RISK-CONTROLLED ADAPTIVE NATIVE-SCALE FUSION: A BACKBONE-AGNOSTIC FORECASTING FRAMEWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-scale modeling is widely used in time series forecasting to capture complementary temporal structures, from fine-grained variations to slower, long-range dynamics. However, temporal aggregation changes the information available at each scale: coarse views suppress high-frequency variation while losing local detail. Requiring coarse branches to reconstruct the finest-grid future, or allowing coarse predictions to modify fine-scale forecasts without sufficient control, can therefore introduce mismatched supervision and harmful cross-scale corrections. To address these, we propose a backbone-agnostic multi-scale forecasting framework that separates native-scale forecasting from prediction-space fusion. Each branch first predicts the future on its native temporal grid, and cross-scale interaction is deferred until these forecasts are completed. We instantiate the framework with two complementary fusion mechanisms: *Safe-linear*, which combines deterministic alignment with learned participation gates, and *Risk-Controlled Adaptive Native-Scale Fusion (RC-ANSF)*, which introduces bounded adaptive alignment for greater flexibility. A two-stage training procedure further separates scale-specific forecasting from cross-scale fusion. Across three backbone families, eight datasets, four forecasting horizons, the proposed framework consistently improves aggregate forecasting performance. These results support native-scale prediction followed by controlled prediction-space fusion as a reusable strategy for multi-scale forecasting. Our framework is available at `https://anonymous.4open.science/r/RC-ANSF-Time-Seires-Forecasting-Framework`.

1 INTRODUCTION

Time series forecasting aims to infer future trajectories from historical observations and serves as a core predictive tool in applications such as energy management (Hong & Fan, 2016), traffic forecasting (Li et al., 2018), and weather prediction (Bi et al., 2023). Despite rapid progress in forecasting architectures, real-world temporal dynamics remain difficult to model because informative patterns coexist over substantially different time spans: short-lived fluctuations and abrupt local changes may be superimposed on persistent trends, cycles, and slowly evolving regimes. A forecasting model therefore needs not only to represent temporal dependencies, but also to determine how information from different temporal resolutions should be modeled and combined for the final forecast.

Multi-scale modeling provides a natural way to expose these complementary temporal structures. At a fine resolution, observations preserve precise timing, local variation, short-period oscillations, and transient events; at coarser resolutions, temporal aggregation attenuates within-window fluctuations and emphasizes slower, more persistent dynamics over a longer effective time span. This coarse–fine complementarity has motivated multi-rate sampling, hierarchical interpolation, and cross-scale mixing in forecasting models such as N-HiTS (Challu et al., 2023), TimeMixer (Wang et al., 2024a), and TimeMixer++ (Wang et al., 2025). In this view, coarse scales are not merely compressed versions of the fine series: they emphasize slowly varying dynamics, while fine scales preserve local temporal details.

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.

Existing methods integrate scale interaction into forecasting: N-HiTS uses hierarchical interpolation (Challu et al., 2023), TimeMixer and TimeMixer++ mix multi-scale representations (Wang et al., 2024a; 2025), and Scaleformer couples successive forecasts through coarse-to-fine refinement (Shabani et al., 2023). This coupling makes it difficult to distinguish how a coarse branch learns useful predictive information from how that information should influence the final forecast. It raises two questions: *(i) how should coarse forecasts match the information preserved at their resolution, and (ii) how should they guide fine-scale predictions while limiting harmful corrections?*

Temporal aggregation is many-to-one: distinct fine-scale trajectories can produce the same coarse sequence. Native-scale supervision matches each branch's target to the information preserved at its resolution, without assuming that coarse forecasting is inherently easier. Even an accurate coarse forecast does not uniquely determine the fine-scale future, and alignment alone cannot recover details removed by aggregation. Moreover, the usefulness of coarse information varies across samples and variables, so unconditional fusion may distort reliable fine-scale predictions. Thus, learning scale-consistent forecasts and controlling their contribution to the final prediction are distinct problems.

We propose a backbone-agnostic framework that separates *native-scale forecasting* from *prediction-space fusion*. Each branch first learns to predict on its own temporal grid; the predictors are then frozen, and a separate fusion stage learns how to selectively combine the aligned forecasts. Two complementary instantiations realize this design: **Safe-linear** uses linear interpolation and gated fusion, while **Risk-Controlled Adaptive Native-Scale Fusion (RC-ANSF)** adds bounded adaptive alignment. This prediction-space interface accommodates heterogeneous backbones without requiring shared internal representations.

Our contributions are as follows:

- We formulate a native-scale prediction interface that matches supervision to temporal resolution before cross-scale fusion.
- We introduce Safe-linear and RC-ANSF as lightweight and bounded-adaptive fusion mechanisms within the same framework.
- Experiments on eight datasets and four horizons demonstrate consistent improvements over the corresponding single-scale baselines across all three backbone families.

2 Related Work

Time-series forecasting models. Forecasting architectures differ in their treatment of temporal and cross-variable dependencies. Informer develops efficient long-sequence attention (Zhou et al., 2021), Autoformer combines decomposition with auto-correlation (Wu et al., 2021), and FEDformer introduces frequency-domain modeling (Zhou et al., 2022). PatchTST combines temporal patching with channel independence (Nie et al., 2023), whereas iTransformer attends over variate tokens (Liu et al., 2024); TimeXer further incorporates exogenous variables through patch-wise self-attention and variate-wise cross-attention (Wang et al., 2024b). Alternatives include decomposition-based linear prediction in DLinear (Zeng et al., 2023), temporal two-dimensional representations in TimesNet (Wu et al., 2023), lightweight MLP mixing in TSMixer (Ekambaram et al., 2023), and centralized series-core interaction in SOFTS (Han et al., 2024). FITS offers another compact design, extending the input sequence through learned interpolation in the complex frequency domain (Xu et al., 2024). More recent designs refine how dependencies are selected: TimeFilter filters patch-specific spatial–temporal graph connections (Hu et al., 2025b), while TimeMosaic adapts patch granularity and decoding to temporal heterogeneity (Ding et al., 2026). These approaches improve the forecasting operator itself. Our framework complements this direction by operating on completed forecasts rather than requiring a shared hidden representation.

Multi-scale forecasting. NHITS combines multi-rate sampling with hierarchical interpolation (Challu et al., 2023), while SCINet recursively downsamples sequences and models interactions between the resulting subsequences (Liu et al., 2022). TimeMixer mixes decomposed components across scales and combines multiple predictors (Wang et al., 2024a), while TimeMixer++ integrates multi-scale and multi-resolution representations for general time-series analysis (Wang et al., 2025). Pathformer learns input-dependent pathways through multi-scale attention blocks (Chen et al., 2024), and AMD combines multi-scale decomposition with adaptive predictor synthesis (Hu et al., 2025a).

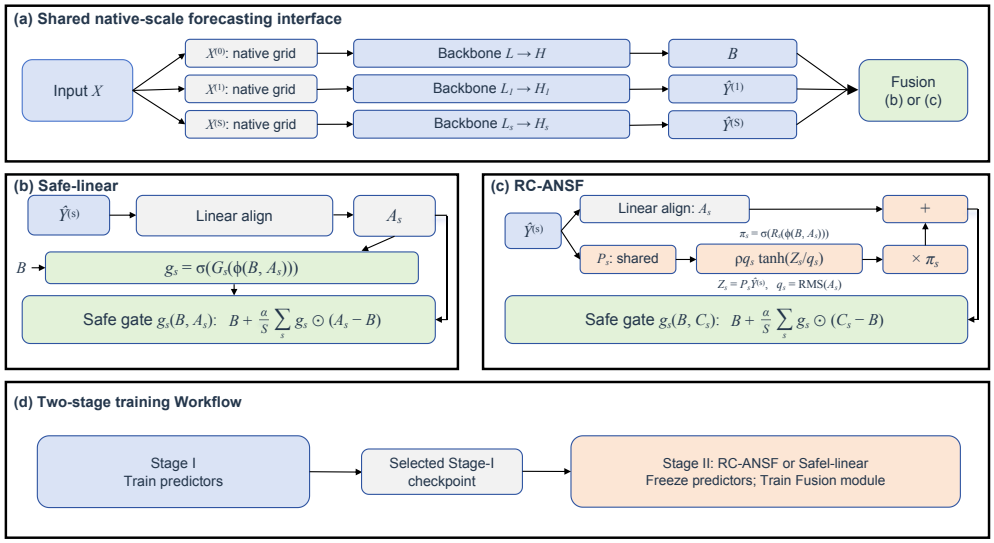

Figure 1: (a) Native-scale forecasting, where each branch predicts and is supervised on its own temporal grid. (b) Safe-linear—our lightweight fusion method. (c) RC-ANSF—our correction fusion method. (d) The two-stage training paradigm used in our framework workflow.

Scaleformer provides a reusable coarse-to-fine refinement procedure with architectural adaptations and cross-scale normalization (Shabani et al., 2023). These neural methods organize scale interaction within representations or successive forecasting. A related statistical approach, temporal hierarchies, combines forecasts across aggregation levels through reconciliation (Athanasopoulos et al., 2017). In contrast to enforcing coherence across resolutions, our objective is a finest-grid forecast with controlled coarse-scale contributions. Our focus is not adaptive weighting alone, but separating native-grid forecasting from subsequent alignment and fusion: each branch first completes its forecast, after which a separate head learns how the frozen outputs contribute to a finest-scale anchor. Safe-linear and RC-ANSF instantiate deterministic and bounded-adaptive alignment within this interface.

3 Method

3.1 Overview

Let $X \in \mathbb{R}^{L\times C}$ be an input sequence and $Y \in \mathbb{R}^{H\times C}$ the future target, where C is the number of variables. Figure 1 summarizes the overall framework and workflow. Our framework separates three operations: forecasting on native temporal grids via existing backbones, aligning completed forecasts, and selectively fusing the aligned forecast candidates. To accommodate different backbone forecasting architectures, our framework workflow consists of two stages: Stage I trains native-scale predictors; Stage II freezes their selected checkpoints and learns a prediction-space fusion head. We consider two such prediction-space heads: Safe-linear uses deterministic alignment, while RC-ANSF adds a bounded learned residual to each coarse-scale forecast candidate.

3.2 Native-scale forecasting

We construct $S+1$ scales, indexed by $s = 0, \ldots, S$, where $s = 0$ is the finest resolution and $S \geq 1$. Let $\mathcal{D}_0$ be the identity mapping and $\mathcal{D}_s$ a temporal downsampling operator. Each scale uses matched inputs and targets,

$$X^{(s)} = \mathcal{D}_s(X), \qquad Y^{(s)} = \mathcal{D}_s(Y), \tag{1}$$

with $X^{(s)} \in \mathbb{R}^{L_s \times C}$ and $Y^{(s)} \in \mathbb{R}^{H_s \times C}$. Its forecasting branch produces

$$\widehat{Y}^{(s)} = f_{\theta_s}(X^{(s)}; L_s, H_s) \in \mathbb{R}^{H_s \times C}. \tag{2}$$

Average pooling with window size W is used between adjacent scales. Pooling changes the temporal resolution, and each branch predicts a future sequence at the corresponding temporal resolution. This prediction process is referred to as the backbone performing forecasting on the native grid of each branch.

All downsampled coarse scales perform native-scale forecasting through their respective backbones: they do not need to generate the finest-grained details inside their forecasting backbones; alignment and fusion are applied only after that prediction is complete. This allows forecasting scales to share a common prediction-space interface while retaining their respective internal temporal and channel operations. We use the finest-scale forecast

$$B = \widehat{Y}^{(0)} \in \mathbb{R}^{H \times C} \tag{3}$$

as the base prediction anchor for further refinement using coarse-scale information.

3.3 SAFE-LINEAR: DETERMINISTIC ALIGNMENT AND ANCHORED CORRECTION FUSION

For each coarse scale $s > 0$, fixed linear interpolation maps the native forecast to the finest-grained grid:

$$A_s = \mathcal{A}_s^{\text{lin}}(\widehat{Y}^{(s)}) \in \mathbb{R}^{H \times C}. \tag{4}$$

This establishes a common-length time series for subsequent fusion correction. Then, a participation gate determines the weight by which each candidate finally acts on the base prediction:

$$g_s = \sigma\big(G_s\big(\phi(B, A_s)\big)\big) \in (0,1)^C. \tag{5}$$

Here G_s is a learnable MLP structure, and $\phi(B, A_s)$ summarizes forecast magnitudes, temporal variability, discrepancy, and directional consistency; see the appendix for details. Such gating can adaptively screen and filter coarse-grained information. We also introduce a global correction budget $\alpha \in [0,1]$ to control the upper bound of the coarse-grained influence on the final prediction. Safe-linear outputs

$$\widehat{Y}_{\text{SL}} = B + \frac{\alpha}{S} \sum_{s=1}^{S} g_s \odot (A_s - B), \tag{6}$$

where $\odot$ denotes channel-wise multiplication broadcast over time. The equivalent candidate weights are $w_{s,c} = \alpha g_{s,c}/S$, with the remaining weight assigned to B. They are nonnegative and sum to one. Thus, the total coarse contribution is at most α, each coarse branch receives at most α/S, and the base weight is at least $1 - \alpha$. As all gates approach zero, the forecast approaches B. Safe-linear provides the deterministic-alignment reference for the adaptive extension below.

3.4 RC-ANSF: BOUNDED RESIDUAL ALIGNMENT

Fixed interpolation cannot learn systematic corrections to forecast shape or amplitude. RC-ANSF retains A_s as an alignment reference and learns a residual proposal from the native forecast:

$$Z_{s,:,c} = P_s \widehat{Y}^{(s)}_{:,c}, \qquad P_s \in \mathbb{R}^{H \times H_s}, \qquad P_s\big|_{\text{init}} = 0. \tag{7}$$

The bias-free temporal matrix is shared across samples and channels within each scale. Zero initialization makes the initial adaptive candidate coincide with deterministic interpolation.

To make the correction budget relative to forecast magnitude, define

$$q_{s,c} = \max\left(\sqrt{\frac{1}{H}\sum_{t=1}^{H} A_{s,t,c}^2}, \epsilon\right), \qquad \epsilon > 0. \tag{8}$$

For a trust radius $\rho \in [0,1]$, the applied residual is parameterized as

$$\Delta_{s,t,c} = \rho q_{s,c} \tanh\left(\frac{Z_{s,t,c}}{q_{s,c}}\right), \qquad |\Delta_{s,t,c}| \le \rho q_{s,c}. \tag{9}$$

The RMS reference gives variables with different forecast magnitudes correspondingly different absolute budgets. The radius controls the admissible residual amplitude in prediction space.

