per-variate corrections produced by the base adapter, CoRe consists of two components: (i) **Shared-anchor Correction Refinement (SCR)** combines each variate's correction with a shared correction anchor through a parameter-efficient bottleneck; and (ii) **Input-Conditioned Spectral Gating** uses spectral descriptors of the current input window to modulate the contribution of the SCR refinement. The resulting refined corrections are added to the backbone prediction to form the final forecast. The complete leakage-free streaming procedure is given in Appendix B.

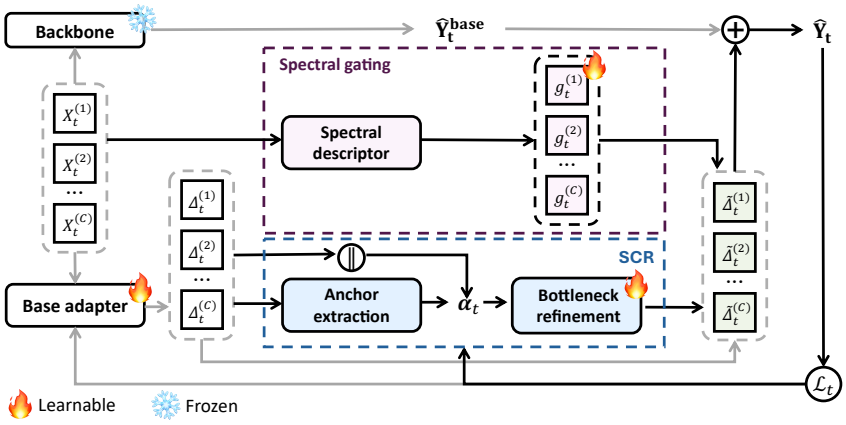

Figure 4: Framework of CoRe for multivariate test-time adaptation. A frozen backbone produces a base forecast, while a base adapter produces per-variate corrections. SCR performs cross-variate interaction in the *correction space*. Spectral descriptor of the current input window provides per-variate gates that modulate the SCR refinement. The refined corrections are added to the unchanged backbone prediction to form the final forecast.

3.4 Shared-anchor Correction Refinement (SCR)

SCR refines each variate's base-adapter correction by combining variate-specific and shared correction information. It consists of a shared correction anchor and a parameter-efficient bottleneck.

Shared correction anchor. Related distribution shifts may induce common patterns in the corrections required across variates. We approximate this shared information by averaging the corrections across variates:

$$\boldsymbol{\alpha}_t = \frac{1}{C}\sum_{j=1}^{C} \boldsymbol{\Delta}_t^{(j)} \in \mathbb{R}^H. \tag{3}$$

For each variate c, its own correction is concatenated with the shared anchor:

$$\mathbf{z}_t^{(c)} = [\boldsymbol{\Delta}_t^{(c)} \| \boldsymbol{\alpha}_t] \in \mathbb{R}^{2H}, \tag{4}$$

where $\|$ denotes concatenation. The concatenation preserves each variate's own correction while adding shared information.

Bottleneck refinement. The joint representation is mapped to a cross-variate refinement through a shared two-layer bottleneck:

$$\boldsymbol{\delta}_t^{(c)} = \mathbf{W}_{\uparrow} \tanh\left(\mathbf{W}_{\downarrow}\mathbf{z}_t^{(c)}\right), \quad \mathbf{W}_{\downarrow} \in \mathbb{R}^{r\times 2H}, \quad \mathbf{W}_{\uparrow} \in \mathbb{R}^{H\times r}. \tag{5}$$

where $\boldsymbol{\delta}_t^{(c)} \in \mathbb{R}^H$ is the refinement for variate c, and r is the bottleneck dimension. Both projections $\mathbf{W}_{\downarrow}, \mathbf{W}_{\uparrow}$ are shared across variates. We simply set $r = C$ by default, giving $O(HC)$ SCR parameters against $O(H^2)$ for unconstrained dense mixing over the horizon. Sharing the anchor and constraining the bottleneck keep the interaction parameter-efficient, which mitigates overfitting under limited test-time supervision.

3.5 Input-Conditioned Spectral Gating

SCR defines how corrections interact across variates. Because temporal dynamics shift over time, the usefulness of shared correction information may also vary. Spectral energy distribution provides useful cues for such changes (Yi et al., 2023; Fan et al., 2024; Piao et al., 2024) and can be computed directly from the current input window. We therefore use a lightweight spectral descriptor to modulate the SCR refinement:

$$\widetilde{\boldsymbol{\Delta}}_t^{(c)} = \boldsymbol{\Delta}_t^{(c)} + g_t^{(c)}\boldsymbol{\delta}_t^{(c)}, \tag{6}$$

where $\mathbf{\Delta}_t^{(c)}$ is the base-adapter correction, $\boldsymbol{\delta}_t^{(c)}$ is the SCR refinement from Equation 5, and $g_t^{(c)} \in (-1, 1)$ provides signed modulation of the refinement for variate c.

Spectral descriptor. For each variate in the input window $\mathbf{X}_t \in \mathbb{R}^{L\times C}$, we subtract its mean and compute a one-sided power spectrum via Fast Fourier Transform (FFT). We then average these across variates to obtain a shared spectral distribution $\overline{\mathbf{P}}_t$ over $K_f = \lfloor L/2 \rfloor + 1$ frequency bins. $\overline{\mathbf{P}}_t$ thus summarizes the window-level temporal dynamics across variates. From $\overline{\mathbf{P}}_t$ we extract four spectral statistics. One is the spectral entropy SE_t, measuring how concentrated the spectral energy is (Powell and Percival, 1979; Inouye et al., 1991). The other three are relative band powers over low-, mid-, and high-frequency bands $(\mathrm{LBR}_t, \mathrm{MBR}_t, \mathrm{HBR}_t)$, measuring how energy is distributed across frequency ranges (Keil et al., 2022). Stacking these gives the spectral descriptor:

$$\mathbf{s}_t = [\mathrm{SE}_t, \mathrm{LBR}_t, \mathrm{MBR}_t, \mathrm{HBR}_t] \in \mathbb{R}^4. \tag{7}$$

Normalization and frequency-band definitions are provided in Appendix B.1.

Per-variate gating. We map the spectral descriptor to a C-dimensional gate:

$$\mathbf{g}_t = \tanh(\mathbf{W}_g \mathbf{s}_t + \mathbf{b}_g) \in (-1, 1)^C. \tag{8}$$

Here $\mathbf{W}_g \in \mathbb{R}^{C\times 4}$ and $\mathbf{b}_g \in \mathbb{R}^C$ are learnable. The descriptor $\mathbf{s}_t$ is shared across variates, so the gates differ only through the rows of $\mathbf{W}_g$ and entries of $\mathbf{b}_g$. The tanh bounds each coefficient to $(-1, 1)$, allowing $g_t^{(c)}$ to provide signed modulation of the SCR refinement for variate c.

3.6 Adaptation Objective

The final adapted forecast is thus $\hat{\mathbf{Y}}_t = \hat{\mathbf{Y}}_t^{\text{base}} + \tilde{\mathbf{\Delta}}_t$. The base-adapter parameters ϕ and CoRe parameters $\Theta = \{\mathbf{W}_\downarrow, \mathbf{W}_\uparrow, \mathbf{W}_g, \mathbf{b}_g\}$ are updated jointly online. At time t, let $\mathcal{B}_t$ denote the adaptation batch whose corresponding targets are available. The parameters are updated by minimizing

$$\mathcal{L}_t = \frac{1}{|\mathcal{B}_t|} \sum_{i\in\mathcal{B}_t} \|\hat{\mathbf{Y}}_i - \mathbf{Y}_i\|_F^2 + \lambda_{\text{reg}} \left(\|\phi\|_2^2 + \|\Theta\|_2^2 \right), \tag{9}$$

where λ_{reg} controls weight decay to mitigate overfitting under limited supervision.

4 Experiments

We evaluate CoRe from four perspectives: overall accuracy and efficiency, the choice of the interaction space, the design of correction-space refinement, and the reliability of online adaptation.

4.1 Experimental Setup

Datasets and evaluation. We evaluate on six multivariate forecasting benchmarks: ETTh1, ETTh2, ETTm1, ETTm2 (Zhou et al., 2021), Weather,[1] and Exchange Rate (Lai et al., 2018), with lookback L=96 and horizons $H \in \{96, 192, 336, 720\}$. We additionally evaluate on high-dimensional datasets Electricity and Traffic (Lai et al., 2018) to assess scalability. We report MSE over the full test stream.

Backbones and baselines. We use seven forecasting backbones: DLinear (Zeng et al., 2023), FreTS (Yi et al., 2023), OLS (Toner and Darlow, 2024), PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024), Informer (Zhou et al., 2021), and MICN (Wang et al., 2023). We compare with the frozen backbone and TSF-TTA methods TAFAS (Kim et al., 2025), and COSA (Im and Kwon, 2026); PETSA (Medeiros et al., 2025) results are in Appendix D.1. All backbones use the same frozen checkpoints, and all TTA modules adapt under delayed supervision. Unless stated otherwise, CoRe uses COSA as the base adapter, r=C, and the same hyperparameters across datasets and backbones. Experiments run on one NVIDIA A100; Appendix B gives implementation details.

[1] `https://www.bgc-jena.mpg.de/wetter/`

4.2 Overall Effectiveness

Table 1: Prediction MSE comparison (lower is better). TAFAS, COSA and CoRe results are averaged over 10 random seeds. For CoRe, standard deviations across seeds are small (maximum: 0.0130), with full standard deviations reported in Table 6 (Appendix B.4). **Best**/second-best highlighting applies to TAFAS vs. COSA vs. CoRe.

		iTransformer				DLinear				FreTS				Informer			
	H	Base	TAFAS	COSA	CoRe	Base	TAFAS	COSA	CoRe	Base	TAFAS	COSA	CoRe	Base	TAFAS	COSA	CoRe
ETTh1	96	0.4580	**0.4381**	0.4638	0.4545	0.4695	**0.4604**	0.4922	0.4916	0.4462	**0.4395**	0.4623	0.4625	0.7466	0.6879	0.6589	**0.6501**
	192	0.5070	0.4911	0.5024	**0.4589**	0.5213	0.5099	0.5371	**0.4961**	0.5021	0.4944	0.5084	**0.4692**	0.9127	0.7905	0.6790	**0.6415**
	336	0.5788	0.5696	0.4956	**0.4661**	0.5659	0.5622	0.5208	**0.4722**	0.5544	0.5560	0.4970	**0.4596**	1.1265	0.9582	0.6727	**0.6302**
	720	0.7067	0.6687	0.5381	**0.5055**	0.6992	0.6685	0.5433	**0.4891**	0.7145	0.6838	0.5616	**0.5052**	1.3596	0.9986	0.7123	**0.6400**
ETTh2	96	0.2731	**0.2701**	0.3050	0.2803	0.2323	**0.2302**	0.2552	0.2426	0.2386	**0.2362**	0.2560	0.2462	0.3510	**0.3419**	0.3747	0.3760
	192	0.3088	0.3042	0.3116	**0.2722**	0.2862	0.2836	0.2739	**0.2275**	0.2867	0.2829	0.2676	**0.2371**	0.4390	0.4311	0.4197	**0.3567**
	336	0.3567	0.3397	0.2552	**0.2245**	0.3252	0.3193	0.2521	**0.2171**	0.3317	0.3221	0.2548	**0.2196**	0.4884	0.4900	0.4127	**0.3463**
	720	0.4297	0.4128	0.2695	**0.2545**	0.4149	0.3939	0.2589	**0.2324**	0.4198	0.3976	0.2516	**0.2336**	0.5206	0.5172	0.3297	**0.3113**
ETTm1	96	0.3740	**0.3538**	0.3799	0.3776	0.3715	**0.3488**	0.3800	0.3881	0.3676	**0.3575**	0.3836	0.3862	0.6475	0.4987	0.4956	**0.4929**
	192	0.4421	**0.4175**	0.4346	0.4184	0.4438	**0.4159**	0.4755	0.4282	0.4325	0.4206	0.4615	**0.4200**	0.6950	0.5472	0.5571	**0.5385**
	336	0.5096	0.4727	0.4825	**0.4585**	0.5183	0.4787	0.5358	**0.4665**	0.5006	0.4816	0.5195	**0.4616**	0.7708	0.6517	0.6031	**0.5771**
	720	0.6083	0.5539	0.5615	**0.5171**	0.5920	0.5495	0.5920	**0.5424**	0.5590	0.5413	0.5773	**0.5200**	0.8308	0.7888	0.6713	**0.6454**
ETTm2	96	0.1791	**0.1722**	0.2226	0.1831	0.1598	**0.1588**	0.1915	0.1644	0.1582	**0.1573**	0.1900	0.1707	0.2839	0.2176	0.2253	**0.2027**
	192	0.2171	0.2153	0.2393	**0.1905**	0.1930	0.1927	0.1961	**0.1652**	0.1922	0.1911	0.1885	**0.1628**	0.4291	0.2732	0.2408	**0.1964**
	336	0.2722	0.2644	0.2473	**0.2058**	0.2324	0.2337	0.1969	**0.1725**	0.2319	0.2308	0.1944	**0.1686**	0.4123	0.3676	0.2660	**0.2114**
	720	0.3321	0.3216	0.2329	**0.1955**	0.3062	0.3053	0.2066	**0.1778**	0.3010	0.2984	0.2110	**0.1799**	0.5985	0.4416	0.2804	**0.2307**
Exchange Rate	96	0.0918	0.0918	0.0913	**0.0832**	0.0913	0.0878	0.0816	**0.0706**	0.0828	0.0790	0.0835	**0.0738**	0.1720	0.1605	0.1187	**0.0985**
	192	0.2086	0.1997	0.1093	**0.0988**	0.1827	0.1736	0.0844	**0.0733**	0.1733	0.1637	0.0959	**0.0806**	0.3080	0.2783	0.1265	**0.1183**
	336	0.3414	0.3119	0.1192	**0.1054**	0.3277	0.3067	0.0867	**0.0804**	0.3241	0.3005	0.0978	**0.0933**	0.6753	0.5693	0.1902	**0.1901**
	720	0.9137	0.9062	0.2228	**0.1957**	0.8383	0.8281	0.1469	**0.1239**	0.8267	0.8142	0.1537	**0.1289**	1.5502	1.5732	0.3524	**0.3403**
Weather	96	0.1708	**0.1625**	0.1753	0.1635	0.1954	0.1828	0.1814	**0.1719**	0.1857	0.1743	0.1739	**0.1617**	0.2168	**0.1762**	0.1940	0.1807
	192	0.2208	0.2086	0.2000	**0.1697**	0.2403	0.2219	0.1970	**0.1711**	0.2311	0.2143	0.1944	**0.1690**	0.3344	0.2567	0.2627	**0.2261**
	336	0.2807	0.2619	0.2439	**0.1736**	0.2918	0.2695	0.2465	**0.1782**	0.2844	0.2639	0.2341	**0.1785**	0.3793	0.3076	0.2773	**0.2202**
	720	0.3572	0.3510	0.2193	**0.1876**	0.3643	0.3550	0.2231	**0.1777**	0.3555	0.3509	0.2084	**0.1763**	0.5370	0.4057	0.2674	**0.2371**

