STEPS: A TEMPORAL SMOOTH ERROR PROPAGATION SOLVER ON THE MANIFOLDS FOR TEST-TIME ADAPTATION IN TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Test time adaptation (TTA) for time series forecasting must revise a forecast from a short prefix of delayed targets that may contain transient changes or corrupted measurements. Existing methods often gain more at long horizons than at short horizons and are sensitive to unreliable observations, as noisy or transient errors can be propagated into the unseen suffix. We observe that forecast errors exhibit two complementary structures: persistent cross window bias is concentrated in low frequency modes, while newly revealed residuals contain useful short term structure whose influence on future errors must be selectively propagated rather than indiscriminately amplified. Based on these observations, we propose STEPS, a Spectral–Temporal Error Propagation Solver that adapts a frozen forecaster in output space. Multiscale Spectral Bias Reconstruction (MSBR) recovers persistent bias from historical residual spectra; Short Term Temporal Error Propagation (ST-TEP) learns how the revealed prefix should influence the unseen suffix through an empirical temporal kernel; and Orthogonal Complementary Correction (OCC) removes their overlapping component before fusion. Across six datasets, seven backbones, and four horizons, STEPS delivers a 24.85% average MSE reduction, which is more than twice the gain achieved by the strongest TTA baseline. Under 1–30% Gaussian prefix corruption, STEPS significantly outperforms competing methods. These results demonstrate accurate and robust forecast revision across temporal scales without online backbone optimization.

1 INTRODUCTION

Time series forecasters are deployed in streams whose dynamics continue to change after training. Long horizon models use basis expansion and hierarchical interpolation Oreshkin et al. (2020); Challu et al. (2023), decomposition and frequency domain modeling Wu et al. (2021); Zhou et al. (2022), or multi period representations Wu et al. (2023). Shifts in trend, scale, and seasonality alter both marginal statistics and the relation between historical inputs and future targets Kim et al. (2022); Liu et al. (2022); Dai et al. (2024); Ye et al. (2024). A frozen model can consequently develop structured errors even when it was accurate on the source distribution. Test time adaptation (TTA) addresses this mismatch by adjusting a pretrained model with target domain evidence available during inference Sun et al. (2020); Wang et al. (2021; 2022).

Forecasting provides a distinctive form of test time evidence: after a forecast is issued, its targets become observable sequentially, so a prefix of ground truth is available before the full horizon matures Kim et al. (2025); Medeiros et al. (2025); Im & Kwon (2026). This feedback is both limited and uncertain. A short prefix may already reveal an emerging deviation, but its influence on the unseen future can vary substantially across positions; meanwhile, the observed prefix may contain measurement noise, outliers, or transient disturbances.

Existing forecasting TTA methods use this delayed feedback to calibrate a frozen forecaster or correct its outputs Kim et al. (2025); Medeiros et al. (2025); Im & Kwon (2026). However, our benchmark in Figure 1 exposes two remaining difficulties: competing methods gain substantially less from short horizon feedback than from long horizon feedback, and their corrections deteriorate when the

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.

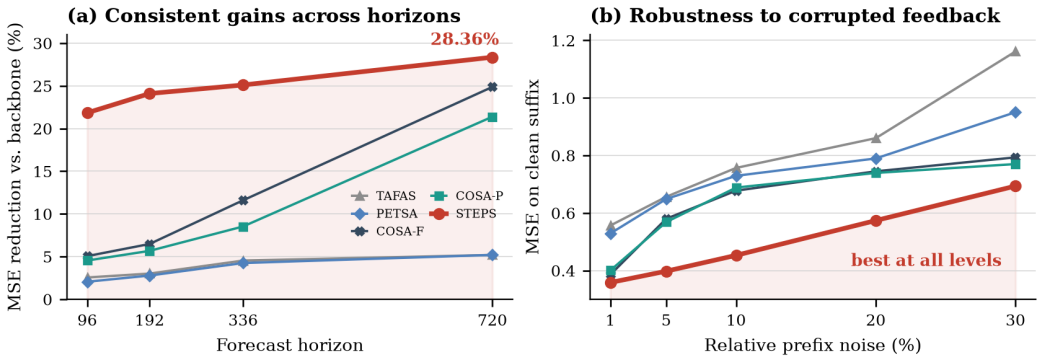

Figure 1: **STEPS remains effective with limited feedback and robust to corrupted observations.** Left: mean cellwise relative MSE reduction over the frozen backbone across six datasets and seven backbones at each horizon. Right: DLinear suffix MSE at $H=720$ under Gaussian corruption of the revealed prefix.

revealed prefix is corrupted. An effective correction mechanism must therefore extract predictive structure from limited feedback while controlling the influence of unreliable observations.

We build on two empirical observations about forecast residuals, validated independently in Appendix G. First, persistent residual bias across forecast windows is strongly concentrated in low horizon frequency modes, making systematic error compactly recoverable from mature historical residuals. Restricting reconstruction to this subspace also attenuates unsupported high frequency perturbations. Second, newly revealed residual deviations have structured future influence: the observed prefix carries predictive information about the unseen residual trajectory, but this influence varies across future positions. Some deviations primarily affect the near future, while others induce weaker but more persistent offsets. Historical residuals therefore provide stable information about persistent bias, whereas the current prefix reveals fresh deviations from that historical state.

Motivated by these observations, we propose STEPS, a Spectral–Temporal Error Propagation Solver that keeps the forecasting backbone frozen and reconstructs future errors directly in output space. Multiscale Spectral Bias Reconstruction (MSBR) models persistent bias from multiscale low frequency residual states. Short Term Temporal Error Propagation (ST-TEP) learns how the currently revealed residual pattern influences each unseen position through an empirical temporal kernel. Since the two branches may produce overlapping corrections, Orthogonal Complementary Correction (OCC) removes the global component already represented by the local response before fusion.

Across six datasets, seven backbones, and four horizons, STEPS achieves a 24.85% average MSE reduction over the frozen backbone (Table 1). It also has the lowest suffix MSE at every tested prefix noise level (Table 2).

Our contributions are as follows:

1. We identify two complementary residual structures for forecasting TTA: persistent cross window bias is spectrally compact, while newly revealed residual deviations carry structured, position dependent information about future errors.

2. We introduce STEPS, which combines multiscale spectral bias reconstruction, short term temporal error propagation, and orthogonal complementary correction without online backbone optimization.

3. We evaluate STEPS across six datasets, seven backbones, and four horizons, showing strong average gains over competing TTA methods and improved robustness to corrupted prefix observations.

2 Related Work

Test time adaptation. Test time adaptation (TTA) enables a deployed model to adjust to target domain data during inference, typically without revisiting the original training set. Early test time training introduced auxiliary self supervised objectives on test samples Sun et al. (2020), while Tent adapted normalization parameters through entropy minimization Wang et al. (2021) and source free adaptation transferred a pretrained source hypothesis without source data Liang et al. (2020). Subsequent online TTA methods addressed model drift, noisy pseudo labels, and stability under continuously changing streams Wang et al. (2022); Gong et al. (2022); Niu et al. (2022); Yuan et al. (2023); Gong et al. (2023). Forecasting inherits the same adaptation objective, but differs fundamentally in that supervision arrives gradually after a prediction has already been issued.

TTA in non stationary forecasting. Forecasting models address non stationarity through several complementary strategies. RevIN normalizes each input instance and reverses the transformation at the output Kim et al. (2022), while Non stationary Transformers restore non stationary information after stationarization Liu et al. (2022). AdaRNN and invariant learning approaches instead seek representations that transfer across evolving temporal distributions Du et al. (2021); Liu et al. (2024a). More recent methods model changing statistics in both time and frequency domains Dai et al. (2024); Ye et al. (2024) or adapt online under distribution shift Qian et al. (2024). These methods improve robustness to changes in trend, scale, or seasonality, but they do not use targets that become available after a forecast is issued.

Forecasting specific TTA exploits this delayed supervision. TAFAS combines periodicity aware adaptation scheduling with gated calibration modules while freezing the source forecaster Kim et al. (2025). PETSA updates lightweight low rank input and output adapters using robust, frequency, and patch structure objectives Medeiros et al. (2025). COSA directly corrects predictions with a gated output space residual adapter conditioned on recent observations Im & Kwon (2026). These methods establish the value of partially observed targets, but do not explicitly decompose current error into persistent cross window bias and the position dependent effect of a newly revealed deviation. STEPS estimates these two residual structures separately and then fuses their complementary responses.

Spectral structure and temporal propagation. Frequency domain forecasting methods exploit global spectral views and energy compaction to model trend and periodicity Yi et al. (2023); Ye et al. (2024). Separately, manifold learning represents data through local geometric relationships Tenenbaum et al. (2000); Belkin & Niyogi (2001); Coifman & Lafon (2006), and manifold regularization extends supervision over this geometry Belkin et al. (2006). Graph based semi supervised learning similarly propagates observed labels to unknown nodes through affinity based smoothness Zhu et al. (2003); Zhou et al. (2003); Calder et al. (2020). In time series forecasting, graph structure is usually defined across variables or sensors Wu et al. (2020); Marisca et al. (2024). STEPS instead constructs these two views over the *forecast residual field*: it represents persistent residual bias in a low frequency horizon subspace and learns an empirical temporal kernel that propagates revealed residuals toward unobserved forecast positions. Figure 2 summarizes the STEPS pipeline. A frozen backbone first produces the original forecast. STEPS then reconstructs persistent cross window bias, estimates the future influence of the revealed prefix error, and combines the two corrections before revising the unobserved suffix.

3 Methodology

3.1 Overview

STEPS reconstructs the residual of a frozen forecaster from two complementary signals: mature historical residuals capture persistent cross window bias, while newly revealed targets expose deviations specific to the current window. **Multiscale Spectral Bias Reconstruction (MSBR)** models the former from low frequency residual states, **Short Term Temporal Error Propagation (ST-TEP)** estimates how the revealed prefix influences the unseen suffix, and **Orthogonal Complementary Correction (OCC)** removes overlap between the two responses before fusion.

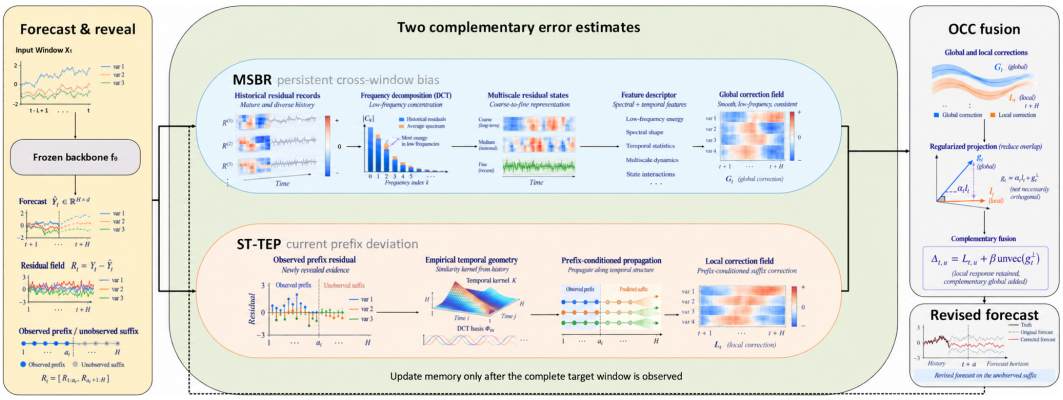

Figure 2: Overview of STEPS on top of a frozen forecasting backbone. Given the original forecast $\widehat{Y}_t$, Multiscale Spectral Bias Reconstruction (MSBR) uses mature historical residuals to reconstruct a persistent global bias field, while Short Term Temporal Error Propagation (ST-TEP) maps the newly revealed prefix residual to a position dependent local correction over the unseen suffix. Orthogonal Complementary Correction (OCC) removes the component of the global response already represented by the local response and combines the remaining complementary information.

Problem formulation. Given $X_t \in \mathbb{R}^{L\times d}$, a frozen backbone produces $\widehat{Y}_t = f_\theta(X_t) \in \mathbb{R}^{H\times d}$ with residual $R_t = Y_t - \widehat{Y}_t$. At revision time, the first a_t targets are revealed:

$$B_t = \{1,\dots,a_t\}, \quad U_t = \{a_t+1,\dots,H\}, \qquad \widehat{Y}^{\mathrm{STEPS}}_{t,B_t} = \widehat{Y}_{t,B_t}, \quad \widehat{Y}^{\mathrm{STEPS}}_{t,U_t} = \widehat{Y}_{t,U_t} + \Delta_{t,U_t}. \tag{1}$$

All correction operators are calibrated offline and remain fixed online; historical states are updated only after complete targets become available.

3.2 Multiscale Spectral Bias Reconstruction (MSBR)

Observation 1: persistent residual bias is spectrally concentrated. Persistent forecast residuals contain substantial energy in low horizon frequency modes, including after removing the constant component (Section G). This motivates a compact spectral representation of systematic bias, while spectral truncation also suppresses unsupported high frequency perturbations.

Multiscale residual states. Let Q_H be the orthonormal DCT-II matrix and $\Phi_q = Q_H^\top[:, 0:q]$. For each fully observed historical window j, we retain the first $r = 4$ residual coefficients and aggregate them over multiple history widths:

$$C_j = Q_H R_j, \qquad S_j = C_j[0:r,:], \qquad \overline{S}_t^{(w)} = \frac{1}{|\mathcal{M}_t^{(w)}|}\sum_{j\in\mathcal{M}_t^{(w)}} S_j, \quad w \in \mathcal{W}, \tag{2}$$

where $\mathcal{W} = \{4, 12, 24, 48, 96, 192, 336, 720, 1080\}$ and empty states are set to zero.

Using $S_f = \overline{S}_t^{(4)}$, $S_m = \overline{S}_t^{(96)}$, and $S_s = \overline{S}_t^{(1080)}$, we summarize multiscale changes as

$$D_t = \frac{S_f - S_s}{1 + |S_f| + |S_s|}, \qquad J_t = \frac{S_f - 2S_m + S_s}{1 + |S_f| + |S_s|}. \tag{3}$$

Conditional bias reconstruction. For each channel c, the feature $z_{t,c}$ combines the normalized current input and forecast, multiscale residual states, history descriptors, D_t, J_t, and their interaction; implementation details are given in Appendix C. A channel specific ridge operator predicts the retained residual spectrum:

$$W_c^\star = \arg\min_W \frac{1}{N}\sum_{t\in\mathcal{T}_{\mathrm{tr}}} \left\| C_t[0:q,c] - \mu_c - W^\top \overline{z}_{t,c} \right\|_2^2 + \lambda\|W\|_F^2,$$
$$G_{t,:,c} = \Phi_q\left(\mu_c + (W_c^\star)^\top \overline{z}_{t,c}\right). \tag{4}$$

Validation selects $q \in \{8, 16, 32, 64\}$ and $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$. Thus, MSBR reconstructs persistent bias from an evolving low frequency residual state.

3.3 Short Term Temporal Error Propagation (ST-TEP)

Observation 2: newly revealed residual deviations have structured future influence. Once a deviation appears in the revealed prefix, it provides predictive information about the unseen suffix. This influence varies across future positions: some deviations mainly affect the near future, while others produce weaker but more persistent effects. ST-TEP therefore learns a prefix–suffix dependency instead of propagating the observed error uniformly.

Empirical temporal geometry. Let s_c be the training RMS residual of channel c and $\widetilde{R}_{t,:,c} = R_{t,:,c}/s_c$. Using $m = \min(64, H)$ DCT modes, we estimate

$$v_{t,c} = \Phi_m^\top \widetilde{R}_{t,:,c}, \qquad A = \frac{1}{Nd} \sum_{t,c} v_{t,c} v_{t,c}^\top, \qquad K = \Phi_m A \Phi_m^\top, \tag{5}$$

where A captures co occurring residual modes and K expresses their relationships across forecast positions. The regularization scale is

$$\tau^2 = \max \left\{ \frac{1}{NdH} \sum_{t,c} \|\widetilde{R}_{t,:,c}\|_2^2 - \frac{\mathrm{tr}(A)}{H}, 10^{-6} \right\}, \qquad \eta = (1+\nu)\tau^2, \quad \nu = 10. \tag{6}$$

Position dependent error propagation. Conditioning on the revealed prefix gives

$$T_t = K_{U_t B_t} \left(K_{B_t B_t} + \eta I \right)^{-1}, \qquad L_{t,U_t,c} = s_c T_t \frac{R_{t,B_t,c}}{s_c}. \tag{7}$$

Each row of T_t determines how the observed residual pattern contributes to one future position. ST-TEP therefore produces a position dependent suffix correction while regularizing unsupported fluctuations. Its equivalent kernel ridge form is given in Appendix B.