A projection router determines the fraction of this bounded residual used for a sample and channel:

$$\pi_s = \sigma\big(R_s\big(\phi(B, A_s)\big)\big) \in (0,1)^C, \tag{10}$$

where R_s is an MLP structure similar to G_s. Then we produce the adapted candidate

$$C_s = A_s + \pi_s \odot \Delta_s. \tag{11}$$

The participation gate is then evaluated on that candidate,

$$g_s = \sigma\big(G_s\big(\phi(B, C_s)\big)\big), \tag{12}$$

and the final prediction is

$$\widehat{Y}_{\mathrm{RC}} = B + \frac{\alpha}{S} \sum_{s=1}^{S} g_s \odot (C_s - B). \tag{13}$$

The two weights act at different points: π_s modulates the alignment residual, while g_s weights the complete candidate. Suppressing the former approaches deterministic alignment; suppressing the latter approaches the fine-scale base prediction.

Structural properties. The adaptive candidates satisfy $|C_{s,t,c} - A_{s,t,c}| \le \rho q_{s,c}$ independently of the magnitude of P_s, and the final prediction retains the convex-combination structure of Safe-linear. Appendix A derives RMS-deviation and squared-loss envelopes. These properties bound departures from deterministic alignment and constrain mixture weights; they do not ensure that the deterministic candidates are accurate. Throughout the paper, *risk control* refers to these structural restrictions together with the empirical excess-loss penalty below, rather than a guarantee of non-increasing test error.

3.5 Two-stage training

The training procedure separates learning the native forecasts from learning how to use them. Fusion is bypassed in Stage I; the forecasting branches are frozen in Stage II.

Stage I: learning native-scale predictions. For training sample i, the native prediction loss at scale s is

$$\ell_{i,s}^{\mathrm{pred}} = \frac{1}{H_s C} \left\| \widehat{Y}_i^{(s)} - Y_i^{(s)} \right\|_F^2. \tag{14}$$

We train the branches using

$$\mathcal{L}_{\mathrm{pred}} = \frac{1}{N} \sum_{i=1}^{N} \sum_{s=0}^{S} \omega_s \ell_{i,s}^{\mathrm{pred}}, \qquad \omega_s \ge 0, \tag{15}$$

where ω_s determines each scale's contribution. The backbone-specific weighting and validation criteria are given in Appendix B. Once a checkpoint is selected, its forecasting and normalization modules are frozen and kept in evaluation mode.

Stage II: learning alignment and fusion. The forecasting and normalization modules selected in Stage I are frozen and kept in evaluation mode. Let $\widehat{Y}_i^F$ denote the fused output and define

$$\ell_i^F = \frac{\|\widehat{Y}_i^F - Y_i\|_F^2}{HC}, \qquad \ell_i^B = \frac{\|B_i - Y_i\|_F^2}{HC}. \tag{16}$$

The primary objective is the final-prediction MSE,

$$\mathcal{L}_{\mathrm{MSE}} = \frac{1}{N} \sum_{i=1}^{N} \ell_i^F. \tag{17}$$

We supplement it with an empirical relative excess-loss penalty,

$$\mathcal{L}_r = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\ell_i^F - \mathrm{sg}(\ell_i^B)}{\max(\mathrm{sg}(\ell_i^B), \epsilon_r)} + m_r\right]_+, \tag{18}$$

where $[x]_+ = \max(x, 0)$ and sg denotes stop-gradient. At $m_r = 0$, the penalty is applied only to samples whose fused loss exceeds the base loss; a positive m_r requires an improvement over the base to incur zero penalty. Because the penalty is computed per sample, improvements on other samples do not offset the penalty assigned to a violating sample. The positive denominator floor ϵ_r stabilizes comparisons when the base error is very small.

Two auxiliary terms provide more specific training preferences. The candidate-quality loss $\mathcal{L}_u$ supervises the participation gate g_s according to whether the deterministic candidate A_s has lower channel-wise squared error than the base prediction B. Both fusion heads use A_s as the reference, providing a fixed auxiliary target during Stage II rather than one that changes with the learned adaptive correction.

For RC-ANSF, the projection penalty $\mathcal{L}_p$ penalizes the magnitude of the applied adaptive residual relative to the deterministic candidate, favoring smaller corrections within the hard amplitude bound. The complete definitions and interpretation of these auxiliary terms are provided in Appendix C.2.

The two Stage-II objectives are

$$\mathcal{L}_{\mathrm{SL}} = \mathcal{L}_{\mathrm{MSE}} + \lambda_r\mathcal{L}_r + \lambda_u\mathcal{L}_u, \tag{19}$$

$$\mathcal{L}_{\mathrm{RC}} = \mathcal{L}_{\mathrm{MSE}} + \lambda_r\mathcal{L}_r + \lambda_u\mathcal{L}_u + \lambda_p\mathcal{L}_p, \tag{20}$$

with nonnegative auxiliary weights. Safe-linear learns the participation gates, while RC-ANSF additionally learns the residual projectors and projection routers. The two fusion heads are trained separately from the same selected Stage-I checkpoint within each matched setting. Ground-truth targets are used only to construct the training objectives; inference requires only the native forecasts and the learned alignment and fusion modules.

4 Experiments

We ask three questions: does the framework improve forecasting across backbone families, does adaptive fusion help when the predictors are fixed, and when is restricting alignment corrections beneficial?

4.1 Experimental setup

We evaluate multivariate forecasting on ETTh1, ETTh2, ETTm1, ETTm2, Weather, Electricity, Traffic, and Solar-Energy, with $H \in \{96, 192, 336, 720\}$. DLinear (Zeng et al., 2023), iTransformer (Liu et al., 2024), and TimeMixer-Core, derived from TimeMixer (Wang et al., 2024a), represent linear, attention-based, and MLP-based forecasting families. We report MSE and MAE, with lower values indicating better forecasts. In the main tables, bold and underline mark the best and second-best unrounded errors among the methods shown for each dataset, backbone, horizon (when reported), and metric.

The primary study compares Vanilla, Scaleformer with the same backbones (SF), Safe-linear, and RC-ANSF at each of the 96 backbone–dataset–horizon settings. SF uses the coarse-to-fine procedure from Scaleformer's official release with these backbones; integration details are in Appendix B.1. Main-table entries average the four horizon means; Appendix D reports per-horizon means and standard deviations. Macro errors weight settings equally, and improvement counts use unrounded metrics. These counts summarize configurations, not independent significance tests.

Fusion comparisons cover 64 DLinear/iTransformer settings; component and radius studies cover all 96 settings. These diagnostic studies use matched frozen Stage-I predictors, with separately fitted fusion heads. Safe-linear and RC also share their selected Stage-I checkpoint within each primary setting and run. The reference radius is $\rho = 0.1$. Data processing and training details are in Appendices B and C; fusion and component definitions are in Appendices E and F.

Table 1: Forecasting results averaged over four horizons. V: Vanilla; SF: Scaleformer; S: Safe-linear; RC: RC-ANSF.

Dataset	DLinear								iTransformer								TimeMixer-Core							
	V		SF		S		RC		V		SF		S		RC		V		SF		S		RC	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.6426	0.5587	**0.4602**	**0.4632**	0.4821	0.4768	0.4698	0.4691	0.5211	0.5091	0.5094	0.5024	0.4850	0.4877	**0.4819**	**0.4856**	0.4219	0.4386	0.4425	0.4561	0.4211	0.4372	**0.4206**	**0.4366**
ETTh2	0.3804	0.4139	0.3995	0.4254	0.3585	**0.3992**	**0.3521**	0.3998	0.4143	0.4446	0.5685	0.5097	0.4000	**0.4326**	**0.3953**	0.4345	0.3589	0.4020	0.3924	0.4196	0.3512	0.3973	**0.3448**	**0.3941**
ETTm1	0.5460	0.5000	0.4216	0.4296	0.4153	0.4218	**0.3983**	**0.4104**	0.4300	0.4394	0.4216	0.4228	0.4062	0.4235	**0.4016**	**0.4208**	0.3503	0.3835	0.3567	0.3857	0.3458	0.3806	**0.3438**	**0.3787**
ETTm2	0.3228	0.3650	0.2828	0.3382	**0.2790**	**0.3326**	0.2799	0.3384	0.3220	0.3715	0.3896	0.4056	0.3133	**0.3626**	**0.3124**	0.3684	0.2581	0.3162	0.3088	0.3515	0.2570	**0.3147**	**0.2564**	0.3209
Weather	0.2538	0.2855	0.2445	0.2921	0.2472	0.2799	**0.2429**	**0.2789**	0.2342	0.2716	0.2296	0.2719	0.2253	0.2644	**0.2219**	**0.2626**	0.2294	0.2623	0.2250	0.2790	0.2281	0.2616	**0.2244**	**0.2608**
Electricity	0.1995	0.3014	0.2054	0.3071	0.1884	0.2904	**0.1836**	**0.2851**	0.2386	0.3347	**0.1674**	**0.2652**	0.2149	0.3181	0.2090	0.3128	0.1718	**0.2670**	**0.1686**	0.2675	0.1726	0.2694	0.1718	0.2691
Traffic	0.4718	0.3468	**0.4455**	**0.3146**	0.4590	0.3342	0.4533	0.3265	0.4624	0.3412	**0.4187**	**0.2989**	0.4661	0.3422	0.4617	0.3372	**0.4302**	**0.2972**	0.4305	0.2979	0.4332	0.3007	0.4320	0.2988
Solar-Energy	0.2939	0.3677	0.2899	0.3599	0.2816	0.3509	**0.2751**	**0.3403**	0.2796	0.3180	**0.1914**	**0.2500**	0.2614	0.3043	0.2555	0.2974	0.2204	0.2949	0.2166	**0.2801**	0.2156	0.2915	**0.2125**	0.2838

Table 2: Shared-predictor fusion comparison, averaged over four horizons.

Dataset	DLinear														iTransformer													
	Base		Mean		Global		MLP		Context		Safe-linear		RC		Base		Mean		Global		MLP		Context		Safe-linear		RC	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.645	0.558	0.482	0.475	0.478	0.474	0.479	0.474	0.542	0.514	0.480	0.475	**0.468**	**0.467**	0.523	0.508	0.499	0.498	0.483	0.488	0.489	0.490	0.513	0.499	0.486	0.487	**0.482**	**0.485**
ETTh2	0.376	0.412	0.353	0.396	0.353	**0.396**	0.355	0.398	0.365	0.403	0.354	0.397	**0.348**	0.398	0.419	0.447	0.402	0.433	0.399	**0.430**	0.400	0.433	0.412	0.441	0.400	0.432	**0.398**	0.436
ETTm1	0.561	0.507	0.428	0.431	0.404	0.414	0.404	0.414	0.457	0.451	0.415	0.422	**0.398**	**0.410**	0.454	0.455	0.412	0.428	**0.406**	**0.424**	0.406	0.424	0.432	0.441	0.411	0.427	0.406	0.424
ETTm2	0.327	0.367	0.284	0.336	**0.274**	**0.328**	0.274	0.330	0.294	0.347	0.281	0.333	0.282	0.339	0.321	0.371	0.310	0.360	0.310	0.361	0.310	0.362	0.315	0.367	**0.309**	**0.360**	0.309	0.367
Weather	0.254	0.286	0.248	0.280	0.247	0.279	0.247	0.282	**0.242**	0.285	0.247	0.280	0.242	**0.279**	0.233	0.271	0.227	0.264	0.227	0.265	0.226	0.270	0.226	0.270	0.225	0.264	**0.222**	**0.262**
Electricity	0.199	0.302	0.217	0.328	0.188	0.289	0.187	0.288	0.190	0.294	0.188	0.291	**0.183**	**0.285**	0.239	0.336	0.243	0.351	0.208	0.311	**0.204**	**0.308**	0.228	0.328	0.211	0.315	0.206	0.311
Traffic	0.471	0.347	0.511	0.370	0.457	0.329	0.456	**0.325**	0.464	0.342	0.459	0.334	**0.453**	0.326	0.466	0.344	0.518	0.372	0.460	0.336	0.459	0.334	0.461	0.338	0.462	0.338	**0.458**	**0.334**
Solar-Energy	0.293	0.368	0.281	0.351	0.281	0.350	0.276	0.342	0.276	0.347	0.280	0.348	**0.274**	**0.338**	0.276	0.318	0.251	0.299	0.248	0.298	0.245	0.298	0.245	0.306	0.252	0.299	**0.245**	**0.292**

4.2 Forecasting performance across backbones

Framework gains and adaptive refinement. Relative to Vanilla, RC reduces macro MSE by 14.6%, 5.6%, and 1.4% for DLinear, iTransformer, and TimeMixer-Core, respectively. The framework therefore benefits all three families in aggregate, with larger gains for DLinear. Relative to Safe-linear, RC lowers MSE in 86/96 settings and MAE in 74/96, with a mean matched relative MSE reduction of 1.33%. The adaptive extension yields smaller but recurrent improvements over deterministic alignment. Safe-linear remains competitive, however, retaining lower average MAE on ETTm2 for all three backbones.

Comparison with an alternative multi-scale framework. RC has lower four-horizon average MSE than SF in 17/24 backbone–dataset pairs, and lower MAE in 15/24. Its advantage is more frequent at longer horizons: RC has lower MSE in 20/24 settings at $H = 720$, compared with 11/24 at $H = 96$. SF retains lower average errors on Electricity, Traffic, and Solar-Energy with iTransformer. Thus, the results favor RC overall and particularly at long horizons in this comparison, without establishing universal superiority over multi-scale forecasting methods.

4.3 Comparison of fusion strategies

Table 2 compares Base, Mean, Global, MLP, Context, Safe-linear, and RC using the same native-scale forecasting checkpoint within each setting. Base is its finest-scale branch, not the independently trained Vanilla model. This comparison tests how completed forecasts are used without retraining their forecasting branches.

Unconditional averaging can degrade a useful fine-scale prediction: with DLinear, Mean increases average MSE from 0.199 to 0.217 on Electricity and from 0.471 to 0.511 on Traffic. RC instead lowers these errors to 0.183 and 0.453, respectively, illustrating the value of controlling coarse-scale participation.

RC lowers MSE relative to Base in all 64 settings and relative to Safe-linear in 58/64. Its macro MSE/MAE is 0.3359/0.3595, the lowest among the seven heads. Against Global and MLP, RC improves MSE in 50/64 and 47/64 settings, with macro MSE reductions of 0.89% and 0.82%, respectively. These modest but recurrent gains support the complete adaptive alignment-and-fusion

head beyond changing only the mixing rule. Component interventions examine which parts of this head account for the observed behavior.

4.4 Ablation study

No TR removes the bounded residual transformation. No gate fixes participation to one and disables candidate-quality supervision. Table 3 presents ETTh2 and ETTm2, the two datasets with the largest relative macro MSE reductions over No TR in the completed study. Appendix F reports all datasets, including settings where the ablations perform better.

Table 3: Component ablations on ETTh2 and ETTm2.