		PatchTST				OLS				MICN			
	H	Base	TAFAS	COSA	CoRe	Base	TAFAS	COSA	CoRe	Base	TAFAS	COSA	CoRe
ETTh1	96	0.4329	**0.4264**	0.4392	0.4371	0.4511	**0.4394**	0.4729	0.4723	0.6244	**0.5836**	0.5837	0.5863
	192	0.4910	0.4820	0.4738	**0.4540**	0.5046	0.4918	0.5173	**0.4763**	0.6543	0.5902	0.5802	**0.5203**
	336	0.5554	0.5463	0.4971	**0.4458**	0.5510	0.5453	0.5041	**0.4568**	0.7352	0.6767	0.5889	**0.5386**
	720	0.7009	0.6826	0.5449	**0.4983**	0.6996	0.6619	0.5443	**0.4886**	0.9336	0.7926	0.6177	**0.5591**
ETTh2	96	0.2355	**0.2349**	0.2506	0.2490	0.2305	**0.2290**	0.2443	0.2389	0.2537	0.2498	0.2538	**0.2476**
	192	0.2855	0.2765	0.2767	**0.2405**	0.2838	0.2824	0.2740	**0.2290**	0.3082	0.3077	0.2923	**0.2430**
	336	0.3174	0.3108	0.2451	**0.2167**	0.3257	0.3202	0.2448	**0.2122**	0.3738	0.3653	0.2828	**0.2486**
	720	0.4063	0.4039	0.2514	**0.2331**	0.4165	0.3978	0.2499	**0.2301**	0.4611	0.4591	0.3058	**0.2746**
ETTm1	96	0.4032	**0.3908**	0.4057	0.4026	0.3706	**0.3495**	0.3780	0.3874	0.4352	**0.3975**	0.4240	0.4232
	192	0.4508	0.4363	0.4417	**0.4183**	0.4438	**0.4148**	0.4730	0.4235	0.4940	0.4587	0.4822	**0.4551**
	336	0.5040	0.4855	0.4989	**0.4509**	0.5177	0.4769	0.5360	**0.4658**	0.5605	0.5136	0.5480	**0.5018**
	720	0.5607	0.5441	0.5696	**0.5225**	0.5927	0.5481	0.5938	**0.5367**	0.6169	0.5726	0.6061	**0.5602**
ETTm2	96	0.1593	**0.1598**	0.1953	0.1721	0.1602	**0.1596**	0.1923	0.1655	0.1697	**0.1724**	0.1909	0.1845
	192	0.2066	0.2052	0.2023	**0.1694**	0.1935	0.1940	0.1978	**0.1671**	0.2198	0.2125	0.2307	**0.1832**
	336	0.2480	0.2530	0.2100	**0.1765**	0.2332	0.2348	0.1985	**0.1726**	0.2603	0.2590	0.2238	**0.1871**
	720	0.3057	0.3045	0.2058	**0.1728**	0.3066	0.3069	0.2067	**0.1787**	0.3352	0.3305	0.2381	**0.1991**
Exchange Rate	96	0.0858	0.0817	0.0798	**0.0739**	0.0814	0.0789	0.0795	**0.0741**	0.1114	0.1100	0.0920	**0.0779**
	192	0.1880	0.1760	0.1029	**0.0897**	0.1727	0.1643	0.0953	**0.0812**	0.2081	0.1864	0.0855	**0.0739**
	336	0.3383	0.3088	0.1094	**0.1000**	0.3225	0.3014	0.1003	**0.0947**	0.3941	0.3644	0.1299	**0.1091**
	720	0.8327	0.8330	0.1743	**0.1348**	0.8356	0.8202	0.1490	**0.1259**	1.4477	0.8503	0.2320	**0.2211**
Weather	96	0.1727	0.1702	0.1664	**0.1622**	0.1958	0.1831	0.1812	**0.1704**	0.1724	0.1723	0.1701	**0.1646**
	192	0.2197	0.2145	0.1939	**0.1724**	0.2405	0.2220	0.1960	**0.1719**	0.2235	0.2215	0.2078	**0.1608**
	336	0.2763	0.2655	0.2834	**0.1810**	0.2922	0.2699	0.2440	**0.1777**	0.2815	0.2775	0.2107	**0.1698**
	720	0.3517	0.3350	0.2056	**0.1765**	0.3645	0.3550	0.2252	**0.1785**	0.3507	0.3682	0.2013	**0.1711**

Table 1 reports MSE across all 168 settings (7 backbones $\times$ 6 datasets $\times$ 4 horizons). Averaged across all settings, CoRe reduces MSE by 25.82% relative to the frozen backbone and by 10.57% relative to COSA. The largest dataset-level reductions over COSA occur on Weather, ETTm2, and Exchange Rate. Gains are more pronounced at medium-to-long horizons: 5.27% at H=96, 13.34% at H=336, and 11.53% at H=720. The few underperforming cases occur mainly at H=96, where the benefit of correction-space interaction may be smaller. CoRe applies larger corrections where the frozen backbone is less accurate (167/168 settings) and yields larger relative improvements there (166/168); see Appendix C.3.

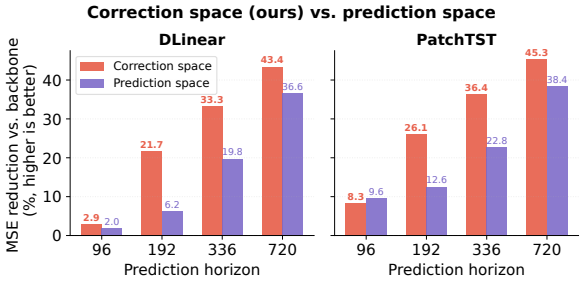

Figure 5: Relative MSE reduction over the frozen backbone for correction-space and prediction-space interaction under the same bottleneck architecture.

Table 2: Parameter-matched comparison (DLinear, 6 datasets × 4 horizons; seed = 42). Exchange Rate reported separately due to its high non-stationarity.

Method	Params	Avg. MSE	Exch. MSE
COSA	1,627,930	0.2974	0.0983
COSA+MLP	1,637,575	0.2945	0.0986
CoRe	1,637,898	0.2660	0.0849

Table 3: Results on large-scale datasets Electricity and Traffic, averaged over 4 horizons and 10 seeds. All methods use a fixed bottleneck rank r=16.

Model	Dataset	Backbone	COSA	CoRe
DLinear	Electricity	0.2258	0.2008	**0.1860**
	Traffic	0.6502	0.6352	**0.6325**
MICN	Electricity	0.1962	0.1619	**0.1458**
	Traffic	0.5275	0.4875	**0.4694**

Runtime overhead. Averaged over six datasets and four horizons with DLinear, CoRe increases the per-adaptation-step runtime from 6.28 ms to 6.76 ms over COSA (+0.48 ms; +7.7%). Thus, the accuracy gains come with modest additional computation; Appendix G.3 provides the full timing breakdown.

4.3 Why Correction-Space Interaction?

Correction vs. prediction space. Correction-space interaction shows a clear advantage over prediction-space interaction at medium-to-long horizons. Figure 5 controls for the interaction architecture, changing only the space in which cross-variate interaction is performed. While the two variants are close at H=96, at H=336 correction-space interaction improves DLinear by 33.3%, compared with 19.8% for prediction-space interaction. The same overall horizon-dependent pattern holds across all seven backbones (Appendix C.1). Together with Proposition 1, this controlled comparison supports the empirical advantage of correction-space over prediction-space interaction, particularly at medium-to-long horizons.

Parameter-matched comparison. The advantage of CoRe is not explained by additional trainable capacity alone. Table 2 compares CoRe with a parameter-matched independent COSA+MLP, which adds per-variate capacity without cross-variate interaction. COSA+MLP improves over COSA by only 0.97% on average (single seed; Table 2), compared with 10.55% for CoRe. The contrast is particularly clear on Exchange Rate, where cross-variate relationships vary substantially over time (Appendix G.1). These results support the benefit of structured cross-variate refinement under a comparable parameter budget.

4.4 Structured and Selective Refinement

Shared and low-dimensional correction structure. As shown in Table 4, the full SCR design improves over COSA by 9.7%, compared with 7.1% without the shared anchor and 7.2% with unconstrained $C \times C$ mixing. These results support the benefit of the shared-anchor structure over the evaluated alternatives. Moreover, the leading singular mode captures 73% of the correction energy on average, revealing pronounced low-dimensional structure. With r=16 $\ll C$, SCR retains improvements over COSA on both Electricity (C=321) and Traffic (C=862), showing that its effectiveness does not require the bottleneck rank to scale with the number of variates (Table 3). Together, these results support the shared-anchor and parameter-efficient bottleneck designs in SCR. See Appendix E.1 for the rank and scalability analysis.

Table 4: Ablation study of CoRe components. Average MSE over three backbones (DLinear, PatchTST, MICN), six datasets, four horizons, and 10 seeds. **vs. COSA**: relative MSE reduction over COSA. *Loss-trend gate*: COSA's adaptive learning-rate signal repurposed as the SCR gate. *Fixed gate* (g=1): uniform cross-variate refinement without input-dependent modulation. $\mathbf{s}_t$: Spectral descriptors for gating (Section 3.5).

Method	DLinear	PatchTST	MICN	Avg. MSE	vs. COSA
COSA	0.2981	0.2927	0.3241	0.3050	–
+SCR (no anchor; per-variate bottleneck)	0.2760	0.2721	0.3018	0.2833	−7.1%
+SCR (no anchor/bottleneck; full $C\times C$ mixing)	0.2745	0.2706	0.3044	0.2832	−7.2%
+SCR (loss-trend gate)	0.2712	0.2657	0.2963	0.2777	−9.0%
+SCR (fixed gate, g=1)	0.2703	0.2651	0.2954	0.2769	−9.2%
CoRe	0.2675	0.2646	0.2942	0.2754	−9.7%

Spectral gating. The spectral signal is computed directly from the current input window and is therefore available from the first test batch, without requiring the historical loss or model-state statistics used by loss-trend or embedding-drift signals. Table 4 shows that with spectral gating, CoRe improves over COSA by 9.7%, compared with 9.2% for a fixed gate and 9.0% for loss-trend gating. Spectral gating thus provides a modest additional gain, while most of the improvement comes from SCR. Appendix E.2 contains the full ablations.

4.5 Adaptation Reliability

Online updates aim to improve performance but can also degrade it (Niu et al., 2023). Reliable adaptation therefore matters alongside average accuracy, particularly under delayed supervision where corrective feedback is not immediately available. CoRe's accuracy gains are accompanied by more reliable online adaptation. We assess reliability on three representative backbones (DLinear, PatchTST, and MICN) using four complementary measures: average correction regret ($\overline{\mathcal{R}}_T$), the frequency of harmful adaptation (p_{neg}), the severity of harmful adaptation ($\overline{\mathcal{R}}_T^+$), and the number of settings in which adaptation underperforms the frozen backbone (n_{worse}). Table 5 shows that CoRe improves over COSA across all four reliability measures. Together, these results show that CoRe makes online adaptation more consistently beneficial relative to the frozen backbone. Appendix F provides the metric definitions and detailed breakdowns.

Table 5: Online adaptation reliability over 18 backbone-dataset combinations (3 backbones × 6 datasets), each averaged over $H \in \{96, 192, 336, 720\}$. Lower is better; darker blue indicates better performance within each column.

Method	$\overline{\mathcal{R}}_T\downarrow$	$p_{\text{neg}}\downarrow$	$\overline{\mathcal{R}}_T^+\downarrow$	$n_{\text{worse}}\downarrow$
TAFAS	−0.010	0.446	0.0138	8
COSA	−0.087	0.277	0.0127	5
CoRe	**−0.117**	**0.232**	**0.0108**	**2**

5 Conclusion

We presented CoRe, a multivariate TSF-TTA framework centered on cross-variate interaction in the *correction space*. By performing interaction on adapter-produced corrections rather than backbone predictions, CoRe avoids directly mixing uncorrected backbone errors during cross-variate interaction. The principle is instantiated through structured Shared-anchor Correction Refinement, complemented by lightweight input-conditioned spectral gating that modulates the SCR refinement. Our analysis establishes the structural distinction between prediction-space and correction-space interaction. Controlled comparisons demonstrate the empirical benefits of this design choice, while parameter-matched controls and experiments across alternative adapters further support the effectiveness of CoRe. Across seven backbones, six datasets, and four horizons, CoRe improves forecasting accuracy in most evaluated settings with modest computational overhead.

AI Use Statement

In this work, we used generative AI tools to polish the writing and do a pre-submission review of the manuscript, which included feedback on the soundness and sufficiency of the experiments (e.g., whether the controls, baselines, and ablations adequately support the claims) and the mathematical proof. This feedback was used only to revise the presentation. All research ideas, methods, theoretical results, experiments, and interpretations are the authors' own. The authors take full responsibility for the final content of this work.

Reproducibility Statement

We provide detailed documentation to support the reproducibility of our method, experiments, and theoretical analysis. The CoRe architecture and adaptation objective are described in Section 3. The datasets, baselines, and evaluation protocol are specified in Section 4.1. Appendix B provides the implementation details, including the online adaptation procedure, initialization, hyperparameters, baseline configurations, and random-seed evaluation. Appendix C.1 provides the controlled comparison of interaction spaces, and additional baseline experiments are documented in Appendix D. Appendix C.2 details the implementations with alternative base adapters. Additional ablations and scalability experiments are reported in Appendix E.1. Appendix F provides the definitions and detailed evaluation of online adaptation reliability. Appendix H.2 presents the formal analysis of Proposition 1, and Appendix H.2.1 provides the corresponding proof. The accompanying data and code are available at `https://anonymous.4open.science/r/CoRe-TTA-CD98`.

Ethics Statement

This work studies test-time adaptation for multivariate time-series forecasting using publicly available benchmark datasets. Our experiments do not involve human subjects or the collection of new personal data. While improved forecasting methods may support applications in consequential domains, their reliability under distribution shifts should be carefully evaluated before deployment. We report both the benefits and limitations of our method to support transparent interpretation of the results.

References

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

L. Bai, L. Yao, C. Li, X. Wang, and C. Wang. Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in neural information processing systems*, 33: 17804–17815, 2020.

S.-J. Bu and S.-B. Cho. Time series forecasting with multi-headed attention-based deep learning for residential energy consumption. *Energies*, 13(18):4722, 2020.

M. Chen, L. Shen, H. Fu, Z. Li, J. Sun, and C. Liu. Calibration of time-series forecasting: Detecting and adapting context-driven distribution shift. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 341–352, 2024.

D. Cheng, F. Yang, S. Xiang, and J. Liu. Financial time series forecasting with multi-modality graph neural network. *Pattern Recognition*, 121:108218, 2022.

P. Christou, S. Chen, X. Chen, and P. Dube. Test time learning for time series forecasting. *arXiv preprint arXiv:2409.14012*, 2024.

S. Deng, Z. Xiao, and M. de Rijke. Adaptive latent decomposition for domain generalization in time series forecasting. *ACM Transactions on Knowledge Discovery from Data*, 2026.

V. Ekambaram, A. Jati, N. Nguyen, P. Sinthong, and J. Kalagnanam. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining*, pages 459–469, 2023.

W. Fan, P. Wang, D. Wang, D. Wang, Y. Zhou, and Y. Fu. Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 7522–7529, 2023.

W. Fan, K. Yi, H. Ye, Z. Ning, Q. Zhang, and N. An. Deep frequency derivative learning for non-stationary time series forecasting. *arXiv preprint arXiv:2407.00502*, 2024.

S. Grover and A. Etemad. Shift-aware test time adaptation and benchmarking for time-series forecasting. In *Second Workshop on Test-Time Adaptation: Putting Updates to the Test! at ICML 2025*, 2025.

L. Han, H.-J. Ye, and D.-C. Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7129–7142, 2024.

H. He, Q. Zhang, K. Yi, X. Xue, S. Wang, L. Hu, and L. Cao. Robust multivariate time series forecasting against intraseries and interseries transitional shift. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

J. Im and H.-Y. Kwon. COSA: Context-aware output-space adapter for test-time adaptation in time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=L7Z5wBMPrW`.