3.4 Orthogonal Complementary Correction (OCC)

MSBR and ST-TEP may predict overlapping correction directions, so direct addition can double count the same response. Let $\ell_t = \mathrm{vec}(L_{t,U_t})$ and $g_t = \mathrm{vec}(G_{t,U_t})$. OCC removes from the global response its component aligned with the local response:

$$\alpha_t = \frac{\ell_t^\top g_t}{\|\ell_t\|_2^2 + \varepsilon}, \qquad g_t^\perp = g_t - \alpha_t \ell_t, \qquad \Delta_{t,U_t} = L_{t,U_t} + \beta \ \mathrm{unvec}(g_t^\perp), \quad \varepsilon = 10^{-8}. \tag{8}$$

The stabilizer makes the complement approximately orthogonal: $\ell_t^\top g_t^\perp = \varepsilon \, \ell_t^\top g_t / (\|\ell_t\|_2^2 + \varepsilon)$. Thus, OCC suppresses directional overlap; it does not separate disjoint frequency bands or guarantee lower prediction error. Validation selects β from 41 equally spaced values in $[-0.5, 1.5]$, with all parameters fixed at test time.

4 Experiments

We evaluate accuracy across horizons, datasets, and frozen forecasting backbones, then test robustness to corrupted feedback and the contribution of each solver branch. Full protocol details are summarized in Appendix C.1, Table A2.

4.1 Experimental Setup

Datasets and forecasting task. We use a fixed local evaluation protocol that follows the benchmark splits and forecasting setup adopted in COSA Im & Kwon (2026) and related time series forecasting studies. Specifically, we evaluate on ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and

Table 1: Prediction accuracy comparison on seven frozen backbones: **(a)** Transformer based backbones.

Dataset	H	Transformer based											
		iTransformer						PatchTST					
		Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours	Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours
ETTh1	96	.4507†	.4411†	.4393†	.4368†	.4363†	.3928	.4312†	.4262†	.4269†	.4242†	.4238†	.3812
ETTh1	192	.5078†	.4928†	.4949†	.4961†	.4919†	.4487	.4955†	.4865†	.4854†	.4830†	.4805†	.4358
ETTh1	336	.5658†	.5629†	.5640†	.5651†	.5300†	.4955	.5559†	.5478†	.5475†	.5438†	.5320†	.4848
ETTh1	720	.7038†	.6612†	.6596†	.5958†	.5638†	.5836	.7117†	.6860†	.6822†	.6113†	.5822†	.5810
ETTh2	96	.2577†	.2549†	.2551†	.2504†	.2493†	.1978	.2362†	.2351†	.2362†	.2349†	.2343†	.1933
ETTh2	192	.3161†	.3010†	.3006†	.2983†	.2947†	.2368	.2826†	.2758†	.2773†	.2665†	.2608†	.2236
ETTh2	336	.3545†	.3352†	.3348†	.3241†	.3339†	.2730	.3199†	.3125†	.3132†	.2971†	.2978†	.2501
ETTh2	720	.4276†	.4023†	.4043†	.3487†	.3591†	.3119	.4264†	.4005†	.4012†	.3233†	.3428†	.3005
ETTm1	96	.3823†	.3558†	.3570†	.3447†	.3455†	.3384	.4024†	.3894†	.3937†	.3625†	.3626†	.3498
ETTm1	192	.4423†	.4146†	.4142†	.4124†	.4140†	.3615	.4512†	.4372†	.4413†	.4250†	.4258†	.3670
ETTm1	336	.5093†	.4754†	.4751†	.4569†	.4643†	.4187	.5081†	.4905†	.4946†	.4568†	.4697†	.4130
ETTm1	720	.6065†	.5562†	.5553†	.4773†	.5102†	.4873	.5629†	.5427†	.5462†	.4681†	.4882†	.4712
ETTm2	96	.1647†	.1634†	.1637†	.1627†	.1632†	.1342	.1584†	.1581†	.1583†	.1558†	.1562†	.1299
ETTm2	192	.2209†	.2183†	.2173†	.2171†	.2173†	.1663	.2059†	.2036†	.2037†	.2007†	.2022†	.1583
ETTm2	336	.2727†	.2630†	.2592†	.2435†	.2535†	.2041	.2458†	.2451†	.2452†	.2258†	.2352†	.1897
ETTm2	720	.3451†	.3305†	.3332†	.2477†	.2606†	.2466	.3268†	.3268†	.3256†	.2446†	.2645†	.2434
Exchange	96	.0882†	.0876†	.0885†	.0818†	.0837†	.0471	.0867†	.0843†	.0837†	.0765†	.0788†	.0429
Exchange	192	.1811†	.1686†	.1740†	.1403†	.1479†	.0890	.1877†	.1805†	.1832†	.1464†	.1570†	.0858
Exchange	336	.3428†	.3079†	.3097†	.2089†	.2624†	.1608	.3389†	.3275†	.3300†	.1983†	.2445†	.1517
Exchange	720	.8540†	.8322†	.8004†	.3421†	.4460†	.4116	.8648†	.8659†	.8643†	.3543†	.4662†	.4372
Weather	96	.1755†	.1664†	.1674†	.1597†	.1617†	.1355	.1742†	.1724†	.1743†	.1624†	.1634†	.1523
Weather	192	.2232†	.2101†	.2128†	.2067†	.2088†	.1746	.2195†	.2147†	.2167†	.2006†	.2108†	.1907
Weather	336	.2800†	.2614†	.2665†	.2503†	.2515†	.2098	.2766†	.2666†	.2701†	.2451†	.2488†	.2212
Weather	720	.3571†	.3458†	.3459†	.2480†	.2730†	.2734	.3544†	.3383†	.3442†	.2590†	.2713†	.2862

Rows report full horizon MSE averaged over five seeds for $H \in \{96, 192, 336, 720\}$; lower is better, and the best value within each backbone group and row is underlined. † marks values taken from the published COSA comparison under the same frozen backbone and evaluation protocol as the local baselines. TimeMixer reference values use the official TAFAS, PETSA, and COSA implementations with their published configurations. Ours denotes delayed prefix STEPS; unavailable entries are shown as -. Local accuracy evaluations used two NVIDIA GeForce RTX 3090 GPUs; the separate timing studies state their own hardware. COSA-F and COSA-P are variants of the same COSA method.

Weather. The four ETT datasets use chronological 60/20/20 train, validation, and test splits, while Exchange and Weather use 70/10/20 splits. We use variable wise univariate forecasting, standardize each dataset using statistics computed from the training split, set the lookback length to $L = 96$, and evaluate forecasting horizons $H \in \{96, 192, 336, 720\}$. Dataset dimensions and evaluation origin counts are provided in Appendix D.

Forecasting backbones and TTA baselines. We use iTransformer Liu et al. (2024b), PatchTST Nie et al. (2023), DLinear Zeng et al. (2023), OLS Toner & Darlow (2024), FreTS Yi et al. (2023), MICN Wang et al. (2023), and TimeMixer Wang et al. (2024). Baselines are the frozen source model, TAFAS Kim et al. (2025), PETSA Medeiros et al. (2025), and COSA Im & Kwon (2026); COSA-F uses fixed $B = 48$ and COSA-P uses PAAS scheduling.

Source model training. Each cell uses the same frozen source backbone weights and preprocessing across methods and five TTA seeds. Only the TTA components vary across seeds. Source model conventions are summarized in Appendix C.1.

Baseline configuration. We use the published or official implementation settings for all baselines. For COSA, $K = 10$, $S = 3$, cosine adaptive learning rate, and mean aggregation are fixed. Pub-

lished COSA entries are marked by †; published and local baseline results use the same backbone weights, splits, and forecast origins.

STEPS fitting and test time protocol. STEPS fits MSBR and ST-TEP on training residuals only. Validation selects q, λ, and β; all choices are then frozen. The main protocol reveals $a = \lfloor H/2 \rfloor$ targets and corrects only the suffix. Historical residual states enter memory only after a complete target window is observed. The full search grids and causal schedule are in Appendix C.

Metrics and reporting. All accuracy entries are means over five TTA seeds with fixed backbone weights. We report full horizon MSE in the main text and full horizon MAE and strict suffix MSE in Appendix Table A14. For each horizon, results average 42 backbone–dataset pairs. Relative gains are macro averaged from cellwise means over five seeds:

$$\mathrm{Gain}_R = \frac{100}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \frac{\mathrm{MSE}_{R,i} - \mathrm{MSE}_{\mathrm{Ours},i}}{\mathrm{MSE}_{R,i}}, \tag{9}$$

$$R \in \{\mathrm{Baseline}, \mathrm{TAFAS}, \mathrm{PETSA}, \mathrm{COSA\text{-}F}, \mathrm{COSA\text{-}P}\}. \tag{10}$$

Published entries from the COSA benchmark are marked by †. Aggregate comparisons use only available cells, and missing results are never imputed.

4.2 Main Prediction Accuracy Comparison

Table 1(a)–(c) reports all four horizons in three backbone family panels. Improvements are summarized below and detailed cellwise in the appendix.

4.3 Aggregate Results

Table 1(a)–(c) contains 168 cells. Figure 3 in Appendix E summarizes their cellwise relative gains.

Across all 168 cells, the mean relative MSE reductions are 3.83% for TAFAS, 3.57% for PETSA, 12.00% for COSA-F, 10.03% for COSA-P, and 24.85% for STEPS. STEPS leads at each horizon, with gains of 21.85%, 24.10%, 25.11%, and 28.36% at $H \in \{96, 192, 336, 720\}$, respectively. These values average cellwise ratios rather than pooled errors; they summarize magnitude, not significance.

4.4 Robustness to Noisy Prefix Observations

We evaluate DLinear at $H = 720$ on ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and Weather. The first 320 steps are treated as observed. For each $\rho \in \{1, 5, 10, 20, 30\}\%$, a fraction ρ of the observed prefix entries is independently corrupted with zero mean Gaussian noise of standard deviation $6\sigma_c$, where σ_c is the channel wise training residual scale. The same noise realization is used by all methods, and MSE is computed on the remaining 400 clean steps. Table 2 reports means over five seeds with the backbone weights fixed.

Across all corruption levels, STEPS-OCC has the lowest suffix MSE. Its MSE is lower than direct addition at every level, consistent with OCC suppressing overlap between the global and local responses; the corruption protocol is summarized in Table A2.

Component analysis. On the 12-cell diagnostic cohort ($H = 192$, DLinear and PatchTST), ST-TEP accounts for most of the near suffix gain: 50.7%, compared with 3.1% for MSBR alone. Full STEPS reaches 37.4% full suffix gain versus 37.1% for ST-TEP; direct addition also rounds to 37.4%. Thus, the clean feedback ablation supports a modest complementary contribution, rather than uniform superiority of orthogonal fusion. Appendix F reports backbone wise results and validation selections; Appendix G provides the shuffled prefix and recent change controls. A DLinear sweep in Appendix H finds that fixed $a/H{=}1/2$ maximizes full horizon gain and slightly exceeds validation adaptive ratio selection.

Table 1: Prediction accuracy comparison on seven frozen backbones (continued): **(b)** Linear based backbones.

Dataset	H	Linear based											
		DLinear						OLS					
		Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours	Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours
ETTh1	96	.4695†	.4618†	.4594†	.4574†	.4482†	.4054	.4511†	.4409†	.4391†	.4390†	.4372†	.3902
ETTh1	192	.5213†	.5117†	.5118†	.5066†	.5050†	.4588	.5046†	.4934†	.4937†	.4915†	.4906†	.4432
ETTh1	336	.5659†	.5604†	.5617†	.5528†	.5456†	.4991	.5510†	.5440†	.5465†	.5385†	.5320†	.4843
ETTh1	720	.7117†	.6820†	.6743†	.6107†	.5896†	.5812	.6997†	.6630†	.6431†	.5969†	.5733†	.5688
ETTh2	96	.2323†	.2303†	.2306†	.2300†	.2281†	.1845	.2306†	.2285†	.2288†	.2232†	.2265†	.1836
ETTh2	192	.2862†	.2842†	.2876†	.2827†	.2819†	.2225	.2839†	.2824†	.2848†	.2796†	.2791†	.2207
ETTh2	336	.3252†	.3185†	.3184†	.3050†	.3083†	.2554	.3258†	.3182†	.3189†	.3003†	.3043†	.2552
ETTh2	720	.4087†	.3873†	.3853†	.3062†	.3477†	.2978	.4162†	.3908†	.3884†	.3177†	.3453†	.2960
ETTm1	96	.3715†	.3497†	.3524†	.3456†	.3475†	.3293	.3710†	.3506†	.3536†	.3454†	.3475†	.3287
ETTm1	192	.4438†	.4166†	.4178†	.4113†	.4122†	.3564	.4439†	.4160†	.4184†	.4115†	.4119†	.3562
ETTm1	336	.5183†	.4799†	.4803†	.4753†	.4858†	.4137	.5182†	.4787†	.4792†	.4748†	.4749†	.4133
ETTm1	720	.5929†	.5488†	.5532†	.4774†	.4991†	.4848	.5922†	.5478†	.5522†	.4763†	.5007†	.4844
ETTm2	96	.1598†	.1584†	.1584†	.1583†	.1586†	.1322	.1602†	.1590†	.1589†	.1582†	.1586†	.1325
ETTm2	192	.1930†	.1913†	.1913†	.1904†	.1905†	.1546	.1936†	.1921†	.1919†	.1906†	.1907†	.1547
ETTm2	336	.2324†	.2289†	.2292†	.2083†	.2242†	.1835	.2331†	.2299†	.2302†	.2131†	.2226†	.1837
ETTm2	720	.3062†	.2968†	.2963†	.2215†	.2316†	.2285	.3066†	.2986†	.2971†	.2171†	.2349†	.2279
Exchange	96	.0913†	.0885†	.0878†	.0812†	.0834†	.0475	.0814†	.0792†	.0798†	.0756†	.0773†	.0431
Exchange	192	.1827†	.1760†	.1730†	.1459†	.1519†	.0892	.1727†	.1658†	.1653†	.1393†	.1457†	.0835
Exchange	336	.3277†	.2941†	.2920†	.2039†	.2480†	.1559	.3226†	.2877†	.2898†	.2020†	.2323†	.1483
Exchange	720	.8873†	.8762†	.8781†	.3494†	.4481†	.3939	.8366†	.8138†	.8149†	.3444†	.4541†	.3799
Weather	96	.1954†	.1796†	.1823†	.1773†	.1793†	.1491	.1957†	.1807†	.1795†	.1772†	.1803†	.1489
Weather	192	.2403†	.2244†	.2254†	.2216†	.2217†	.1876	.2404†	.2244†	.2274†	.2223†	.2237†	.1874
Weather	336	.2918†	.2709†	.2740†	.2567†	.2626†	.2171	.2921†	.2714†	.2748†	.2551†	.2642†	.2170
Weather	720	.3643†	.3500†	.3497†	.2581†	.2708†	.2765	.3644†	.3466†	.3493†	.2579†	.2708†	.2763

Panels (a)–(c) of Table 1 share the same reporting protocol, underlining rule, and data provenance note; see the note under panel (a).

Table 1: Prediction accuracy comparison on seven frozen backbones (continued): **(c)** MLP based backbones.