Dataset	H	DLinear						iTransformer						TimeMixer-Core					
		RC		No TR		No gate		RC		No TR		No gate		RC		No TR		No gate	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh2	96	**0.2868**	**0.3499**	0.2977	0.3586	0.2916	0.3572	**0.3505**	**0.4002**	0.3582	0.4094	0.3545	0.4064	**0.2751**	**0.3406**	0.2799	0.3448	0.2812	0.3474
	192	**0.3369**	**0.3882**	0.3500	0.4000	0.3426	0.3960	**0.3892**	**0.4269**	0.4457	0.4710	0.3967	0.4348	**0.3440**	**0.3920**	0.3603	0.4037	0.3582	0.4023
	336	**0.3634**	**0.4115**	0.3913	0.4295	0.3661	0.4153	**0.4088**	**0.4484**	0.5190	0.5197	0.4141	0.4525	**0.3630**	**0.4099**	0.4158	0.4485	0.3760	0.4216
	720	**0.4048**	**0.4424**	0.6228	0.5487	0.4087	0.4473	**0.4429**	**0.4677**	0.9632	0.6883	0.4513	0.4761	**0.4060**	**0.4415**	0.7636	0.6143	0.4067	0.4447
ETTm2	96	0.1758	0.2652	0.1724	0.2620	**0.1721**	**0.2618**	0.2321	0.3209	**0.2311**	**0.3182**	0.2366	0.3238	0.1637	0.2550	**0.1637**	**0.2547**	0.1645	0.2556
	192	0.2439	0.3213	**0.2378**	**0.3103**	0.2429	0.3192	0.2758	0.3477	**0.2752**	**0.3437**	0.2853	0.3543	0.2216	0.2984	**0.2202**	**0.2946**	0.2234	0.2999
	336	0.3075	0.3554	**0.2951**	**0.3448**	0.3039	0.3576	**0.3263**	**0.3757**	0.3375	0.3858	0.3371	0.3906	**0.2762**	0.3361	0.2775	**0.3350**	0.2868	0.3475
	720	0.4003	**0.4154**	0.4302	0.4365	**0.3980**	0.4159	**0.4024**	**0.4220**	0.4200	0.4397	0.4171	0.4370	**0.3544**	**0.3892**	0.3712	0.3993	0.3582	0.3957

Bounded alignment: protection with an accuracy trade-off. On ETTh2, RC improves both metrics over both ablations for every backbone and horizon. For iTransformer at $H = 720$, No TR increases MSE from 0.4429 to 0.9632. On ETTm2, RC reduces macro MSE by 1.51% but improves MSE in only 5/12 settings. Across all 96 settings, RC lowers macro MSE from 0.3340 to 0.3245 while winning 27 settings and losing 69. Weather illustrates the opposite trade-off: No TR has lower MSE in 11/12 settings, whereas RC has lower MAE in 10/12. Thus, the aggregate gain reflects avoiding large errors in a subset of configurations, not uniform accuracy improvements or a guarantee on sample-wise test risk.

Learned participation. Relative to No gate, RC lowers MSE in 69/96 settings and MAE in 78/96; macro MSE decreases from 0.3459 to 0.3245. This supports selective participation overall, but the intervention also removes its supervision and changes the base coefficient, so it does not isolate the gate network alone. The additional No router comparison in Appendix F yields nearly unchanged macro error; the results do not establish that every component is necessary.

4.5 Parameter sensitivity

We vary $\rho \in \{0.05, 0.10, 0.25, 0.50, 1.00\}$ with matched frozen predictors. In Figure 2, larger radii increase long-horizon error on ETTh2, whereas Traffic generally benefits from a larger correction budget. This contrast is consistent with $|\Delta_{s,t,c}| \leq \rho q_{s,c}$: ρ limits the relative alignment deviation but does not identify a universally optimal correction strength. The reference $\rho = 0.1$ is not presented as a per-dataset optimum or as a setting selected by this test-set sweep. These curves describe the implemented parameterization; varying ρ can affect optimization as well as the amplitude limit. Complete MSE/MAE results are in Appendix G.

4.6 Discussion

What the comparisons establish. The primary study shows gains from the complete framework across three forecasting families, while the shared-predictor study attributes further gains to how completed forecasts are aligned and combined. These comparisons support a reusable prediction-space interface rather than a claim of state-of-the-art accuracy across all forecasting architectures.

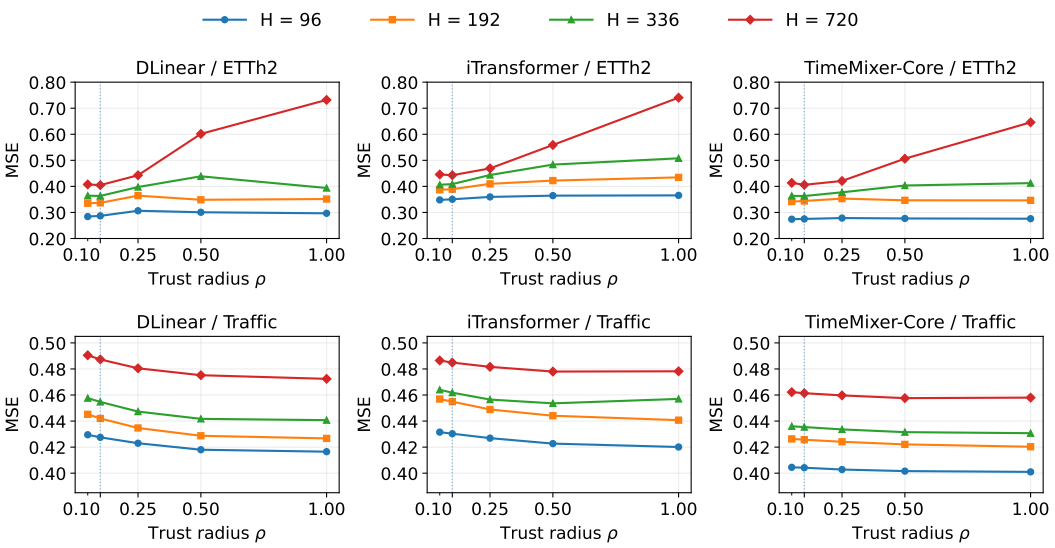

Figure 2: Trust-radius sensitivity in raw MSE on ETTh2 (top) and Traffic (bottom). Columns correspond to DLinear, iTransformer, and TimeMixer-Core; curves show all four horizons. The dotted line marks the reference $\rho = 0.1$. Predictors are fixed and a fusion head is fitted for each radius.

Since the Vanilla comparison also changes branch capacity and training, it does not isolate native-scale supervision as the sole source of improvement.

Interpreting risk control. The component and radius studies reveal a trade-off: bounded adaptation avoids large errors in some configurations but restricts useful corrections in others. The structural bound controls alignment deviation; the empirical risk penalty discourages excess training loss. Neither guarantees lower test error. The evidence supports task-dependent control of cross-scale corrections, not unconditional improvements from tighter constraints.

5 Conclusion

We presented a framework that separates native-scale forecasting from prediction-space fusion. Its two instantiations offer complementary choices: Safe-linear combines deterministic alignment with learned participation, while RC-ANSF adds bounded adaptive corrections. Across three backbone families, eight datasets, and four horizons, the framework improves aggregate forecasting accuracy, and shared-predictor comparisons support the adaptive extension. The component studies also expose the trade-off between limiting harmful corrections and retaining useful flexibility. These findings support native-scale forecasting followed by controlled fusion as a reusable multi-scale design.

Reproducibility statement

The framework, equations and training procedure are in Section 3; evaluation and implementation details are in Section 4 and Appendix B. The experiments can be reproduced using the source code in the abstract and the parameter settings and seeds specified in the appendix.

Ethics statement

This work uses established public time-series benchmarks and does not involve human subjects or personally identifiable information.

AI use statement

AI tools are used for grammar checking and polishing.

References

George Athanasopoulos, Rob J. Hyndman, Nikolaos Kourentzes, and Fotios Petropoulos. Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74, 2017. doi: 10.1016/j.ejor.2017.02.046.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619:533–538, 2023. doi: 10.1038/s41586-023-06185-3.

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.

Peng Chen, Yingying Zhang, Yunyao Cheng, Yang Shu, Yihang Wang, Qingsong Wen, Bin Yang, and Chenjuan Guo. Pathformer: Multi-scale transformers with adaptive pathways for time series forecasting. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=lJkOCMP2aW.

Kuiye Ding, Fanda Fan, Chunyi Hou, Zheya Wang, Lei Wang, Zhengxin Yang, and Jianfeng Zhan. TimeMosaic: Temporal heterogeneity guided time series forecasting via adaptive granularity patch and segment-wise decoding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20790–20798, 2026. doi: 10.1609/aaai.v40i25.39218.

Vijay Ekambaram, Arindam Jati, Nam Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. TSMixer: Lightweight MLP-mixer model for multivariate time series forecasting. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023. doi: 10.1145/3580305.3599533. URL https://doi.org/10.1145/3580305.3599533.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. SOFTS: Efficient multivariate time series forecasting with series-core fusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/754612bde73a8b65ad8743f1f6d8ddf6-Abstract-Conference.html.

Tao Hong and Shu Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016. doi: 10.1016/j.ijforecast.2015.11.011.

Yifan Hu, Peiyuan Liu, Peng Zhu, Dawei Cheng, and Tao Dai. Adaptive multi-scale decomposition framework for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17359–17367, 2025a. doi: 10.1609/aaai.v39i16.33908.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. TimeFilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *Proceedings of the 42nd International Conference on Machine Learning*, volume

267 of *Proceedings of Machine Learning Research*, pp. 24893–24911, 2025b. URL https://proceedings.mlr.press/v267/hu25ac.html.

Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=SJiHXGWAZ.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/2ea18fdc667e0ef2ad82b2b4d65147ad-Abstract-Conference.html.

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.

Mohammad Amin Shabani, Amir Abdi, Lili Meng, and Tristan Sylvain. Scaleformer: Iterative multi-scale refining transformers for time series forecasting. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=sCrnllCtjoE.

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*, 2024a. URL https://openreview.net/forum?id=7oLshfEIC2.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. TimeMixer++: A general time series pattern machine for universal predictive analysis. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=1CLzLXSFNn.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. TimeXer: Empowering transformers for time series forecasting with exogenous variables. In *Advances in Neural Information Processing Systems*, volume 37, 2024b. URL https://papers.nips.cc/paper_files/paper/2024/hash/0113ef4642264adc2e6924a3cbbdf532-Abstract-Conference.html.

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, pp. 22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *International Conference on Learning Representations*, 2023.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/701251e1db4a2e4dd2ef23f5265d5936-Paper-Conference.pdf.

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. doi: 10.1609/aaai.v37i9.26317.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

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

Appendix guide. Appendix A proves the structural bounds. Appendices B and C specify the experimental protocol, backbone interfaces, and implementation details. Appendices D–G provide the complete primary, fusion, component, and radius results in the same order as the main experiments. Appendix H provides selected qualitative examples and a scale-participation diagnostic, not additional evidence of average performance.

A STRUCTURAL BOUNDS

We derive the structural properties of the alignment and fusion parameterization in Section 3.4. The statements hold for each sample; sample indices are omitted. For $V \in \mathbb{R}^{H\times C}$, define $\|V\|_{\mathrm{rms}} = \|V\|_F/\sqrt{HC}$.

Proposition 1 (bounded alignment deviation). For every coarse scale, the adapted candidate satisfies

$$\|C_s - A_s\|_{\mathrm{rms}} \le \rho\sqrt{\|A_s\|_{\mathrm{rms}}^2 + \epsilon^2}. \tag{21}$$

The bound does not depend on the magnitude of the learned projector P_s.

Proof. Equations 9 and 11, together with $0 \le \pi_{s,c} \le 1$, give $|C_{s,t,c} - A_{s,t,c}| \le \rho q_{s,c}$. Consequently,

$$\|C_s - A_s\|_{\mathrm{rms}}^2 \le \frac{\rho^2}{C}\sum_c q_{s,c}^2. \tag{22}$$

Writing $a_{s,c}^2 = H^{-1}\sum_t A_{s,t,c}^2$, we have $q_{s,c}^2 = \max(a_{s,c}^2, \epsilon^2) \le a_{s,c}^2 + \epsilon^2$. Averaging over channels and taking square roots gives Equation 21. □

Proposition 2 (convex fusion and coefficient budget). For $0 \le \alpha \le 1$, the fused prediction can be written channel-wise as

$$\widehat{Y}_{:,c} = w_{0,c}B_{:,c} + \sum_{s=1}^{S} w_{s,c}C_{s,:,c}, \tag{23}$$

where

$$w_{s,c} = \frac{\alpha}{S}g_{s,c}, \qquad w_{0,c} = 1 - \sum_{s=1}^{S} w_{s,c} \ge 1 - \alpha. \tag{24}$$

The coefficients are nonnegative and sum to one, and $w_{s,c} \le \alpha/S$ for every coarse branch.

Proof. Expanding Equation 13 yields

$$\widehat{Y}_{:,c} = \left(1 - \frac{\alpha}{S}\sum_s g_{s,c}\right)B_{:,c} + \sum_s \frac{\alpha}{S}g_{s,c}C_{s,:,c}. \tag{25}$$

Since $g_{s,c} \in [0,1]$, the coarse weights sum to at most α. The claimed bounds and unit total follow directly. □

At $\alpha < 1$, the base has a strictly positive uniform coefficient floor. At $\alpha = 1$, the uniform lower bound is zero. Finite sigmoid logits produce gates in $(0,1)$: exact zero-gate fallback is an algebraic boundary case, approached as the logits decrease. Similarly, $P_s = 0$ recovers the deterministic candidate, not a separately trained Safe-linear head with potentially different gate parameters.

Corollary 1 (prediction and squared-loss envelopes). Define

$$d = \frac{\alpha}{S}\sum_{s=1}^{S}\left(\|A_s - B\|_{\mathrm{rms}} + \rho\sqrt{\|A_s\|_{\mathrm{rms}}^2 + \epsilon^2}\right). \tag{26}$$

Then

$$\|\widehat{Y}_{\mathrm{RC}} - B\|_{\mathrm{rms}} \le d. \tag{27}$$

For $\ell(U,Y) = \|U - Y\|_{\mathrm{rms}}^2$,

$$\ell(\widehat{Y}_{\mathrm{RC}}, Y) - \ell(B,Y) \le 2\|B - Y\|_{\mathrm{rms}}d + d^2. \tag{28}$$

Proof. Multiplication by channel weights in $[0, 1]$ cannot increase the RMS norm. Applying the triangle inequality to Equation 13 gives

$$\|\widehat{Y}_{\mathrm{RC}} - B\|_{\mathrm{rms}} \le \frac{\alpha}{S} \sum_s \|C_s - B\|_{\mathrm{rms}} \tag{29}$$

$$\le \frac{\alpha}{S} \sum_s (\|A_s - B\|_{\mathrm{rms}} + \|C_s - A_s\|_{\mathrm{rms}}) \le d, \tag{30}$$

where the last step uses Proposition 1. Next set $E = B - Y$ and $\delta = \widehat{Y}_{\mathrm{RC}} - B$. The squared-loss difference is

$$\ell(\widehat{Y}_{\mathrm{RC}}, Y) - \ell(B, Y) = \frac{2}{HC}\langle E, \delta\rangle + \|\delta\|^2_{\mathrm{rms}}. \tag{31}$$

Cauchy–Schwarz and $\|\delta\|_{\mathrm{rms}} \le d$ imply Equation 28. □

Interpretation. The radius bounds the additional alignment correction, while the prediction envelope also depends on $\|A_s - B\|_{\mathrm{rms}}$. This discrepancy need not be small even when $\rho = 0$. The squared-loss envelope consequently bounds the size of a possible increase but does not force it to be nonpositive. These are pointwise properties of the parameterization, distinct from both the empirical training penalty and a statistical guarantee on test risk.