T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano. Quantification of eeg irregularity by use of the entropy of the power spectrum. *Electroencephalography and clinical neurophysiology*, 79(3):204–210, 1991.

X. Jin, Y. Park, D. Maddix, H. Wang, and Y. Wang. Domain adaptation for time series forecasting via attention sharing. In *International Conference on Machine Learning*, pages 10280–10297. PMLR, 2022.

Z. Karevan and J. A. Suykens. Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9, 2020.

A. Keil, E. M. Bernat, M. X. Cohen, M. Ding, M. Fabiani, G. Gratton, E. S. Kappenman, E. Maris, K. E. Mathewson, R. T. Ward, et al. Recommendations and publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series. *Psychophysiology*, 59(5):e14052, 2022.

A. Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109(1):604–615, 1934.

H. Kim, S. Kim, J. Mok, and S. Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 17868–17876, 2025.

T. Kim, J. Kim, Y. Tae, C. Park, J.-H. Choi, and J. Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International conference on learning representations*, 2021.

G. Lai, W.-C. Chang, Y. Yang, and H. Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pages 95–104, 2018.

J. Langford, A. Smola, and M. Zinkevich. Slow learners are fast. *arXiv preprint arXiv:0911.0491*, 2009.

J. Liang, R. He, and T. Tan. A comprehensive survey on test-time adaptation under distribution shifts. *International Journal of Computer Vision*, 133(1):31–64, 2025.

X. Lin, Z. Zeng, T. Wei, Z. Liu, H. Tong, et al. Cats: Mitigating correlation shift for multivariate time series classification. *arXiv preprint arXiv:2504.04283*, 2025.

Y. Liu, H. Wu, J. Wang, and M. Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in neural information processing systems*, 35:9881–9893, 2022.

Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=JePfAI8fah.

Z. Liu, M. Cheng, Z. Li, Z. Huang, Q. Liu, Y. Xie, and E. Chen. Adaptive normalization for non-stationary time series forecasting: A temporal slice perspective. *Advances in Neural Information Processing Systems*, 36:14273–14292, 2023.

V. Losing, B. Hammer, and H. Wersing. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 275:1261–1274, 2018.

H. R. Medeiros, H. Sharifi-Noghabi, G. L. Oliveira, and S. Irandoust. Accurate parameter-efficient test-time adaptation for time series forecasting. *arXiv preprint arXiv:2506.23424*, 2025.

Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=Jbdc0vTOcol.

S. Niu, J. Wu, Y. Zhang, Z. Wen, Y. Chen, P. Zhao, and M. Tan. Towards stable test-time adaptation in dynamic wild world. *arXiv preprint arXiv:2302.12400*, 2023.

X. Piao, Z. Chen, Y. Dong, Y. Matsubara, and Y. Sakurai. Frednormer: Frequency domain normalization for non-stationary time series forecasting. *arXiv preprint arXiv:2410.01860*, 2024.

G. Powell and I. Percival. A spectral entropy method for distinguishing regular and irregular motion of hamiltonian systems. *Journal of Physics A: Mathematical and General*, 12(11): 2053–2071, 1979.

X. Qiu, H. Cheng, X. Wu, J. Lu, J. Hu, C. Guo, C. S. Jensen, and B. Yang. A comprehensive survey of deep learning for multivariate time series forecasting: A channel strategy perspective. *arXiv preprint arXiv:2502.10721*, 2025a.

X. Qiu, X. Wu, Y. Lin, C. Guo, J. Hu, and B. Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pages 1185–1196, 2025b.

Y. Sun, X. Wang, Z. Liu, J. Miller, A. Efros, and M. Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020.

W. Toner and L. Darlow. An analysis of linear time series forecasting models. *arXiv preprint arXiv:2403.14587*, 2024.

D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

H. Wang, J. Peng, F. Huang, J. Wang, J. Chen, and Y. Xiao. Micn: Multi-scale local and global context modeling for long-term series forecasting. In *The eleventh international conference on learning representations*, 2023.

H. Wang, R. Xu, G. Kementzidis, K. Cho, S. R. Villarreal, and Y. Deng. Towards principled test-time adaptation for time series forecasting. *arXiv preprint arXiv:2605.17250*, 2026.

M. Wang and W. Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312: 135–153, 2018.

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. *Advances in neural information processing systems*, 37:469–498, 2024.

Q. Wen, W. Chen, L. Sun, Z. Zhang, L. Wang, R. Jin, T. Tan, et al. Onenet: Enhancing time series forecasting models under concept drift by online ensembling. *Advances in Neural Information Processing Systems*, 36:69949–69980, 2023.

N. Wiener. Generalized harmonic analysis. *Acta mathematica*, 55(1):117–258, 1930.

G. Woo, C. Liu, A. Kumar, C. Xiong, S. Savarese, and D. Sahoo. Unified training of universal time series forecasting transformers. In *Forty-first International Conference on Machine Learning*, 2024.

F. Xiong, Z. Xie, Y. Sun, H. Wang, and J. Lin. Seed: Spectral entropy-guided evaluation of spatial-temporal dependencies for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 27153–27161, 2026.

W. Ye, S. Deng, Q. Zou, and N. Gui. Frequency adaptive normalization for non-stationary time series forecasting. *Advances in Neural Information Processing Systems*, 37:31350–31379, 2024.

K. Yi, Q. Zhang, W. Fan, S. Wang, P. Wang, H. He, N. An, D. Lian, L. Cao, and Z. Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

A. Zeng, M. Chen, L. Zhang, and Q. Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11121–11128, 2023.

Y. Zhang, W. Chen, Z. Zhu, D. Qin, L. Sun, X. Wang, Q. Wen, Z. Zhang, L. Wang, and R. Jin. Addressing concept shift in online time series forecasting: Detect-then-adapt. *arXiv preprint arXiv:2403.14949*, 2024.

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. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 11106–11115, 2021.

A Extended Related Work

Cross-variate modeling. Multivariate forecasting approaches are commonly described as channel-independent, channel-dependent, or partially coupled architectures (Qiu et al., 2025a). Channel-independent models avoid explicit cross-variate interaction, whereas channel-dependent models couple variables through mechanisms such as attention or frequency-domain mixing (Nie et al., 2023; Zeng et al., 2023; Zhou et al., 2021; Liu et al., 2024; Yi et al., 2023). Recent work also explores structured or selective interaction (Ekambaram et al., 2023; Han et al., 2024; Qiu et al., 2025b). Empirical studies show that the effectiveness of cross-variate modeling varies across data characteristics (Toner and Darlow, 2024; Chen et al., 2024; He et al., 2025; Lin et al., 2025).

Distribution shift and adaptation. Normalization-based methods such as RevIN, Non-Stationary Transformer, Dish-TS, SAN, and FAN modify inputs or representations to improve robustness to non-stationarity (Kim et al., 2021; Liu et al., 2022; Fan et al., 2023; Liu et al., 2023; Ye et al., 2024). Other work studies online learning (Wen et al., 2023; Zhang et al., 2024), domain adaptation (Jin et al., 2022), domain generalization (Deng et al., 2026), and test-time learning (Christou et al., 2024; Medeiros et al., 2025) under different supervision and optimization settings. CoRe focuses specifically on the interaction space for cross-variate adaptation under delayed-supervision TSF-TTA.

B Implementation Details

B.1 Spectral Descriptor Computation

For an input window $\mathbf{X}_t \in \mathbb{R}^{L\times C}$, we first remove the temporal mean of each variate and compute its one-sided power spectrum using the real-valued Fast Fourier Transform:

$$P_t^{(c)}(f) = \left|\mathrm{rFFT}\left(X_t^{(c)} - \bar{X}_t^{(c)}\right)[f]\right|^2, \qquad f = 0, \ldots, K_f - 1, \tag{10}$$

where $K_f = \lfloor L/2 \rfloor + 1$. Mean subtraction removes the DC component so that the descriptors characterize temporal variation rather than the absolute level of the series.

We average the spectra across variates and normalize them into a distribution over frequency bins:

$$P_t(f) = \frac{1}{C}\sum_{c=1}^{C} P_t^{(c)}(f), \qquad \bar{P}_t(f) = \frac{P_t(f) + \varepsilon}{\sum_{f'} P_t(f') + K_f\varepsilon}. \tag{11}$$

The smoothing constant $\varepsilon > 0$ avoids undefined logarithms, and $\sum_f \bar{P}_t(f) = 1$.

Spectral entropy. We compute normalized spectral entropy as

$$\mathrm{SE}_t = -\frac{\sum_{f=0}^{K_f-1} \bar{P}_t(f)\log \bar{P}_t(f)}{\log K_f} \in [0, 1]. \tag{12}$$

Spectral entropy summarizes how concentrated or dispersed spectral energy is across frequencies (Powell and Percival, 1979; Inouye et al., 1991).

Band-energy ratios. To complement spectral concentration with coarse information about where the energy is located, we partition the frequency range into low, middle, and

high thirds:

$$\mathrm{LBR}_t = \sum_{f=0}^{\lfloor K_f/3 \rfloor} \bar{P}_t(f), \tag{13}$$

$$\mathrm{MBR}_t = \sum_{f=\lfloor K_f/3 \rfloor+1}^{\lfloor 2K_f/3 \rfloor} \bar{P}_t(f), \tag{14}$$

$$\mathrm{HBR}_t = \sum_{f=\lfloor 2K_f/3 \rfloor+1}^{K_f-1} \bar{P}_t(f). \tag{15}$$

By construction,

$$\mathrm{LBR}_t + \mathrm{MBR}_t + \mathrm{HBR}_t = 1. \tag{16}$$

The complete descriptor is

$$\mathbf{s}_t = [\mathrm{SE}_t, \mathrm{LBR}_t, \mathrm{MBR}_t, \mathrm{HBR}_t] \in \mathbb{R}^4. \tag{17}$$

The entropy term describes spectral concentration, while the three band ratios provide complementary information about coarse frequency allocation.

B.2 Full Online Adaptation Procedure

Algorithm 1 summarizes the complete streaming procedure. The forecasting backbone remains frozen throughout deployment. Bias terms in the SCR projections are omitted for notational simplicity. The base-adapter parameters ϕ and the parameters introduced by CoRe,

$$\Theta = \{\mathbf{W}_\downarrow, \mathbf{W}_\uparrow, \mathbf{W}_g, \mathbf{b}_g\}, \tag{18}$$

are updated only using forecast–target pairs whose targets have already been revealed.

At adaptation time t, let $\mathcal{B}_t$ denote the adaptation batch whose corresponding targets are available. This definition is identical to the leakage-free objective in Section 3.6.

Initialization. The SCR projection matrices $\mathbf{W}_\downarrow$ and $\mathbf{W}_\uparrow$ use Xavier-uniform initialization with gain 0.01 and zero biases. The spectral projection $\mathbf{W}_g$ is initialized to zero and $\mathbf{b}_g = -\mathbf{1}$, giving

$$g_t^{(c)} = \tanh(-1) \approx -0.76 \tag{19}$$

at initialization. Thus, the initial SCR contribution is negatively modulated rather than initialized at a large unconstrained value.

Hyperparameters. Unless otherwise specified, we follow the optimization protocol and hyperparameters of the corresponding base adapter. The main CoRe-specific architectural choice is the SCR bottleneck rank r. We set $r = C$ on the six standard datasets. For the high-dimensional datasets (Electricity, Traffic) we use $r = 16 \ll C$, with discussion in Appendix E.1. The spectral-gate bias is initialized as $\mathbf{b}_g = -\mathbf{1}$.

B.3 Baseline Configuration Details

For reproducibility, we summarize the baseline configurations used in our experiments. Unless noted otherwise, we follow the default settings in the official repositories.

TAFAS (Kim et al., 2025) Following the official configuration,[2] we employ OPTIMIZING_METHOD = adam with BASE_LR = 0.005, WEIGHT_DECAY = 0.0001, MOMENTUM = 0.9, NESTEROV = True, and DAMPENING = 0.0. Method settings are PAAS = True, PERIOD_N = 1, BATCH_SIZE = 64, STEPS = 1, ADJUST_PRED = True, CALI_MODULE = True, GATING_INIT = 0.01, HIDDEN_DIM = 128, and GCM_VAR_WISE = True. The adapted module is the calibration module only (TTA.MODULE_NAMES_TO_ADAPT = 'cali').

[2] https://github.com/kimanki/TAFAS/config.py

Algorithm 1: CoRe Full Procedure

Input : test stream $\{\mathbf{X}_t\}$; frozen backbone f_θ; base adapter $\mathcal{A}_\phi$; CoRe parameters Θ; buffer size B; adaptation steps S

Output : adapted forecasts $\{\hat{\mathbf{Y}}_t\}$

Initialize ϕ and Θ

Initialize optional auxiliary state used by the base adapter

for *each forecast time* t **do**

- $\hat{\mathbf{Y}}_t^{\text{base}} \leftarrow f_\theta(\mathbf{X}_t)$
- Compute spectral descriptor $\mathbf{s}_t$ from $\mathbf{X}_t$
- $\mathbf{g}_t \leftarrow \tanh\left(\mathbf{W}_g \mathbf{s}_t + \mathbf{b}_g\right)$
- $\boldsymbol{\Delta}_t \leftarrow \mathcal{A}_\phi\left(\hat{\mathbf{Y}}_t^{\text{base}}, \mathbf{I}_t\right)$
- $\boldsymbol{\alpha}_t \leftarrow \frac{1}{C}\sum_{c=1}^{C} \boldsymbol{\Delta}_t^{(c)}$
- **for** $c = 1$ **to** C **do**
  - $\mathbf{z}_t^{(c)} \leftarrow [\boldsymbol{\Delta}_t^{(c)} \| \boldsymbol{\alpha}_t]$
  - $\boldsymbol{\delta}_t^{(c)} \leftarrow \mathbf{W}_\uparrow \tanh\left(\mathbf{W}_\downarrow \mathbf{z}_t^{(c)}\right)$
- $\widetilde{\boldsymbol{\Delta}}_t^{(c)} \leftarrow \boldsymbol{\Delta}_t^{(c)} + g_t^{(c)} \boldsymbol{\delta}_t^{(c)} \quad \forall c$
- $\hat{\mathbf{Y}}_t \leftarrow \hat{\mathbf{Y}}_t^{\text{base}} + \widetilde{\boldsymbol{\Delta}}_t$
- Form the adaptation batch $\mathcal{B}_t$ from samples whose corresponding targets are available
- Update the optional base-adapter state $\mathbf{I}_t$ using only revealed observations
- **if** $|\mathcal{B}_t| > 0$ **then**
  - **for** $s = 1$ **to** S **do**
    - $\mathcal{L}_t \leftarrow \frac{1}{|\mathcal{B}_t|} \sum_{\tau \in \mathcal{B}_t} \left\| \hat{\mathbf{Y}}_\tau - \mathbf{Y}_\tau \right\|_F^2 + \lambda_{\text{reg}} \left(\|\phi\|_2^2 + \|\Theta\|_2^2 \right)$
    - Clip gradients to norm 1.0
    - Update ϕ and Θ according to the base-adapter optimization protocol

COSA (Im and Kwon, 2026) Following the official configuration,[3] we use OPTIMIZING_METHOD = adam with BASE_LR = 0.005, WEIGHT_DECAY = 0.0001, MOMENTUM = 0.9, NESTEROV = True, and DAMPENING = 0.0. Method settings are BATCH_SIZE = 25, STEPS = 20, BUFFER_SIZE = 10, BUFFER_CONTEXT_SIZE = 5, ADAPT_FREQUENCY = 50, FAST_ADAPTATION = True, ADAPTIVE_LR = True, MAX_LR = 0.005, MIN_LR = 0.0001, MOMENTUM_FACTOR = 0.9, CONVERGENCE_THRESHOLD = 1e-4, VAR_WISE_GATING = True, ADAPTER_LAYERS = 1, HIDDEN_DIM = 64, PER_BATCH_LR_RESET = True, PAAS = False, and PERIOD_N = 1. The adapted module is the calibration module only (TTA.MODULE_NAMES_TO_ADAPT = 'cali').