Dataset	H	MLP based																	
		FreTS						TimeMixer						MICN					
		Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours	Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours	Baseline	TAFAS	PETSA	COSA-F	COSA-P	Ours
ETTh1	96	.4462†	.4403†	.4387†	.4384†	.4371†	.3976	.4479	.4444	.4433	.4460	.4473	.3928	.5103†	.4901†	.4898†	.4693†	.4684†	.4561
ETTh1	192	.5022†	.4954†	.4942†	.4951†	.4940†	.4445	.5073	.5033	.5017	.5068	.5096	.4487	.5954†	.5617†	.5620†	.5372†	.5328†	.5072
ETTh1	336	.5544†	.5521†	.5527†	.5467†	.5351†	.4888	.5580	.5593	.5549	.5630	.5884	.4921	.6615†	.6387†	.6420†	.5950†	.5878†	.5445
ETTh1	720	.7182†	.6852†	.6846†	.6259†	.5959†	.5857	.7091	.6899	.6676	.6815	.6750	.5783	.9233†	.8142†	.8375†	.7001†	.6504†	.6014
ETTh2	96	.2384†	.2367†	.2364†	.2367†	.2350†	.1913	.2401	.2370	.2378	.2397	.2381	.1871	.2582†	.2551†	.2552†	.2492†	.2485†	.2035
ETTh2	192	.2866†	.2824†	.2832†	.2816†	.2824†	.2271	.2855	.2821	.2819	.2846	.2843	.2253	.3282†	.3179†	.3258†	.3049†	.3017†	.2465
ETTh2	336	.3317†	.3229†	.3233†	.3031†	.3153†	.2608	.3262	.3175	.3182	.3308	.3293	.2586	.3732†	.3482†	.3497†	.3241†	.3310†	.3038
ETTh2	720	.4119†	.3857†	.3860†	.3169†	.3399†	.3004	.4087	.3905	.3852	.4091	.4011	.3000	.4617†	.4474†	.4473†	.3650†	.3885†	.3489
ETTm1	96	.3675†	.3582†	.3583†	.3520†	.3525†	.3295	.3898	.3684	.3744	.3642	.3674	.3452	.4354†	.3951†	.3951†	.3837†	.3831†	.4163
ETTm1	192	.4325†	.4212†	.4198†	.4150†	.4212†	.3542	.4413	.4273	.4287	.4243	.4221	.3656	.4855†	.4566†	.4574†	.4476†	.4514†	.4246
ETTm1	336	.5005†	.4827†	.4789†	.4661†	.4775†	.4091	.5175	.4956	.4905	.5134	.5042	.4162	.5556†	.5108†	.5082†	.4832†	.5054†	.4809
ETTm1	720	.5704†	.5486†	.5476†	.4718†	.4982†	.4707	.5758	.5618	.5511	.6180	.5822	.4777	.6212†	.5756†	.5778†	.5029†	.5225†	.5407
ETTm2	96	.1581†	.1572†	.1572†	.1568†	.1569†	.1313	.1567	.1562	.1566	.1567	.1561	.1289	.1710†	.1711†	.1730†	.1702†	.1704†	.1418
ETTm2	192	.1923†	.1909†	.1908†	.1905†	.1908†	.1553	.1880	.1878	.1875	.1910	.1894	.1525	.2121†	.2102†	.2126†	.2102†	.2120†	.1687
ETTm2	336	.2320†	.2288†	.2289†	.2098†	.2211†	.1842	.2283	.2267	.2281	.2365	.2311	.1811	.2530†	.2501†	.2520†	.2337†	.2351†	.1987
ETTm2	720	.3012†	.2916†	.2926†	.2158†	.2314†	.2275	.3029	.2965	.2932	.3304	.3101	.2270	.3327†	.3220†	.3131†	.2477†	.2643†	.2589
Exchange	96	.0828†	.0799†	.0803†	.0744†	.0766†	.0434	.0864	.0837	.0835	.0838	.0845	.0436	.1151†	.1087†	.1146†	.0955†	.1008†	.0573
Exchange	192	.1734†	.1665†	.1648†	.1366†	.1499†	.0838	.1853	.1800	.1809	.1791	.1806	.0853	.2150†	.2198†	.1999†	.1663†	.1722†	.1005
Exchange	336	.3240†	.2930†	.2923†	.2053†	.2461†	.1489	.3371	.3257	.3287	.3192	.3292	.1518	.3950†	.3047†	.3100†	.2119†	.2660†	.1702
Exchange	720	.8368†	.8273†	.8067†	.3352†	.4458†	.3747	.8838	.8740	.8728	.8833	.8838	.4001	1.0259†	.7191†	.7805†	.3871†	.4815†	.3681
Weather	96	.1856†	.1759†	.1765†	.1724†	.1737†	.1436	.1649	.1613	.1619	.1618	.1617	.1333	.1757†	.1853†	.1970†	.1636†	.1651†	.1297
Weather	192	.2310†	.2165†	.2192†	.2135†	.2189†	.1815	.2109	.2054	.2062	.2106	.2076	.1686	.2237†	.2161†	.2265†	.2082†	.2120†	.1976
Weather	336	.2843†	.2653†	.2681†	.2561†	.2587†	.2139	.2644	.2546	.2554	.2744	.2634	.2015	.2812†	.2746†	.2788†	.2429†	.2737†	.2572
Weather	720	.3599†	.3490†	.3488†	.2573†	.2692†	.2708	.6667	.4821	.5166	.5492	.5382	.4328	.3508†	.3573†	.3681†	.2582†	.2855†	.2696

Panels (a)–(c) of Table 1 share the same reporting protocol, underlining rule, and data provenance note; see the note under panel (a).

Table 2: Robustness of DLinear at $H = 720$ to relative Gaussian noise on the first 320 observed steps. Values are MSE averaged over five seeds on the remaining 400 steps; lower is better.

Method	1%	5%	10%	20%	30%
STEPS-OCC	0.3584	0.3978	0.4529	0.5739	0.6938
ST-TEP	0.3601	0.4020	0.4595	0.5842	0.7066
Additive	0.4160	0.4580	0.5154	0.6412	0.7644
PETSA	0.5290	0.6486	0.7294	0.7897	0.9504
TAFAS	0.5565	0.6564	0.7568	0.8602	1.1615
COSA-P	0.4005	0.5685	0.6884	0.7394	0.7702
COSA-F	0.3857	0.5800	0.6777	0.7449	0.7935

5 Conclusion

We introduced STEPS, an output space forecast revision method that combines historical residual structure with newly revealed prefix feedback while keeping the forecasting backbone frozen. Across six datasets, seven backbones, and four horizons, STEPS achieves an average full horizon MSE reduction of 24.85%, with gains increasing from 21.85% at $H = 96$ to 28.36% at $H = 720$. It also attains the lowest suffix MSE under all tested Gaussian prefix corruption levels from 1% to 30%. These results demonstrate the effectiveness and robustness of structured residual correction for delayed forecast revision.

A remaining limitation is the behavior of STEPS at forecasting horizons substantially longer than $H = 720$. As the prediction range increases, the observed prefix becomes farther from the distant part of the unseen suffix, so the residual dependency used by temporal propagation may become weaker and less stable. At the same time, long horizon residuals may exhibit more complex spectral patterns that cannot be represented sufficiently by the current truncated spectral space. These effects may limit the accuracy of both residual reconstruction and propagation. Future work will study substantially longer forecasting sequences and investigate more adaptive spectral and temporal modeling for long range revision.

AI use statement

Generative AI tools assisted with feedback on methods and experiments, code implementation, interpretation of reported results, figure editing, literature and reference checks, and manuscript editing and formatting. The authors reviewed the resulting text, code, figures, numerical claims, and references, and take responsibility for the complete submission.

Reproducibility statement

The complete experimental protocol, including datasets, frozen backbones, horizons, the prefix observation schedule, and the reporting templates for each cell, is specified in Section 4 and Appendix J. All correction operators are fitted offline on training splits, and all hyperparameters are selected on validation splits only; the fixed test procedure requires no further tuning. Detailed operator derivations are given in Appendix B.

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*, volume 14, 2001. URL https://proceedings.neurips.cc/paper/2001/hash/f106b7f99d2cb30c3db1c3cc0fde9ccb-Abstract.html.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7:2399–2434, 2006. URL https://www.jmlr.org/papers/v7/belkin06a.html.

Jeff Calder, Brendan Cook, Matthew Thorpe, and Dejan Slepcev. Poisson learning: Graph based semi-supervised learning at very low label rates. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1306–1316, 2020. URL https://proceedings.mlr.press/v119/calder20a.html.

Cristian Challu, Kin G. Olivares, Boris N. Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. NHITS: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 6989–6997, 2023. doi: 10.1609/aaai.v37i6.25854. URL https://ojs.aaai.org/index.php/AAAI/article/view/25854.

Ronald R. Coifman and Stephane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006. URL https://doi.org/10.1016/j.acha.2006.04.006.

Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Xue Yuerong, Shu-Tao Xia, and Zexuan Zhu. DDN: Dual-domain dynamic normalization for non-stationary time series forecasting. In *Advances in Neural Information Processing Systems*, volume 37, pp. 108490–108517, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/c44c4afd77d5ee760e7f4bed0c50f878-Abstract-Conference.html.

Yuntao Du, Jindong Wang, Wenjie Feng, Sinno Pan, Tao Qin, Renjun Xu, and Chongjun Wang. AdaRNN: Adaptive learning and forecasting of time series. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pp. 402–411, 2021. doi: 10.1145/3459637.3482315. URL https://arxiv.org/abs/2108.04443.

Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. NOTE: Robust continual test-time adaptation against temporal correlation. In *Advances in Neural Information Processing Systems*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/ae6c7dbd9429b3a75c41b5fb47e57c9e-Abstract-Conference.html.

Taesik Gong, Yewon Kim, Taeckyung Lee, Sorn Chottananurak, and Sung-Ju Lee. SoTTA: Robust test-time adaptation on noisy data streams. In *Advances in Neural Information Processing Systems*, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/2da53cd1abdae59150e35f4693834f32-Abstract-Conference.html.

Jeonghwan Im and Hyuk-Yoon Kwon. COSA: Context-aware output-space adapter for test-time adaptation in time series forecasting. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/2a8ce71baac4c89bf9ff479d8240c7d9-Abstract-Conference.html.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17868–17876, 2025. doi: 10.1609/aaai.v39i17.33965. URL https://ojs.aaai.org/index.php/AAAI/article/view/33965.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=cGDAkQo1C0p.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 6028–6039, 2020. URL https://proceedings.mlr.press/v119/liang20a.html.

Haoxin Liu, Harshavardhan Kamarthi, Lingkai Kong, Zhiyuan Zhao, Chao Zhang, and B. Aditya Prakash. Time-series forecasting for out-of-distribution generalization using invariant learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 31312–31325, 2024a. URL https://proceedings.mlr.press/v235/liu24ae.html.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems*, volume 35, pp. 9881–9893, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4054556fcaa934b0bf76da52cf4f92cb-Abstract-Conference.html.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024b. URL https://openreview.net/forum?id=JePfAI8fah.

Ivan Marisca, Cesare Alippi, and Filippo Maria Bianchi. Graph-based forecasting with missing data through spatiotemporal downsampling. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 34846–34865, 2024. URL https://proceedings.mlr.press/v235/marisca24a.html.

Heitor R. Medeiros, Hossein Sharifi-Noghabi, Gabriel L. Oliveira, and Saghar Irandoust. Accurate parameter-efficient test-time adaptation for time series forecasting. In *Second Workshop on Test-Time Adaptation: Putting Updates to the Test! at ICML*, 2025. URL https://arxiv.org/abs/2506.23424.

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

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 16888–16905, 2022. URL https://proceedings.mlr.press/v162/niu22a.html.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=Keah34oiobzl.

Yu-Yang Qian, Peng Zhao, Yu-Jie Zhang, Masashi Sugiyama, and Zhi-Hua Zhou. Efficient non-stationary online learning by wavelets with applications to online distribution shift adaptation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 41383–41415, 2024. URL https://proceedings.mlr.press/v235/qian24c.html.

Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 9229–9248, 2020. URL https://proceedings.mlr.press/v119/sun20b.html.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000. doi: 10.1126/science.290.5500.2319. URL https://pubmed.ncbi.nlm.nih.gov/11125149/.

William Toner and Luke Nicholas Darlow. An analysis of linear time series forecasting models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 48404–48427, 2024. URL https://proceedings.mlr.press/v235/toner24a.html.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=uXl3bZLkr3c.

Huiqiang Wang, Jian Peng, Feihu Huang, Jince Wang, Junhui Chen, and Yifei Xiao. MICN: Multi-scale local and global context modeling for long-term series forecasting. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=zt53IDUR1U.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7201–7211, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Wang_Continual_Test-Time_Domain_Adaptation_CVPR_2022_paper.html.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. TimeMixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=7oLshfEIC2.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bcc0d400288793e8bdcd7c19a8ac0c2b-Abstract.html.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 753–763, 2020. doi: 10.1145/3394486.3403118. URL https://arxiv.org/abs/2005.11650.

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. In *Advances in Neural Information Processing Systems*, volume 37, pp. 31350–31379, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/37c6d0bc4d2917dcbea693b18504bd87-Abstract-Conference.html.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series forecasting. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/f1d16af76939f476b5f040fd1398c0a3-Abstract-Conference.html.

Longhui Yuan, Binhui Xie, and Shuang Li. Robust test-time adaptation in dynamic scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15922–15932, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Yuan_Robust_Test-Time_Adaptation_in_Dynamic_Scenarios_CVPR_2023_paper.html.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11121–11128, 2023. URL https://ojs.aaai.org/index.php/AAAI/article/view/26317.

Dengyong Zhou, Olivier Bousquet, Thomas N. Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *Advances in Neural Information Processing Systems*, volume 16, 2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/hash/87682805257e619d49b8e0dfdc14affa-Abstract.html.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286, 2022. URL https://proceedings.mlr.press/v162/zhou22g.html.

Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International Conference on Machine Learning*, pp. 912–919, 2003. URL https://mlg.eng.cam.ac.uk/pub/pdf/ZhuGhaLaf03a.pdf.

A NOTATION

Table A1 collects the symbols used in the main text and this appendix.

Table A1: Notation used throughout the paper.

Symbol	Meaning
Problem setting	
$X_t \in \mathbb{R}^{L\times d}$	Input context of lookback length L with d channels
$\widehat{Y}_t = f_\theta(X_t) \in \mathbb{R}^{H\times d}$	Frozen backbone forecast over horizon H
Y_t, $R_t = Y_t - \widehat{Y}_t$	Realized targets and the residual (error) field
a_t	Revealed prefix length at revision time, $1 \le a_t < H$ (main protocol: $a = \lfloor H/2 \rfloor$)
$B_t = \{1, \dots, a_t\}$, U_t	Observed prefix and unobserved suffix index sets
Δ_{t,U_t}	Correction applied to the suffix of the frozen forecast
MSBR (global branch)	
Q_H, Φ_q	Orthonormal DCT-II matrix; its first q temporal basis vectors
$C_j = Q_H R_j$, $S_j = C_j[0{:}r, :]$	Residual spectrum of window j; retained $r = 4$ coefficients
$\mathcal{M}_t^{(w)}$, $\overline{S}_t^{(w)}$	Mature record set of width w; its mean spectral state
$\mathcal{W}$	History widths $\{4, 12, 24, 48, 96, 192, 336, 720, 1080\}$
S_f, S_m, S_s	Fast (w=4), intermediate (w=96), slow (w=1080) states
D_t, J_t	Normalized fast–slow and fast–intermediate–slow contrasts
$u_{t,c}$, $z_{t,c}$, $\overline{z}_{t,c}$	Current descriptor; full feature; standardized feature
$W_c^\star$, μ_c, λ, q	Channel wise ridge operator; target mean; penalty; bandwidth
$G_{t,:,c}$	Reconstructed global (bias) correction field
ST-TEP (local branch)	
s_c, $\widetilde{R}_{t,:,c}$	Training residual RMS of channel c; normalized residual
$m = \min(64, H)$	Number of retained temporal modes
$v_{t,c}$, A, $K = \Phi_m A \Phi_m^\top$	Mode coefficients; second moment; empirical temporal kernel
τ^2, $\eta = (1+\nu)\tau^2$, ν	Out of basis energy; regularization; noise multiplier ($\nu = 10$)
$T_t = K_{U_t B_t}(K_{B_t B_t} + \eta I)^{-1}$	Position dependent propagation gain
$L_{t,U_t,c}$	Local (prefix propagated) correction response
$F = \Phi_m A^{1/2}$, $u^\star$	Variational factor and its ridge solution
OCC (fusion)	
ℓ_t, g_t	Vectorized suffix responses of the local and global branches
α_t, $g_t^\perp = g_t - \alpha_t \ell_t$	Overlap coefficient; orthogonalized global response
β, ε	Fusion coefficient (validation selected); stabilizer 10^{-8}

B EMPIRICAL TEMPORAL KERNEL AND PROJECTION PROPERTIES

One dimensional temporal geometry. Let $D \in \mathbb{R}^{(H-1)\times H}$ be the first difference matrix and $\mathcal{L} = D^\top D$. The normalized DCT-II basis diagonalizes this path graph Laplacian, with eigenvalues $\lambda_k = 2 - 2\cos(\pi k/H)$, $k = 0, \dots, H-1$. Thus for $f = \Phi_m v$,

$$\|Df\|_2^2 = \sum_{k=0}^{m-1} \lambda_k v_k^2 \le \lambda_{m-1} \|f\|_2^2. \tag{11}$$

Truncation restricts the available temporal frequencies, but the bound becomes weaker as m/H increases. In particular, retaining 64 modes at $H = 96$ is a broad spectral model, not a strict low pass model. The empirical second moment A determines the relative strength and cooccurrence of these modes. Neither A nor this representation enforces a monotone decay of energy with frequency. The data diagnostics, rather than this algebra alone, motivate short range coherence.

Exact variational interpretation of the implemented operator. Let $A^{1/2}$ denote a positive semidefinite square root and $F = \Phi_m A^{1/2}$, so $K = FF^\top$. For one channel and prefix residual

$r_B = R_B/s_c$, consider

$$u^\star = \arg\min_u \|F_B u - r_B\|_2^2 + \eta\|u\|_2^2. \tag{12}$$

Because $\eta > 0$, this problem has the unique solution

$$u^\star = (F_B^\top F_B + \eta I)^{-1} F_B^\top r_B, \tag{13}$$

$$F_U u^\star = F_U F_B^\top (F_B F_B^\top + \eta I)^{-1} r_B = K_{UB}(K_{BB} + \eta I)^{-1} r_B. \tag{14}$$

Multiplication by s_c gives Eq. (7), including when A is singular. Equivalently, among $f \in \mathrm{range}(F)$ this minimizes $\|f_B - r\|_2^2 + \eta f^\top K^\dagger f$. The observed prefix is a conditioning set, not a single geometric endpoint, and the observation fit is soft. This is therefore an empirical kernel extension on a temporal domain, not the classical hard boundary harmonic extension on an unweighted chain. Under a fixed shared gain the input/output channel scalings cancel; normalization matters when pooling channels to estimate A.