B Experimental coverage and training protocol

The primary study evaluates Vanilla, Scaleformer with the same backbones (SF), Safe-linear, and RC-ANSF with DLinear, iTransformer, and TimeMixer-Core on eight datasets and four horizons. Each of the 96 backbone–dataset–horizon settings uses seeds 2021–2023. The shared-predictor fusion study compares seven heads on the 64 DLinear and iTransformer settings at seed 2021. The core component and radius studies use all three backbones at seed 2021. Qualitative visualizations use saved seed-2021 checkpoints, while their full-test reference means follow the primary three-seed protocol. A reference result reused in more than one comparison remains the same fitted model.

B.1 Backbone integration

Native-scale forecasting in our framework. At scale s, an independently parameterized backbone maps the pooled history of length L_s to a forecast of length H_s. Except on Solar-Energy, the wrapper applies reversible normalization at each scale and reverses it before alignment and fusion; dataset-level standardization is used throughout. Backbone weights are not shared across scales. Safe-linear and RC-ANSF use the same native-scale forecasting interface and selected Stage-I checkpoint; only their Stage-II heads differ. Vanilla uses a single instance of the corresponding backbone at the original input and output lengths, without cross-scale fusion.

DLinear and iTransformer. We wrap the official DLinear and iTransformer implementations, setting their input and output lengths separately for each scale. DLinear retains its moving-average decomposition and two temporal linear maps, with weights shared across channels (`individual=False`). iTransformer retains its inverted variable embedding, attention encoder, and forecast projection; its internal normalization is disabled because normalization is handled by the wrapper. Calendar covariates are not passed to either backbone.

Construction of TimeMixer-Core. TimeMixer-Core is a channel-independent, single-scale specialization of the TimeMixer forecasting path. We remove the internal input pyramid, cross-scale seasonal/trend mixing, and summation of multiple predictors, leaving multi-scale organization to the external framework. Each channel is processed with shared weights: a circular one-dimensional convolution (kernel size 3) embeds scalar observations into d_{model} features, followed by embedding dropout. No positional or calendar embedding is added.

With only one internal scale, each decomposition block produces seasonal and trend features whose sum is its input E. The retained block therefore reduces to a feature-wise residual MLP,

$$E_{\mathrm{season}} + E_{\mathrm{trend}} = E, \qquad E' = E + W_2\,\mathrm{GELU}(W_1 E + b_1) + b_2. \tag{32}$$

After two such blocks, a temporal linear map changes the sequence length from L_s to H_s, and a feature projection maps d_{model} to one value per channel. Thus, TimeMixer-Core consists of value embedding, residual MLP blocks, and a temporal prediction head; it is not the full multi-scale TimeMixer model. Its width and dropout settings are given in Appendix C.1.

The same backbones in Scaleformer. SF adapts the coarse-to-fine procedure in Scaleformer's official release to the same three backbone implementations. Following the FiLMMS direct-prediction interface, each scale concatenates its history with a forecast initialization: zeros at the coarsest scale, or the linearly upsampled, stop-gradient prediction from the preceding coarser scale. The concatenated sequence is jointly normalized with scale-specific learnable affine parameters, passed to the backbone, and the forecast is denormalized. Each backbone instance consequently maps $L_s + H_s$ inputs to H_s outputs; the additional H_s entries are predicted values, not extra observed history. DLinear's temporal maps, iTransformer's input embedding, and TimeMixer-Core's temporal head are sized to this interface. Instances have independent weights across scales, and the finest-scale prediction is the final output. The comparison matches backbone families and forecasting tasks; parameter counts can differ.

B.2 Two-stage training

Stage-I objectives and model selection. For DLinear, Equation 15 uses $\omega_0 = 1$ and $\omega_s = 1/S$ for $s > 0$: the fine loss plus the mean coarse loss. These coefficients sum to two, and checkpoint selection uses only fine-path validation MSE. For iTransformer and TimeMixer-Core, training and validation use $\omega_s = 1/(S+1)$, the uniform mean over native-scale losses. Vanilla has one scale and therefore uses finest-grid MSE. Within every matched setting and seed, Safe-linear and RC-ANSF use the same selected forecasting checkpoint. Component variants likewise use their matched Stage-I checkpoint.

Stage-II training. The selected forecasting and normalization modules are frozen and kept in evaluation mode. Safe-linear fits the participation gates; RC-ANSF additionally fits the residual projectors and projection routers. Early stopping selects the smallest complete validation objective, including the enabled auxiliary penalties, rather than MSE alone. For both stages, validation selection averages minibatch objectives, including the final incomplete batch; reported test MSE/MAE instead accumulate error over all test elements. Separately fitted heads share the forecasting inputs to fusion, while their objectives differ as specified in Section 3.5.

Scaleformer training and selection. SF uses the same histories and native scale lengths, but its own coarse-to-fine training procedure. Its objective is $\frac{1}{2}\sum_{s=0}^{S} 2^{-s}\mathcal{L}_s$, with the finest grid indexed by zero. The epoch cap is the sum of our two stage caps, with early-stopping patience five. Checkpoints are selected by finest-grid validation MSE accumulated over all elements. For each setting, learning rates are searched over $\eta\{1, 0.3, 0.1\}$ using seed 2021; the selected rate is reused for seeds 2022 and 2023. Tuning uses validation data only, and the selected first-run checkpoint is reused for testing. This protocol matches forecasting tasks, not parameter counts or training compute. The smallest rate is selected in 65/96 settings, so the comparison is limited to the evaluated search range rather than exhaustive tuning.

C Implementation details and auxiliary objectives

C.1 Data and optimization

The ETT loaders use chronological 12/4/4-month training, validation, and test blocks, with hourly or 15-minute sampling as appropriate. The custom and Solar loaders use chronological 70/10/20 percent splits. Standardization is fitted on training data. Training windows are shuffled and the last incomplete batch is dropped; validation and test windows are not shuffled and all windows are retained. The task is multivariate forecasting with label length zero, and the wrapper's forward path does not consume calendar features. Scales are constructed by factor-two average pooling, applied to both the input and the future target.

Training uses AdamW with weight decay 0.01 and no learning-rate schedule. Early-stopping patience is 5 epochs in Stage I and 3 epochs in Stage II. The maximum epoch counts for the two stages are 40/40 for ETTh1, ETTm1, and ETTm2; 10/10 for ETTh2; and 20/20 for Weather, Electricity, Traffic, and Solar-Energy. Actual stopping epochs depend on validation performance.

The iTransformer configurations use two layers, width 64, eight attention heads, feed-forward width 128, and dropout 0.1. TimeMixer-Core uses two layers, width 16, feed-forward width 32, a moving-average window of 25, and dropout 0.1. The gate and router are two-layer MLPs with hidden width 16, GELU activation, and no dropout; their final-layer weights are initialized to zero and biases to -2.

Table 4: Training configurations.

Dataset	L	S	Batch sizes	η: $H = 96, 192$	η: $H = 336, 720$
ETTh1	720	3	32 / 16 / 32	0.01	0.01
ETTh2	336	3	128 / 32 / 64	0.01	0.01
ETTm1	720	2	32 / 16 / 32	0.01	0.01
ETTm2	336	2	32 / 32 / 64	0.01	0.01
Weather	336	3	32 / 16 / 32	0.001	0.005
Electricity	336	3	16 / 4 / 8	0.01	0.01
Traffic	336	3	16 / 2 / 4	0.005	0.005
Solar-Energy	192	2	32 / 8 / 16	0.01	0.01

Table 4 lists the history length L, coarse-branch count S ($S + 1$ scales in total), batch sizes in DLinear / iTransformer / TimeMixer-Core order, and learning rate η shared by both stages. Vanilla uses the corresponding batch size, learning rate, and Stage-I budget. For both fusion heads, $\alpha = 1$, $\lambda_r = \lambda_u = 0.1$, and $m_r = 0$. The utility-label margin is zero. RC additionally uses $\rho = 0.1$ and $\lambda_p = 0.01$, except where explicitly varied. Training uses float32 without mixed precision. Scale-level normalization has learnable affine parameters, except on Solar-Energy where this wrapper normalization is disabled. SF retains its own normalization of the concatenated history and forecast on every dataset.

Prediction-summary features and numerical constants. For one sample and channel, let U denote a candidate, $D = U - B$, and $r_B = \max(\mathrm{RMS}(B), \epsilon)$. The eight-dimensional feature map is

$$\phi(B, U) = \left[\frac{\mu(B)}{r_B}, \frac{\sigma(B)}{r_B}, \frac{\mu(U)}{r_B}, \frac{\sigma(U)}{r_B}, \frac{\mu(D)}{r_B}, \frac{\sigma(D)}{r_B}, \frac{\mu(|D|)}{r_B}, \kappa(B, U) \right], \tag{33}$$

where μ and σ are the temporal mean and population standard deviation, and

$$\kappa(B, U) = \mathrm{clip}_{[-1,1]} \left(\frac{\mu(B \odot U)}{\max(\mathrm{RMS}(B)\,\mathrm{RMS}(U), \epsilon)} \right). \tag{34}$$

Safe-linear gates and RC routers use $U = A_s$; RC participation gates use $U = C_s$. Features and resulting weights are per sample and channel, and the weights are broadcast over the horizon. The RMS and feature floors use float32 machine epsilon, $\epsilon \approx 1.19 \times 10^{-7}$. The relative-risk floor is $\epsilon_r = 10^{-8}$, and probabilities used in binary cross-entropy are clipped to $[10^{-6}, 1 - 10^{-6}]$.

C.2 Candidate-quality supervision and projection regularization

Candidate-quality supervision. For sample i, coarse scale s, and channel c, define the deterministic candidate and base losses

$$\ell^A_{i,s,c} = \frac{1}{H} \sum_{t=1}^{H} (A_{i,s,t,c} - Y_{i,t,c})^2, \qquad \ell^B_{i,c} = \frac{1}{H} \sum_{t=1}^{H} (B_{i,t,c} - Y_{i,t,c})^2. \tag{35}$$

The binary target is

$$u_{i,s,c} = \mathbf{1}\{\ell^A_{i,s,c} < \ell^B_{i,c}\}, \tag{36}$$

and the auxiliary loss is

$$\mathcal{L}_u = -\frac{1}{NSC} \sum_{i=1}^{N} \sum_{s=1}^{S} \sum_{c=1}^{C} \left[u_{i,s,c} \log g_{i,s,c} + (1 - u_{i,s,c}) \log(1 - g_{i,s,c}) \right]. \tag{37}$$

Choice and interpretation of the quality reference. Frozen Stage-I predictors make A_s and its quality label fixed for a given sample. Both heads use this reference so that target construction does not depend on the evolving residual projector and router. In RC-ANSF, the gate input uses C_s, whose accuracy can differ from that of A_s even under a bounded correction. Thus, $\mathcal{L}_u$ supplies a stable proxy for the underlying coarse forecast, not a label for the adapted candidate's accuracy. The final-prediction MSE directly supervises the fused output. A fixed target does not isolate gradients: $\mathcal{L}_u$ can backpropagate through the gate's input into the projector and router. Nor does standalone candidate accuracy measure marginal contribution, since errors may cancel after fusion.

Projection regularization. For RC-ANSF, the applied residual is penalized by

$$\mathcal{L}_p = \frac{1}{NS} \sum_{i=1}^{N} \sum_{s=1}^{S} \frac{\mathrm{mean}_{t,c} \, |\pi_{i,s,c} \Delta_{i,s,t,c}|}{\max(\mathrm{mean}_{t,c} \, |A_{i,s,t,c}|, \epsilon_p)}. \tag{38}$$

This is a soft preference for smaller deviations, whereas Equation 9 imposes a hard amplitude limit. Its mean-absolute-magnitude denominator differs from the channel-wise RMS reference used in the hard bound. We use ϵ_p equal to machine epsilon for the prediction dtype. The reference component and radius configurations use $\lambda_p = 0.01$; the core component study does not separately remove this regularizer.

D Full primary results and aggregation

The full primary table reports means and sample standard deviations over the repeated runs specified in Appendix B for each backbone, dataset, horizon, method, and metric. The standard deviation describes the variation across the three runs at a fixed setting; it is not variation across horizons. The full-horizon format complements the dataset-level averages in the main paper.

Aggregation rules. Let $e_{k,m,r}$ be an error metric for setting k, method m, and run r. The primary setting-level mean is $\bar{e}_{k,m} = \frac{1}{3} \sum_{r=1}^{3} e_{k,m,r}$, and its sample standard deviation is

$$\mathrm{std}_{k,m} = \sqrt{\frac{1}{2} \sum_{r=1}^{3} (e_{k,m,r} - \bar{e}_{k,m})^2}. \tag{39}$$

Dataset-level main-table entries average the four horizon means within a backbone. Macro errors average $\bar{e}_{k,m}$ over all included settings. The reported average relative MSE reduction from Safe-linear to RC-ANSF is

$$\frac{100}{K} \sum_{k=1}^{K} \frac{\bar{e}_{k,\mathrm{SL}} - \bar{e}_{k,\mathrm{RC}}}{\bar{e}_{k,\mathrm{SL}}}, \tag{40}$$

where e denotes MSE and $K = 96$ for the primary study. This average of matched ratios differs from the relative change between two macro MSEs. Primary improvement counts compare unrounded run-averaged metrics within each setting. Because settings share datasets and forecasting horizons, these counts should not be treated as independent statistical trials.

Table 5: Full forecasting results: mean MSE and MAE.