We configure the adaptation steps in CoRe following the established settings of the base adapter, which we also empirically found to be optimal.

B.4 Variance Across Random Seeds

Table 6 reports the standard deviation of CoRe across 10 random seeds for the 168 main experimental settings. The observed variance is small relative to the average performance differences reported in the main results.

[3] https://github.com/bigbases/COSA_ICLR2026/config.py

Table 6: Standard deviation ($\pm 1\sigma$, over 10 random seeds) of CoRe MSE.

		iTransformer				DLinear				FreTS				Informer			
	H	96	192	336	720	96	192	336	720	96	192	336	720	96	192	336	720
Dataset	ETTh1	.0034	.0036	.0032	.0024	.0032	.0066	.0024	.0015	.0036	.0050	.0042	.0041	.0032	.0048	.0029	.0023
	ETTh2	.0054	.0088	.0042	.0033	.0036	.0036	.0027	.0019	.0053	.0074	.0033	.0024	.0122	.0115	.0081	.0039
	ETTm1	.0034	.0041	.0036	.0040	.0042	.0093	.0070	.0093	.0028	.0032	.0038	.0053	.0128	.0055	.0046	.0047
	ETTm2	.0046	.0043	.0029	.0011	.0016	.0020	.0026	.0031	.0033	.0025	.0025	.0027	.0061	.0034	.0044	.0030
	Exchange	.0047	.0018	.0023	.0036	.0041	.0012	.0028	.0041	.0033	.0021	.0024	.0048	.0021	.0029	.0096	.0058
	Weather	.0020	.0027	.0024	.0020	.0040	.0039	.0024	.0014	.0025	.0012	.0025	.0025	.0012	.0047	.0028	.0022

		PatchTST				OLS				MICN			
	H	96	192	336	720	96	192	336	720	96	192	336	720
Dataset	ETTh1	.0021	.0040	.0019	.0038	.0021	.0040	.0031	.0026	.0055	.0037	.0036	.0028
	ETTh2	.0033	.0053	.0031	.0017	.0044	.0048	.0028	.0024	.0054	.0024	.0027	.0031
	ETTm1	.0033	.0046	.0041	.0052	.0047	.0130	.0050	.0054	.0041	.0043	.0055	.0116
	ETTm2	.0023	.0014	.0017	.0015	.0029	.0019	.0032	.0036	.0033	.0020	.0023	.0026
	Exchange	.0032	.0025	.0023	.0019	.0023	.0020	.0024	.0039	.0037	.0008	.0030	.0046
	Weather	.0021	.0030	.0019	.0016	.0043	.0030	.0027	.0022	.0031	.0020	.0031	.0012

C Additional Analysis of Correction-Space Interaction

C.1 Full Correction-Space vs. Prediction-Space Comparison

We provide the full comparison between correction-space and prediction-space interaction across all seven backbones. This controlled experiment isolates the interaction space while keeping the refinement architecture and base-adapter correction unchanged. Figure 6 reports the resulting MSE reductions.

At $H = 96$, the two interaction schemes are generally comparable. The separation becomes clearer at medium-to-long horizons, where correction-space interaction obtains larger MSE reductions across the evaluated backbones. At $H = 192$, $H = 336$, and $H = 720$, the largest observed margins between the two variants are 16.2%, 15.0%, and 9.5%, respectively. Together with Proposition 1, these results show that the choice of interaction space has a substantial empirical consequence across backbone architectures.

C.2 Generality across Base Adapters

We examine whether correction-space interaction remains effective beyond the COSA base adapter used in the main experiments. We evaluate CoRe with two structurally different correction interfaces: TAFAS and a standalone MLP adapter. Together with COSA, these experiments cover a context-conditioned adapter, a gated calibration module, and a minimal additive correction adapter.

C.2.1 CoRe on TAFAS

TAFAS applies an output Gated Calibration Module (GCM):

$$\mathrm{GCM}(\hat{\mathbf{Y}}_t) = \hat{\mathbf{Y}}_t + \tanh(\boldsymbol{\alpha}) \circ \underbrace{\left(\{\mathbf{W}^c \hat{\mathbf{Y}}_t^c\}_{c=1}^{C} + \mathbf{b}\right)}_{\text{correction}}, \tag{20}$$

where $\mathbf{W}^c \in \mathbb{R}^{H \times H}$, $\mathbf{b} \in \mathbb{R}^{H \times C}$, and $\boldsymbol{\alpha} \in \mathbb{R}^C$.

We apply SCR to the additive correction inside the GCM before its original multiplicative gate is applied. Across six datasets and four horizons using DLinear, this yields a 4.49% average MSE reduction over TAFAS (Table 7).

C.2.2 CoRe on a Standalone MLP Adapter

We also construct a minimal standalone correction adapter:

$$\boldsymbol{\Delta}_t = \mathrm{MLP}_{\text{base}}\left(\hat{\mathbf{Y}}_t^{\text{base}}\right) \in \mathbb{R}^{H \times C}. \tag{21}$$

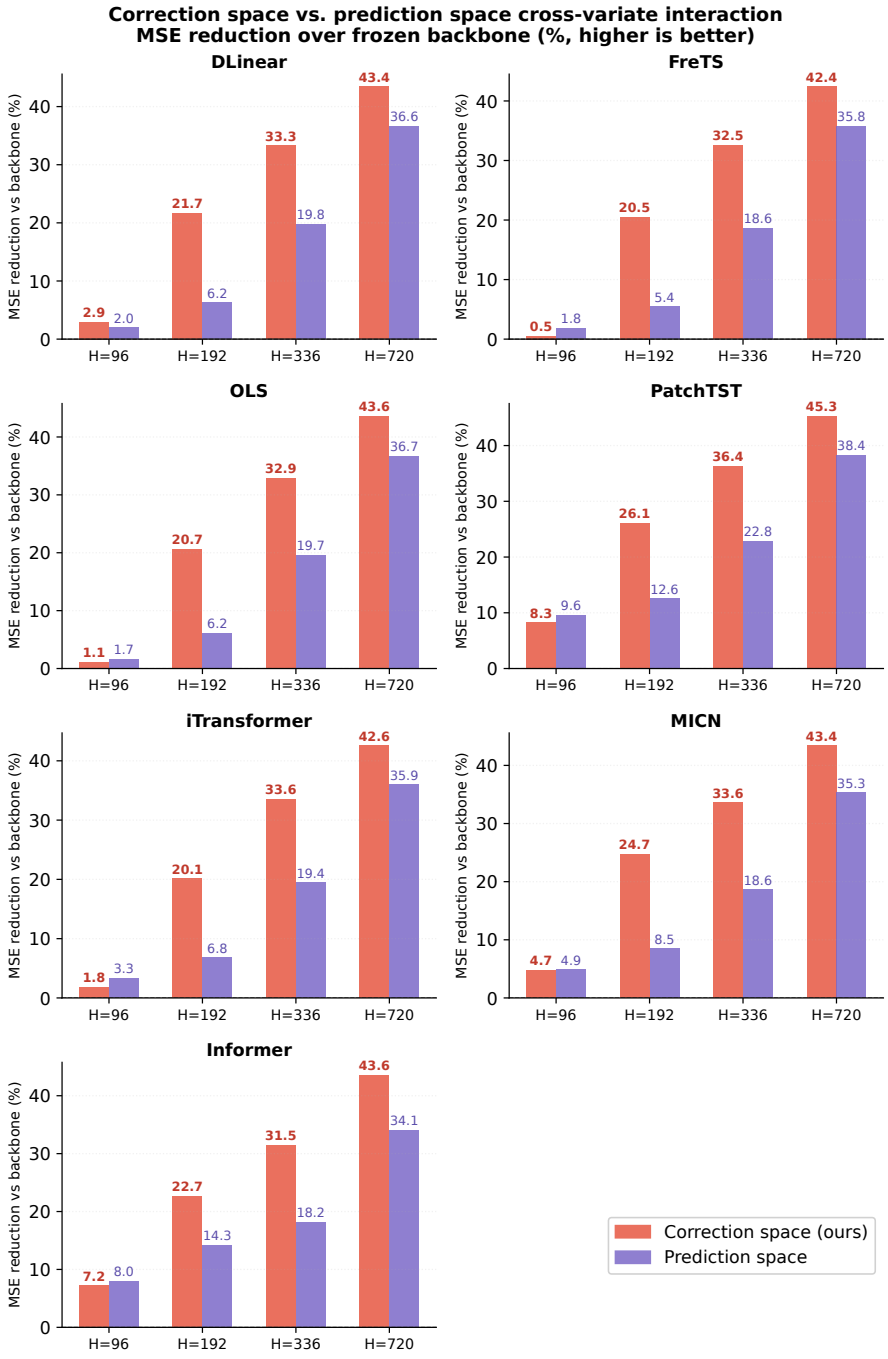

Figure 6: Controlled comparison of correction-space and prediction-space interaction across all seven backbones.

The base adapter is a two-layer MLP with GELU activation and a zero-initialized output layer. It contains no context buffer, adaptive learning-rate mechanism, or periodicity-aware scheduling.

SCR then operates on $\mathbf{\Delta}_t$ exactly as in Section 3.4. Across seven backbones, six datasets, and four horizons, the resulting method improves 156/168 settings over the standalone adapter, with a 3.70% average MSE reduction (Table 8). The remaining negative differences are small in absolute magnitude ($|\Delta\text{MSE}| < 0.005$).

Table 7: MSE of TAFAS and TAFAS+CoRe (ours). **Bold** = better result. Avg. Gain = relative MSE reduction averaged over $H \in \{96, 192, 336, 720\}$.

Backbone	Dataset	$H = 96$		$H = 192$		$H = 336$		$H = 720$		Avg. Gain
		TAFAS	+CoRe	TAFAS	+CoRe	TAFAS	+CoRe	TAFAS	+CoRe	
DLinear	ETTh1	0.4604	**0.4604**	0.5099	**0.5020**	0.5622	**0.5515**	0.6685	**0.6548**	+1.83%
	ETTh2	0.2302	**0.2039**	0.2836	**0.2710**	**0.3193**	0.3229	**0.3939**	0.4140	+2.41%
	ETTm1	0.3488	**0.3342**	0.4159	**0.3924**	0.4787	**0.4681**	**0.5495**	0.5539	+2.81%
	ETTm2	0.1588	**0.1505**	0.1927	**0.1841**	0.2337	**0.2298**	0.3053	**0.3008**	+3.21%
	Weather	0.1828	**0.1531**	0.2219	**0.2136**	0.2695	**0.2454**	0.3550	**0.3164**	+9.95%
	Exchange	0.0878	**0.0789**	0.1736	**0.1599**	0.3067	**0.2820**	0.8281	**0.8217**	+6.71%
	Avg.									+4.49%

Table 8: MSE of CoRe-MLP without and with SCR, averaged over 4 horizons and 10 seeds. **w/o SCR**: standalone MLP correction only. **CoRe-MLP**: MLP correction + SCR. Gain: relative MSE reduction from SCR.

	ETTh1		ETTh2		ETTm1		ETTm2	
Backbone	w/o SCR	CoRe-MLP	w/o SCR	CoRe-MLP	w/o SCR	CoRe-MLP	w/o SCR	CoRe-MLP
DLinear	0.4842	**0.4767**	0.2519	**0.2378**	0.5285	**0.5101**	0.2096	**0.1949**
FreTS	0.4722	**0.4681**	0.2556	**0.2439**	0.5243	**0.4957**	0.2056	**0.1923**
iTransformer	0.4685	**0.4673**	0.2710	**0.2605**	0.5337	**0.5145**	0.2296	**0.2131**
PatchTST	0.4595	**0.4526**	0.2534	**0.2494**	0.5265	**0.5038**	0.2121	**0.1980**
OLS	0.4706	**0.4661**	0.2503	**0.2383**	0.5270	**0.5087**	0.2086	**0.1966**
MICN	0.5486	**0.5420**	0.2697	**0.2601**	0.5561	**0.5443**	0.2124	**0.2017**
Informer	0.6207	**0.6180**	0.3841	**0.3768**	0.6913	**0.6642**	0.2618	**0.2533**

	Exchange		Weather		Avg. MSE		Gain / n	
Backbone	w/o SCR	CoRe-MLP	w/o SCR	CoRe-MLP	w/o SCR	CoRe-MLP		
DLinear	0.1087	**0.0978**	0.2124	**0.2009**	0.2992	**0.2864**	+4.28%	23/24
FreTS	0.1161	**0.1050**	0.2077	**0.1972**	0.2969	**0.2838**	+4.41%	22/24
iTransformer	0.1549	**0.1354**	0.2037	**0.1929**	0.3102	**0.2973**	+4.16%	23/24
PatchTST	0.1178	**0.1122**	0.2022	**0.1896**	0.2953	**0.2843**	+3.73%	22/24
OLS	0.1071	**0.1026**	0.2136	**0.1993**	0.2962	**0.2853**	+3.68%	23/24
MICN	0.1498	**0.1362**	0.1955	**0.1870**	0.3053	**0.2952**	+3.31%	22/24
Informer	**0.2405**	0.2427	0.2629	**0.2491**	0.4102	**0.4007**	+2.32%	21/24

Adding SCR to TAFAS and to a standalone two-layer MLP adapter reduces MSE by 4.49% and 3.70% on average, respectively. The gains across two distinct base-adapter designs support correction-space interaction beyond the particular COSA instantiation used in the main experiments.

C.3 Backbone-Error Diagnostics

We examine whether the correction signal underlying correction-space interaction varies systematically with frozen-backbone error. Across all 168 main settings, we relate the frozen-backbone error e_t for each test window to (i) the correction magnitude $\|\boldsymbol{\Delta}_t\|$ and (ii) the relative improvement

$$\mathrm{RI}_t = \frac{e_t - e_t^{\mathrm{CoRe}}}{e_t}. \tag{22}$$

Across settings, backbone error is positively associated with correction magnitude, with mean Pearson correlation $r = 0.59$ and positive correlation in 167/168 settings. After normalization by local signal scale, the error–correction association remains similar ($r = 0.60$).

Backbone error is also positively associated with relative improvement, with mean $r = 0.34$ and positive correlation in 166/168 settings. Restricting the analysis to non-overlapping windows gives mean $r = 0.36$, positive in 138/161 measurable settings; seven Exchange-Rate settings at $H = 720$ contain too few non-overlapping windows for reliable estimation.

Overall, larger frozen-backbone errors are associated with stronger learned corrections across nearly all settings, showing that the correction signal varies systematically with the amount of frozen-backbone error observed in the stream.

D Additional Comparisons and Generalization Results

D.1 Comparison with Other TSF-TTA Methods

PETSA. PETSA (Medeiros et al., 2025) is a parameter-efficient TSF-TTA method that uses low-rank, dynamically gated calibration modules at its input and output.

We compare CoRe with PETSA using the same frozen backbone checkpoints and delayed-supervision forecasting protocol, providing a controlled comparison with a recent TSF-TTA method.

Table 9 reports the full comparison across all 168 settings (seven backbones × six datasets × four horizons). CoRe reduces MSE relative to PETSA by 23.39% on average and outperforms it in 145/168 settings. The average gain is positive for all seven backbones, ranging from 20.86% on iTransformer to 27.74% on Informer.

Table 9: Comparison with PETSA (Medeiros et al., 2025). We report the MSE reduction of CoRe relative to PETSA for each backbone, averaged over six datasets and four prediction horizons. Both methods use the same frozen backbone checkpoints. Higher is better.