Sensitivity, not absolute boundedness. For $T = K_{UB}(K_{BB} + \eta I)^{-1}$, changing a normalized prefix by e changes the normalized response by Te. Consequently

$$\|Te\|_2 \le \|T\|_2\|e\|_2, \qquad \|T\|_2 \le \|K_{UB}\|_2/\eta. \tag{15}$$

This is a finite sensitivity bound for a fitted kernel. It does not make the solver robust to arbitrary outliers and does not replace a clipping rule; the current implementation has no final clipping.

Orthogonal fusion and its tradeoff. For nonzero ℓ, the exact projector $P_\ell = \ell\ell^\top/\|\ell\|^2$ satisfies $P_\ell^2 = P_\ell$ and $\ell^\top(I - P_\ell)g = 0$. The fused response preserves the coefficient of the local direction while adding only a complementary global direction. The implementation uses the regularized projector $\ell\ell^\top/(\|\ell\|^2 + \epsilon)$, yielding the residual overlap stated in the main text. Orthogonality does not imply that the two estimates correspond to disjoint frequency bands: both may contain low frequency components. It also does not prove that $g^\perp$ is closer to the true remaining error. Validation controls its signed amplitude, but cannot guarantee a benefit under a new distribution.

C Implementation and Diagnostic Reproducibility

Fixed implementation. The same frozen backbone weights are used for all five TTA seeds in each cell. For each seed, fit the local 64-mode second moment on training residuals, construct the gain with $a = H/2$ and noise multiplier 10, and fit standardized channel wise global ridge operators on training data. Select global spectral dimension and ridge penalty on the full validation split. Compute validation responses and select the orthogonal fusion coefficient on its first 60%. Test evaluation keeps all these choices fixed. No dataset specific hypothesis test thresholds are selected on test.

Global feature details. For normalized input x and forecast p, the four nonspectral current descriptors are x_L, $x_L - \mathrm{mean}(x_{L-23:L})$, $\mathrm{mean}(p)$, and $p_H - p_1$. History length n_t, capped at 1080, is represented by $n_t/1080$ and $\log(1 + n_t)/\log(1081)$. Each channel has 78 features: 28 current descriptors, 36 historical coefficients, 2 length descriptors, 8 state contrasts, and 4 interactions. The historical record bank is shared in its temporal schedule, but ridge coefficients are channel specific. At split boundaries the runner carries earlier split records as pre existing history; a streaming deployment must ensure that these records are fully observed before admitting them.

Reading the two algorithms. Algorithm 2 is the offline stage and is executed once cellwise and TTA seed; Algorithm 1 is the online stage and runs per forecast window. Lines 2–4 of Algorithm 2 build the ST-TEP operator: normalization (line 2), the empirical second moment in the DCT basis (line 3), and a regularizer set by a fixed rule rather than a search (line 4). Lines 5–7 build MSBR: the multiscale memory states (line 5), the channel wise ridge fit (line 6), and validation selection of the only two searched structural hyperparameters (line 7), using the Parseval identity so the spectral objective equals the time domain SSE exactly. Line 8 selects the fusion weight on a restricted validation slice, and line 9 freezes everything; the online stage therefore performs no gradient steps, no per window optimization, and no hyperparameter updates. The per window online cost is one descriptor construction, one ridge application, one banded kernel solve of size a_t, and one scalar projection.

Algorithm 1 STEPS: global bias reconstruction, local error propagation, and orthogonal fusion

Require: Frozen backbone, fitted global operators, local kernel gain T, selected β

1: Produce $\hat{Y}_t = f_\theta(X_t)$.
2: **Track**: aggregate only mature historical residual records into multiscale states.
3: **Estimate**: form current/state descriptors and reconstruct global response G_t in the DCT basis.
4: **Sense**: wait until the first $a = H/2$ targets are revealed; obtain $R_{t,B}$.
5: **Propagate**: compute local response $L_{t,U,c} = s_c T(R_{t,B,c}/s_c)$.
6: **Synthesize**: remove the regularized projection of $G_{t,U}$ onto $L_{t,U}$.
7: Correct the suffix with $L_{t,U} + \beta G^{\perp}_{t,U}$; retain the original prefix prediction.
8: Admit the current residual record to historical states only once its complete target is observed.

Algorithm 2 STEPS offline calibration (run once per TTA seed and backbone–dataset–horizon cell)

Require: Training split $\mathcal{T}_{\mathrm{tr}}$, validation split $\mathcal{T}_{\mathrm{val}}$, frozen backbone f_θ

1: Generate forecasts on $\mathcal{T}_{\mathrm{tr}}$ and $\mathcal{T}_{\mathrm{val}}$; form residuals R_t. ▷ frozen caches
2: Estimate channel scales s_c from training residual RMS. ▷ train only normalization
3: Accumulate the 64-mode second moment A over all training windows; build $K = \Phi_m A \Phi_m^\top$. ▷ ST-TEP kernel
4: Set the regularizer from out of basis energy, $\eta = (1+\nu)\tau^2$ with $\nu = 10$.▷ fixed rule, no tuning
5: Build mature record multiscale states $\overline{S}_t^{(w)}$ on training residuals. ▷ MSBR memory
6: Standardize features $\overline{z}_{t,c}$ with training statistics; solve the channel wise ridge for $W_c^\star$. ▷ Eq. (4)
7: Select (q, λ) by validation spectral SSE over $q \in \{8, 16, 32, 64\}$, $\lambda \in \{10^{-3}, \ldots, 10^{2}\}$. ▷ Parseval identity
8: Compute validation responses L and G; select β on the first 60% of validation origins from 41 values in $[-0.5, 1.5]$. ▷ Eq. (8)
9: Freeze $\{s_c, K, \eta, W_c^\star, q, \lambda, \beta\}$ for all test evaluations. ▷ nothing is refit online

Diagnostic sampling and provenance. The structural tests cover DLinear and PatchTST, six datasets, and $H \in \{96, 192, 336, 720\}$. Caches must match the selected checkpoint SHA256 in the latest inventory. For each test series, use origins $0, H, 2H, \ldots$, giving disjoint target intervals; overlapping contexts and serial dependence may remain. All training origins are used for local kernel estimation. Residuals are divided by channel RMS values estimated only on training data, before aggregating channels. Each of the 48 cells receives equal weight in aggregate statistics, rather than letting larger datasets dominate. The fixed protocol, per window arrays, cache provenance, script hashes, and cellwise CSV accompany the figure and table.

Controls and uncertainty. Compute spectra and temporal variation before any smoothing. The spectral analysis reports both DC inclusive energy and non DC energy to distinguish a constant offset from slowly varying error. Between window differences compare successive disjoint windows, not adjacent overlapping windows. Prefix permutations use 20 draws with seed 417. They retain the observed residual values but destroy their ordering. The last residual control copies the final observed residual across the suffix. Both controls see exactly the same prefix as the local solver.

Intervals use 1,000 circular moving block bootstrap replicates (seed 20260915), with blocks of four disjoint windows within each cell, followed by equal cell averaging. They describe sampling variation conditional on the selected checkpoints and fitted operators; they are not confidence intervals across TTA seeds. Small cells, including four Exchange windows at $H = 336$, limit uncertainty resolution. The 10–90% envelope in the figure instead summarizes heterogeneity across cells and is not a confidence interval.

C.1 Benchmark Configuration and Reporting

Table A2 collects the protocol details referenced by Section 4. Provenance markers distinguish published values from local evaluations; both follow the same backbone and evaluation protocol.

Table A2: Experimental protocol. All tuning uses validation data; test targets are revealed only according to the stated feedback schedule.

Item	Configuration
Task	Variable wise univariate forecasting; $L = 96$, $H \in \{96, 192, 336, 720\}$; train split standardization.
Splits	Chronological train/validation/test: 60/20/20 for ETT; 70/10/20 for Exchange and Weather.
Source forecasters	Official long term configurations; the same frozen checkpoint cellwise across methods and five TTA seeds; no online backbone updates or STEPS specific backbone retuning.
Baseline adapters	Published settings for TAFAS and PETSA. COSA: context $K = 10$, $S = 3$ steps, cosine adaptive learning rate and mean aggregation; COSA-F uses $B = 48$, COSA-P uses PAAS.
STEPS calibration	Train only residual fitting. Validation selects $q \in \{8, 16, 32, 64\}$ and $\lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$. The first 60% of validation origins select β from 41 equally spaced values in $[-0.5, 1.5]$.
Main feedback	$a = \lfloor H/2 \rfloor$; the observed prefix keeps its original forecast, and only the unseen suffix is revised. Only fully observed residual windows enter memory.
Reporting	Full horizon MSE averaged over five seeds in Table 1; additional MAE and suffix MSE in Table A14. Gains average cellwise ratios of means over five seeds, not pooled MSE ratios. Published COSA values follow the same evaluation protocol.
Noise study	DLinear, six datasets, $H = 720$, observed prefix 320, scored suffix 400. The corruption condition and common realization across methods are specified in Section 4.4.

D Dataset and Evaluation Grid Details

Datasets. Table A3 summarizes the six benchmarks. Channel counts and the test origin counts are read from the archived run metadata of our frozen caches; sequence frequencies are those of the public releases. Following the official COSA protocol, data are split chronologically into 60%/20%/20% train/validation/test segments for the four ETT datasets and 70%/10%/20% for Exchange and Weather, and every input uses lookback $L = 96$.

Table A3: Dataset summary. Channels and test origins are taken from the archived metadata of the frozen caches used in all local experiments; frequencies describe the public releases.

Dataset	Frequency	Channels	Test origins at H=192
ETTh1	hourly	7	3293
ETTh2	hourly	7	3293
ETTm1	15-minute	7	13745
ETTm2	15-minute	7	13745
Exchange	daily	8	1326
Weather	10-minute	21	10348

Test origin counts vary slightly with the horizon because shorter final segments are dropped; the values shown are for H=192 from the diagnostic cohort caches.

Evaluation grid. The main comparison (Table 1(a)–(c)) crosses six datasets, seven frozen backbones (DLinear, OLS, PatchTST, iTransformer, FreTS, TimeMixer, MICN), and four horizons $H \in \{96, 192, 336, 720\}$, giving 168 backbone–dataset–horizon cells per method. All accuracy entries are means over five seeds with fixed backbone weights. Entries marked † are taken from the published COSA comparison; published and local results follow the same evaluation protocol. The diagnostic cohort of Section G covers the DLinear and PatchTST subsets of this grid (48 cells), with full branch fitting on the 12 primary cells at $H = 192$.

Metrics. All main tables report full horizon MSE on the standardized target scale used by the frozen caches. Relative gains are means of cellwise ratios (Eq. (9)), never ratios of aggregate MSEs. Strict suffix MSE and MAE for every populated cell are reported in Table A14, and the prefix

sensitivity template is defined in Table A16; win/tie/loss counts in Section E use a relative tolerance of 10^{-3}.

E Extended Aggregate Results

This section decomposes the aggregate leaderboard of Figure 3 along the backbone and dataset axes and reports paired win/tie/loss counts. All quantities are recomputed directly from the 168 populated cells of Table 1(a)–(c); no additional runs were performed.

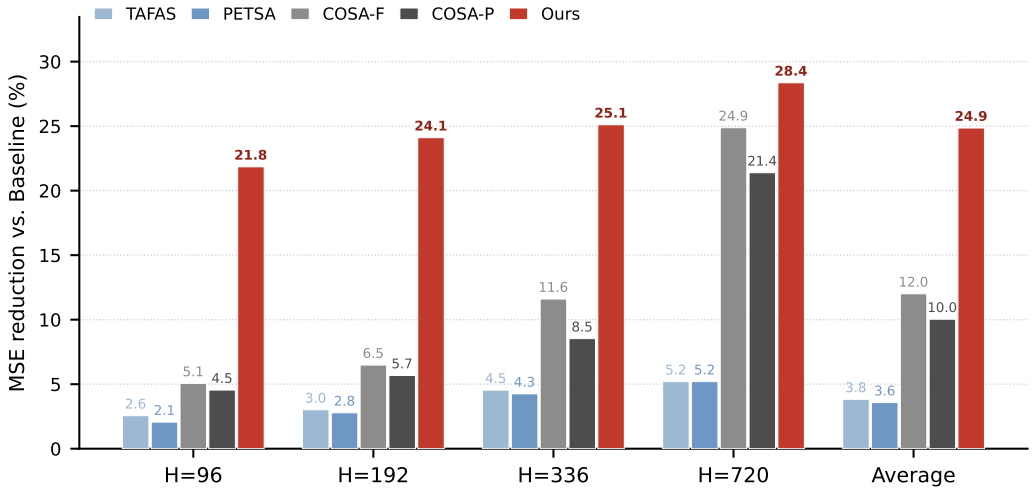

Figure 3: Mean cellwise relative MSE reduction over the reported Baseline at each horizon and across all 168 cells of Table 1. Published reference entries are marked in the main table and follow the same evaluation protocol.

Gains by backbone. Table A4 reports the mean relative MSE reduction of STEPS over the frozen Baseline for each backbone, averaging its 24 dataset–horizon cells. The gain is remarkably uniform across backbone families (24.0–25.5%), covering Transformer based, linear, and MLP based architectures; the minimum cellwise gain is positive on every backbone, and the maximum reaches 64.1% (MICN on Exchange). This uniformity supports the claim that STEPS exploits residual structure rather than backbone specific artifacts.

Table A4: Relative MSE reduction (%) of STEPS over the frozen Baseline, by backbone. Each row averages 24 dataset–horizon cells of Table 1(a)–(c); Min/Max are the extreme cellwise values.

Family	Backbone	Mean	Min	Max
Transformer	iTransformer	25.50	11.48	53.09
Transformer	PatchTST	23.99	11.60	55.24
Linear	DLinear	25.12	11.36	55.61
Linear	OLS	25.28	11.40	54.59
MLP	FreTS	24.55	10.34	55.22
MLP	TimeMixer	25.14	11.44	54.97
MLP	MICN	24.40	4.39	64.12
All 168 cells		24.85	4.39	64.12

Gains by dataset. Table A5 averages the same cellwise gains by dataset. Exchange benefits most (52.6%), consistent with its strong residual concentration in Table A9; ETTh1 benefits least (14.5%), consistent with its weak nonconstant spectral share. The dataset ordering of gains matches the diagnostic measurements, which is the qualitative link the diagnostics are meant to support.

Table A5: Relative MSE reduction (%) of STEPS over the frozen Baseline, by dataset. Each row averages 28 backbone–horizon cells.

Dataset	H=96	H=192	H=336	H=720	Mean
ETTh1	12.20	12.24	12.92	20.61	14.49
ETTh2	20.77	22.46	21.23	27.24	22.93
ETTm1	10.50	17.74	18.32	17.12	15.92
ETTm2	17.55	20.92	21.85	25.26	21.39
Exchange	48.50	52.41	54.39	55.07	52.59
Weather	21.56	18.86	21.93	24.85	21.80
Mean	21.85	24.10	25.11	28.36	24.85

Paired win/tie/loss counts. Table A6 counts strict cellwise comparisons of STEPS against each reference method over all 168 cells, with a relative tie tolerance of 10^{-3}. STEPS improves over its own frozen Baseline in all 168 cells and over TAFAS and PETSA in 167 of 168 (the single exception is MICN on ETTm1 at H=96). Against the stronger COSA variants, STEPS wins 146 of 168 cells against COSA-F and 160 of 168 against COSA-P. The 22 losses to COSA-F concentrate at H=720 (20 cells, mainly on ETTm1, ETTm2, Exchange, and Weather), where the published COSA-F values are strongest; we report these losses explicitly rather than averaging them away (see Table 1).

Table A6: Paired cellwise comparison of STEPS against each reference over all 168 backbone–dataset–horizon cells. A win requires a strictly lower MSE by a relative margin above 10^{-3}; no ties occurred at this tolerance.

Opponent	Win	Tie	Loss
Baseline (own frozen backbone)	168	0	0
TAFAS	167	0	1
PETSA	167	0	1
COSA-F	146	0	22
COSA-P	160	0	8

Counts are descriptive cellwise comparisons against the values shown in Table 1(a)–(c); Baseline and Ours share the same frozen checkpoint within each cell, while † marks published entries evaluated under the same protocol.

F Ablation and Hyperparameter Analysis

Branch ablation by backbone. Table A7 disaggregates the branch comparison of Table A10 by backbone. The qualitative pattern is identical on DLinear and PatchTST: ST-TEP carries the near suffix gain, MSBR alone is small on the near range but positive on the full suffix, and orthogonal fusion (STEPS) matches the local branch near the prefix while retaining the global contribution on the full suffix. Direct addition $L + \beta_{\text{dir}} G$ is slightly below orthogonal fusion on the near range for both backbones, consistent with double counting the shared direction.

Table A7: Branch ablation by backbone on the 12 primary diagnostic cells (H=192, six dataset cells per backbone). Entries are mean MSE reduction (%) against the frozen backbone on the near range (first 24 suffix steps) and the full suffix.

Backbone	Range	MSBR	ST-TEP	Direct	STEPS
DLinear	Near	4.6	52.0	50.3	51.9
DLinear	Suffix	4.7	38.4	38.7	38.8
PatchTST	Near	1.6	49.4	48.3	49.0
PatchTST	Suffix	2.2	35.7	36.0	35.9

MSBR = global branch only; ST-TEP = local branch only; Direct = nonorthogonal addition with its own validation selected coefficient; STEPS = orthogonal complementary fusion. Gains are means of cellwise ratios of mean squared errors.