H	Method	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Traffic		Solar-Energy	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DLinear																	
96	V	0.5960	0.5335	0.3081	0.3617	0.4882	0.4702	0.2132	0.2995	0.1769	0.2259	0.1677	0.2743	0.4420	0.3291	0.2451	0.3376
	SF	0.3991	0.4225	0.2981	0.3552	0.3799	0.4073	0.1792	0.2712	0.1734	0.2297	0.1506	0.2528	0.4148	0.2915	0.2347	0.3218
	S	0.4324	0.4414	0.2881	0.3454	0.3606	0.3906	0.1772	0.2657	0.1744	0.2238	0.1574	0.2632	0.4333	0.3207	0.2366	0.3248
	RC	0.4191	0.4326	0.2869	0.3493	0.3466	0.3805	0.1774	0.2681	0.1730	0.2239	0.1538	0.2585	0.4289	0.3144	0.2313	0.3160
192	V	0.6013	0.5364	0.3678	0.4016	0.5081	0.4811	0.2832	0.3437	0.2185	0.2612	0.1814	0.2869	0.4638	0.3445	0.2833	0.3522
	SF	0.4311	0.4370	0.3865	0.4163	0.4008	0.4142	0.2296	0.3045	0.2129	0.2670	0.1968	0.3030	0.4290	0.3008	0.2727	0.3386
	S	0.4717	0.4644	0.3456	0.3869	0.3969	0.4116	0.2405	0.3104	0.2168	0.2593	0.1729	0.2778	0.4497	0.3305	0.2716	0.3393
	RC	0.4586	0.4551	0.3431	0.3908	0.3798	0.4002	0.2422	0.3173	0.2136	0.2588	0.1691	0.2729	0.4439	0.3223	0.2654	0.3293
336	V	0.6465	0.5596	0.4068	0.4338	0.5452	0.5015	0.3420	0.3780	0.2759	0.3035	0.2023	0.3044	0.4746	0.3477	0.3244	0.3935
	SF	0.4746	0.4675	0.4128	0.4373	0.4149	0.4246	0.2917	0.3450	0.2590	0.3042	0.2086	0.3113	0.4474	0.3160	0.3121	0.3737
	S	0.4996	0.4841	0.3786	0.4167	0.4217	0.4267	0.3004	0.3477	0.2654	0.2951	0.1916	0.2940	0.4608	0.3337	0.3099	0.3705
	RC	0.4878	0.4771	0.3712	0.4155	0.4033	0.4142	0.3015	0.3538	0.2602	0.2931	0.1867	0.2888	0.4552	0.3261	0.3019	0.3581
720	V	0.7265	0.6054	0.4388	0.4586	0.6423	0.5474	0.4529	0.4388	0.3438	0.3514	0.2464	0.3400	0.5068	0.3661	0.3229	0.3877
	SF	0.5358	0.5259	0.5005	0.4928	0.4909	0.4723	0.4307	0.4322	0.3325	0.3676	0.2656	0.3611	0.4906	0.3502	0.3400	0.4054
	S	0.5248	0.5172	0.4220	0.4477	0.4822	0.4584	0.3981	0.4067	0.3324	0.3415	0.2318	0.3265	0.4921	0.3518	0.3082	0.3691
	RC	0.5139	0.5116	0.4074	0.4436	0.4634	0.4467	0.3984	0.4143	0.3248	0.3397	0.2250	0.3200	0.4853	0.3430	0.3016	0.3578
iTransformer																	
96	V	0.4571	0.4677	0.3686	0.4108	0.3880	0.4181	0.2480	0.3303	0.1527	0.2053	0.1836	0.2922	0.4400	0.3317	0.2346	0.2948
	SF	0.4077	0.4340	0.3837	0.4253	0.3326	0.3663	0.1817	0.2787	0.1559	0.2086	0.1324	0.2308	0.3782	0.2781	0.1688	0.2278
	S	0.4308	0.4497	0.3484	0.3948	0.3671	0.4022	0.2349	0.3180	0.1489	0.2008	0.1763	0.2857	0.4397	0.3314	0.2339	0.2837
	RC	0.4274	0.4474	0.3485	0.3996	0.3616	0.3987	0.2351	0.3211	0.1482	0.2006	0.1732	0.2824	0.4355	0.3264	0.2269	0.2760
192	V	0.4897	0.4927	0.4081	0.4394	0.4002	0.4239	0.2912	0.3570	0.1968	0.2459	0.1972	0.3021	0.4556	0.3400	0.2704	0.3149
	SF	0.4789	0.4790	0.4577	0.4729	0.3872	0.3994	0.2929	0.3525	0.1956	0.2451	0.1561	0.2534	0.4144	0.2977	0.1887	0.2514
	S	0.4723	0.4782	0.3904	0.4231	0.3844	0.4109	0.2822	0.3449	0.1916	0.2406	0.2065	0.3127	0.4577	0.3387	0.2596	0.3083
	RC	0.4686	0.4749	0.3903	0.4281	0.3801	0.4085	0.2824	0.3512	0.1894	0.2392	0.1987	0.3050	0.4538	0.3342	0.2507	0.2977
336	V	0.5248	0.5101	0.4192	0.4507	0.4288	0.4390	0.3302	0.3757	0.2562	0.2919	0.2878	0.3741	0.4657	0.3425	0.3010	0.3247
	SF	0.4922	0.4937	0.9245	0.6294	0.4291	0.4304	0.5423	0.4830	0.2508	0.2899	0.1757	0.2758	0.4206	0.2982	0.2042	0.2621
	S	0.5000	0.4966	0.4095	0.4438	0.4128	0.4289	0.3270	0.3707	0.2453	0.2837	0.2071	0.3105	0.4731	0.3450	0.2710	0.3068
	RC	0.4949	0.4932	0.4043	0.4450	0.4088	0.4266	0.3258	0.3774	0.2412	0.2813	0.2038	0.3077	0.4678	0.3391	0.2677	0.3024
720	V	0.6129	0.5656	0.4615	0.4775	0.5029	0.4764	0.4187	0.4231	0.3309	0.3431	0.2857	0.3704	0.4881	0.3506	0.3122	0.3377
	SF	0.6586	0.6028	0.5079	0.5110	0.5373	0.4952	0.5416	0.5083	0.3162	0.3440	0.2052	0.3006	0.4617	0.3215	0.2040	0.2587
	S	0.5369	0.5263	0.4516	0.4685	0.4604	0.4521	0.4091	0.4167	0.3152	0.3323	0.2696	0.3634	0.4939	0.3537	0.2808	0.3185
	RC	0.5367	0.5269	0.4382	0.4654	0.4560	0.4494	0.4064	0.4239	0.3089	0.3293	0.2603	0.3562	0.4899	0.3493	0.2768	0.3133
TimeMixer-Core																	
96	V	0.3718	0.4020	0.2791	0.3447	0.3083	0.3554	0.1650	0.2537	0.1545	0.2021	0.1423	0.2409	0.4056	0.2856	0.2003	0.2870
	SF	0.3725	0.4036	0.2783	0.3425	0.3079	0.3527	0.1645	0.2560	0.1509	0.2113	0.1414	0.2412	0.4064	0.2827	0.1940	0.2665
	S	0.3726	0.4022	0.2739	0.3385	0.3037	0.3516	0.1655	0.2539	0.1520	0.1996	0.1431	0.2430	0.4087	0.2879	0.1978	0.2788
	RC	0.3715	0.4009	0.2728	0.3393	0.3011	0.3491	0.1653	0.2570	0.1509	0.1994	0.1427	0.2429	0.4077	0.2862	0.1942	0.2722
192	V	0.4065	0.4239	0.3466	0.3865	0.3305	0.3712	0.2203	0.2920	0.1960	0.2396	0.1581	0.2556	0.4213	0.2926	0.2188	0.2917
	SF	0.4141	0.4322	0.3598	0.3964	0.3336	0.3700	0.2193	0.2952	0.1927	0.2548	0.1552	0.2525	0.4195	0.2889	0.2139	0.2718
	S	0.4112	0.4267	0.3437	0.3875	0.3316	0.3690	0.2202	0.2915	0.1941	0.2384	0.1590	0.2577	0.4237	0.2958	0.2110	0.2849
	RC	0.4109	0.4259	0.3435	0.3897	0.3291	0.3666	0.2211	0.2979	0.1916	0.2380	0.1585	0.2575	0.4225	0.2940	0.2069	0.2756
336	V	0.4382	0.4463	0.3762	0.4144	0.3570	0.3895	0.2792	0.3303	0.2473	0.2785	0.1728	0.2683	0.4351	0.2993	0.2325	0.3084
	SF	0.4489	0.4590	0.3992	0.4268	0.3639	0.3911	0.3186	0.3689	0.2431	0.2996	0.1707	0.2697	0.4331	0.3002	0.2227	0.2815
	S	0.4405	0.4449	0.3679	0.4118	0.3527	0.3867	0.2756	0.3270	0.2465	0.2779	0.1756	0.2742	0.4367	0.3026	0.2209	0.2973
	RC	0.4386	0.4437	0.3631	0.4101	0.3507	0.3847	0.2762	0.3362	0.2421	0.2765	0.1747	0.2737	0.4355	0.3007	0.2174	0.2861
720	V	0.4712	0.4821	0.4338	0.4624	0.4052	0.4179	0.3677	0.3890	0.3201	0.3290	0.2139	0.3032	0.4589	0.3114	0.2299	0.2925
	SF	0.5345	0.5295	0.5321	0.5126	0.4213	0.4290	0.5327	0.4857	0.3132	0.3502	0.2071	0.3067	0.4630	0.3196	0.2359	0.3004
	S	0.4601	0.4746	0.4192	0.4515	0.3951	0.4153	0.3668	0.3864	0.3195	0.3304	0.2128	0.3028	0.4639	0.3168	0.2325	0.3050
	RC	0.4612	0.4760	0.3999	0.4375	0.3945	0.4143	0.3629	0.3926	0.3130	0.3294	0.2112	0.3022	0.4621	0.3144	0.2314	0.3012

V: Vanilla; SF: Scaleformer; S: Safe-linear; RC: RC-ANSF.

Table 6: Full forecasting results: sample standard deviations of MSE and MAE.

H	Method	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Traffic		Solar-Energy	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DLinear																	
96	V	0.0243	0.0116	0.0028	0.0023	0.0427	0.0212	0.0026	0.0002	0.0015	0.0012	0.0023	0.0021	0.0023	0.0032	0.0153	0.0231
	SF	0.0021	0.0031	0.0061	0.0043	0.0371	0.0268	0.0021	0.0005	0.0020	0.0018	0.0025	0.0026	0.0007	0.0018	0.0022	0.0049
	S	0.0036	0.0043	0.0018	0.0004	0.0018	0.0020	0.0035	0.0049	0.0012	0.0010	0.0019	0.0023	0.0021	0.0026	0.0109	0.0203
	RC	0.0036	0.0047	0.0018	0.0006	0.0017	0.0025	0.0015	0.0025	0.0011	0.0009	0.0018	0.0022	0.0020	0.0024	0.0093	0.0190
192	V	0.0259	0.0116	0.0089	0.0042	0.0167	0.0084	0.0055	0.0013	0.0015	0.0009	0.0028	0.0052	0.0034	0.0055	0.0012	0.0097
	SF	0.0061	0.0052	0.0093	0.0057	0.0144	0.0093	0.0035	0.0064	0.0018	0.0021	0.0335	0.0338	0.0004	0.0008	0.0061	0.0144
	S	0.0017	0.0008	0.0079	0.0032	0.0153	0.0090	0.0031	0.0048	0.0010	0.0007	0.0024	0.0042	0.0032	0.0043	0.0043	0.0129
	RC	0.0027	0.0016	0.0095	0.0046	0.0123	0.0081	0.0021	0.0051	0.0007	0.0006	0.0022	0.0039	0.0024	0.0033	0.0041	0.0121
336	V	0.0098	0.0035	0.0115	0.0053	0.0121	0.0070	0.0167	0.0069	0.0034	0.0029	0.0054	0.0069	0.0005	0.0016	0.0038	0.0122
	SF	0.0052	0.0045	0.0314	0.0208	0.0054	0.0043	0.0033	0.0034	0.0009	0.0025	0.0215	0.0204	0.0013	0.0024	0.0048	0.0086
	S	0.0056	0.0042	0.0074	0.0036	0.0083	0.0064	0.0069	0.0034	0.0011	0.0010	0.0032	0.0044	0.0001	0.0008	0.0037	0.0041
	RC	0.0051	0.0037	0.0068	0.0036	0.0070	0.0060	0.0067	0.0024	0.0019	0.0009	0.0025	0.0036	0.0006	0.0002	0.0015	0.0036
720	V	0.0055	0.0006	0.0018	0.0008	0.0186	0.0083	0.0158	0.0078	0.0016	0.0011	0.0006	0.0014	0.0016	0.0022	0.0040	0.0063
	SF	0.0164	0.0083	0.0554	0.0179	0.0048	0.0048	0.0224	0.0142	0.0049	0.0058	0.0036	0.0037	0.0002	0.0012	0.0033	0.0071
	S	0.0119	0.0072	0.0029	0.0018	0.0080	0.0033	0.0067	0.0034	0.0009	0.0005	0.0017	0.0016	0.0019	0.0020	0.0057	0.0118
	RC	0.0119	0.0077	0.0029	0.0024	0.0045	0.0016	0.0030	0.0017	0.0006	0.0005	0.0016	0.0013	0.0018	0.0021	0.0060	0.0121
iTransformer																	
96	V	0.0024	0.0014	0.0107	0.0060	0.0116	0.0079	0.0088	0.0055	0.0006	0.0008	0.0015	0.0021	0.0052	0.0055	0.0075	0.0022
	SF	0.0037	0.0045	0.0372	0.0209	0.0085	0.0076	0.0079	0.0089	0.0008	0.0014	0.0008	0.0005	0.0027	0.0018	0.0024	0.0067
	S	0.0028	0.0031	0.0010	0.0019	0.0031	0.0015	0.0029	0.0006	0.0013	0.0019	0.0040	0.0036	0.0108	0.0118	0.0123	0.0091
	RC	0.0022	0.0028	0.0019	0.0026	0.0022	0.0010	0.0032	0.0012	0.0014	0.0018	0.0036	0.0032	0.0093	0.0102	0.0084	0.0055
192	V	0.0222	0.0139	0.0061	0.0068	0.0083	0.0040	0.0066	0.0064	0.0020	0.0017	0.0068	0.0069	0.0036	0.0031	0.0340	0.0116
	SF	0.0518	0.0267	0.0238	0.0163	0.0179	0.0018	0.0413	0.0231	0.0016	0.0013	0.0005	0.0005	0.0009	0.0018	0.0048	0.0074
	S	0.0049	0.0026	0.0026	0.0023	0.0024	0.0008	0.0073	0.0051	0.0005	0.0002	0.0203	0.0221	0.0019	0.0005	0.0172	0.0130
	RC	0.0065	0.0035	0.0011	0.0019	0.0018	0.0010	0.0063	0.0048	0.0005	0.0005	0.0154	0.0171	0.0022	0.0007	0.0173	0.0124
336	V	0.0052	0.0017	0.0096	0.0075	0.0082	0.0042	0.0029	0.0025	0.0015	0.0016	0.0641	0.0488	0.0058	0.0052	0.0071	0.0108
	SF	0.0097	0.0062	0.4782	0.1132	0.0208	0.0075	0.0307	0.0296	0.0026	0.0007	0.0050	0.0061	0.0022	0.0009	0.0009	0.0021
	S	0.0109	0.0081	0.0029	0.0017	0.0072	0.0045	0.0020	0.0019	0.0018	0.0013	0.0007	0.0013	0.0066	0.0065	0.0124	0.0049
	RC	0.0073	0.0054	0.0046	0.0032	0.0068	0.0041	0.0013	0.0025	0.0015	0.0010	0.0011	0.0006	0.0063	0.0070	0.0151	0.0062
720	V	0.0788	0.0359	0.0070	0.0066	0.0705	0.0396	0.0013	0.0008	0.0013	0.0012	0.0598	0.0450	0.0072	0.0069	0.0091	0.0104
	SF	0.0449	0.0313	0.0294	0.0157	0.0559	0.0332	0.0449	0.0211	0.0021	0.0038	0.0082	0.0040	0.0061	0.0051	0.0030	0.0032
	S	0.0132	0.0066	0.0016	0.0007	0.0100	0.0079	0.0056	0.0034	0.0012	0.0013	0.0268	0.0247	0.0083	0.0078	0.0243	0.0136
	RC	0.0107	0.0055	0.0041	0.0020	0.0095	0.0074	0.0035	0.0018	0.0013	0.0015	0.0233	0.0221	0.0083	0.0080	0.0235	0.0129
TimeMixer-Core																	
96	V	0.0022	0.0007	0.0038	0.0064	0.0021	0.0023	0.0008	0.0007	0.0011	0.0012	0.0025	0.0051	0.0023	0.0035	0.0067	0.0106
	SF	0.0039	0.0042	0.0082	0.0051	0.0056	0.0047	0.0021	0.0031	0.0013	0.0037	0.0006	0.0016	0.0038	0.0048	0.0027	0.0053
	S	0.0028	0.0008	0.0022	0.0006	0.0032	0.0021	0.0015	0.0014	0.0004	0.0003	0.0010	0.0018	0.0033	0.0034	0.0092	0.0112
	RC	0.0030	0.0010	0.0027	0.0021	0.0022	0.0013	0.0014	0.0018	0.0004	0.0001	0.0010	0.0016	0.0031	0.0031	0.0074	0.0101
192	V	0.0012	0.0011	0.0096	0.0012	0.0022	0.0022	0.0008	0.0011	0.0011	0.0012	0.0023	0.0040	0.0013	0.0044	0.0086	0.0030
	SF	0.0052	0.0037	0.0130	0.0125	0.0038	0.0012	0.0040	0.0054	0.0012	0.0027	0.0017	0.0030	0.0012	0.0023	0.0060	0.0040
	S	0.0046	0.0039	0.0028	0.0013	0.0049	0.0016	0.0016	0.0002	0.0005	0.0004	0.0029	0.0057	0.0027	0.0009	0.0026	0.0031
	RC	0.0045	0.0039	0.0036	0.0029	0.0040	0.0014	0.0013	0.0009	0.0004	0.0006	0.0030	0.0057	0.0028	0.0012	0.0035	0.0069
336	V	0.0130	0.0129	0.0048	0.0035	0.0026	0.0018	0.0029	0.0015	0.0006	0.0006	0.0008	0.0017	0.0027	0.0026	0.0072	0.0115
	SF	0.0181	0.0154	0.0510	0.0358	0.0083	0.0040	0.0524	0.0340	0.0033	0.0027	0.0015	0.0032	0.0009	0.0017	0.0047	0.0070
	S	0.0064	0.0032	0.0013	0.0032	0.0013	0.0014	0.0013	0.0007	0.0002	0.0001	0.0023	0.0043	0.0019	0.0009	0.0042	0.0064
	RC	0.0056	0.0031	0.0019	0.0011	0.0005	0.0013	0.0005	0.0002	0.0001	0.0007	0.0023	0.0043	0.0018	0.0010	0.0048	0.0047
720	V	0.0039	0.0040	0.0106	0.0075	0.0083	0.0024	0.0022	0.0014	0.0005	0.0009	0.0031	0.0043	0.0031	0.0016	0.0040	0.0033
	SF	0.0346	0.0289	0.0742	0.0376	0.0077	0.0046	0.1336	0.0547	0.0040	0.0016	0.0051	0.0042	0.0061	0.0064	0.0140	0.0136
	S	0.0065	0.0022	0.0071	0.0005	0.0026	0.0018	0.0032	0.0018	0.0010	0.0010	0.0016	0.0027	0.0020	0.0021	0.0071	0.0092
	RC	0.0069	0.0026	0.0146	0.0086	0.0026	0.0019	0.0076	0.0036	0.0005	0.0014	0.0015	0.0028	0.0018	0.0018	0.0073	0.0091