Backbone	**CoRe vs. PETSA**
DLinear	22.64%
FreTS	21.90%
iTransformer	20.86%
MICN	24.83%
OLS	21.71%
PatchTST	24.03%
Informer	27.74%
Average	**23.39%**

DynaTTA. We additionally compare with DynaTTA (Grover and Etemad, 2025). Following the evaluation protocol of Im and Kwon (2026), we re-implement DynaTTA within the common forecasting codebase for a consistent comparison.

The selected configuration uses ALPHA_MIN = 10^{-4}, ALPHA_MAX = 10^{-3}, KAPPA = 1.0, ETA = 0.1, EPS = 10^{-6}, WARMUP_FACTOR = 1, MSE_BUFFER_SIZE = 256, RTAB_SIZE = 360, RDB_SIZE = 100, METRIC_HISTORY_SIZE = 256, UPDATE_BUFFERS_INTERVAL = 1, and UPDATE_METRICS_INTERVAL = 1.

Table 10 reports the comparison across the evaluated settings. CoRe outperforms DynaTTA and COSA in the majority of cases, with particularly large differences in several Exchange-Rate and Weather settings.

D.2 Comparison with Normalization-Based Methods

Table 11 compares CoRe with RevIN (Kim et al., 2021), which applies reversible instance normalization, and FAN (Ye et al., 2024), which uses frequency-aware normalization, using DLinear. RevIN and FAN improve several forecasting settings, while CoRe achieves stronger aggregate performance under the evaluated test-time adaptation protocol.

D.3 Generalization to an Additional Backbone: TimeXer

We further evaluate CoRe on another forecasting backbone, TimeXer (Wang et al., 2024) to test whether correction-space interaction remains effective with a modern backbone that already models cross-variate dependencies. We integrate the official TimeXer implementation

Table 10: Prediction MSE comparison with DynaTTA across five backbones (lower is better). **Bold**: best among TTA methods. Underline: second-best among TTA methods. DynaTTA results use the COSA re-implementation with prediction-drift signal.

		DLinear			FreTS			OLS			iTransformer			MICN		
	H	DynaTTA	COSA	CoRe	DynaTTA	COSA	CoRe	DynaTTA	COSA	CoRe	DynaTTA	COSA	CoRe	DynaTTA	COSA	CoRe
ETTh1	96	**0.4708**	0.4922	0.4916	**0.4511**	0.4623	0.4625	**0.4486**	0.4729	0.4723	**0.4509**	0.4638	0.4545	**0.5804**	0.5837	0.5863
	192	0.5321	0.5371	**0.4961**	0.5138	0.5084	**0.4692**	0.5082	0.5173	**0.4763**	0.5156	0.5024	**0.4589**	0.6032	0.5802	**0.5203**
	336	0.5792	0.5208	**0.4722**	0.5838	0.4970	**0.4596**	0.5626	0.5041	**0.4568**	0.6052	0.4956	**0.4661**	0.7037	0.5889	**0.5386**
	720	0.7047	0.5433	**0.4891**	0.7086	0.5616	**0.5052**	0.6933	0.5443	**0.4886**	0.7118	0.5381	**0.5055**	0.8282	0.6177	**0.5591**
ETTh2	96	**0.2338**	0.2552	0.2426	**0.2397**	0.2560	0.2462	**0.2325**	0.2443	0.2389	**0.2767**	0.3050	0.2803	0.2612	0.2538	**0.2476**
	192	0.2888	0.2739	**0.2275**	0.2915	0.2676	**0.2371**	0.2909	0.2740	**0.2290**	0.3077	0.3116	**0.2722**	0.3454	0.2923	**0.2430**
	336	0.3380	0.2521	**0.2171**	0.3476	0.2548	**0.2196**	0.3428	0.2448	**0.2122**	0.3703	0.2552	**0.2245**	0.4000	0.2828	**0.2486**
	720	0.4360	0.2589	**0.2324**	0.4310	0.2516	**0.2336**	0.4273	0.2499	**0.2301**	0.4574	0.2695	**0.2545**	0.5259	0.3058	**0.2746**
ETTm1	96	0.3814	**0.3800**	0.3881	**0.3731**	0.3836	0.3862	0.3865	**0.3780**	0.3874	**0.3694**	0.3799	0.3776	**0.4186**	0.4240	0.4232
	192	0.4809	0.4755	**0.4282**	0.4826	0.4615	**0.4200**	0.4786	0.4730	**0.4235**	0.4790	0.4346	**0.4184**	0.5433	0.4822	**0.4551**
	336	0.5711	0.5358	**0.4665**	0.5818	0.5195	**0.4616**	0.5677	0.5360	**0.4658**	0.5760	0.4825	**0.4585**	0.6056	0.5480	**0.5018**
	720	0.6271	0.5920	**0.5424**	0.6903	0.5773	**0.5200**	0.6557	0.5938	**0.5367**	0.7122	0.5615	**0.5171**	0.6977	0.6061	**0.5602**
ETTm2	96	**0.1640**	0.1915	0.1644	**0.1667**	0.1900	0.1707	0.1683	0.1923	**0.1655**	0.2149	0.2226	**0.1831**	0.2027	0.1909	**0.1845**
	192	0.2040	0.1961	**0.1652**	0.2069	0.1885	**0.1628**	0.2113	0.1978	**0.1671**	0.3044	0.2393	**0.1905**	0.2471	0.2307	**0.1832**
	336	0.2908	0.1969	**0.1725**	0.2743	0.1944	**0.1686**	0.2914	0.1985	**0.1726**	0.4027	0.2473	**0.2058**	0.3212	0.2238	**0.1871**
	720	0.3771	0.2066	**0.1778**	0.4174	0.2110	**0.1799**	0.4574	0.2067	**0.1787**	0.4922	0.2329	**0.1955**	0.4980	0.2381	**0.1991**
Exchange	96	0.0948	0.0816	**0.0706**	0.0907	0.0835	**0.0738**	0.0918	0.0795	**0.0741**	0.1015	0.0913	**0.0832**	0.1207	0.0920	**0.0779**
	192	0.1975	0.0844	**0.0733**	0.1868	0.0959	**0.0806**	0.1748	0.0953	**0.0812**	0.2375	0.1093	**0.0988**	0.2070	0.0855	**0.0739**
	336	0.3001	0.0867	**0.0804**	0.3108	0.0978	**0.0933**	0.3024	0.1003	**0.0947**	0.3328	0.1192	**0.1054**	0.3579	0.1299	**0.1091**
	720	0.8364	0.1469	**0.1239**	0.8253	0.1537	**0.1289**	0.8287	0.1490	**0.1259**	0.9572	0.2228	**0.1957**	0.8618	0.2320	**0.2211**
Weather	96	0.1950	0.1814	**0.1719**	0.1865	0.1739	**0.1617**	0.1993	0.1812	**0.1704**	0.1738	0.1753	**0.1635**	0.2105	0.1701	**0.1646**
	192	0.3311	0.1970	**0.1711**	0.2785	0.1944	**0.1690**	0.3054	0.1960	**0.1719**	0.2683	0.2000	**0.1697**	0.3536	0.2078	**0.1608**
	336	0.4103	0.2465	**0.1782**	0.3846	0.2341	**0.1785**	0.5927	0.2440	**0.1777**	0.4036	0.2439	**0.1736**	0.5471	0.2107	**0.1698**
	720	0.5105	0.2231	**0.1777**	0.5684	0.2084	**0.1763**	0.4995	0.2252	**0.1785**	0.4934	0.2193	**0.1876**	0.6519	0.2013	**0.1711**

Table 11: Prediction MSE comparison with RevIN and FAN on DLinear (lower is better).

Dataset	H	RevIN	FAN	CoRe
ETTh1	96	**0.4591**	0.4620	0.4916
	192	0.5121	0.5159	**0.4961**
	336	0.5587	0.5427	**0.4722**
	720	0.7063	0.6593	**0.4891**
ETTh2	96	**0.2302**	0.2512	0.2426
	192	0.2834	0.2973	**0.2275**
	336	0.3228	0.3249	**0.2171**
	720	0.4111	0.4188	**0.2324**
ETTm1	96	0.3704	**0.3695**	0.3881
	192	0.4431	0.8090	**0.4282**
	336	0.5167	0.6710	**0.4665**
	720	0.5904	0.5628	**0.5424**
ETTm2	96	**0.1599**	0.1835	0.1644
	192	0.1934	0.2436	**0.1652**
	336	0.2326	0.2869	**0.1725**
	720	0.3061	0.3395	**0.1778**
Exchange	96	0.0864	0.0968	**0.0706**
	192	0.1779	0.2079	**0.0733**
	336	0.3257	0.3718	**0.0804**
	720	0.8461	0.8001	**0.1239**
Weather	96	0.1933	0.1827	**0.1719**
	192	0.2371	0.2235	**0.1711**
	336	0.2881	0.2705	**0.1782**
	720	0.3598	0.3402	**0.1777**

and evaluate the same six datasets and four prediction horizons as in the main experiments. Architecture hyperparameters follow the official configuration for each dataset and horizon. Because no official Exchange-Rate configuration is provided, we use the ETTm2 configuration, which is closest in dimensionality.

Table 12 compares frozen TimeXer with TAFAS, PETSA, COSA, and CoRe. All TTA results are averaged over 10 seeds. Across six datasets and four horizons, CoRe improves over COSA in 23 of 24 settings, reducing MSE by 9.13% on average. It also reduces average MSE

Table 12: Evaluation on TimeXer (Wang et al., 2024). MSE is reported for the frozen TimeXer backbone and four TTA methods over six datasets and four horizons. TTA results are mean ± standard deviation over 10 seeds. Lower is better. Bold denotes the best TTA result in each setting.

Dataset	H	Base (TimeXer)	TAFAS	COSA	CoRe	PETSA
ETTh1	96	0.4301	**0.4304±0.0004**	0.4407±0.0001	0.4367±0.0019	0.4349±0.0002
	192	0.4895	0.4996±0.0001	0.4867±0.0004	**0.4513±0.0049**	0.4922±0.0003
	336	0.5605	0.5795±0.0113	0.4776±0.0003	**0.4329±0.0025**	0.5544±0.0005
	720	0.6931	0.6810±0.0010	0.5195±0.0001	**0.4813±0.0046**	0.6698±0.0014
ETTh2	96	0.2367	**0.2317±0.0001**	0.2492±0.0008	0.2469±0.0034	0.2367±0.0003
	192	0.3000	0.3021±0.0001	0.2907±0.0035	**0.2420±0.0039**	0.2986±0.0002
	336	0.3223	0.3254±0.0007	0.2494±0.0022	**0.2261±0.0026**	0.3166±0.0003
	720	0.4432	0.4460±0.0030	0.2701±0.0007	**0.2446±0.0035**	0.4135±0.0003
ETTm1	96	0.3979	**0.3546±0.0008**	0.3728±0.0010	0.3783±0.0039	0.3991±0.0015
	192	0.4345	0.4080±0.0001	0.4314±0.0002	**0.4061±0.0045**	0.4307±0.0004
	336	0.4890	0.4704±0.0003	0.4701±0.0030	**0.4378±0.0043**	0.4715±0.0005
	720	0.5525	0.5547±0.0013	0.5433±0.0009	**0.4966±0.0143**	0.5321±0.0007
ETTm2	96	0.1479	**0.1476±0.0003**	0.1750±0.0007	0.1580±0.0021	0.1506±0.0001
	192	0.1818	0.1853±0.0006	0.1794±0.0012	**0.1513±0.0019**	0.1828±0.0002
	336	0.2342	0.2365±0.0008	0.1880±0.0014	**0.1664±0.0025**	0.2339±0.0005
	720	0.3088	0.3165±0.0022	0.2141±0.0011	**0.1855±0.0047**	0.3011±0.0007
Exchange	96	0.0954	0.0922±0.0008	0.0892±0.0009	**0.0783±0.0024**	0.0948±0.0002
	192	0.1982	0.1904±0.0008	0.1095±0.0007	**0.0949±0.0031**	0.1934±0.0002
	336	0.3680	0.3218±0.0002	0.1132±0.0019	**0.1102±0.0023**	0.3601±0.0004
	720	0.9037	0.9291±0.0000	0.1935±0.0046	**0.1859±0.0060**	0.8866±0.0011
Weather	96	0.1506	0.1504±0.0009	0.1544±0.0001	**0.1483±0.0019**	0.1525±0.0001
	192	0.2011	0.2005±0.0007	0.1829±0.0003	**0.1571±0.0017**	0.1969±0.0002
	336	0.2566	0.2550±0.0006	0.2219±0.0009	**0.1673±0.0036**	0.2482±0.0004
	720	0.3384	0.3428±0.0017	0.2008±0.0001	**0.1750±0.0028**	0.3385±0.0021

Table 13: Average MSE reduction of CoRe over COSA by backbone family, averaged over six datasets and four horizons.

Family	Backbones	MSE reduction vs. COSA
CI	DLinear, FreTS, OLS, PatchTST	10.87%
CD	iTransformer, Informer, MICN	10.18%

by 24.42% relative to frozen TimeXer and 23.62% relative to PETSA. The only regression relative to COSA occurs on ETTm1 at $H = 96$.

Implementation note. The official TimeXer implementation performs an in-place division during input normalization. This operation conflicts with repeated backpropagation through cached inputs during TTA. For TAFAS and PETSA, we replace it with the mathematically equivalent non-in-place operation. The numerical forward computation is unchanged. COSA and CoRe are unaffected by this implementation issue.

D.4 Channel-Independent and Channel-Dependent Backbones

We group the seven main backbones according to whether they explicitly model cross-variate dependencies in their forecasting architecture. DLinear, FreTS, OLS, and PatchTST are treated as channel-independent (CI), while iTransformer, Informer, and MICN are treated as channel-dependent (CD) under our experimental configurations.

CoRe improves both groups by comparable margins, showing that correction-space adaptation remains useful for both channel-independent and channel-dependent forecasting backbones. A summary is shown in Table 13.

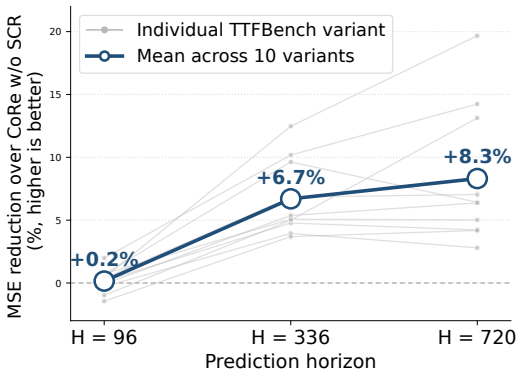

Figure 7: Robustness on TTFBench (ETTh1) across ten distribution-shift variants. Grey curves denote individual variants and the solid curve denotes their mean.

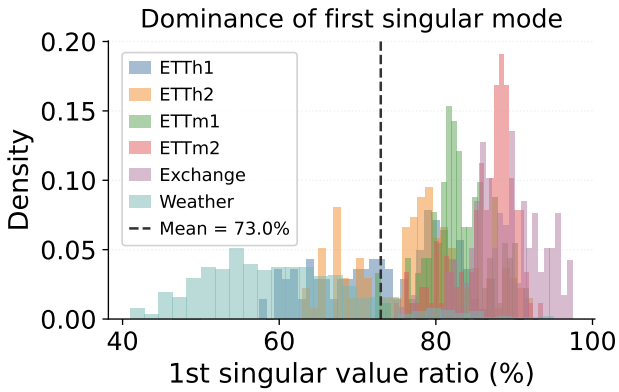

Figure 8: Distribution of the leading singular-mode energy ratio of $\mathbf{\Delta}_t$. The first mode captures 73% of correction energy on average.