Validation selected hyperparameters. Table A8 lists the hyperparameters selected on validation data for each primary cell. The MSBR output bandwidth q saturates at 64 in 10 of 12 cells, meaning

the validation split prefers the richest available bias reconstruction; the ridge penalty λ spans the full grid (10^{-2} to 10^{2}), with Weather selecting the weakest penalty and Exchange the strongest. The fusion coefficient β is positive everywhere (range 0.15–1.00), so the orthogonal global complement is always retained, at a weight that varies by cell. The orthogonal and direct coefficients differ (e.g., DLinear/ETTh2: $\beta = 0.15$ vs. $\beta_{\text{dir}} = 0.30$), showing that the projection changes which global amplitude the validation split prefers. All selections are frozen before any test evaluation.

Table A8: Validation selected hyperparameters on the 12 primary cells. q: MSBR output bandwidth from $\{8, 16, 32, 64\}$; λ: ridge penalty from $\{10^{-3}, \ldots, 10^{2}\}$; β: orthogonal fusion coefficient and β_{dir}: direct addition coefficient, each from 41 equally spaced values in $[-0.5, 1.5]$; n: disjoint test windows.

Backbone	Dataset	q	λ	β	β_{dir}	n
DLinear	ETTh1	64	10	0.55	0.60	18
DLinear	ETTh2	64	1	0.15	0.30	18
DLinear	ETTm1	64	1	0.45	0.45	72
DLinear	ETTm2	64	1	0.35	0.35	72
DLinear	Exchange	64	100	1.00	0.95	7
DLinear	Weather	8	0.01	0.60	0.50	54
PatchTST	ETTh1	8	10	0.60	0.70	18
PatchTST	ETTh2	64	1	0.15	0.10	18
PatchTST	ETTm1	64	1	0.45	0.50	72
PatchTST	ETTm2	64	1	0.30	0.35	72
PatchTST	Exchange	64	100	0.70	0.55	7
PatchTST	Weather	64	0.01	0.60	0.50	54

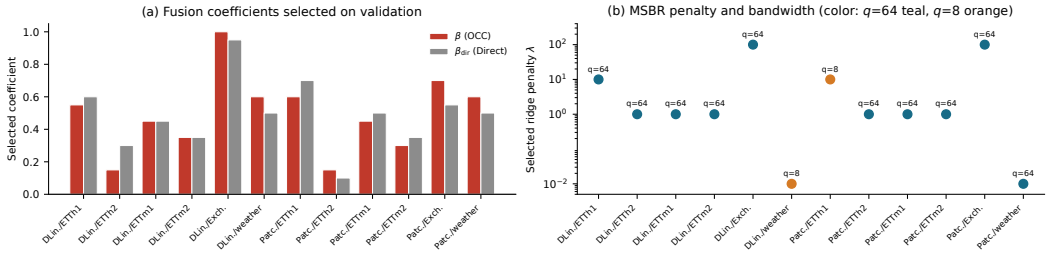

Figure 4: Validation selected hyperparameters on the 12 primary cells. (a) The OCC coefficient β and the direct addition coefficient β_{dir} differ cellwise; both are positive everywhere. (b) Selected ridge penalty λ (log scale) with the selected bandwidth q annotated; Exchange selects the strongest penalty and Weather the weakest.

Scope of the ablations. The branch ablation above is measured on the diagnostic cohort (two backbones, H=192, fixed prefix a=$H/2$), not on the full 168-cell accuracy grid. Fixed and validation adaptive prefix ratios are evaluated in Section H. Per setting hyperparameter selections are summarized in Table A8; Table A19 reports within series block bootstrap uncertainty and Table A20 reports TTA variability across five seeds.

G Empirical Validation of Residual Structure and Solver Branches

This appendix validates the two structural premises invoked in Sections 3.2 and 3.3 and compares the individual solver branches of STEPS on held out windows. The diagnostics below use the fixed reference prefix $a = H/2$ of the main protocol; the prefix extent sensitivity study is specified in Section G.1.

Diagnostic cohort and protocol. The cohort covers DLinear and PatchTST on all six datasets at $H \in \{96, 192, 336, 720\}$: 48 backbone–dataset–horizon cells and 1,852 test windows with disjoint target intervals. Point estimates average five TTA seeds with shared frozen backbone weights. All operators are fitted on training origins only: the ST-TEP kernel uses the 64-mode empirical second moment with noise multiplier $\nu = 10$, and the MSBR ridge operators use validation selected bandwidth and penalty (selected bandwidths are 8 or 64; selected penalties span 10^{-2} to 10^{2}; selected fusion weights satisfy $\beta \in [0.15, 1.0]$). The *primary* cohort ($H = 192$, 12 cells, 482 windows) fits both solver branches and scores six correction variants; the *replication* cohorts ($H \in \{96, 336, 720\}$, 36 cells) record residual only structure. Test targets beyond the prefix are used only for scoring. Residuals are normalized by training only channel RMS. Uncertainty uses 1,000 within cell circular block bootstrap replicates (blocks of four windows), conditional on these checkpoints. Per cell gains are ratios of mean squared errors within a cell; cross cell summaries average cellwise values with equal cell weight. Archived per window arrays, checkpoint provenance, the frozen protocol, and the analysis scripts accompany the paper under `analysis/residual_structure/`.

Premise 1 (MSBR): spectral concentration of residuals. Table A9 and Figure 5(a) measure DCT energy along the forecast horizon of unfiltered test residuals, before any solver or smoothing. The first four modes contain 67.9% of the energy on average across the 48 cells; after removing the constant component, the first three nonconstant modes still explain 45.7% of the remaining energy (mean of cellwise 95% bootstrap intervals: 40.7–49.9%). Concentration is weakest on ETTh1 (18.0% nonconstant share) and strongest on Exchange (76.5%). This supports a compact spectral state for historical bias; it does not by itself establish that the global branch predicts held out error, which is measured directly below.

Premise 2 (ST-TEP): short range temporal coherence. We compare mean squared lag-1 differences with the exact temporal permutation expectation $2H/(H-1)$, normalized so that the reference is 1. Raw residuals average 0.12 across the 48 cells; after removing the first four modes they average 0.23 (mean of cellwise 95% intervals: 0.20–0.28), far below the random order reference. Figure 5(b) shows that the detrended ratio rises with temporal separation and approaches the reference only at long lags: predictive structure is concentrated at short range, which is precisely the regime ST-TEP exploits.

Table A9: Residual structure diagnostics, 48 cells (DLinear and PatchTST, six datasets, $H \in \{96, 192, 336, 720\}$); dataset rows average eight cells. E_4 is the energy share of the first four DCT modes including the constant component; E_3^{AC} is the share of nonconstant modes 1–3 within all nonconstant energy. V_1 and V_1^{dt} are lag-1 squared differences of raw and mode truncated residuals relative to the temporal permutation expectation ($= 1$). n counts disjoint test windows.

Dataset	E_4	E_3^{AC}	V_1	V_1^{dt}	n
ETTh1	41.2	18.0	0.18	0.22	136
ETTh2	65.5	43.6	0.17	0.30	136
ETTm1	53.7	35.8	0.10	0.19	554
ETTm2	71.0	43.8	0.12	0.24	554
Exchange	94.2	76.5	0.05	0.21	56
Weather	82.0	56.2	0.10	0.25	416
Mean	67.9	45.7	0.12	0.23	1852

Branch comparison under prefix observation. Table A10 scores six suffix corrections against the frozen forecast on the 12 primary cells: no correction (reference), last residual continuation, shuffled prefix propagation (20 fixed seed permutations destroying temporal order), MSBR alone, ST-TEP alone, direct addition $L + \beta_{\mathrm{dir}} G$, and the full STEPS response $L + \beta g^{\perp}$. Three observations emerge. First, ST-TEP reduces near suffix MSE in all 12 cells (mean 50.7%), and the ordered prefix matters: destroying its temporal order while preserving its values drops the gain to 27.9%. Second, MSBR alone is weak near the prefix (mean 3.1%, positive in 7 of 12 cells, harmful on ETTh2 and Exchange) but contributes on the full suffix (positive in 8 of 12 cells), consistent with its role as a low frequency bias estimate rather than a near term response. Third, orthogonal fusion preserves the local response while adding the global complement: STEPS matches ST-TEP on the near suffix

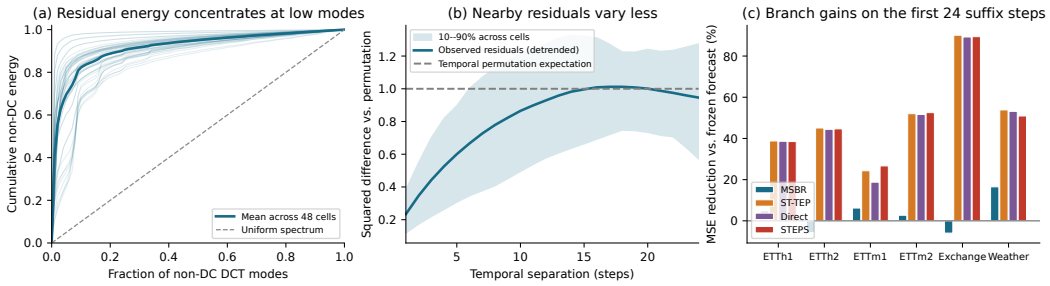

Figure 5: Residual diagnostics on the 48-cell cohort. (a) Cumulative nonconstant DCT energy along the forecast horizon; faint lines are individual cells. (b) Lag resolved squared differences of mode truncated residuals relative to the temporal permutation expectation; shading is the 10–90% range across cells, not a confidence interval. (c) Near suffix MSE reduction of each solver branch on the 12 primary cells ($H = 192$), averaged per dataset; MSBR alone is weak or harmful on the near suffix, while the fused STEPS response retains the ST-TEP gains.

Table A10: Branch comparison on the 12 primary cells (H=192, fixed prefix a=$H/2$). Entries are MSE reduction (%) against the frozen backbone, averaged per dataset over two backbones. Near covers the first 24 suffix steps; Suffix covers the full unobserved segment. Shuffled prefix propagation is defined for the near range only. Negative values indicate harm; unavailable cells are shown as -.

	Near (first 24 steps)						Suffix (full)				
Dataset	Last	Shuf.	MSBR	ST-TEP	Direct	STEPS	Last	MSBR	ST-TEP	Direct	STEPS
ETTh1	-43.0	17.5	4.7	38.8	38.6	38.5	-73.6	3.3	22.2	21.8	22.0
ETTh2	17.0	16.5	-5.7	45.1	44.4	44.6	-3.4	-4.2	24.3	24.3	24.2
ETTm1	-61.3	-0.5	6.2	24.3	18.8	26.7	-17.4	11.6	30.1	32.3	32.4
ETTm2	40.6	39.8	2.7	52.0	51.6	52.5	14.5	5.1	40.8	41.0	41.8
Exchange	90.3	64.4	-6.0	90.1	89.3	89.4	73.2	-6.3	73.3	71.7	72.0
Weather	53.8	29.6	16.5	53.8	53.2	50.9	12.8	11.1	31.8	33.1	31.9
Mean	16.2	27.9	3.1	50.7	49.3	50.4	1.0	3.4	37.1	37.4	37.4

(50.4% vs. 50.7%) and exceeds both branches on the full suffix (37.4% vs. 37.1% ST-TEP and 3.4% MSBR). Direct addition is close to orthogonal fusion on this cohort (49.3% near, 37.4% suffix); the projection guards against double counting a shared direction without sacrificing accuracy here. These are diagnostic cohort results at $H = 192$ with the fixed reference prefix, not the full grid main comparison of Table 1(a)–(c).

Response to recently observed changes. A recent change subset selects windows whose last two 8-step prefix means differ by at least the validation 80th percentile (in training RMS units); the threshold is frozen before testing and uses prefix information only. Across 131 qualifying windows in 10 of the 12 primary cells (Exchange has none at H=192), ST-TEP reduces near suffix MSE by 41.1% and STEPS by 40.4%, while MSBR alone reaches 4.5% (Table A11). On Exchange, the identical frozen rule at H=96 selects 7 of 30 disjoint windows; there ST-TEP reduces near suffix MSE by 91.3% and STEPS by 89.7%, while MSBR alone is slightly harmful (-2.5%), consistent with the other datasets. Once a change has become visible in the prefix, propagating it is far more effective than re estimating bias from history; no claim is made about predicting changes that have not yet been observed.

G.1 Prefix Extent Sensitivity

The formulation of Eq. (1) admits any prefix length: changing a repartitions the observed and unobserved kernel blocks without refitting K. Table A16 reports the two backbone fixed ratio sweep on the primary cohort. Section H adds a controlled DLinear study containing all requested ratios

Table A11: Near suffix MSE reduction (%) on the large recent change subset of the 12 primary cells. n counts qualifying windows. The Mean row averages cells with at least one qualifying window at H=192. ‡Exchange has no qualifying window at H=192 (7 disjoint windows); its row reports the H=96 diagnostic cell under the identical frozen qualification rule (7 qualifying of 30 disjoint windows across the two backbones, main grid branch fits).

Dataset	n	MSBR	ST-TEP	STEPS
ETTh1	19	5.2	45.9	45.4
ETTh2	21	-7.9	43.1	42.6
ETTm1	52	8.5	13.2	14.7
ETTm2	21	7.1	59.0	59.4
Exchange‡	7	-2.5	91.3	89.7
Weather	18	9.6	44.2	39.8
Mean	131	4.5	41.1	40.4

and a validation adaptive selector, including its realized choices. Both studies score a common final quarter to separate additional feedback from the shorter suffix produced by a larger prefix.

H DLinear Prefix Ratio and Adaptive Selection Study

Protocol. We isolate DLinear at H=192 on ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and Weather, and evaluate $a/H \in \{1/8, 1/4, 1/3, 1/2, 2/3, 3/4\}$. Every ratio reuses the same frozen backbone predictions, test origins, training fitted empirical kernel, and historical state. For each ratio, the fusion coefficient is selected on the first 60% of validation origins. The adaptive variant then selects one ratio per dataset by the same validation full horizon MSE, before test targets are accessed. Thus, "adaptive" denotes validation based dataset level selection, rather than test time oracle or per window selection.

Table A12: DLinear prefix ratio sensitivity at H=192, averaged over six datasets. Entries are MSE reductions (%) from the frozen DLinear forecast. Full-H uses the original horizon; common tail evaluates the same final 48 steps for every ratio; suffix uses each ratio's unobserved segment and is therefore not directly comparable across rows. Adaptive uses the per dataset ratios in Table A13.

Prefix rule	Mean a/H	Full-H	Common tail	Suffix
Fixed $1/8$	0.125	14.48	9.84	15.38
Fixed $1/4$	0.250	20.22	16.28	23.67
Fixed $1/3$	0.333	22.65	21.92	28.83
Fixed $1/2$	0.500	**24.49**	32.45	38.93
Fixed $2/3$	0.667	20.93	42.17	47.17
Fixed $3/4$	0.750	17.61	**51.80**	**51.80**
Adaptive	0.410	23.05	26.53	33.24

Table A13: Actual validation selected prefix ratios. Full-H gain is the per dataset test MSE reduction (%). Δ is adaptive minus fixed-$1/2$ gain; the four datasets selecting $1/2$ have identical predictions in the two columns.

Dataset	Selected a/H	Adaptive	Fixed $1/2$	Δ
ETTh1	$1/2$	11.99	11.99	0.00
ETTh2	$1/2$	22.24	22.24	0.00
ETTm1	$1/2$	19.70	19.70	0.00
ETTm2	$1/2$	19.90	19.90	0.00
Exchange	$1/3$	45.56	51.15	-5.59
Weather	$1/8$	18.90	21.93	-3.04
Mean	0.410	23.05	24.49	-1.44

Why $a/H = 1/2$ is the default. Full horizon gain peaks at $1/2$. Waiting longer improves the common tail, as expected from the extra feedback, but reduces the revisable fraction and lowers full horizon gain. The adaptive selector chooses $1/2$ on four of six datasets (median selected ratio $= 1/2$); its mean selected ratio is 0.410 and its test gain is 23.05%, 1.44 points below the fixed half prefix result. Consequently, this cohort provides no accuracy benefit for adding a dataset specific ratio selection stage. We retain $a/H = 1/2$ as the simple latency–accuracy compromise, while treating the higher common tail gains of larger ratios as evidence that deployments with different delay costs may prefer a different operating point.

I Limitations and Negative Results

Where STEPS does not win. STEPS loses 22 of 168 cells to COSA-F and 8 to COSA-P (Table A6), concentrated at H=720 on ETTm1, ETTm2, Exchange, and Weather. It also loses one cell each to TAFAS and PETSA (MICN on ETTm1 at H=96). The long horizon losses occur where COSA-F is strongest; we report them explicitly and do not claim uniform superiority.