E FULL SHARED-PREDICTOR FUSION COMPARISON

All seven heads use the same native-scale forecasting checkpoint within a setting. Table 7 summarizes their implementations; B is the fine-scale forecast and A_s is the linearly aligned coarse forecast.

Table 7: Shared-predictor fusion methods.

Method	Implementation
Base	Return the finest-scale prediction B without using coarse forecasts.
Mean	Equally average B and all linearly aligned coarse forecasts A_s.
Global	Learn one softmax-normalized weight per scale and form a weighted sum of the aligned forecasts. Weights are shared across samples, time steps, and channels.
MLP	Stack the aligned differences $A_s - B$ at each time step and channel, pass them through a shared two-layer MLP, and add its scalar output to B. The final layer is initialized to zero.
Context	Use the mean, standard deviation, and last value of the base and native coarse forecasts to predict channel-wise gain and bias corrections to B. Coarse forecasts are not temporally aligned.
Safe-linear	Learn sample–channel participation gates for A_s and apply the anchored correction in Equation 6.
RC (RC-ANSF)	Add a bounded, routed residual to each A_s with $\rho = 0.1$, then apply participation-gated fusion as in Equation 13.

Results cover both DLinear and iTransformer on eight datasets and four horizons. Main-table entries average the four horizons within a dataset and backbone; the full table retains each horizon. Base is the fine branch of the shared forecasting checkpoint, not the independently trained Vanilla system. This comparison evaluates fusion choices, with variability across runs addressed separately by the primary study.

Table 8: Full shared-predictor fusion comparison.

H	Fusion	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Traffic		Solar-Energy	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DLinear																	
96	Base	0.6036	0.5339	0.3097	0.3632	0.5373	0.4944	0.2108	0.2994	0.1756	0.2247	0.1651	0.2721	0.4394	0.3254	0.2363	0.3290
	Mean	0.4356	0.4402	0.2848	0.3428	0.3737	0.4015	0.1752	0.2628	0.1732	0.2230	0.1849	0.3027	0.4884	0.3607	0.2289	0.3162
	Global	0.4260	0.4359	0.2844	0.3425	0.3543	0.3858	0.1695	0.2567	0.1731	0.2230	0.1548	0.2598	0.4309	0.3161	0.2289	0.3157
	MLP	0.4253	0.4363	0.2861	0.3440	0.3544	0.3859	0.1696	0.2565	0.1734	0.2252	0.1548	0.2596	0.4299	0.3131	0.2275	0.3138
	Context	0.4944	0.4758	0.3064	0.3606	0.4228	0.4322	0.1825	0.2745	0.1729	0.2289	0.1546	0.2602	0.4331	0.3215	0.2254	0.3142
	Safe-linear	0.4310	0.4380	0.2874	0.3450	0.3595	0.3913	0.1732	0.2601	0.1730	0.2227	0.1555	0.2611	0.4316	0.3182	0.2284	0.3140
	RC	0.4169	0.4287	0.2868	0.3499	0.3464	0.3817	0.1758	0.2652	0.1717	0.2229	0.1521	0.2566	0.4275	0.3124	0.2236	0.3028
192	Base	0.5885	0.5285	0.3621	0.3987	0.4924	0.4736	0.2769	0.3424	0.2195	0.2614	0.1819	0.2884	0.4637	0.3452	0.2823	0.3454
	Mean	0.4735	0.4648	0.3386	0.3843	0.3882	0.4068	0.2465	0.3177	0.2177	0.2596	0.2043	0.3204	0.4984	0.3633	0.2680	0.3290
	Global	0.4737	0.4649	0.3394	0.3845	0.3670	0.3921	0.2371	0.3089	0.2175	0.2595	0.1713	0.2757	0.4468	0.3239	0.2669	0.3282
	MLP	0.4742	0.4647	0.3392	0.3838	0.3672	0.3924	0.2375	0.3087	0.2175	0.2629	0.1699	0.2728	0.4438	0.3183	0.2620	0.3192
	Context	0.5001	0.4819	0.3546	0.3919	0.4187	0.4302	0.2479	0.3223	0.2130	0.2628	0.1754	0.2801	0.4533	0.3371	0.2686	0.3347
	Safe-linear	0.4730	0.4642	0.3401	0.3847	0.3800	0.4012	0.2441	0.3157	0.2175	0.2595	0.1727	0.2782	0.4481	0.3291	0.2669	0.3274
	RC	0.4615	0.4562	0.3369	0.3882	0.3660	0.3909	0.2439	0.3213	0.2139	0.2587	0.1689	0.2731	0.4420	0.3204	0.2613	0.3191
336	Base	0.6574	0.5635	0.3938	0.4278	0.5487	0.5041	0.3612	0.3860	0.2794	0.3063	0.2048	0.3079	0.4748	0.3488	0.3266	0.4036
	Mean	0.5044	0.4871	0.3699	0.4125	0.4375	0.4393	0.3109	0.3524	0.2669	0.2967	0.2239	0.3352	0.5123	0.3685	0.3117	0.3736
	Global	0.5038	0.4870	0.3711	0.4128	0.4216	0.4288	0.2925	0.3404	0.2657	0.2953	0.1948	0.2980	0.4588	0.3286	0.3108	0.3695
	MLP	0.5017	0.4847	0.3730	0.4144	0.4215	0.4289	0.2926	0.3453	0.2660	0.2978	0.1937	0.2965	0.4580	0.3243	0.3041	0.3590
	Context	0.5537	0.5191	0.3777	0.4149	0.4658	0.4580	0.3181	0.3654	0.2585	0.3011	0.1976	0.3042	0.4676	0.3450	0.3074	0.3736
	Safe-linear	0.5032	0.4863	0.3705	0.4128	0.4276	0.4327	0.3061	0.3500	0.2656	0.2959	0.1946	0.2981	0.4607	0.3343	0.3098	0.3698
	RC	0.4910	0.4791	0.3634	0.4115	0.4091	0.4202	0.3075	0.3554	0.2597	0.2934	0.1894	0.2927	0.4547	0.3259	0.3015	0.3567
720	Base	0.7293	0.6056	0.4382	0.4583	0.6638	0.5570	0.4576	0.4408	0.3431	0.3511	0.2460	0.3393	0.5075	0.3673	0.3269	0.3950
	Mean	0.5132	0.5095	0.4180	0.4449	0.5123	0.4757	0.4044	0.4095	0.3333	0.3415	0.2563	0.3551	0.5463	0.3867	0.3171	0.3852
	Global	0.5078	0.5073	0.4174	0.4444	0.4744	0.4512	0.3967	0.4055	0.3320	0.3401	0.2308	0.3240	0.4909	0.3476	0.3167	0.3852
	MLP	0.5150	0.5096	0.4220	0.4483	0.4736	0.4508	0.3970	0.4084	0.3322	0.3411	0.2300	0.3236	0.4913	0.3444	0.3117	0.3772
	Context	0.6205	0.5774	0.4212	0.4458	0.5190	0.4853	0.4257	0.4254	0.3237	0.3491	0.2308	0.3315	0.5037	0.3641	0.3039	0.3643
	Safe-linear	0.5121	0.5105	0.4188	0.4456	0.4914	0.4622	0.4018	0.4076	0.3324	0.3411	0.2300	0.3249	0.4943	0.3540	0.3147	0.3827
	RC	0.5014	0.5047	0.4048	0.4424	0.4686	0.4484	0.4003	0.4154	0.3245	0.3393	0.2232	0.3187	0.4873	0.3454	0.3084	0.3716
iTransformer																	
96	Base	0.4488	0.4591	0.3692	0.4095	0.3803	0.4129	0.2423	0.3273	0.1543	0.2078	0.1769	0.2860	0.4357	0.3292	0.2404	0.2897
	Mean	0.4403	0.4546	0.3455	0.3906	0.3631	0.4017	0.2314	0.3173	0.1495	0.2009	0.2002	0.3189	0.4941	0.3638	0.2273	0.2794
	Global	0.4217	0.4457	0.3430	0.3884	0.3637	0.4017	0.2319	0.3177	0.1496	0.2010	0.1720	0.2818	0.4325	0.3242	0.2274	0.2796
	MLP	0.4215	0.4472	0.3426	0.3889	0.3635	0.4017	0.2324	0.3201	0.1496	0.2078	0.1709	0.2792	0.4311	0.3215	0.2255	0.2768
	Context	0.4384	0.4542	0.3664	0.4072	0.3740	0.4094	0.2360	0.3224	0.1500	0.2054	0.1703	0.2781	0.4308	0.3224	0.2191	0.2741
	Safe-linear	0.4282	0.4474	0.3488	0.3935	0.3637	0.4017	0.2316	0.3174	0.1498	0.2017	0.1719	0.2816	0.4330	0.3241	0.2271	0.2781
	RC	0.4256	0.4459	0.3505	0.4002	0.3601	0.3990	0.2321	0.3209	0.1491	0.2012	0.1695	0.2791	0.4303	0.3211	0.2213	0.2729
192	Base	0.5268	0.5165	0.4040	0.4370	0.4219	0.4341	0.2937	0.3580	0.1982	0.2472	0.3142	0.3944	0.4637	0.3465	0.2655	0.3250
	Mean	0.4733	0.4773	0.3875	0.4210	0.3857	0.4108	0.2746	0.3390	0.1928	0.2406	0.2470	0.3575	0.5103	0.3688	0.2520	0.3065
	Global	0.4579	0.4693	0.3897	0.4200	0.3818	0.4090	0.2752	0.3402	0.1932	0.2410	0.2117	0.3188	0.4573	0.3372	0.2519	0.3052
	MLP	0.4596	0.4705	0.3890	0.4227	0.3824	0.4097	0.2742	0.3396	0.1922	0.2485	0.2026	0.3103	0.4546	0.3339	0.2476	0.3054
	Context	0.4818	0.4814	0.4003	0.4328	0.4120	0.4300	0.2880	0.3528	0.1924	0.2470	0.2925	0.3799	0.4563	0.3393	0.2479	0.3170
	Safe-linear	0.4681	0.4754	0.3881	0.4220	0.3869	0.4119	0.2739	0.3391	0.1920	0.2406	0.2221	0.3287	0.4587	0.3393	0.2530	0.3074
	RC	0.4629	0.4715	0.3892	0.4269	0.3821	0.4094	0.2758	0.3477	0.1894	0.2388	0.2113	0.3183	0.4549	0.3349	0.2431	0.2941
336	Base	0.5197	0.4990	0.4536	0.4758	0.4632	0.4621	0.3388	0.3777	0.2525	0.2907	0.2160	0.3186	0.4725	0.3469	0.3161	0.3361
	Mean	0.5196	0.5122	0.4146	0.4468	0.4231	0.4349	0.3295	0.3716	0.2468	0.2838	0.2426	0.3497	0.5179	0.3711	0.2692	0.3063
	Global	0.5014	0.4959	0.4097	0.4428	0.4148	0.4312	0.3300	0.3720	0.2476	0.2849	0.2049	0.3079	0.4647	0.3366	0.2563	0.3019
	MLP	0.5068	0.4958	0.4107	0.4460	0.4151	0.4314	0.3289	0.3735	0.2469	0.2905	0.2028	0.3073	0.4645	0.3361	0.2540	0.3023
	Context	0.5076	0.4943	0.4380	0.4628	0.4362	0.4454	0.3320	0.3741	0.2479	0.2917	0.2078	0.3129	0.4670	0.3401	0.2529	0.3172
	Safe-linear	0.4964	0.4919	0.4115	0.4448	0.4206	0.4339	0.3279	0.3708	0.2442	0.2827	0.2078	0.3113	0.4661	0.3394	0.2697	0.3067
	RC	0.4936	0.4899	0.4088	0.4484	0.4165	0.4312	0.3263	0.3757	0.2405	0.2801	0.2045	0.3084	0.4618	0.3342	0.2640	0.3009
720	Base	0.5970	0.5589	0.4509	0.4673	0.5513	0.5106	0.4102	0.4198	0.3256	0.3396	0.2507	0.3441	0.4904	0.3518	0.2804	0.3209
	Mean	0.5625	0.5484	0.4591	0.4722	0.4755	0.4630	0.4044	0.4135	0.3177	0.3319	0.2827	0.3798	0.5500	0.3838	0.2555	0.3039
	Global	0.5502	0.5398	0.4533	0.4684	0.4632	0.4539	0.4047	0.4136	0.3172	0.3313	0.2421	0.3370	0.4866	0.3451	0.2581	0.3050
	MLP	0.5675	0.5466	0.4596	0.4757	0.4630	0.4537	0.4045	0.4144	0.3172	0.3329	0.2387	0.3346	0.4870	0.3435	0.2546	0.3059
	Context	0.6225	0.5665	0.4424	0.4620	0.5052	0.4811	0.4050	0.4170	0.3138	0.3377	0.2418	0.3394	0.4901	0.3497	0.2614	0.3155
	Safe-linear	0.5515	0.5338	0.4528	0.4682	0.4710	0.4605	0.4028	0.4129	0.3158	0.3316	0.2436	0.3389	0.4883	0.3476	0.2578	0.3031
	RC	0.5477	0.5332	0.4429	0.4677	0.4664	0.4575	0.4024	0.4220	0.3099	0.3290	0.2400	0.3362	0.4849	0.3441	0.2522	0.2985