D.5 Robustness under Diverse Distribution Shifts

TTFBench (Grover and Etemad, 2025) contains multiple types and intensities of distribution shift for TSF-TTA evaluation. Using ETTh1 with DLinear, we evaluate ten representative shift variants with the same hyperparameter configuration as in the main experiments.

Figure 7 reports the MSE reduction of CoRe relative to the corresponding variant without SCR. The improvement remains positive across the evaluated shift types and is generally larger at medium-to-long prediction horizons.

E Additional Ablation and Scalability Analysis

E.1 SCR Rank, Structure, and Scalability

Motivated by the low-dimensional correction structure observed in Appendix G.1, we examine whether effective cross-variate refinement requires dense interaction across all C variates.

Correction-space structure. We first examine the singular-value spectrum of the learned correction matrix $\mathbf{\Delta}_t \in \mathbb{R}^{H\times C}$. Figure 8 reports the energy captured by its dominant singular mode. Across settings, the leading mode captures approximately 73% of the correction energy on average, revealing pronounced low-dimensional structure in the correction space. This observation motivates the reduced-rank parameterizations evaluated below.

Structured versus dense interaction. We compare dense mixing, fixed-rank SCR ($r = 16$), and the variate-matched configuration ($r = C$) over the 168 main settings. Dense mixing uses approximately 34× more parameters on average than SCR with $r = C$, yet obtains a lower win rate (128/168 versus 162/168) and a smaller average MSE reduction over the frozen backbone (23.09% versus 25.82%). Table 14 summarizes the parameter cost and aggregate performance, with full results reported in Table 15.

Table 14: Adapter-only parameter counts and accuracy for three cross-variate interaction designs, averaged over 168 settings.

	Dense	**SCR** r=C	**SCR** r=16
Avg params	339,024	9,922	16,480
Avg imp. vs. baseline	+23.09%	+25.82%	+20.37%
Win rate vs. baseline	128/168	162/168	147/168

Parameter scaling. The SCR bottleneck uses

$$\mathbf{W}_{\downarrow} \in \mathbb{R}^{r\times 2H}, \qquad \mathbf{W}_{\uparrow} \in \mathbb{R}^{H\times r}.$$

Ignoring bias terms, SCR therefore contains

$$2Hr + Hr = 3Hr$$

weights. With the default $r = C$, its parameter count scales as $O(HC)$; with a fixed $r \ll C$, it scales as $O(Hr)$.

For comparison, the dense interaction used in the rank/capacity ablation is implemented as a single `Linear`$(2H \to H)$ layer (Table 14), containing $2H^2$ weights (and H bias parameters).

Rank sensitivity and default choice. We additionally compare smaller fixed ranks $r \in \{4, 8, 16, 32\}$ with the default $r = C$. Among the evaluated configurations on the six standard datasets (Table 16), $r = C$ provides the strongest aggregate performance. Smaller ranks remain effective in many settings and are useful when reducing the parameter cost is important. We therefore use $r = C$ as the default for the main experiments and $r = 16 \ll C$ for the high-dimensional experiments.

Table 16: Ablation on SCR pooling strategy and bottleneck rank (DLinear backbone, 24 settings). *avg_ imp*: average MSE improvement over backbone.

Configuration	**avg_ imp**
Mean pooling, $r = 4$	+18.09%
Mean pooling, $r = 8$	+18.84%
Mean pooling, $r = 16$	+19.66%
Mean pooling, $r = 32$	+20.24%
Mean pooling, $r = C$	**+25.45%**
Max pooling, $r = 16$	+19.36%

High-Dimensional Datasets. The default SCR configuration uses $r = C$, so its parameter cost increases with the number of variates. We therefore examine whether correction-space interaction remains useful with a substantially smaller bottleneck rank in high-dimensional settings.

Table 17 reports results on Electricity ($C = 321$) and Traffic ($C = 862$) using $r = 16 \ll C$. Despite the substantial rank reduction, CoRe retains positive aggregate improvements over COSA. The gains are particularly clear on Electricity. On Traffic, the improvements are smaller and several individual settings show negative differences, including DLinear at $H = 720$.

Together with the singular-value analysis in Appendix E.1, these results show that useful cross-variate correction structure can be retained with a bottleneck substantially smaller

than C in high-dimensional settings. These results show that aggregate improvements can be retained with a bottleneck rank substantially smaller than C, while also highlighting the limits of a single fixed rank at very large C.

E.2 Spectral Descriptor and Gate Ablations

Table 18 evaluates alternative spectral descriptors and gating variants. Differences among the spectral-descriptor variants are small when averaged across settings. We use the full four-dimensional descriptor

$$[\mathrm{SE}, \mathrm{LBR}, \mathrm{MBR}, \mathrm{HBR}]$$

as the default because it jointly summarizes spectral concentration and coarse frequency allocation at negligible additional cost.

The fixed-gate and input-conditioned variants further isolate the contribution of adaptive modulation. The full spectral gate provides a smaller additional average gain beyond the structured SCR interaction.

E.3 Spectral Gate Analysis

Shared versus per-variate descriptors. The main implementation computes a shared spectral descriptor by averaging the power spectra across variates before constructing $\mathbf{s}_t$. We additionally evaluate per-variate spectral descriptors. The two variants perform comparably (average MSE 0.2752 versus 0.2754), so we retain the shared descriptor as the simpler default.

Gate behavior. Table 19 summarizes statistics of the learned gates, and Figure 9 shows representative gate trajectories on Weather and Exchange Rate. The gate varies substantially across datasets and test batches, with approximately 10.5% of activations near zero. This shows that the spectral controller produces input-dependent modulation rather than applying SCR with a fixed contribution. Together with the fixed-gate ablation in Appendix E.2, these statistics show that the gate is input-dependent and provides an empirical benefit over fixed modulation.

Table 19: SCR spectral gate statistics across datasets (DLinear backbone, averaged over 4 horizons and 10 seeds). Mean: time-averaged gate value across all variates and test batches; Std: standard deviation across variates and batches; Near zero ($|g|<0.1$): fraction of gate activations close to zero, indicating suppressed cross-variate refinement; Near init ($g<-0.8$): fraction of gate activations close to the initialisation bias $\mathbf{b}_g=-1$, indicating gates that have not moved significantly from initialisation. The gate is bounded in $(-1, 1)$ by tanh.

Dataset	**Mean**	**Std**	**Near zero** ($\lvert g\rvert<0.1$)	**Near init** ($g<-0.8$)
ETTh1	−0.583	0.325	5.9%	34.5%
ETTh2	−0.631	0.285	9.3%	37.5%
ETTm1	−0.357	0.421	9.9%	15.1%
ETTm2	−0.498	0.354	15.5%	28.4%
Exchange	−0.676	0.194	2.4%	23.9%
Weather	−0.352	0.371	20.3%	9.0%
Average	−0.516	0.325	10.5%	24.7%

Gate-bias sensitivity. Table 20 evaluates different spectral-gate bias initializations over the reported DLinear settings. The default $\mathbf{b}_g = -\mathbf{1}$ achieves the strongest aggregate performance. Relative to this choice, initialization at $\mathbf{b}_g = \mathbf{0}$ increases MSE by 1.56%.

Table 20: Gate bias sensitivity.

$\mathbf{b}_g$	−2.0	−1.0	0.0
vs. default ↓	+0.46%	+0.00%	+1.56%

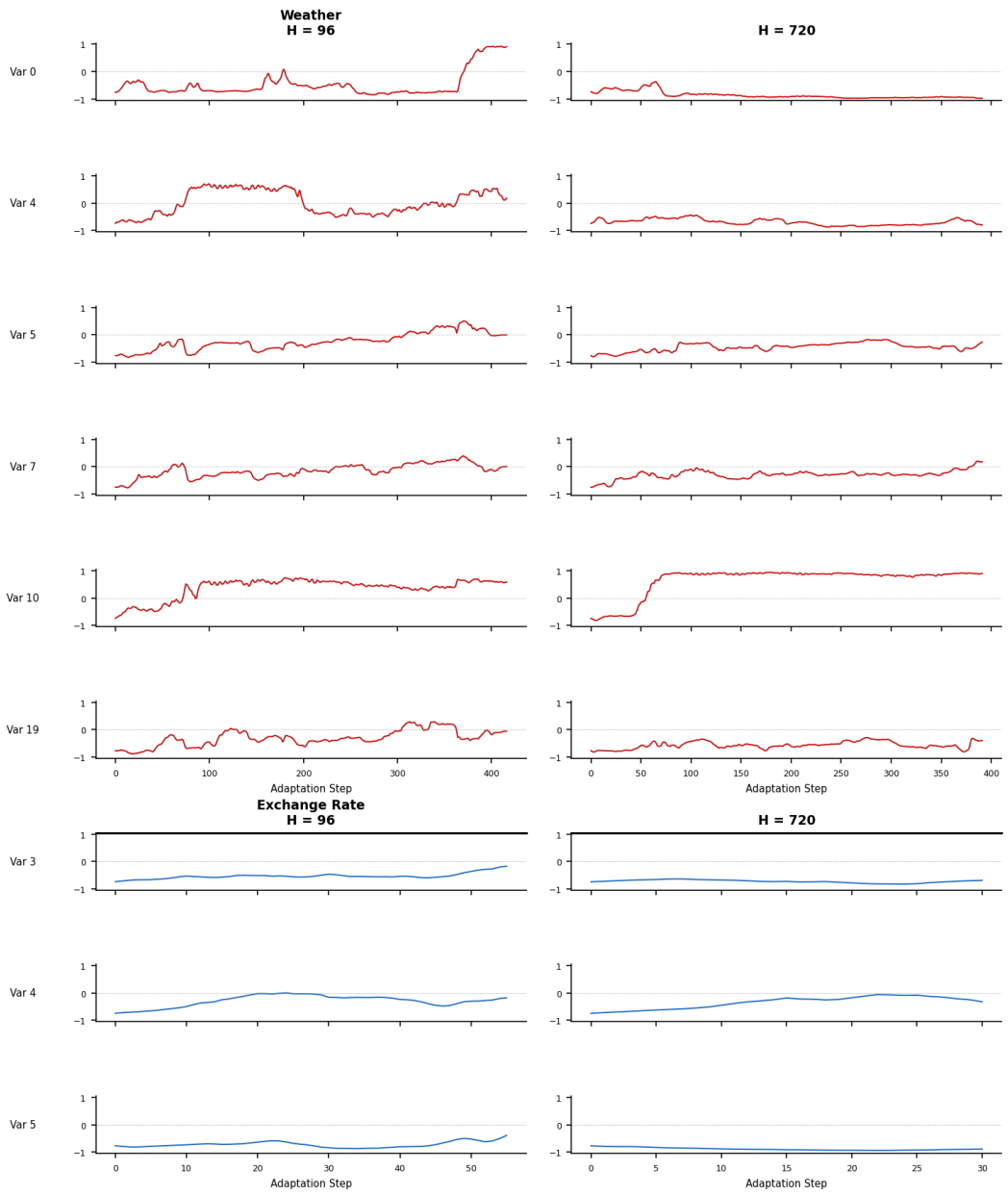

Figure 9: Example spectral-gate trajectories over adaptation steps on Weather and Exchange Rate.

F Online Adaptation Reliability

We evaluate adaptation reliability relative to the frozen backbone using the mean per-batch regret

$$\overline{\mathcal{R}}_T(m) = \frac{1}{T}\sum_{t=1}^{T}\left(\ell_t^m - \ell_t^{\text{frozen}}\right), \tag{23}$$

where m denotes the adapted method and negative values indicate an average improvement over the frozen backbone.

We additionally report three complementary quantities. The positive-part regret $\overline{\mathcal{R}}_T^+$ measures the magnitude of harmful adaptation when it occurs; p_{neg} denotes the frequency with which

the adapted model underperforms the frozen backbone; and n_{worse} counts the 18 horizon-averaged backbone–dataset combinations with positive mean regret. Together, these metrics characterize two complementary aspects of reliability: how often online adaptation is harmful and how severe the harm is when it occurs.

Table 21 reports the full reliability results across the 18 backbone–dataset combinations.

G Additional Diagnostics

G.1 Dataset and Correction Statistics

Cross-variate correlations. Table 22 reports descriptive statistics of cross-variate correlations across the evaluated datasets. Before computing static correlations, we apply the Augmented Dickey–Fuller test to each variate. Differencing is applied when the unit-root null cannot be rejected at the chosen threshold: order 1 for Exchange Rate and order 2 for Electricity. Temporal correlation statistics are computed on the raw series.

The six main datasets have broadly comparable mean static absolute correlations, while their temporal variation differs substantially. Exchange Rate has the largest shifting standard deviation among the reported datasets (std = 0.479), whereas Traffic and Electricity exhibit more stable correlation statistics. These measurements provide descriptive context for the diversity of cross-variate structure represented by the benchmark suite.

Table 22: Pairwise Pearson correlation statistics across variates. N: number of variates; Window: rolling window size (approx. 1 week in physical time). **Mean** $|r|$: time-averaged absolute pairwise correlation computed on stationary series after ADF test. **Shifting (std)**: standard deviation of windowed correlations over time, measuring temporal instability of cross-variate dependencies (computed on raw series to preserve trend-switching signals). **CV**: coefficient of variation (std/mean). **Range**: peak-to-trough variation of windowed correlations. **Strong** ($|r|>0.7$) and **Weak** ($|r|\leq 0.3$): fraction of variate pairs by static $|r|$ strength. Datasets above the divider are used in main experiments.

Dataset	N	Window	Mean $\lvert r\rvert$	Shifting (std)	CV	Range	Strong	Weak
ETTh1	7	168	0.222	0.266	0.919	0.771	9.5%	81.0%
ETTh2	7	168	0.325	0.237	0.846	0.865	4.8%	57.1%
ETTm1	7	672	0.224	0.264	0.913	0.771	9.5%	81.0%
ETTm2	7	672	0.325	0.234	0.845	0.872	4.8%	57.1%
Weather	21	1008	0.296	0.180	0.736	0.637	21.0%	64.3%
Exchange Rate†	8	30	0.305	**0.479**	**0.957**	**0.992**	3.6%	42.9%
Traffic	862	168	0.631	0.111	0.207	0.651	25.7%	10.1%
Electricity††	321	168	0.119	0.114	0.340	0.524	0.0%	91.0%

† Static Mean $|r|$ computed after first-order differencing (ADF $p < 0.05$);
Shifting metrics computed on raw series to preserve trend-switching signals.
†† Second-order differencing applied (ADF $p < 0.05$).

Correction heterogeneity. Table 23 additionally reports variation among per-variate correction signals. The coefficient of variation and cosine-similarity statistics show that the learned corrections can differ across variates in both magnitude and direction. We use these measurements as descriptive properties of the correction space rather than as causal explanations of the gating results.

Table 23: Per-variate correction heterogeneity and the benefit of spectral gating over fixed gating (DLinear backbone). *CV*: coefficient of variation of per-variate correction norms across variates; *Cos-sim std*: standard deviation of cosine similarity between each variate's correction and the cross-variate mean anchor; *Fixed gate* and *Spectral gate*: average MSE under each gating strategy (10 seeds); *Gain*: relative improvement from spectral over fixed gating. Datasets ordered by CV.

Dataset	CV	Cos-sim std	Fixed gate	Spectral gate	Gain
ETTh1	0.217	0.097	0.4889	**0.4873**	+0.34%
Exchange	0.322	0.223	0.0890	**0.0871**	+2.19%
ETTm1	0.428	0.117	0.4598	**0.4563**	+0.76%
ETTm2	0.514	0.137	0.1716	**0.1700**	+0.95%
ETTh2	0.554	0.119	0.2329	**0.2299**	+1.29%
Weather	0.696	0.151	0.1797	**0.1747**	+2.77%
Average	0.455	0.141	0.2703	**0.2675**	+1.03%

Low-dimensional correction structure. Across settings, the leading singular mode captures approximately 73% of the correction energy on average, revealing pronounced low-dimensional structure in the correction space. This observation motivates the reduced-rank parameterization examined in Appendix E.1.