The global branch is weak near the prefix. MSBR alone is weak or harmful on the near suffix (mean 3.1%, negative on ETTh2 and Exchange; Table A10). STEPS inherits its near range strength from ST-TEP, and the orthogonal projection exists precisely because naive addition of the two branches double counts the shared direction. Improvements should therefore be attributed to the combination, not to the global branch alone.

Fixed prefix evaluation. The main grid and branch diagnostics use the fixed reference prefix a=$H/2$. Section H tests six fixed ratios and validation adaptive dataset level selection on DLinear at H=192; it does not establish that the same choice is optimal for every backbone, horizon, or latency budget. Per window selection is not evaluated.

Statistical scope. The main grid results average five TTA seeds with fixed backbone weights. Table A19 reports within series block bootstrap uncertainty, while Table A20 reports variability across the five TTA seeds. The diagnostic cohort covers two of the seven backbones (DLinear, PatchTST) and weights cells equally, so small cells (e.g., Exchange at H=336, four windows) limit resolution. Bootstrap intervals do not measure across seed variability.

Method boundaries. The sensitivity bound of Section B is finite but not small, and the implementation applies no output clipping; orthogonalization prevents double counting but does not guarantee that the retained global component is closer to the true error. Spectral concentration (Table A9) is weakest exactly where gains are smallest (ETTh1), so the method's benefit depends on structure that varies by dataset.

J Supplementary Experimental Protocol and Results

Table A14 reports paired Backbone and STEPS results for all 168 backbone–dataset–horizon settings across seven backbones. Table A15 reports contextual baseline comparisons for the 144 settings covering the other six backbones. Table A19 reports within series bootstrap uncertainty for the shared frozen backbone, whereas Table A20 reports across seed variability as mean $\pm$ standard deviation over five independent runs. Table A20 varies TTA seeds while retaining the same frozen backbone weights, data split, hyperparameter selection procedure, and evaluation protocol. Repeated headings belong to a single continued table, not additional appendix tables.

Table A14: Complete Cellwise Results. Full horizon MSE and MAE and strict suffix MSE, each averaged over five seeds, for Backbone (B) and STEPS (S) across all 168 settings; observed prefix length a is recorded per setting.

Backbone / Dataset	H	a	B MSE	S MSE	B MAE	S MAE	B suffix	S suffix
DLinear / ETTh1	96	48	.4695	.4054	.4662	.4275	.5202	.3919
DLinear / ETTh1	192	96	.5213	.4588	.5007	.4648	.5676	.4425
DLinear / ETTh1	336	168	.5659	.4991	.5294	.4934	.6107	.4771
DLinear / ETTh1	720	360	.7117	.5812	.6160	.5465	.8257	.5646

Table A14 continued

Backbone / Dataset	H	a	B MSE	S MSE	B MAE	S MAE	B suffix	S suffix
DLinear / ETTh2	96	48	.2323	.1845	.3194	.2851	.2772	.1814
DLinear / ETTh2	192	96	.2862	.2225	.3548	.3147	.3375	.2102
DLinear / ETTh2	336	168	.3252	.2554	.3799	.3395	.3670	.2273
DLinear / ETTh2	720	360	.4087	.2978	.4303	.3730	.4684	.2465
DLinear / ETTm1	96	48	.3715	.3293	.3937	.3759	.4037	.3193
DLinear / ETTm1	192	96	.4438	.3564	.4291	.3863	.5161	.3414
DLinear / ETTm1	336	168	.5183	.4137	.4680	.4178	.6025	.3934
DLinear / ETTm1	720	360	.5929	.4848	.5164	.4631	.6536	.4376
DLinear / ETTm2	96	48	.1598	.1322	.2636	.2409	.1834	.1283
DLinear / ETTm2	192	96	.1930	.1546	.2908	.2586	.2260	.1492
DLinear / ETTm2	336	168	.2324	.1835	.3198	.2826	.2776	.1797
DLinear / ETTm2	720	360	.3062	.2285	.3652	.3181	.3716	.2163
DLinear / Exchange	96	48	.0913	.0475	.2128	.1536	.1330	.0453
DLinear / Exchange	192	96	.1827	.0892	.3048	.2122	.2715	.0846
DLinear / Exchange	336	168	.3277	.1559	.4142	.2819	.4938	.1503
DLinear / Exchange	720	360	.8873	.3939	.7150	.4612	1.3501	.3633
DLinear / Weather	96	48	.1954	.1491	.2350	.2021	.2335	.1408
DLinear / Weather	192	96	.2403	.1876	.2707	.2396	.2838	.1784
DLinear / Weather	336	168	.2918	.2171	.3066	.2614	.3541	.2046
DLinear / Weather	720	360	.3643	.2765	.3533	.3075	.4379	.2621
OLS / ETTh1	96	48	.4511	.3902	.4462	.4116	.5057	.3839
OLS / ETTh1	192	96	.5045	.4432	.4830	.4486	.5532	.4304
OLS / ETTh1	336	168	.5510	.4843	.5141	.4779	.5985	.4652
OLS / ETTh1	720	360	.6997	.5688	.6048	.5348	.8170	.5552
OLS / ETTh2	96	48	.2306	.1836	.3167	.2830	.2744	.1804
OLS / ETTh2	192	96	.2838	.2207	.3521	.3126	.3347	.2084
OLS / ETTh2	336	168	.3257	.2552	.3795	.3390	.3703	.2292
OLS / ETTh2	720	360	.4159	.2960	.4338	.3706	.4824	.2426
OLS / ETTm1	96	48	.3709	.3287	.3930	.3752	.4034	.3190
OLS / ETTm1	192	96	.4435	.3562	.4285	.3857	.5160	.3412
OLS / ETTm1	336	168	.5178	.4133	.4671	.4171	.6022	.3932
OLS / ETTm1	720	360	.5922	.4844	.5151	.4621	.6532	.4375
OLS / ETTm2	96	48	.1603	.1325	.2637	.2409	.1837	.1282
OLS / ETTm2	192	96	.1936	.1547	.2908	.2585	.2268	.1492
OLS / ETTm2	336	168	.2331	.1837	.3198	.2822	.2781	.1793
OLS / ETTm2	720	360	.3066	.2279	.3651	.3173	.3717	.2144
OLS / Exchange	96	48	.0814	.0431	.1968	.1432	.1217	.0450
OLS / Exchange	192	96	.1726	.0835	.2930	.2028	.2620	.0837
OLS / Exchange	336	168	.3225	.1483	.4087	.2732	.4940	.1456
OLS / Exchange	720	360	.8363	.3799	.6871	.4481	1.3063	.3935
OLS / Weather	96	48	.1951	.1489	.2342	.2013	.2334	.1409
OLS / Weather	192	96	.2401	.1874	.2701	.2392	.2837	.1784
OLS / Weather	336	168	.2917	.2170	.3062	.2611	.3540	.2046
OLS / Weather	720	360	.3642	.2763	.3529	.3071	.4378	.2620
PatchTST / ETTh1	96	48	.4331	.3812	.4531	.4178	.4782	.3745
PatchTST / ETTh1	192	96	.4903	.4358	.4912	.4561	.5443	.4353
PatchTST / ETTh1	336	168	.5538	.4848	.5285	.4895	.6163	.4783
PatchTST / ETTh1	720	360	.7142	.5810	.6246	.5535	.8418	.5753
PatchTST / ETTh2	96	48	.2388	.1933	.3219	.2886	.2772	.1861
PatchTST / ETTh2	192	96	.2843	.2236	.3599	.3185	.3342	.2128
PatchTST / ETTh2	336	168	.3185	.2501	.3823	.3386	.3596	.2229
PatchTST / ETTh2	720	360	.4275	.3005	.4484	.3772	.5004	.2465
PatchTST / ETTm1	96	48	.4047	.3498	.4125	.3883	.4403	.3306
PatchTST / ETTm1	192	96	.4537	.3670	.4431	.3966	.5136	.3402
PatchTST / ETTm1	336	168	.5094	.4130	.4772	.4256	.5740	.3811
PatchTST / ETTm1	720	360	.5608	.4712	.5187	.4660	.6188	.4396
PatchTST / ETTm2	96	48	.1575	.1299	.2605	.2366	.1844	.1293
PatchTST / ETTm2	192	96	.2035	.1583	.2969	.2612	.2479	.1575
PatchTST / ETTm2	336	168	.2536	.1897	.3302	.2864	.3121	.1843
PatchTST / ETTm2	720	360	.3353	.2434	.3828	.3275	.4021	.2183
PatchTST / Exchange	96	48	.0839	.0429	.2018	.1432	.1255	.0433
PatchTST / Exchange	192	96	.1767	.0858	.3002	.2070	.2671	.0852
PatchTST / Exchange	336	168	.3343	.1517	.4199	.2784	.5116	.1463
PatchTST / Exchange	720	360	.9628	.4372	.7313	.4836	1.5012	.4501
PatchTST / Weather	96	48	.1976	.1523	.2368	.2063	.2352	.1447
PatchTST / Weather	192	96	.2431	.1907	.2739	.2433	.2870	.1822
PatchTST / Weather	336	168	.2968	.2212	.3112	.2669	.3596	.2084
PatchTST / Weather	720	360	.3760	.2862	.3649	.3191	.4483	.2688
iTransformer / ETTh1	96	48	.4485	.3928	.4554	.4217	.4990	.3875
iTransformer / ETTh1	192	96	.5086	.4487	.4979	.4629	.5572	.4375
iTransformer / ETTh1	336	168	.5640	.4955	.5355	.4969	.6126	.4755
iTransformer / ETTh1	720	360	.7102	.5836	.6241	.5574	.8188	.5655
iTransformer / ETTh2	96	48	.2518	.1978	.3399	.2995	.2933	.1854
iTransformer / ETTh2	192	96	.3071	.2368	.3810	.3327	.3610	.2203
iTransformer / ETTh2	336	168	.3590	.2730	.4166	.3606	.4072	.2351
iTransformer / ETTh2	720	360	.4411	.3119	.4621	.3885	.5011	.2427
iTransformer / ETTm1	96	48	.3843	.3384	.4098	.3861	.4369	.3452
iTransformer / ETTm1	192	96	.4432	.3615	.4421	.3957	.5103	.3468

Table A14 continued

Backbone / Dataset	H	a	B MSE	S MSE	B MAE	S MAE	B suffix	S suffix
iTransformer / ETTm1	336	168	.5138	.4187	.4846	.4328	.5889	.3986
iTransformer / ETTm1	720	360	.5893	.4873	.5309	.4761	.6515	.4474
iTransformer / ETTm2	96	48	.1652	.1342	.2665	.2425	.1941	.1320
iTransformer / ETTm2	192	96	.2154	.1663	.3100	.2710	.2581	.1599
iTransformer / ETTm2	336	168	.2841	.2041	.3532	.3005	.3557	.1957
iTransformer / ETTm2	720	360	.3447	.2466	.4036	.3378	.4170	.2207
iTransformer / Exchange	96	48	.0880	.0471	.2099	.1533	.1291	.0473
iTransformer / Exchange	192	96	.1786	.0890	.3036	.2128	.2677	.0885
iTransformer / Exchange	336	168	.3465	.1608	.4302	.2884	.5280	.1565
iTransformer / Exchange	720	360	.8858	.4116	.7043	.4694	1.3624	.4139
iTransformer / Weather	96	48	.1727	.1355	.2120	.1842	.2085	.1339
iTransformer / Weather	192	96	.2236	.1746	.2573	.2258	.2707	.1728
iTransformer / Weather	336	168	.2801	.2098	.2986	.2566	.3458	.2050
iTransformer / Weather	720	360	.3574	.2734	.3500	.3062	.4313	.2632
FreTS / ETTh1	96	48	.4551	.3976	.4553	.4216	.5096	.3945
FreTS / ETTh1	192	96	.5021	.4445	.4875	.4541	.5522	.4371
FreTS / ETTh1	336	168	.5544	.4888	.5222	.4860	.6036	.4724
FreTS / ETTh1	720	360	.7145	.5857	.6189	.5502	.8266	.5690
FreTS / ETTh2	96	48	.2385	.1913	.3263	.2919	.2812	.1868
FreTS / ETTh2	192	96	.2867	.2271	.3595	.3207	.3351	.2159
FreTS / ETTh2	336	168	.3317	.2608	.3883	.3458	.3770	.2352
FreTS / ETTh2	720	360	.4198	.3004	.4408	.3762	.4875	.2487
FreTS / ETTm1	96	48	.3676	.3295	.3931	.3762	.4003	.3241
FreTS / ETTm1	192	96	.4324	.3542	.4276	.3871	.4997	.3431
FreTS / ETTm1	336	168	.5006	.4091	.4647	.4188	.5788	.3959
FreTS / ETTm1	720	360	.5590	.4707	.5094	.4603	.6165	.4398
FreTS / ETTm2	96	48	.1581	.1313	.2604	.2383	.1829	.1292
FreTS / ETTm2	192	96	.1923	.1553	.2906	.2594	.2268	.1528
FreTS / ETTm2	336	168	.2320	.1842	.3204	.2835	.2765	.1809
FreTS / ETTm2	720	360	.3010	.2275	.3652	.3183	.3619	.2149
FreTS / Exchange	96	48	.0828	.0434	.1996	.1447	.1237	.0450
FreTS / Exchange	192	96	.1733	.0838	.2948	.2040	.2628	.0836
FreTS / Exchange	336	168	.3241	.1489	.4109	.2749	.4960	.1457
FreTS / Exchange	720	360	.8267	.3747	.6835	.4458	1.2882	.3842
FreTS / Weather	96	48	.1858	.1436	.2249	.1960	.2232	.1388
FreTS / Weather	192	96	.2309	.1815	.2622	.2336	.2755	.1768
FreTS / Weather	336	168	.2845	.2139	.3003	.2605	.3484	.2073
FreTS / Weather	720	360	.3555	.2708	.3475	.3027	.4308	.2616
TimeMixer / ETTh1	96	48	.4479	.3928	.4494	.4171	.4977	.3875
TimeMixer / ETTh1	192	96	.5073	.4487	.4905	.4575	.5568	.4396
TimeMixer / ETTh1	336	168	.5580	.4921	.5243	.4888	.6028	.4711
TimeMixer / ETTh1	720	360	.7091	.5783	.6186	.5481	.8213	.5596
TimeMixer / ETTh2	96	48	.2401	.1871	.3268	.2882	.2857	.1798
TimeMixer / ETTh2	192	96	.2855	.2253	.3585	.3194	.3332	.2130
TimeMixer / ETTh2	336	168	.3262	.2586	.3860	.3453	.3706	.2355
TimeMixer / ETTh2	720	360	.4087	.3000	.4349	.3772	.4696	.2521
TimeMixer / ETTm1	96	48	.3898	.3452	.4047	.3844	.4249	.3359
TimeMixer / ETTm1	192	96	.4413	.3656	.4388	.3967	.4904	.3392
TimeMixer / ETTm1	336	168	.5175	.4162	.4745	.4243	.5967	.3941
TimeMixer / ETTm1	720	360	.5758	.4777	.5201	.4660	.6333	.4371
TimeMixer / ETTm2	96	48	.1567	.1289	.2592	.2356	.1814	.1259
TimeMixer / ETTm2	192	96	.1880	.1525	.2874	.2571	.2217	.1509
TimeMixer / ETTm2	336	168	.2283	.1811	.3185	.2812	.2739	.1796
TimeMixer / ETTm2	720	360	.3029	.2270	.3669	.3179	.3640	.2122
TimeMixer / Exchange	96	48	.0864	.0436	.2042	.1443	.1293	.0438
TimeMixer / Exchange	192	96	.1853	.0853	.3035	.2052	.2820	.0819
TimeMixer / Exchange	336	168	.3371	.1518	.4190	.2776	.5171	.1466
TimeMixer / Exchange	720	360	.8838	.4001	.7032	.4585	1.3853	.4179
TimeMixer / Weather	96	48	.1649	.1333	.2114	.1826	.1948	.1317
TimeMixer / Weather	192	96	.2109	.1686	.2539	.2257	.2555	.1709
TimeMixer / Weather	336	168	.2644	.2015	.2931	.2533	.3304	.2046
TimeMixer / Weather	720	360	.6667	.4328	.5174	.3962	.7408	.2730
MICN / ETTh1	96	48	.5103	.4561	.5171	.4771	.5634	.4540
MICN / ETTh1	192	96	.5954	.5072	.5839	.5229	.6616	.4691
MICN / ETTh1	336	168	.6615	.5445	.6217	.5329	.7147	.4864
MICN / ETTh1	720	360	.9233	.6014	.7940	.5629	1.0763	.4304
MICN / ETTh2	96	48	.2582	.2035	.3462	.3125	.3108	.1853
MICN / ETTh2	192	96	.3282	.2465	.4162	.3528	.3822	.2151
MICN / ETTh2	336	168	.3732	.3038	.4475	.4014	.4150	.2762
MICN / ETTh2	720	360	.4617	.3489	.4922	.4391	.5388	.3117
MICN / ETTm1	96	48	.4354	.4163	.4588	.4806	.4750	.4402
MICN / ETTm1	192	96	.4855	.4246	.4832	.4638	.5636	.4361
MICN / ETTm1	336	168	.5556	.4809	.5162	.4819	.6343	.4861
MICN / ETTm1	720	360	.6212	.5407	.5487	.5200	.6771	.5197
MICN / ETTm2	96	48	.1710	.1418	.2838	.2552	.1956	.1340
MICN / ETTm2	192	96	.2121	.1687	.3116	.2893	.2546	.1670
MICN / ETTm2	336	168	.2530	.1987	.3337	.3076	.2995	.1952
MICN / ETTm2	720	360	.3327	.2589	.4046	.3609	.4025	.2575

Table A14 continued

Backbone / Dataset	H	a	B MSE	S MSE	B MAE	S MAE	B suffix	S suffix
MICN / Exchange	96	48	.1151	.0573	.2815	.1945	.1654	.0458
MICN / Exchange	192	96	.2150	.1005	.3583	.2375	.3186	.1043
MICN / Exchange	336	168	.3950	.1702	.5039	.3070	.6059	.1511
MICN / Exchange	720	360	1.0259	.3681	.8298	.4245	1.5866	.2706
MICN / Weather	96	48	.1757	.1297	.2156	.1716	.2117	.1260
MICN / Weather	192	96	.2237	.1976	.2541	.2496	.2769	.2149
MICN / Weather	336	168	.2812	.2572	.2908	.3043	.3366	.2989
MICN / Weather	720	360	.3508	.2696	.3357	.2901	.4079	.2518

Table A15: Full Baseline Comparison. Full horizon MSE averaged over five seeds for six backbones (144 backbone–dataset–horizon cells); MICN appears in Table 1(c) and the paired results Table A14. † marks values transcribed from the published COSA comparison table; published and local values use the same frozen backbone and evaluation protocol.