Base: finest-scale Stage-I branch; RC: RC-ANSF.

F FULL CORE COMPONENT ANALYSIS

The full core table reports RC-ANSF, No TR, No router, and No gate for the three backbones, eight datasets, and four horizons. The reference RC-ANSF configuration uses $\rho = 0.1$ and $\lambda_p = 0.01$. Table 9 specifies each implementation.

Table 9: Component variants.

Method	Implementation
RC (RC-ANSF)	Use the bounded residual, learned projection router, and learned participation gate with the complete Stage-II objective.
No TR	Replace $\Delta_s = \rho q_s \odot \tanh(Z_s/q_s)$ by $\Delta_s = Z_s$; retain the projector, router, and participation gate.
No router	Fix $\pi_s = 1$, retaining the bounded residual and learned participation gate.
No gate	Fix $g_s = 1$ and disable $\mathcal{L}_u$, retaining the bounded residual and projection router. At $\alpha = 1$, the base has zero direct mixture coefficient.

The interventions should be interpreted according to these changes. In particular, No gate changes participation, supervision, and base retention together. No TR compares the recorded residual parameterizations; attribution solely to the amplitude cap would additionally require matching their local optimization scale. The full results show both settings where the restrictions help and settings where a relaxed variant is more accurate. No router gives macro MSE 0.32449 versus 0.32453 for RC, so these results do not establish a separate accuracy benefit from the router. These task-level comparisons do not measure sample-wise test-risk tails.

Table 10: Full component ablations.

H	Variant	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Traffic		Solar-Energy	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DLinear																	
96	RC	0.4169	0.4287	0.2868	0.3499	0.3464	0.3817	0.1758	0.2652	0.1717	0.2229	0.1521	0.2566	0.4275	0.3124	0.2236	0.3028
	No TR	0.3806	0.4084	0.2977	0.3586	0.3302	0.3679	0.1724	0.2620	0.1649	0.2193	0.1431	0.2425	0.4140	0.2913	0.2153	0.2874
	No router	0.4175	0.4294	0.2868	0.3499	0.3465	0.3819	0.1762	0.2702	0.1717	0.2234	0.1519	0.2561	0.4276	0.3125	0.2237	0.3031
	No gate	0.4181	0.4287	0.2916	0.3572	0.3438	0.3788	0.1721	0.2618	0.1713	0.2238	0.1949	0.3113	0.5195	0.3689	0.2240	0.3035
192	RC	0.4615	0.4562	0.3369	0.3882	0.3660	0.3909	0.2439	0.3213	0.2139	0.2587	0.1689	0.2731	0.4420	0.3204	0.2613	0.3191
	No TR	0.4269	0.4352	0.3500	0.4000	0.3394	0.3726	0.2378	0.3103	0.2056	0.2572	0.1569	0.2562	0.4247	0.2946	0.2328	0.2915
	No router	0.4616	0.4562	0.3369	0.3882	0.3659	0.3908	0.2432	0.3203	0.2139	0.2586	0.1688	0.2731	0.4420	0.3205	0.2618	0.3195
	No gate	0.4693	0.4613	0.3426	0.3960	0.3608	0.3871	0.2429	0.3192	0.2137	0.2599	0.2103	0.3241	0.5281	0.3704	0.2592	0.3133
336	RC	0.4910	0.4791	0.3634	0.4115	0.4091	0.4202	0.3075	0.3554	0.2597	0.2934	0.1894	0.2927	0.4547	0.3259	0.3015	0.3567
	No TR	0.4529	0.4547	0.3913	0.4295	0.3783	0.3974	0.2951	0.3448	0.2516	0.3003	0.1717	0.2713	0.4382	0.3011	0.2570	0.3075
	No router	0.4910	0.4790	0.3635	0.4115	0.4094	0.4204	0.3071	0.3550	0.2600	0.2936	0.1891	0.2924	0.4547	0.3259	0.3014	0.3565
	No gate	0.4951	0.4824	0.3661	0.4153	0.4068	0.4192	0.3039	0.3576	0.2587	0.2949	0.2318	0.3427	0.5404	0.3746	0.2986	0.3480
720	RC	0.5014	0.5047	0.4048	0.4424	0.4686	0.4484	0.4003	0.4154	0.3245	0.3393	0.2232	0.3187	0.4873	0.3454	0.3084	0.3716
	No TR	0.5344	0.5293	0.6228	0.5487	0.4261	0.4277	0.4302	0.4365	0.3124	0.3464	0.2060	0.3037	0.4660	0.3165	0.2726	0.3110
	No router	0.5013	0.5046	0.4048	0.4425	0.4689	0.4486	0.4002	0.4154	0.3245	0.3392	0.2230	0.3186	0.4873	0.3455	0.3086	0.3721
	No gate	0.5024	0.5064	0.4087	0.4473	0.4661	0.4467	0.3980	0.4159	0.3217	0.3398	0.2603	0.3603	0.5719	0.3917	0.3049	0.3613
iTransformer																	
96	RC	0.4256	0.4459	0.3505	0.4002	0.3601	0.3990	0.2321	0.3209	0.1491	0.2012	0.1695	0.2791	0.4303	0.3211	0.2213	0.2729
	No TR	0.4138	0.4409	0.3582	0.4094	0.3558	0.3960	0.2311	0.3182	0.1489	0.2014	0.1623	0.2704	0.4178	0.3080	0.2181	0.2711
	No router	0.4255	0.4459	0.3507	0.3993	0.3598	0.3990	0.2322	0.3205	0.1491	0.2013	0.1695	0.2792	0.4303	0.3213	0.2207	0.2730
	No gate	0.4461	0.4576	0.3545	0.4064	0.3665	0.4046	0.2366	0.3238	0.1491	0.2018	0.2052	0.3228	0.5229	0.3725	0.2220	0.2756
192	RC	0.4629	0.4715	0.3892	0.4269	0.3821	0.4094	0.2758	0.3477	0.1894	0.2388	0.2113	0.3183	0.4549	0.3349	0.2431	0.2941
	No TR	0.4473	0.4611	0.4457	0.4710	0.3783	0.4085	0.2752	0.3437	0.1885	0.2401	0.1858	0.2956	0.4360	0.3185	0.2310	0.2829
	No router	0.4627	0.4714	0.3881	0.4253	0.3822	0.4096	0.2760	0.3478	0.1894	0.2388	0.2113	0.3184	0.4549	0.3349	0.2427	0.2942
	No gate	0.4688	0.4696	0.3967	0.4348	0.3788	0.4074	0.2853	0.3543	0.1892	0.2400	0.2279	0.3405	0.5327	0.3725	0.2369	0.2876
336	RC	0.4936	0.4899	0.4088	0.4484	0.4165	0.4312	0.3263	0.3757	0.2405	0.2801	0.2045	0.3084	0.4618	0.3342	0.2640	0.3009
	No TR	0.4959	0.4933	0.5190	0.5197	0.4150	0.4358	0.3375	0.3858	0.2409	0.2873	0.1954	0.3010	0.4472	0.3230	0.2506	0.2938
	No router	0.4936	0.4899	0.4089	0.4483	0.4165	0.4312	0.3262	0.3755	0.2405	0.2800	0.2043	0.3083	0.4618	0.3342	0.2639	0.3009
	No gate	0.5345	0.5213	0.4141	0.4525	0.4259	0.4378	0.3371	0.3906	0.2419	0.2811	0.2505	0.3565	0.5401	0.3756	0.2494	0.2924
720	RC	0.5477	0.5332	0.4429	0.4677	0.4664	0.4575	0.4024	0.4220	0.3099	0.3290	0.2400	0.3362	0.4849	0.3441	0.2522	0.2985
	No TR	0.5812	0.5568	0.9632	0.6883	0.4665	0.4628	0.4200	0.4397	0.3091	0.3411	0.2257	0.3281	0.4729	0.3355	0.2456	0.2948
	No router	0.5479	0.5331	0.4428	0.4676	0.4663	0.4574	0.4027	0.4225	0.3099	0.3290	0.2397	0.3361	0.4850	0.3443	0.2527	0.2985
	No gate	0.5709	0.5532	0.4513	0.4761	0.4705	0.4594	0.4171	0.4370	0.3080	0.3288	0.2893	0.3867	0.5720	0.3876	0.2490	0.3014
TimeMixer-Core																	
96	RC	0.3738	0.4010	0.2751	0.3406	0.3033	0.3505	0.1637	0.2550	0.1505	0.1995	0.1437	0.2445	0.4042	0.2848	0.1959	0.2678
	No TR	0.3678	0.3975	0.2799	0.3448	0.3045	0.3513	0.1637	0.2547	0.1487	0.2002	0.1419	0.2421	0.4003	0.2778	0.1943	0.2658
	No router	0.3744	0.4014	0.2752	0.3407	0.3033	0.3507	0.1643	0.2557	0.1504	0.1994	0.1436	0.2444	0.4040	0.2844	0.1952	0.2660
	No gate	0.4056	0.4215	0.2812	0.3474	0.3024	0.3494	0.1645	0.2556	0.1502	0.2003	0.2002	0.3182	0.5194	0.3644	0.1900	0.2593
192	RC	0.4140	0.4295	0.3440	0.3920	0.3267	0.3654	0.2216	0.2984	0.1918	0.2385	0.1551	0.2510	0.4257	0.2953	0.2085	0.2752
	No TR	0.4105	0.4301	0.3603	0.4037	0.3256	0.3639	0.2202	0.2946	0.1892	0.2419	0.1539	0.2506	0.4174	0.2836	0.2053	0.2760
	No router	0.4135	0.4293	0.3428	0.3907	0.3269	0.3659	0.2212	0.2975	0.1918	0.2385	0.1550	0.2508	0.4255	0.2950	0.2085	0.2751
	No gate	0.4496	0.4553	0.3582	0.4023	0.3267	0.3654	0.2234	0.2999	0.1921	0.2433	0.2151	0.3303	0.5428	0.3775	0.2108	0.2794
336	RC	0.4425	0.4435	0.3630	0.4099	0.3503	0.3835	0.2762	0.3361	0.2422	0.2767	0.1773	0.2785	0.4354	0.3018	0.2224	0.2808
	No TR	0.4484	0.4541	0.4158	0.4485	0.3514	0.3840	0.2775	0.3350	0.2396	0.2863	0.1708	0.2704	0.4306	0.2953	0.2169	0.2828
	No router	0.4426	0.4434	0.3630	0.4099	0.3498	0.3833	0.2759	0.3355	0.2422	0.2766	0.1773	0.2783	0.4351	0.3016	0.2226	0.2808
	No gate	0.4570	0.4570	0.3760	0.4216	0.3511	0.3845	0.2868	0.3475	0.2411	0.2791	0.2343	0.3475	0.5535	0.3798	0.2161	0.2740
720	RC	0.4635	0.4786	0.4060	0.4415	0.3973	0.4128	0.3544	0.3892	0.3136	0.3299	0.2129	0.3042	0.4614	0.3160	0.2397	0.3115
	No TR	0.5592	0.5488	0.7636	0.6143	0.4591	0.4659	0.3712	0.3993	0.3039	0.3371	0.2033	0.2990	0.4549	0.3060	0.2367	0.3107
	No router	0.4633	0.4785	0.4058	0.4413	0.3972	0.4127	0.3545	0.3893	0.3136	0.3299	0.2129	0.3041	0.4612	0.3158	0.2395	0.3114
	No gate	0.4962	0.5089	0.4067	0.4447	0.3999	0.4166	0.3582	0.3957	0.3112	0.3318	0.2784	0.3814	0.5802	0.3944	0.2885	0.4266

G FULL TRUST-RADIUS SENSITIVITY

The radius study evaluates $\rho \in \{0.05, 0.10, 0.25, 0.50, 1.00\}$ on the same 96 settings, using the matched frozen predictors and $\lambda_p = 0.01$. The $\rho = 0.1$ reference reuses the corresponding primary RC-ANSF fit; the other radii use separately fitted fusion heads. The full table reports MSE and MAE for each backbone, dataset, and horizon.

The sweep describes how performance responds to the available correction budget. It does not provide a test-set rule for choosing hyperparameters. A radius of one still bounds the residual by the RMS reference and is distinct from No TR. Safe-linear is a separately fitted deterministic-alignment head and is not labeled as an independently trained $\rho = 0$ RC configuration.