G.2 Qualitative Forecast Visualization

Figures 10 and 11 compare COSA and CoRe on Exchange Rate and Weather. The differences are visually more pronounced in several medium-to-long-horizon examples, where CoRe more closely follows the selected ground-truth trajectories.

On Exchange Rate, the illustrated windows include reductions in per-window MSE of up to 36% relative to COSA (e.g., $H = 192$, Var. 6: $0.0160 \rightarrow 0.0102$). These examples are consistent with the aggregate quantitative results in Table 1.

G.3 Runtime Breakdown

Table 24 reports the runtime decomposition with a DLinear backbone, averaged over six datasets and four horizons. Relative to COSA, the measured per-adaptation-step runtime increases from 6.28 ms to 6.76 ms (+0.48 ms; +7.7%). Spectral descriptor computation and SCR account for 1.14 ms and 1.45 ms, respectively, while differences in the remaining components partially offset this additional cost. The full-stream runtime increases from 11.42 s to 12.17 s (+6.5%).

Table 24: Detailed per-step timing breakdown averaged over six datasets and four horizons (DLinear). Overhead = CoRe − COSA; "-" indicates a component absent in that method.

Component	COSA(ms)	CoRe(ms)	Overhead(ms)
Backbone prediction	1.27	1.04	−0.22
Spectral FFT	-	1.14	+1.14
SCR forward	-	1.45	+1.45
Loss computation	0.84	1.00	+0.16
Backward + update	4.42	4.11	−0.31
Per adaptation step	6.28	6.76	+0.48 (+7.7%)

H Theoretical Analysis

H.1 Theoretical Connections

The spectral descriptor summarizes local temporal structure in the frequency domain. The Wiener–Khinchin relation connects the power spectrum of a stationary process to its second-order temporal structure (Wiener, 1930; Khintchine, 1934). We use FFT-derived spectra as empirical descriptors of the finite current window rather than as estimates of a globally stationary process.

Normalized spectral entropy provides a classical measure of spectral energy spread (Powell and Percival, 1979; Inouye et al., 1991). Recent multivariate forecasting work has also used spectral entropy to guide dependency-aware modeling (Xiong et al., 2026). Together with the band-energy ratios, it provides the compact four-dimensional descriptor used by the spectral gate.

The structured SCR bottleneck is also consistent with the observed geometry of the correction matrices. As shown in Figure 8, the leading singular mode captures 73% of correction energy on average. This observation motivates examining reduced-rank correction interaction in the high-dimensional experiments.

H.2 Theoretical Analysis of Interaction Space

The main text distinguishes prediction-space and correction-space interaction using generic cross-variate operators. Here we give a nonlinear characterization of the structural distinction.

Let

$$\hat{\mathbf{Y}}_t^{\text{base}} = \mathbf{Y}_t + \mathbf{e}_t^{\text{base}}, \tag{24}$$

where $\mathbf{e}_t^{\text{base}}$ is the frozen-backbone prediction error.

Prediction-space interaction takes the form

$$\hat{\mathbf{Y}}_t^{\text{pred}} = \hat{\mathbf{Y}}_t^{\text{base}} + \boldsymbol{\Delta}_t + \mathcal{M}_{\text{pred}}\left(\hat{\mathbf{Y}}_t^{\text{base}}\right), \tag{25}$$

while correction-space interaction takes the form

$$\hat{\mathbf{Y}}_t^{\text{corr}} = \hat{\mathbf{Y}}_t^{\text{base}} + \boldsymbol{\Delta}_t + \mathcal{M}_{\text{corr}}\left(\boldsymbol{\Delta}_t\right). \tag{26}$$

The operators may be nonlinear. For CoRe, the SCR bottleneck and its spectral modulation are included in $\mathcal{M}_{\text{corr}}$; conditioning on the current input window is fixed when analyzing the interaction at time t.

H.2.1 Proof of Proposition 1

The prediction-space interaction output evaluated on the actual backbone forecast is

$$\mathcal{M}_{\text{pred}}\left(\mathbf{Y}_t + \mathbf{e}_t^{\text{base}}\right). \tag{27}$$

To isolate its direct sensitivity to the backbone error, define

$$\boldsymbol{\Gamma}_t^{\text{pred}} := \mathcal{M}_{\text{pred}}\left(\mathbf{Y}_t + \mathbf{e}_t^{\text{base}}\right) - \mathcal{M}_{\text{pred}}\left(\mathbf{Y}_t\right). \tag{28}$$

If $\mathcal{M}_{\text{pred}}$ is continuously differentiable, let

$$h(s) = \mathcal{M}_{\text{pred}}\left(\mathbf{Y}_t + s\mathbf{e}_t^{\text{base}}\right), \qquad s \in [0,1]. \tag{29}$$

By the chain rule,

$$h'(s) = J_{\mathcal{M}_{\text{pred}}}\left(\mathbf{Y}_t + s\mathbf{e}_t^{\text{base}}\right)\left[\mathbf{e}_t^{\text{base}}\right], \tag{30}$$

where $J_{\mathcal{M}_{\text{pred}}}(\mathbf{A})[\mathbf{E}]$ denotes the action of the Jacobian of $\mathcal{M}_{\text{pred}}$ at $\mathbf{A}$ on perturbation $\mathbf{E}$. Applying the fundamental theorem of calculus gives

$$\boldsymbol{\Gamma}_t^{\text{pred}} = \int_0^1 J_{\mathcal{M}_{\text{pred}}}\left(\mathbf{Y}_t + s\mathbf{e}_t^{\text{base}}\right)\left[\mathbf{e}_t^{\text{base}}\right] ds. \tag{31}$$

Equation 31 makes explicit that prediction-space cross-variate interaction is directly sensitive to the error contained in the frozen-backbone prediction.

In correction space, the interaction operator instead receives $\mathbf{\Delta}_t$:

$$\mathcal{M}_{\text{corr}}\left(\mathbf{\Delta}_t\right). \tag{32}$$

Thus, $\mathbf{e}_t^{\text{base}}$ does not enter the cross-variate interaction operator directly. The correction $\mathbf{\Delta}_t$ may itself depend on the backbone prediction through the base adapter; the structural distinction concerns the quantity on which cross-variate interaction is explicitly performed. This proves Proposition 1.

Linear special case. For a linear prediction-space interaction applied along the variate dimension,

$$\mathcal{M}_{\text{pred}}(\mathbf{A}) = \mathbf{A}\mathbf{W}_{\text{pred}}^{\top}, \tag{33}$$

Equation 28 reduces to

$$\mathbf{\Gamma}_t^{\text{pred}} = \mathbf{e}_t^{\text{base}}\mathbf{W}_{\text{pred}}^{\top}. \tag{34}$$

Hence the explicit coupled-error term used in the linear illustration is a special case of the nonlinear characterization above.

H.3 Implication under Delayed Supervision

The structural distinction above occurs in a delayed-supervision setting. For a forecast issued at time t, the complete target corresponding to that forecast becomes available only after the prediction horizon. Consequently, the loss associated with this forecast cannot be used for a supervised adaptation update before the corresponding target has been observed.

Prediction-space interaction therefore acts on a backbone prediction containing $\mathbf{e}_t^{\text{base}}$ before supervision associated with that forecast becomes available. During this feedback window, prediction-space interaction continues to operate on error-containing backbone outputs before the corresponding supervised feedback becomes available. As H increases, the associated supervision arrives later, motivating us to examine whether the empirical consequence of this structural difference varies with the prediction horizon. We evaluate this empirically in Appendix C.1. Correction-space interaction instead performs its cross-variate operation on the adapter-produced correction while preserving the backbone prediction path.

This argument is a mechanistic motivation, not a guarantee that the performance gap must increase monotonically with H. Standard analyses of online learning with delayed feedback likewise contain delay-dependent optimization terms (Langford et al., 2009), but we do not use them to claim a performance ordering between the two interaction spaces. The analysis isolates a structural difference between the two interaction spaces; the controlled comparison in Appendix C.1 tests its empirical consequence while holding the refinement architecture fixed.

I Limitations and Future Work

Short prediction horizons. The benefit of correction-space interaction is smaller at $H = 96$, with slight degradation in a small number of settings. Adapting the interaction capacity to the available correction signal may improve short-horizon performance.

Interaction capacity and base-adapter dependence. The default choice $r = C$ scales with the number of variates. Although reduced-rank SCR remains effective overall on Electricity and Traffic, its gains are smaller in some high-dimensional settings. Moreover, CoRe operates on corrections produced by an underlying adapter and therefore depends on the information available in this correction interface. Results with TAFAS and a standalone MLP show that the interaction principle transfers beyond COSA, while adaptive rank selection and stronger correction interfaces remain directions for future work.

Scope of the analysis and deployment setting. Our analysis establishes the structural distinction between prediction- and correction-space interaction rather than an accuracy guarantee. Experiments follow the standard delayed-supervision protocol and assume backbones that support an additive correction interface. Extending correction-space interaction to irregular feedback delays and alternative interfaces, including time-series foundation models such as MOIRAI (Woo et al., 2024) and Chronos (Ansari et al., 2024), is an interesting direction.

Table 15: Full rank/capacity ablation results of CoRe. We compare the default variate-matched rank setting (r=C), a fixed-rank variant (r=16), and a dense horizon-space interaction without bottleneck. The dense horizon-space interaction used in this comparison contains $2H^2 + H$ parameters, whereas SCR with rank r contains $3Hr + r + H$ parameters. Lower MSE is better; **bold** and underlining indicate the best and second-best results among the three variants, respectively.

		iTransformer			**DLinear**			**FreTS**			**OLS**			**PatchTST**			**MICN**			**Informer**		
	H	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)	r=C	r=16	Dense (horizon)
ETTh1	96	0.4545	**0.4390**	0.5133	0.4916	**0.4722**	0.5133	0.4625	**0.4500**	0.4638	0.4723	**0.4517**	0.4707	0.4371	**0.4353**	0.4396	0.5863	**0.5461**	0.5839	0.6501	**0.6203**	0.6543
	192	**0.4589**	0.4767	0.4987	**0.4961**	0.4963	0.4987	**0.4692**	0.4806	0.4902	0.4763	0.4786	**0.4758**	0.4540	0.4685	**0.4520**	0.5203	0.5290	**0.5189**	0.6415	**0.6341**	0.6355
	336	**0.4661**	0.4918	0.4686	0.4722	0.5113	**0.4686**	**0.4596**	0.4979	0.4686	0.4568	0.4915	**0.4558**	0.4458	0.4817	**0.4413**	0.5386	0.5394	**0.5380**	0.6302	0.6456	**0.6290**
	720	0.5055	0.5387	**0.4771**	0.4891	0.5330	**0.4771**	0.5052	0.5482	**0.4889**	**0.4886**	0.5305	0.4908	**0.4983**	0.5388	0.5009	**0.5591**	0.5837	0.5614	0.6400	0.6524	**0.6399**
ETTh2	96	0.2803	0.2736	**0.2460**	0.2426	0.2405	**0.2383**	0.2462	0.2455	**0.2422**	0.2389	**0.2385**	0.2434	0.2490	**0.2437**	0.2447	**0.2476**	0.2551	0.2497	0.3760	**0.3476**	0.3571
	192	0.2722	0.3154	**0.2145**	**0.2275**	0.2989	0.2321	**0.2371**	0.2926	0.2388	0.2290	0.3004	**0.2213**	0.2405	0.2590	**0.2341**	0.2430	0.3096	**0.2364**	**0.3567**	0.3917	0.3839
	336	0.2245	0.2683	**0.2031**	0.2171	0.2660	**0.2158**	0.2196	0.2601	**0.2189**	**0.2122**	0.2641	0.2130	0.2167	0.2572	**0.2164**	0.2486	0.2798	**0.2478**	**0.3463**	0.4037	0.3498
	720	0.2545	0.2744	**0.2166**	0.2324	0.2449	**0.2279**	**0.2336**	0.2485	0.2338	**0.2301**	0.2396	0.2317	**0.2331**	0.2465	0.2345	**0.2746**	0.2829	0.2861	0.3113	0.3405	**0.3105**
ETTm1	96	0.3776	**0.3430**	0.4121	0.3881	**0.3560**	0.4121	0.3862	**0.3584**	0.3776	0.3874	**0.3556**	0.3831	0.4026	**0.3656**	0.3979	0.4232	**0.3724**	0.4219	0.4929	**0.4130**	0.4876
	192	0.4184	**0.4052**	0.4843	0.4282	**0.4247**	0.4260	**0.4200**	0.4243	0.4218	0.4235	**0.4230**	0.4280	**0.4183**	0.4215	0.4543	0.4551	**0.4422**	0.4505	0.5385	**0.4782**	0.5301
	336	0.4585	**0.4367**	0.5230	0.4665	**0.4564**	0.5230	0.4616	0.4628	**0.4575**	0.4658	**0.4626**	0.4640	**0.4509**	0.4561	0.5043	0.5018	**0.4770**	0.5026	0.5771	**0.5270**	0.5766
	720	0.5171	**0.4708**	0.5416	0.5424	**0.4864**	0.5290	0.5200	**0.4683**	0.5253	0.5367	**0.4783**	0.5417	0.5225	**0.4692**	0.5339	0.5602	**0.4832**	0.5520	0.6454	**0.5631**	0.6360
ETTm2	96	0.1831	0.1856	**0.1623**	0.1644	0.1684	**0.1623**	0.1707	**0.1645**	0.1710	0.1655	0.1687	**0.1650**	0.1721	**0.1615**	0.1732	0.1845	**0.1687**	0.1723	0.2027	0.2011	**0.2001**
	192	0.1905	0.2074	**0.1658**	0.1652	0.1904	**0.1634**	**0.1628**	0.1898	0.1638	0.1671	0.1930	**0.1644**	0.1694	0.2065	**0.1678**	0.1832	0.2035	**0.1791**	**0.1964**	0.2128	0.1986
	336	0.2058	0.2048	**0.1720**	0.1725	0.1773	**0.1720**	**0.1686**	0.1761	0.1690	**0.1726**	0.1758	0.1742	0.1765	0.1827	**0.1752**	0.1871	0.1988	**0.1836**	0.2114	0.2202	**0.2105**
	720	0.1955	0.1964	**0.1747**	0.1778	0.1776	**0.1769**	0.1799	0.1803	**0.1758**	0.1787	**0.1773**	0.1791	**0.1728**	0.1772	0.1738	0.1991	**0.1973**	0.2013	0.2307	0.2345	**0.2299**
Exchange	96	0.0832	0.0916	**0.0586**	0.0706	0.0886	**0.0631**	0.0738	0.0815	**0.0716**	0.0741	0.0808	**0.0734**	0.0739	0.0800	**0.0709**	0.0779	0.0963	**0.0756**	**0.0985**	0.1482	0.1003
	192	0.0988	0.1453	**0.0728**	0.0733	0.1033	**0.0728**	**0.0806**	0.1116	0.0848	**0.0812**	0.1106	0.0819	0.0897	0.1153	**0.0869**	**0.0739**	0.1078	0.0757	**0.1183**	0.1568	0.1203
	336	0.1054	0.1419	**0.0721**	**0.0804**	0.1124	0.0825	**0.0933**	0.1254	0.0962	0.0947	0.1247	**0.0915**	0.1000	0.1323	**0.0987**	0.1091	0.1528	**0.1083**	0.1901	0.2537	**0.1880**
	720	0.1957	0.2597	**0.1310**	**0.1239**	0.1790	0.1244	**0.1289**	0.1921	0.1371	**0.1259**	0.1833	0.1313	**0.1348**	0.1930	0.1375	0.2211	0.3579	**0.2194**	**0.3403**	0.6156	0.3652
Weather	96	0.1635	**0.1588**	0.3015	**0.1719**	0.1732	0.1801	**0.1617**	0.1703	0.1632	**0.1704**	0.1723	0.1754	0.1622	0.1644	**0.1602**	0.1646	**0.1636**	0.1652	0.1807	**0.1751**	0.1815
	192	**0.1697**	0.1936	0.1823	**0.1711**	0.2012	0.1726	0.1690	0.2030	**0.1689**	**0.1719**	0.2018	0.1777	0.1724	0.2002	**0.1675**	**0.1608**	0.1919	**0.1608**	**0.2261**	0.2336	0.2316
	336	**0.1736**	0.2050	0.1907	0.1782	0.1992	**0.1780**	**0.1785**	0.1998	0.1801	**0.1777**	0.1988	0.1796	0.1810	0.2045	**0.1760**	0.1698	0.2072	**0.1689**	**0.2202**	0.2527	0.2205
	720	**0.1876**	0.2157	0.1994	**0.1777**	0.2127	0.1792	**0.1763**	0.2114	0.1769	**0.1785**	0.2136	0.1810	0.1765	0.2037	**0.1754**	**0.1711**	0.1999	0.1712	**0.2371**	0.2439	0.2399

Table 17: Results on large-scale datasets Electricity (C=321) and Traffic (C=862), averaged over 10 seeds. **COSA→CoRe**: MSE improvement of CoRe over COSA. All methods use a fixed bottleneck rank r=16 for these datasets to keep the parameter cost independent of C.