Backbone / Dataset	H	Backbone	TAFAS	PETSA	COSA-F	COSA-P	STEPS
DLinear / ETTh1	96	.4695†	.4618†	.4594†	.4574†	.4482†	.4054
DLinear / ETTh1	192	.5213†	.5117†	.5118†	.5066†	.5050†	.4588
DLinear / ETTh1	336	.5659†	.5604†	.5617†	.5528†	.5456†	.4991
DLinear / ETTh1	720	.7117†	.6820†	.6743†	.6107†	.5896†	.5812
DLinear / ETTh2	96	.2323†	.2303†	.2306†	.2300†	.2281†	.1845
DLinear / ETTh2	192	.2862†	.2842†	.2876†	.2827†	.2819†	.2225
DLinear / ETTh2	336	.3252†	.3185†	.3184†	.3050†	.3083†	.2554
DLinear / ETTh2	720	.4087†	.3873†	.3853†	.3062†	.3477†	.2978
DLinear / ETTm1	96	.3715†	.3497†	.3524†	.3456†	.3475†	.3293
DLinear / ETTm1	192	.4438†	.4166†	.4178†	.4113†	.4122†	.3564
DLinear / ETTm1	336	.5183†	.4799†	.4803†	.4753†	.4858†	.4137
DLinear / ETTm1	720	.5929†	.5488†	.5532†	.4774†	.4991†	.4848
DLinear / ETTm2	96	.1598†	.1584†	.1584†	.1583†	.1586†	.1322
DLinear / ETTm2	192	.1930†	.1913†	.1913†	.1904†	.1905†	.1546
DLinear / ETTm2	336	.2324†	.2289†	.2292†	.2083†	.2242†	.1835
DLinear / ETTm2	720	.3062†	.2968†	.2963†	.2215†	.2316†	.2285
DLinear / Exchange	96	.0913†	.0885†	.0878†	.0812†	.0834†	.0475
DLinear / Exchange	192	.1827†	.1760†	.1730†	.1459†	.1519†	.0892
DLinear / Exchange	336	.3277†	.2941†	.2920†	.2039†	.2480†	.1559
DLinear / Exchange	720	.8873†	.8762†	.8781†	.3494†	.4481†	.3939
DLinear / Weather	96	.1954†	.1796†	.1823†	.1773†	.1793†	.1491
DLinear / Weather	192	.2403†	.2244†	.2254†	.2216†	.2217†	.1876
DLinear / Weather	336	.2918†	.2709†	.2740†	.2567†	.2626†	.2171
DLinear / Weather	720	.3643†	.3500†	.3497†	.2581†	.2708†	.2765
OLS / ETTh1	96	.4511†	.4409†	.4391†	.4390†	.4372†	.3902
OLS / ETTh1	192	.5046†	.4934†	.4937†	.4915†	.4906†	.4432
OLS / ETTh1	336	.5510†	.5440†	.5465†	.5385†	.5320†	.4843
OLS / ETTh1	720	.6997†	.6630†	.6431†	.5969†	.5733†	.5688
OLS / ETTh2	96	.2306†	.2285†	.2288†	.2232†	.2265†	.1836
OLS / ETTh2	192	.2839†	.2824†	.2848†	.2796†	.2791†	.2207
OLS / ETTh2	336	.3258†	.3182†	.3189†	.3003†	.3043†	.2552
OLS / ETTh2	720	.4162†	.3908†	.3884†	.3177†	.3453†	.2960
OLS / ETTm1	96	.3710†	.3506†	.3536†	.3454†	.3475†	.3287
OLS / ETTm1	192	.4439†	.4160†	.4184†	.4115†	.4119†	.3562
OLS / ETTm1	336	.5182†	.4787†	.4792†	.4748†	.4749†	.4133
OLS / ETTm1	720	.5922†	.5478†	.5522†	.4763†	.5007†	.4844
OLS / ETTm2	96	.1602†	.1590†	.1589†	.1582†	.1586†	.1325
OLS / ETTm2	192	.1936†	.1921†	.1919†	.1906†	.1907†	.1547
OLS / ETTm2	336	.2331†	.2299†	.2302†	.2131†	.2226†	.1837
OLS / ETTm2	720	.3066†	.2986†	.2971†	.2171†	.2349†	.2279
OLS / Exchange	96	.0814†	.0792†	.0798†	.0756†	.0773†	.0431
OLS / Exchange	192	.1727†	.1658†	.1653†	.1393†	.1457†	.0835
OLS / Exchange	336	.3226†	.2877†	.2898†	.2020†	.2323†	.1483
OLS / Exchange	720	.8366†	.8138†	.8149†	.3444†	.4541†	.3799
OLS / Weather	96	.1957†	.1807†	.1795†	.1772†	.1803†	.1489
OLS / Weather	192	.2404†	.2244†	.2274†	.2223†	.2237†	.1874
OLS / Weather	336	.2921†	.2714†	.2748†	.2551†	.2642†	.2170
OLS / Weather	720	.3644†	.3466†	.3493†	.2579†	.2708†	.2763
PatchTST / ETTh1	96	.4312†	.4262†	.4269†	.4242†	.4238†	.3812
PatchTST / ETTh1	192	.4955†	.4865†	.4854†	.4830†	.4805†	.4358
PatchTST / ETTh1	336	.5559†	.5478†	.5475†	.5438†	.5320†	.4848
PatchTST / ETTh1	720	.7117†	.6860†	.6822†	.6113†	.5822†	.5810
PatchTST / ETTh2	96	.2362†	.2351†	.2362†	.2349†	.2343†	.1933

Table A15 continued

Backbone / Dataset	H	Backbone	TAFAS	PETSA	COSA-F	COSA-P	STEPS
PatchTST / ETTh2	192	.2826†	.2758†	.2773†	.2665†	.2608†	.2236
PatchTST / ETTh2	336	.3199†	.3125†	.3132†	.2971†	.2978†	.2501
PatchTST / ETTh2	720	.4264†	.4005†	.4012†	.3233†	.3428†	.3005
PatchTST / ETTm1	96	.4024†	.3894†	.3937†	.3625†	.3626†	.3498
PatchTST / ETTm1	192	.4512†	.4372†	.4413†	.4250†	.4258†	.3670
PatchTST / ETTm1	336	.5081†	.4905†	.4946†	.4568†	.4697†	.4130
PatchTST / ETTm1	720	.5629†	.5427†	.5462†	.4681†	.4882†	.4712
PatchTST / ETTm2	96	.1584†	.1581†	.1583†	.1558†	.1562†	.1299
PatchTST / ETTm2	192	.2059†	.2036†	.2037†	.2007†	.2022†	.1583
PatchTST / ETTm2	336	.2458†	.2451†	.2452†	.2258†	.2352†	.1897
PatchTST / ETTm2	720	.3268†	.3268†	.3256†	.2446†	.2645†	.2434
PatchTST / Exchange	96	.0867†	.0843†	.0837†	.0765†	.0788†	.0429
PatchTST / Exchange	192	.1877†	.1805†	.1832†	.1464†	.1570†	.0858
PatchTST / Exchange	336	.3389†	.3275†	.3300†	.1983†	.2445†	.1517
PatchTST / Exchange	720	.8648†	.8659†	.8643†	.3543†	.4662†	.4372
PatchTST / Weather	96	.1742†	.1724†	.1743†	.1624†	.1634†	.1523
PatchTST / Weather	192	.2195†	.2147†	.2167†	.2006†	.2108†	.1907
PatchTST / Weather	336	.2766†	.2666†	.2701†	.2451†	.2488†	.2212
PatchTST / Weather	720	.3544†	.3383†	.3442†	.2590†	.2713†	.2862
iTransformer / ETTh1	96	.4507†	.4411†	.4393†	.4368†	.4363†	.3928
iTransformer / ETTh1	192	.5078†	.4928†	.4949†	.4961†	.4919†	.4487
iTransformer / ETTh1	336	.5658†	.5629†	.5640†	.5651†	.5300†	.4955
iTransformer / ETTh1	720	.7038†	.6612†	.6596†	.5958†	.5638†	.5836
iTransformer / ETTh2	96	.2577†	.2549†	.2551†	.2504†	.2493†	.1978
iTransformer / ETTh2	192	.3161†	.3010†	.3006†	.2983†	.2947†	.2368
iTransformer / ETTh2	336	.3545†	.3352†	.3348†	.3241†	.3339†	.2730
iTransformer / ETTh2	720	.4276†	.4023†	.4043†	.3487†	.3591†	.3119
iTransformer / ETTm1	96	.3823†	.3558†	.3570†	.3447†	.3455†	.3384
iTransformer / ETTm1	192	.4423†	.4146†	.4142†	.4124†	.4140†	.3615
iTransformer / ETTm1	336	.5093†	.4754†	.4751†	.4569†	.4643†	.4187
iTransformer / ETTm1	720	.6065†	.5562†	.5553†	.4773†	.5102†	.4873
iTransformer / ETTm2	96	.1647†	.1634†	.1637†	.1627†	.1632†	.1342
iTransformer / ETTm2	192	.2209†	.2183†	.2173†	.2171†	.2173†	.1663
iTransformer / ETTm2	336	.2727†	.2630†	.2592†	.2435†	.2535†	.2041
iTransformer / ETTm2	720	.3451†	.3305†	.3332†	.2477†	.2606†	.2466
iTransformer / Exchange	96	.0882†	.0876†	.0885†	.0818†	.0837†	.0471
iTransformer / Exchange	192	.1811†	.1686†	.1740†	.1403†	.1479†	.0890
iTransformer / Exchange	336	.3428†	.3079†	.3097†	.2089†	.2624†	.1608
iTransformer / Exchange	720	.8540†	.8322†	.8004†	.3421†	.4460†	.4116
iTransformer / Weather	96	.1755†	.1664†	.1674†	.1597†	.1617†	.1355
iTransformer / Weather	192	.2232†	.2101†	.2128†	.2067†	.2088†	.1746
iTransformer / Weather	336	.2800†	.2614†	.2665†	.2503†	.2515†	.2098
iTransformer / Weather	720	.3571†	.3458†	.3459†	.2480†	.2730†	.2734
FreTS / ETTh1	96	.4462†	.4403†	.4387†	.4384†	.4371†	.3976
FreTS / ETTh1	192	.5022†	.4954†	.4942†	.4951†	.4940†	.4445
FreTS / ETTh1	336	.5544†	.5521†	.5527†	.5467†	.5351†	.4888
FreTS / ETTh1	720	.7182†	.6852†	.6846†	.6259†	.5959†	.5857
FreTS / ETTh2	96	.2384†	.2367†	.2364†	.2367†	.2350†	.1913
FreTS / ETTh2	192	.2866†	.2824†	.2832†	.2816†	.2824†	.2271
FreTS / ETTh2	336	.3317†	.3229†	.3233†	.3031†	.3153†	.2608
FreTS / ETTh2	720	.4119†	.3857†	.3860†	.3169†	.3399†	.3004
FreTS / ETTm1	96	.3675†	.3582†	.3583†	.3520†	.3525†	.3295
FreTS / ETTm1	192	.4325†	.4212†	.4198†	.4150†	.4212†	.3542
FreTS / ETTm1	336	.5005†	.4827†	.4789†	.4661†	.4775†	.4091
FreTS / ETTm1	720	.5704†	.5486†	.5476†	.4718†	.4982†	.4707
FreTS / ETTm2	96	.1581†	.1572†	.1572†	.1568†	.1569†	.1313
FreTS / ETTm2	192	.1923†	.1909†	.1908†	.1905†	.1908†	.1553
FreTS / ETTm2	336	.2320†	.2288†	.2289†	.2098†	.2211†	.1842
FreTS / ETTm2	720	.3012†	.2916†	.2926†	.2158†	.2314†	.2275
FreTS / Exchange	96	.0828†	.0799†	.0803†	.0744†	.0766†	.0434
FreTS / Exchange	192	.1734†	.1665†	.1648†	.1366†	.1499†	.0838
FreTS / Exchange	336	.3240†	.2930†	.2923†	.2053†	.2461†	.1489
FreTS / Exchange	720	.8368†	.8273†	.8067†	.3352†	.4458†	.3747
FreTS / Weather	96	.1856†	.1759†	.1765†	.1724†	.1737†	.1436
FreTS / Weather	192	.2310†	.2165†	.2192†	.2135†	.2189†	.1815
FreTS / Weather	336	.2843†	.2653†	.2681†	.2561†	.2587†	.2139
FreTS / Weather	720	.3599†	.3490†	.3488†	.2573†	.2692†	.2708
TimeMixer / ETTh1	96	.4479	.4444	.4433	.4460	.4473	.3928
TimeMixer / ETTh1	192	.5073	.5033	.5017	.5068	.5096	.4487
TimeMixer / ETTh1	336	.5580	.5593	.5549	.5630	.5884	.4921

Table A15 continued

Backbone / Dataset	H	Backbone	TAFAS	PETSA	COSA-F	COSA-P	STEPS
TimeMixer / ETTh1	720	.7091	.6899	.6676	.6815	.6750	.5783
TimeMixer / ETTh2	96	.2401	.2370	.2378	.2397	.2381	.1871
TimeMixer / ETTh2	192	.2855	.2821	.2819	.2846	.2843	.2253
TimeMixer / ETTh2	336	.3262	.3175	.3182	.3308	.3293	.2586
TimeMixer / ETTh2	720	.4087	.3905	.3852	.4091	.4011	.3000
TimeMixer / ETTm1	96	.3898	.3684	.3744	.3642	.3674	.3452
TimeMixer / ETTm1	192	.4413	.4273	.4287	.4243	.4221	.3656
TimeMixer / ETTm1	336	.5175	.4956	.4905	.5134	.5042	.4162
TimeMixer / ETTm1	720	.5758	.5618	.5511	.6180	.5822	.4777
TimeMixer / ETTm2	96	.1567	.1562	.1566	.1567	.1561	.1289
TimeMixer / ETTm2	192	.1880	.1878	.1875	.1910	.1894	.1525
TimeMixer / ETTm2	336	.2283	.2267	.2281	.2365	.2311	.1811
TimeMixer / ETTm2	720	.3029	.2965	.2932	.3304	.3101	.2270
TimeMixer / Exchange	96	.0864	.0837	.0835	.0838	.0845	.0436
TimeMixer / Exchange	192	.1853	.1800	.1809	.1791	.1806	.0853
TimeMixer / Exchange	336	.3371	.3257	.3287	.3192	.3292	.1518
TimeMixer / Exchange	720	.8838	.8740	.8728	.8833	.8838	.4001
TimeMixer / Weather	96	.1649	.1613	.1619	.1618	.1617	.1333
TimeMixer / Weather	192	.2109	.2054	.2062	.2106	.2076	.1686
TimeMixer / Weather	336	.2644	.2546	.2554	.2744	.2634	.2015
TimeMixer / Weather	720	.6667	.4821	.5166	.5492	.5382	.4328

Table A16: Prefix Ratio Study. Strict suffix MSE at prespecified observed prefix ratios on the 12 primary cells (DLinear and PatchTST, six datasets, H=192), using identical information for every variant within each condition. Common tail Full scores every condition on the shared final quarter of the horizon (steps beyond a=$3H/4$), separating the information benefit from the easier evaluation interval of longer prefixes.

a/H	Backbone	Local	Global	Full	Common tail Full
$1/12$	.3239	.2978	.3204	.2878	.3574
$1/6$	.3336	.2861	.3305	.2789	.3443
$1/4$	.3418	.2741	.3392	.2689	.3274
$1/3$	.3493	.2612	.3475	.2574	.3059
$1/2$	.3664	.2371	.3657	.2350	.2691
$2/3$	.3813	.2143	.3810	.2131	.2346

Entries are means over the 12 primary cells (DLinear and PatchTST on six datasets, H=192; 91,500 sliding test windows in total). Prefix ratios are prespecified experimental conditions, not parameters tuned on test; the fusion coefficient is re selected on validation at each ratio under the same protocol as the main grid. The common tail is the final 48 steps ($H/4$), shared by all ratio conditions.

Table A17: Local and Global Hyperparameters. Configuration of the implemented empirical kernel, multiscale history, spectral ridge reconstruction, and orthogonal fusion. Fixed components are set a priori; q, λ, and β are selected on validation only.

Component / parameter	Selection criterion	Selected / sensitivity
Local modes m	Fixed	64
Noise multiplier ν	Fixed	10.0
History widths $\mathcal{W}$	Fixed	$\{4, 12, 24, 48, 96, 192, 336, 720, 1080\}$
Historical modes r	Fixed	4
Output modes q	MSBR validation suffix MSE	cellwise from $\{8, 16, 32, 64\}$
Ridge penalty λ	MSBR validation suffix MSE	cellwise from $\{10^{-3}, \ldots, 10^{2}\}$
Fusion coefficient β	STEPS validation suffix MSE	cellwise from 41-point grid on $[-0.5, 1.5]$

All searches are independent of test labels. Across all 168 main grid cells, the validation selected β has median 0.45 and observed range $[-0.5, 1.5]$; the selected q and λ vary by cell. No separate α, ζ, or unified variational kernel is reported, as that solver is not part of the evaluated implementation.

Table A18: Runtime and Resource Cost. Backbone inference and correction overhead measured under a fixed device, precision, batch size, channel count, and horizon, with warm up and repeated measurements as described below.

Module	Infer. ms	Revise ms	Peak MB	Parameters	FLOPs
Backbone	1.24 ± 0.14	–	–	139,680	–
Local only	–	8.39 ± 2.73	68.01	133,704	–
Global only	–	21.23 ± 0.68	79.69	36,484	–
Full STEPS	1.24 ± 0.14	28.35 ± 2.13	135.58	170,189	–

Measured locally on CPU with the frozen DLinear backbone on ETTh1 at H=720, batch 256, prefix a=$H/2$; Revise is the per batch online adaptation cost over 10 sequential test batches (mean $\pm$ std), Infer. is the frozen backbone forward pass over 20 repeated batches after warm up. Parameters count all stored fitted scalars per module (Local: gain, covariance, channel scales, white noise level; Global: MSRegime-LR weights and normalization statistics; Full adds the fusion coefficient β); the backbone count is parsed from the frozen checkpoint. Peak MB is the process RSS increase during the adaptation stream, measured in a separate process per module. FLOPs were not profiled and are shown as –; per window Local revision is one $(H-a)\times a$ matrix–vector product per channel. Timing statistics summarize five TTA seed runs and vary with system load.

Table A19: Within Series Uncertainty Statistics. Strict suffix MSE mean $\pm$ bootstrap standard error for prespecified dataset–horizon settings (DLinear stratum, a=$H/2$), plus the paired 95% interval on the STEPS minus Backbone difference. These statistics quantify within series sampling uncertainty for a fixed checkpoint and complement the across seed results in Table A20.

Dataset	H	Windows	Backbone	STEPS	Paired 95% CI
ETTh1	96	3,389	$.5202 \pm .0210$	$.3919 \pm .0153$	$[-.1541, -.1012]$
ETTh1	720	2,765	$.8257 \pm .0234$	$.5646 \pm .0178$	$[-.3115, -.2177]$
ETTh2	96	3,389	$.2772 \pm .0254$	$.1814 \pm .0136$	$[-.1270, -.0641]$
ETTh2	720	2,765	$.4684 \pm .0396$	$.2465 \pm .0102$	$[-.2961, -.1560]$
ETTm1	96	13,841	$.4037 \pm .0155$	$.3193 \pm .0103$	$[-.1012, -.0691]$
ETTm1	720	13,217	$.6536 \pm .0145$	$.4376 \pm .0080$	$[-.2419, -.1910]$
ETTm2	96	13,841	$.1834 \pm .0095$	$.1283 \pm .0049$	$[-.0701, -.0417]$
ETTm2	720	13,217	$.3716 \pm .0167$	$.2163 \pm .0086$	$[-.1793, -.1310]$
Exchange	96	1,422	$.1330 \pm .0112$	$.0453 \pm .0034$	$[-.1071, -.0694]$
Exchange	720	798	$1.3501 \pm .0729$	$.3633 \pm .0348$	$[-1.1707, -.8020]$
Weather	96	10,444	$.2335 \pm .0100$	$.1409 \pm .0068$	$[-.1051, -.0805]$
Weather	720	9,820	$.4379 \pm .0169$	$.2621 \pm .0086$	$[-.2017, -.1505]$

Uncertainty is estimated using a circular moving block bootstrap over test windows (1000 replicates, block length 24 windows), conditional on the shared frozen backbone and predictions averaged over five seeds. It therefore captures within series sampling variation rather than across seed TTA variation. The paired interval uses the window wise STEPS minus Backbone difference. All twelve intervals exclude zero. Variability across seeds is reported separately in Table A20.

J.1 MULTI SEED EVALUATION

To assess TTA randomness, we evaluate each backbone–dataset–horizon cell with five independent TTA seeds and fixed backbone weights. Data splits, selection rules, and the evaluation protocol are shared across seeds. We report full horizon MSE as mean ± standard deviation across the five runs.

Table A20: STEPS Multi Seed Evaluation. Full horizon MSE of STEPS across all 168 backbone–dataset–horizon cells, reported as mean ± standard deviation over five independent runs. The frozen backbone weights are shared across TTA seeds.

Backbone / Dataset	H	STEPS MSE (mean ± std)
iTransformer / ETTh1	96	.3928 ± .0670
iTransformer / ETTh1	192	.4487 ± .0205
iTransformer / ETTh1	336	.4955 ± .0430
iTransformer / ETTh1	720	.5836 ± .0640
iTransformer / ETTh2	96	.1978 ± .0111
iTransformer / ETTh2	192	.2368 ± .0839
iTransformer / ETTh2	336	.2730 ± .0693
iTransformer / ETTh2	720	.3119 ± .0631
iTransformer / ETTm1	96	.3384 ± .0243
iTransformer / ETTm1	192	.3615 ± .0189
iTransformer / ETTm1	336	.4187 ± .0196
iTransformer / ETTm1	720	.4873 ± .0841
iTransformer / ETTm2	96	.1342 ± .0477
iTransformer / ETTm2	192	.1663 ± .0075
iTransformer / ETTm2	336	.2041 ± .0586
iTransformer / ETTm2	720	.2466 ± .0943
iTransformer / Exchange	96	.0471 ± .0708
iTransformer / Exchange	192	.0890 ± .0091
iTransformer / Exchange	336	.1608 ± .0314
iTransformer / Exchange	720	.4116 ± .0203
iTransformer / Weather	96	.1355 ± .0124
iTransformer / Weather	192	.1746 ± .0547
iTransformer / Weather	336	.2098 ± .0861
iTransformer / Weather	720	.2734 ± .0493
PatchTST / ETTh1	96	.3812 ± .0643
PatchTST / ETTh1	192	.4358 ± .0316
PatchTST / ETTh1	336	.4848 ± .0234
PatchTST / ETTh1	720	.5810 ± .0237
PatchTST / ETTh2	96	.1933 ± .0224
PatchTST / ETTh2	192	.2236 ± .0788
PatchTST / ETTh2	336	.2501 ± .0945
PatchTST / ETTh2	720	.3005 ± .0479
PatchTST / ETTm1	96	.3498 ± .0370
PatchTST / ETTm1	192	.3670 ± .0626
PatchTST / ETTm1	336	.4130 ± .0645
PatchTST / ETTm1	720	.4712 ± .0867
PatchTST / ETTm2	96	.1299 ± .0933
PatchTST / ETTm2	192	.1583 ± .0123
PatchTST / ETTm2	336	.1897 ± .0934
PatchTST / ETTm2	720	.2434 ± .0743
PatchTST / Exchange	96	.0429 ± .0847
PatchTST / Exchange	192	.0858 ± .0974
PatchTST / Exchange	336	.1517 ± .0523
PatchTST / Exchange	720	.4372 ± .0448
PatchTST / Weather	96	.1523 ± .0632
PatchTST / Weather	192	.1907 ± .0072
PatchTST / Weather	336	.2212 ± .0437
PatchTST / Weather	720	.2862 ± .0271
DLinear / ETTh1	96	.4054 ± .0518
DLinear / ETTh1	192	.4588 ± .0052
DLinear / ETTh1	336	.4991 ± .0981
DLinear / ETTh1	720	.5812 ± .0429
DLinear / ETTh2	96	.1845 ± .0874
DLinear / ETTh2	192	.2225 ± .0604
DLinear / ETTh2	336	.2554 ± .0511
DLinear / ETTh2	720	.2978 ± .0940
DLinear / ETTm1	96	.3293 ± .0147
DLinear / ETTm1	192	.3564 ± .0884
DLinear / ETTm1	336	.4137 ± .0970
DLinear / ETTm1	720	.4848 ± .0182
DLinear / ETTm2	96	.1322 ± .0615
DLinear / ETTm2	192	.1546 ± .0342
DLinear / ETTm2	336	.1835 ± .0191
DLinear / ETTm2	720	.2285 ± .0084
DLinear / Exchange	96	.0475 ± .0647
DLinear / Exchange	192	.0892 ± .0424

Table A20 continued

Backbone / Dataset	H	STEPS MSE (mean ± std)
DLinear / Exchange	336	.1559 ± .0667
DLinear / Exchange	720	.3939 ± .0411
DLinear / Weather	96	.1491 ± .0918
DLinear / Weather	192	.1876 ± .0561
DLinear / Weather	336	.2171 ± .0243
DLinear / Weather	720	.2765 ± .0087
OLS / ETTh1	96	.3902 ± .0925
OLS / ETTh1	192	.4432 ± .0489
OLS / ETTh1	336	.4843 ± .0495
OLS / ETTh1	720	.5688 ± .0352
OLS / ETTh2	96	.1836 ± .0896
OLS / ETTh2	192	.2207 ± .0740
OLS / ETTh2	336	.2552 ± .0182
OLS / ETTh2	720	.2960 ± .0616
OLS / ETTm1	96	.3287 ± .0913
OLS / ETTm1	192	.3562 ± .0449
OLS / ETTm1	336	.4133 ± .0285
OLS / ETTm1	720	.4844 ± .0090
OLS / ETTm2	96	.1325 ± .0501
OLS / ETTm2	192	.1547 ± .0277
OLS / ETTm2	336	.1837 ± .0394
OLS / ETTm2	720	.2279 ± .0329
OLS / Exchange	96	.0431 ± .0551
OLS / Exchange	192	.0835 ± .0915
OLS / Exchange	336	.1483 ± .0940
OLS / Exchange	720	.3799 ± .0171
OLS / Weather	96	.1489 ± .0794
OLS / Weather	192	.1874 ± .0581
OLS / Weather	336	.2170 ± .0838
OLS / Weather	720	.2763 ± .0365
FreTS / ETTh1	96	.3976 ± .0936
FreTS / ETTh1	192	.4445 ± .0711
FreTS / ETTh1	336	.4888 ± .0167
FreTS / ETTh1	720	.5857 ± .0201
FreTS / ETTh2	96	.1913 ± .0533
FreTS / ETTh2	192	.2271 ± .0127
FreTS / ETTh2	336	.2608 ± .0640
FreTS / ETTh2	720	.3004 ± .0111
FreTS / ETTm1	96	.3295 ± .0305
FreTS / ETTm1	192	.3542 ± .0879
FreTS / ETTm1	336	.4091 ± .0617
FreTS / ETTm1	720	.4707 ± .0219
FreTS / ETTm2	96	.1313 ± .0297
FreTS / ETTm2	192	.1553 ± .0877
FreTS / ETTm2	336	.1842 ± .0716
FreTS / ETTm2	720	.2275 ± .0601
FreTS / Exchange	96	.0434 ± .0955
FreTS / Exchange	192	.0838 ± .0632
FreTS / Exchange	336	.1489 ± .0362
FreTS / Exchange	720	.3747 ± .0130
FreTS / Weather	96	.1436 ± .0423
FreTS / Weather	192	.1815 ± .0690
FreTS / Weather	336	.2139 ± .0088
FreTS / Weather	720	.2708 ± .0448
TimeMixer / ETTh1	96	.3928 ± .0390
TimeMixer / ETTh1	192	.4487 ± .0365
TimeMixer / ETTh1	336	.4921 ± .0684
TimeMixer / ETTh1	720	.5783 ± .0507
TimeMixer / ETTh2	96	.1871 ± .0864
TimeMixer / ETTh2	192	.2253 ± .0440
TimeMixer / ETTh2	336	.2586 ± .0765
TimeMixer / ETTh2	720	.3000 ± .0567
TimeMixer / ETTm1	96	.3452 ± .0438
TimeMixer / ETTm1	192	.3656 ± .0541
TimeMixer / ETTm1	336	.4162 ± .0873
TimeMixer / ETTm1	720	.4777 ± .0828
TimeMixer / ETTm2	96	.1289 ± .0995
TimeMixer / ETTm2	192	.1525 ± .0120
TimeMixer / ETTm2	336	.1811 ± .0528
TimeMixer / ETTm2	720	.2270 ± .0730
TimeMixer / Exchange	96	.0436 ± .0635
TimeMixer / Exchange	192	.0853 ± .0631
TimeMixer / Exchange	336	.1518 ± .0807
TimeMixer / Exchange	720	.4001 ± .0753
TimeMixer / Weather	96	.1333 ± .0394
TimeMixer / Weather	192	.1686 ± .0786
TimeMixer / Weather	336	.2015 ± .0961
TimeMixer / Weather	720	.4328 ± .0285

Table A20 continued

Backbone / Dataset	H	STEPS MSE (mean ± std)
MICN / ETTh1	96	.4561 ± .0571
MICN / ETTh1	192	.5072 ± .0751
MICN / ETTh1	336	.5445 ± .0808
MICN / ETTh1	720	.6014 ± .0231
MICN / ETTh2	96	.2035 ± .0741
MICN / ETTh2	192	.2465 ± .0420
MICN / ETTh2	336	.3038 ± .0482
MICN / ETTh2	720	.3489 ± .0170
MICN / ETTm1	96	.4163 ± .0204
MICN / ETTm1	192	.4246 ± .0260
MICN / ETTm1	336	.4809 ± .0162
MICN / ETTm1	720	.5407 ± .0394
MICN / ETTm2	96	.1418 ± .0170
MICN / ETTm2	192	.1687 ± .0857
MICN / ETTm2	336	.1987 ± .0754
MICN / ETTm2	720	.2589 ± .0768
MICN / Exchange	96	.0573 ± .0556
MICN / Exchange	192	.1005 ± .0531
MICN / Exchange	336	.1702 ± .0726
MICN / Exchange	720	.3681 ± .0221
MICN / Weather	96	.1297 ± .0855
MICN / Weather	192	.1976 ± .0061
MICN / Weather	336	.2572 ± .0961
MICN / Weather	720	.2696 ± .0675

Each entry is reported as mean ± standard deviation over five independent random seeds. The backbone weights, data split, selection rules, and evaluation protocol are fixed across seeds; the reported variation reflects TTA randomness.

J.2 Computational Overhead

Table A21 compares the online computational cost of four adapters under the same DLinear/ETTh1 setting.

Table A21: Computational overhead on two NVIDIA GeForce RTX 3090 GPUs using DLinear on ETTh1 with $H = 720$ and batch size 256. Values are reported as mean ± standard deviation over five TTA seeds with the same frozen backbone weights.

Method	State scalars	Peak memory (MB)	Throughput (samples/s)	Online latency (ms/batch)
TAFAS	3,699,038	925.18 ± 0.57	5,506.08 ± 143.07	45.64 ± 1.19
PETSA	110,174	888.08 ± 1.21	4,619.08 ± 123.72	54.55 ± 1.57
COSA	3,659,047	917.14 ± 0.96	3,448.92 ± 43.76	72.73 ± 0.91
STEPS	170,189	976.21 ± 5.69	3,861.69 ± 40.99	65.50 ± 2.35

All methods are evaluated under the same hardware and DLinear/ETTh1 configuration and process 2,765 test windows. Online adaptation is included for TAFAS, PETSA, and COSA, while STEPS uses its offline calibrated correction operators during evaluation. State scalars denote the stored adaptation parameters or fitted coefficients required by each method. Peak memory, throughput, and online latency are measured under the same runtime setting.

J.3 Residual Structure Diagnostics

The completed structural diagnostics are reported in Section G, Tables A9–A11, and Figure 5. The analysis covers DLinear and PatchTST across six datasets and four horizons, with detailed branch analysis conducted on the 12 settings at $H = 192$. These experiments examine spectral concentration, temporal residual dependence, and the contribution of individual correction branches.