Model	Dataset	H	Backbone	COSA	CoRe	COSA→CoRe
DLinear	Electricity	96	0.2078	0.2024	**0.1992**	+1.58%
		192	0.2081	0.1922	**0.1771**	+7.86%
		336	0.2228	0.1947	**0.1751**	+10.07%
		720	0.2644	0.2139	**0.1926**	+9.96%
	Traffic	96	0.6710	0.6685	**0.6648**	+0.55%
		192	0.6251	0.6127	**0.6031**	+1.57%
		336	0.6328	0.6133	**0.6074**	+0.96%
		720	0.6717	**0.6464**	0.6545	−1.25%
MICN	Electricity	96	0.1766	0.1634	**0.1526**	+6.61%
		192	0.1915	0.1592	**0.1409**	+11.49%
		336	0.1973	0.1575	**0.1396**	+11.37%
		720	0.2193	0.1674	**0.1501**	+10.33%
	Traffic	96	0.4955	**0.4602**	0.4603	−0.02%
		192	0.5080	0.4737	**0.4449**	+6.08%
		336	0.5306	0.4929	**0.4642**	+5.82%
		720	0.5757	0.5232	**0.5083**	+2.85%

Table 18: Ablation study of CoRe components. Average MSE over three backbones (DLinear, PatchTST, MICN), six datasets, four horizons, and 10 seeds. **vs. COSA**: relative MSE reduction over COSA. *Loss-trend gate*: COSA's adaptive learning-rate signal repurposed as the SCR gate. *Fixed gate* (g=1): uniform cross-variate refinement without input-dependent modulation. SE_t: spectral entropy; $\mathrm{LBR}/\mathrm{MBR}/\mathrm{HBR}_t$: low/mid/high band energy ratios (Section 3.5).

Method	DLinear	PatchTST	MICN	Avg. MSE	vs. COSA
COSA	0.2981	0.2927	0.3241	0.3050	–
+SCR (no anchor; per-variate bottleneck)	0.2760	0.2721	0.3018	0.2833	−7.1%
+SCR (no anchor/bottleneck; full $C\times C$ mixing)	0.2745	0.2706	0.3044	0.2832	−7.2%
+SCR (loss-trend gate)	0.2712	0.2657	0.2963	0.2777	−9.0%
+SCR (fixed gate, g=1)	0.2703	0.2651	0.2954	0.2769	−9.2%
+SCR ($\mathbf{s}_t$=[SE_t])	0.2676	0.2638	0.2946	0.2753	−9.7%
+SCR ($\mathbf{s}_t$=[$\mathrm{SE}_t, \mathrm{LBR}_t$])	0.2677	0.2638	0.2938	0.2751	−9.8%
+SCR ($\mathbf{s}_t$=[$\mathrm{SE}_t, \mathrm{LBR}_t, \mathrm{MBR}_t$])	0.2680	0.2641	0.2941	0.2754	−9.7%
CoRe-PV ($\mathbf{s}_t^{(c)}$=[$\mathrm{SE}_t^{(c)}, \mathrm{LBR}_t^{(c)}, \mathrm{MBR}_t^{(c)}, \mathrm{HBR}_t^{(c)}$])	0.2671	0.2640	0.2945	0.2752	−9.8%
CoRe ($\mathbf{s}_t$=[$\mathrm{SE}_t, \mathrm{LBR}_t, \mathrm{MBR}_t, \mathrm{HBR}_t$])	0.2675	0.2646	0.2942	0.2754	−9.7%

Table 21: Online adaptation consistency over 18 backbone–dataset combinations (3 backbones $\times$ 6 datasets), each averaged over $H \in \{96, 192, 336, 720\}$. $\overline{\mathcal{R}}_T$: mean per-batch regret relative to the frozen backbone (negative = better than backbone on average); p_{neg}: fraction of batches where the adapted model underperforms the backbone. Best per row in **bold**. Overall row also reports $\overline{\mathcal{R}}_T^+$ (average regret on settings where the method is worse than the backbone) and n_{worse}: number of backbone–dataset combinations (out of 18) with positive mean regret.

		TAFAS		**COSA**		**CoRe**	
Model	**Dataset**	$\overline{\mathcal{R}}_T$	p_{neg}	$\overline{\mathcal{R}}_T$	p_{neg}	$\overline{\mathcal{R}}_T$	p_{neg}
DLinear	ETTh1	−0.003	0.519	−0.041	0.402	**−0.080**	**0.300**
	ETTh2	+0.003	0.570	−0.056	0.279	**−0.089**	**0.223**
	ETTm1	−0.024	**0.330**	+0.013	0.383	**−0.035**	0.382
	ETTm2	+0.001	0.546	−0.026	0.391	**−0.055**	**0.330**
	Exchange	−0.013	0.213	−0.262	0.142	**−0.274**	**0.088**
	Weather	−0.011	0.395	−0.061	0.183	**−0.094**	**0.139**
MICN	ETTh1	−0.030	0.419	−0.145	0.190	**−0.187**	**0.146**
	ETTh2	+0.011	0.531	−0.066	0.244	**−0.098**	**0.189**
	ETTm1	−0.020	0.413	−0.013	0.335	**−0.043**	**0.331**
	ETTm2	+0.004	0.549	−0.025	0.387	**−0.060**	**0.308**
	Exchange	−0.127	0.364	−0.403	0.128	**−0.419**	**0.095**
	Weather	+0.037	0.629	−0.059	0.194	**−0.089**	**0.159**
PatchTST	ETTh1	−0.007	0.434	−0.057	0.349	**−0.089**	**0.268**
	ETTh2	+0.000	0.444	−0.057	0.280	**−0.076**	**0.245**
	ETTm1	−0.010	0.348	−0.001	0.350	**−0.026**	**0.352**
	ETTm2	+0.004	0.522	−0.025	0.376	**−0.059**	**0.307**
	Exchange	−0.003	0.389	−0.243	0.161	**−0.261**	**0.131**
	Weather	+0.005	0.417	−0.041	0.207	**−0.082**	**0.173**
Overall avg		−0.010	0.446	−0.087	0.277	**−0.117**	**0.232**
$\overline{\mathcal{R}}_T^+$		0.0138		0.0127		**0.0108**	
n_{worse} / 18		8		5		**2**	

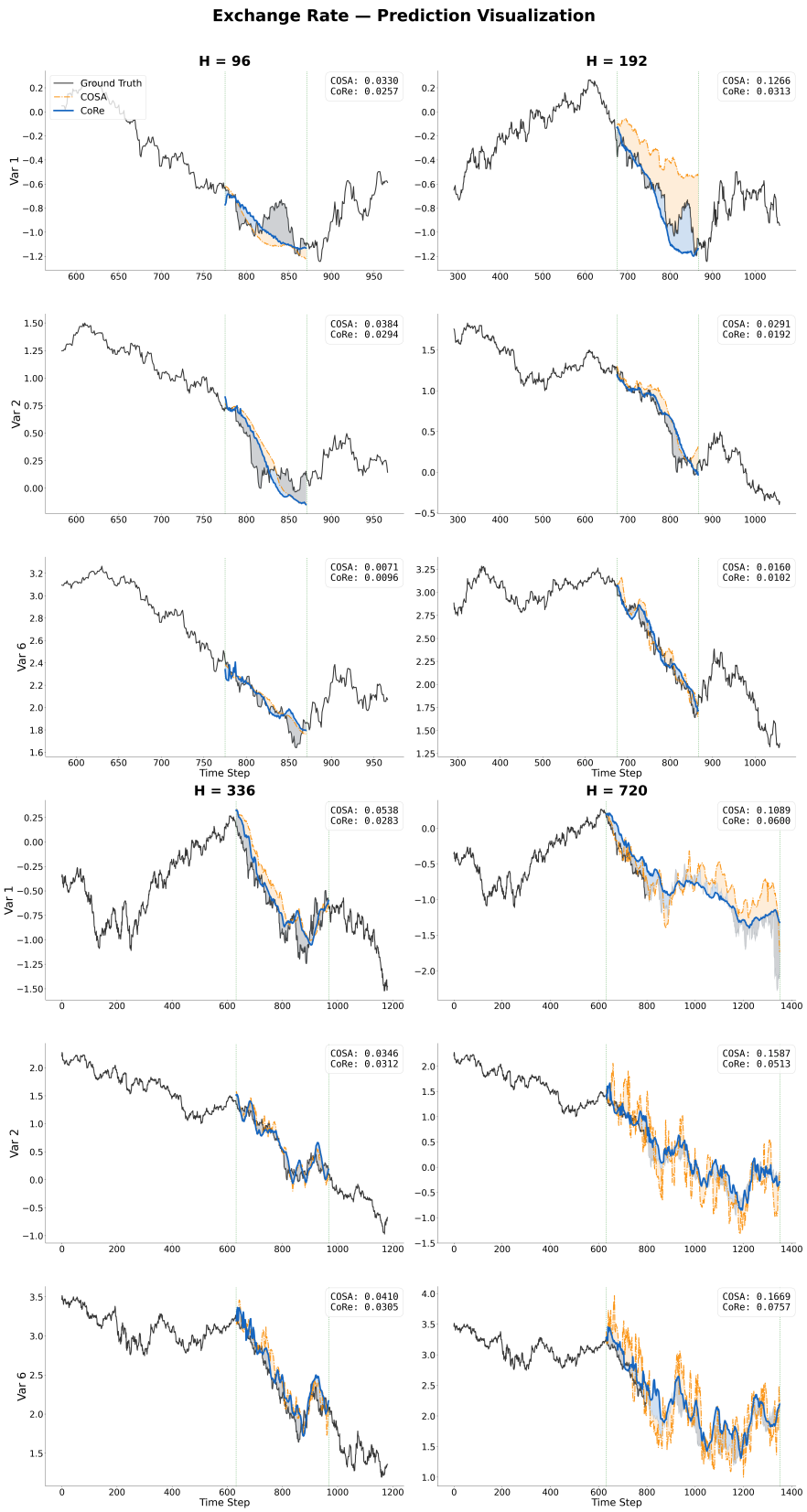

Figure 10: Qualitative comparison between COSA and CoRe on Exchange Rate across four prediction horizons and three selected variates.

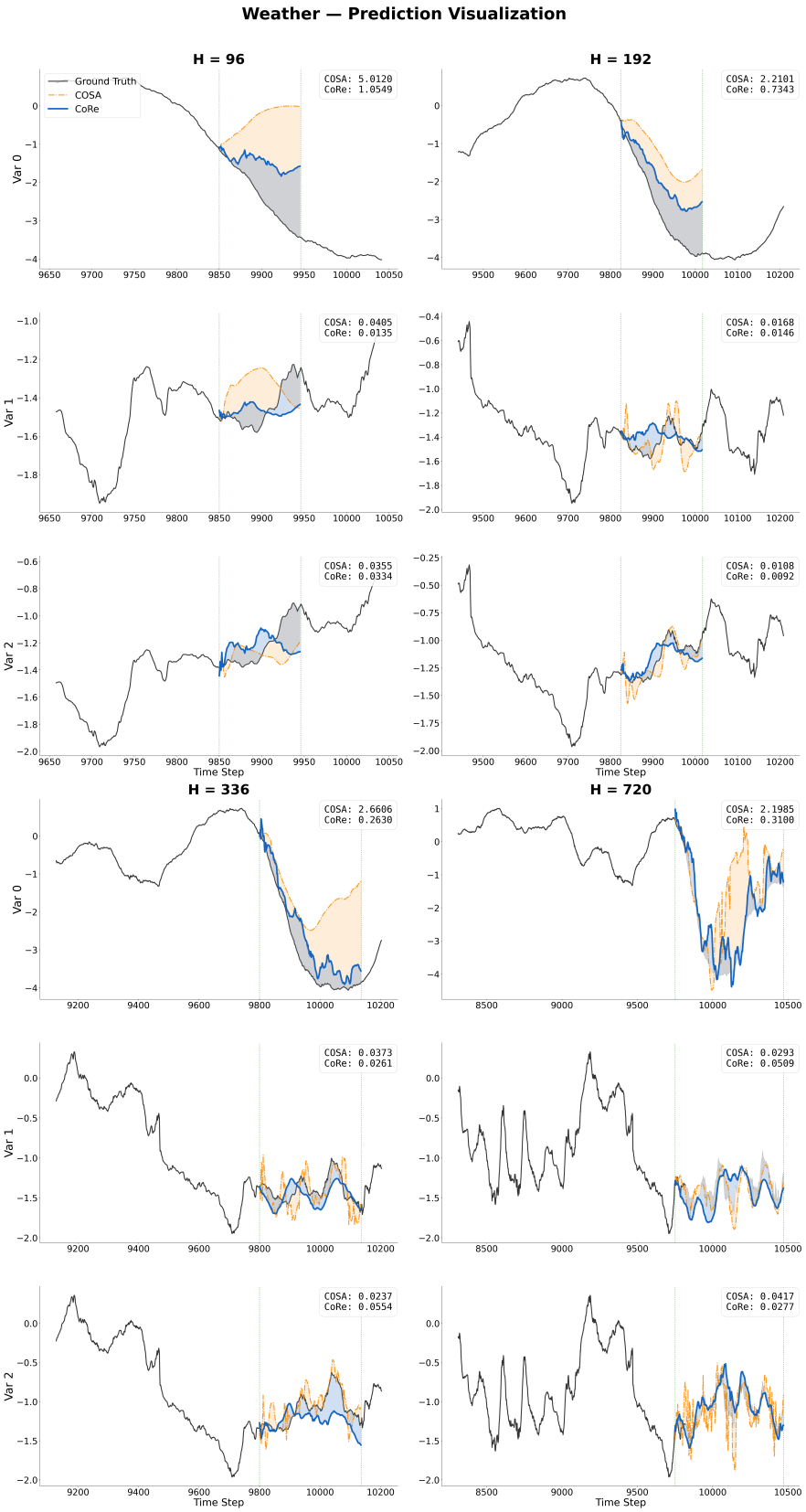

Figure 11: Qualitative comparison between COSA and CoRe on Weather across four prediction horizons and three selected variates.