Table 11: Full trust-radius sensitivity.

H	ρ	ETTh1		ETTh2		ETTm1		ETTm2		Weather		Electricity		Traffic		Solar-Energy	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
DLinear																	
96	0.05	0.4236	0.4330	0.2842	0.3450	0.3528	0.3862	0.1740	0.2651	0.1721	0.2226	0.1539	0.2590	0.4295	0.3152	0.2256	0.3076
	0.10	0.4169	0.4287	0.2868	0.3499	0.3464	0.3817	0.1758	0.2652	0.1717	0.2229	0.1521	0.2566	0.4275	0.3124	0.2236	0.3028
	0.25	0.4025	0.4200	0.3065	0.3685	0.3358	0.3733	0.1747	0.2640	0.1708	0.2233	0.1487	0.2515	0.4230	0.3058	0.2204	0.2973
	0.50	0.3897	0.4122	0.3007	0.3622	0.3308	0.3691	0.1736	0.2625	0.1691	0.2224	0.1455	0.2466	0.4180	0.2982	0.2179	0.2934
	1.00	0.3864	0.4111	0.2967	0.3558	0.3295	0.3680	0.1739	0.2629	0.1681	0.2220	0.1444	0.2444	0.4165	0.2952	0.2136	0.2898
192	0.05	0.4671	0.4597	0.3353	0.3835	0.3711	0.3945	0.2430	0.3177	0.2154	0.2585	0.1703	0.2751	0.4452	0.3251	0.2641	0.3223
	0.10	0.4615	0.4562	0.3369	0.3882	0.3660	0.3909	0.2439	0.3213	0.2139	0.2587	0.1689	0.2731	0.4420	0.3204	0.2613	0.3191
	0.25	0.4485	0.4479	0.3645	0.4128	0.3495	0.3790	0.2434	0.3167	0.2124	0.2601	0.1646	0.2675	0.4347	0.3097	0.2592	0.3169
	0.50	0.4364	0.4404	0.3486	0.3985	0.3438	0.3752	0.2383	0.3110	0.2088	0.2590	0.1608	0.2620	0.4287	0.3011	0.2563	0.3126
	1.00	0.4297	0.4373	0.3515	0.4018	0.3437	0.3759	0.2375	0.3110	0.2055	0.2573	0.1588	0.2584	0.4267	0.2989	0.2488	0.3027
336	0.05	0.4965	0.4820	0.3642	0.4094	0.4175	0.4257	0.3035	0.3493	0.2624	0.2940	0.1922	0.2955	0.4576	0.3298	0.3054	0.3633
	0.10	0.4910	0.4791	0.3634	0.4115	0.4091	0.4202	0.3075	0.3554	0.2597	0.2934	0.1894	0.2927	0.4547	0.3259	0.3015	0.3567
	0.25	0.4822	0.4762	0.3975	0.4397	0.3919	0.4080	0.3093	0.3578	0.2564	0.2959	0.1837	0.2861	0.4473	0.3152	0.2943	0.3484
	0.50	0.4694	0.4654	0.4387	0.4652	0.3845	0.4021	0.3074	0.3559	0.2542	0.2960	0.1788	0.2807	0.4417	0.3072	0.2905	0.3407
	1.00	0.4675	0.4635	0.3938	0.4363	0.3853	0.4025	0.3022	0.3519	0.2536	0.2975	0.1764	0.2775	0.4407	0.3064	0.2777	0.3207
720	0.05	0.5068	0.5073	0.4076	0.4404	0.4783	0.4541	0.3954	0.4082	0.3283	0.3396	0.2265	0.3217	0.4905	0.3496	0.3114	0.3773
	0.10	0.5014	0.5047	0.4048	0.4424	0.4686	0.4484	0.4003	0.4154	0.3245	0.3393	0.2232	0.3187	0.4873	0.3454	0.3084	0.3716
	0.25	0.5000	0.5084	0.4427	0.4747	0.4543	0.4452	0.4255	0.4356	0.3195	0.3444	0.2166	0.3132	0.4805	0.3360	0.3030	0.3612
	0.50	0.5077	0.5143	0.6015	0.5587	0.4488	0.4445	0.4391	0.4410	0.3188	0.3507	0.2121	0.3102	0.4752	0.3283	0.2979	0.3489
	1.00	0.5233	0.5229	0.7319	0.6114	0.4492	0.4442	0.4387	0.4416	0.3177	0.3502	0.2123	0.3114	0.4724	0.3252	0.2903	0.3357
iTransformer																	
96	0.05	0.4269	0.4466	0.3481	0.3955	0.3618	0.4002	0.2315	0.3193	0.1491	0.2010	0.1705	0.2802	0.4315	0.3227	0.2222	0.2738
	0.10	0.4256	0.4459	0.3505	0.4002	0.3601	0.3990	0.2321	0.3209	0.1491	0.2012	0.1695	0.2791	0.4303	0.3211	0.2213	0.2729
	0.25	0.4219	0.4440	0.3596	0.4093	0.3575	0.3970	0.2316	0.3193	0.1491	0.2019	0.1670	0.2765	0.4269	0.3170	0.2211	0.2737
	0.50	0.4184	0.4422	0.3645	0.4139	0.3575	0.3969	0.2313	0.3192	0.1490	0.2015	0.1651	0.2739	0.4227	0.3119	0.2192	0.2734
	1.00	0.4173	0.4420	0.3655	0.4154	0.3567	0.3963	0.2325	0.3217	0.1489	0.2015	0.1639	0.2725	0.4201	0.3093	0.2212	0.2752
192	0.05	0.4653	0.4732	0.3869	0.4228	0.3842	0.4103	0.2730	0.3418	0.1904	0.2392	0.2164	0.3232	0.4569	0.3371	0.2472	0.2990
	0.10	0.4629	0.4715	0.3892	0.4269	0.3821	0.4094	0.2758	0.3477	0.1894	0.2388	0.2113	0.3183	0.4549	0.3349	0.2431	0.2941
	0.25	0.4560	0.4663	0.4098	0.4463	0.3786	0.4073	0.2775	0.3462	0.1887	0.2396	0.2007	0.3083	0.4489	0.3283	0.2388	0.2902
	0.50	0.4519	0.4628	0.4223	0.4555	0.3771	0.4068	0.2744	0.3444	0.1885	0.2402	0.1931	0.3014	0.4441	0.3237	0.2394	0.2907
	1.00	0.4521	0.4628	0.4346	0.4636	0.3768	0.4069	0.2764	0.3468	0.1885	0.2402	0.1911	0.3000	0.4407	0.3208	0.2387	0.2914
336	0.05	0.4945	0.4902	0.4063	0.4436	0.4181	0.4320	0.3254	0.3725	0.2421	0.2810	0.2060	0.3096	0.4640	0.3368	0.2662	0.3017
	0.10	0.4936	0.4899	0.4088	0.4484	0.4165	0.4312	0.3263	0.3757	0.2405	0.2801	0.2045	0.3084	0.4618	0.3342	0.2640	0.3009
	0.25	0.4927	0.4904	0.4436	0.4755	0.4141	0.4304	0.3389	0.3888	0.2385	0.2809	0.2011	0.3056	0.4565	0.3284	0.2634	0.3016
	0.50	0.4882	0.4867	0.4836	0.5012	0.4142	0.4306	0.3400	0.3860	0.2406	0.2854	0.1982	0.3040	0.4536	0.3267	0.2658	0.3021
	1.00	0.4902	0.4881	0.5078	0.5160	0.4135	0.4301	0.3432	0.3909	0.2407	0.2865	0.1982	0.3048	0.4570	0.3305	0.2633	0.3009
720	0.05	0.5487	0.5333	0.4457	0.4661	0.4681	0.4582	0.4003	0.4158	0.3128	0.3299	0.2419	0.3374	0.4865	0.3458	0.2544	0.3001
	0.10	0.5477	0.5332	0.4429	0.4677	0.4664	0.4575	0.4024	0.4220	0.3099	0.3290	0.2400	0.3362	0.4849	0.3441	0.2522	0.2985
	0.25	0.5503	0.5372	0.4690	0.4933	0.4664	0.4583	0.4332	0.4516	0.3061	0.3316	0.2362	0.3338	0.4816	0.3406	0.2504	0.2987
	0.50	0.5525	0.5403	0.5589	0.5453	0.4737	0.4667	0.4409	0.4561	0.3057	0.3354	0.2337	0.3333	0.4780	0.3369	0.2508	0.2983
	1.00	0.5624	0.5443	0.7405	0.6325	0.4719	0.4647	0.4311	0.4485	0.3059	0.3327	0.2331	0.3341	0.4782	0.3377	0.2506	0.2983
TimeMixer-Core																	
96	0.05	0.3746	0.4017	0.2741	0.3379	0.3042	0.3514	0.1637	0.2545	0.1510	0.1994	0.1438	0.2446	0.4045	0.2854	0.1977	0.2695
	0.10	0.3738	0.4010	0.2751	0.3406	0.3033	0.3505	0.1637	0.2550	0.1505	0.1995	0.1437	0.2445	0.4042	0.2848	0.1959	0.2678
	0.25	0.3720	0.3997	0.2785	0.3451	0.3037	0.3507	0.1639	0.2552	0.1499	0.2001	0.1432	0.2439	0.4028	0.2824	0.1948	0.2659
	0.50	0.3708	0.3989	0.2769	0.3429	0.3040	0.3509	0.1637	0.2549	0.1482	0.1991	0.1426	0.2430	0.4016	0.2801	0.1944	0.2640
	1.00	0.3708	0.3990	0.2760	0.3418	0.3038	0.3505	0.1637	0.2546	0.1480	0.1989	0.1426	0.2433	0.4010	0.2790	0.1950	0.2648
192	0.05	0.4141	0.4297	0.3418	0.3869	0.3278	0.3666	0.2196	0.2945	0.1928	0.2382	0.1553	0.2511	0.4263	0.2962	0.2097	0.2784
	0.10	0.4140	0.4295	0.3440	0.3920	0.3267	0.3654	0.2216	0.2984	0.1918	0.2385	0.1551	0.2510	0.4257	0.2953	0.2085	0.2752
	0.25	0.4136	0.4297	0.3532	0.4017	0.3262	0.3649	0.2229	0.2997	0.1911	0.2405	0.1549	0.2512	0.4241	0.2928	0.2076	0.2759
	0.50	0.4173	0.4372	0.3466	0.3933	0.3262	0.3649	0.2227	0.2996	0.1901	0.2404	0.1549	0.2515	0.4221	0.2895	0.2083	0.2771
	1.00	0.4096	0.4267	0.3463	0.3920	0.3266	0.3649	0.2230	0.3000	0.1896	0.2416	0.1548	0.2512	0.4203	0.2870	0.2085	0.2757
336	0.05	0.4438	0.4440	0.3638	0.4091	0.3505	0.3839	0.2738	0.3302	0.2441	0.2767	0.1777	0.2785	0.4360	0.3028	0.2234	0.2842
	0.10	0.4425	0.4435	0.3630	0.4099	0.3503	0.3835	0.2762	0.3361	0.2422	0.2767	0.1773	0.2785	0.4354	0.3018	0.2224	0.2808
	0.25	0.4457	0.4485	0.3773	0.4233	0.3501	0.3833	0.2832	0.3396	0.2396	0.2787	0.1762	0.2781	0.4336	0.2990	0.2226	0.2818
	0.50	0.4442	0.4476	0.4034	0.4426	0.3503	0.3837	0.2791	0.3346	0.2383	0.2787	0.1749	0.2763	0.4315	0.2952	0.2194	0.2787
	1.00	0.4462	0.4507	0.4125	0.4499	0.3504	0.3836	0.2770	0.3308	0.2379	0.2804	0.1752	0.2772	0.4307	0.2941	0.2212	0.2819
720	0.05	0.4628	0.4774	0.4135	0.4446	0.3971	0.4126	0.3561	0.3860	0.3170	0.3300	0.2136	0.3044	0.4622	0.3172	0.2399	0.3124
	0.10	0.4635	0.4786	0.4060	0.4415	0.3973	0.4128	0.3544	0.3892	0.3136	0.3299	0.2129	0.3042	0.4614	0.3160	0.2397	0.3115
	0.25	0.4709	0.4880	0.4209	0.4594	0.4022	0.4198	0.3681	0.4033	0.3086	0.3339	0.2111	0.3039	0.4597	0.3129	0.2389	0.3094
	0.50	0.4820	0.4970	0.5063	0.5135	0.4081	0.4259	0.3771	0.4083	0.3079	0.3406	0.2097	0.3030	0.4576	0.3090	0.2389	0.3101
	1.00	0.4982	0.5052	0.6458	0.5783	0.4107	0.4314	0.3748	0.4046	0.3065	0.3407	0.2091	0.3050	0.4580	0.3101	0.2384	0.3089

H QUALITATIVE FORECASTS AND SCALE PARTICIPATION

We show post-hoc selections from archived test-window scans under the protocol in Appendix B. All methods share the same history, future window, and channel within each comparison. Local gains describe selected cases, not representative performance; captions separately report full-test means from Appendix D. RC improves on Vanilla in 81 of 96 primary settings; the other 15 remain in the full tables.

H.1 VANILLA AND FRAMEWORK-AUGMENTED PREDICTIONS

Figures 3–9 show seven datasets and three backbone families. We retain the ETTm1 comparison and add six cases screened for forecast structure and history/future continuity. All seven have lower selected-window RC error than both alternatives. The first uses shared-axis panels; subsequent plots overlay the methods. Gray shows observed history and blue the future truth; the vertical line marks the forecast origin. Values are unmodified, without smoothing or separate rescaling.

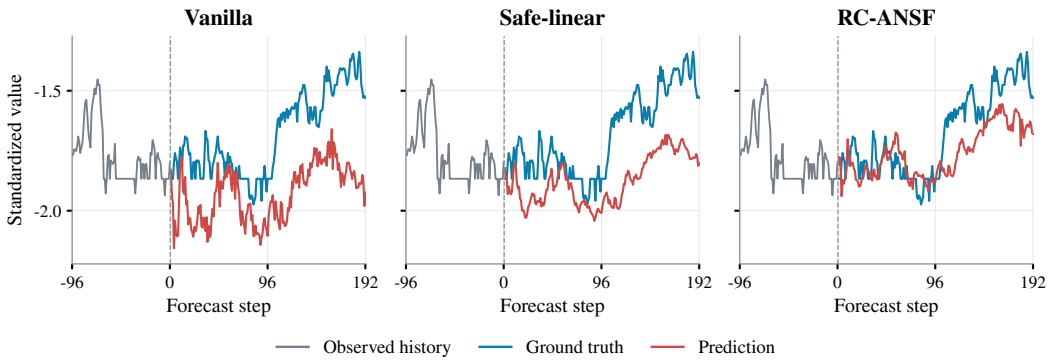

Figure 3: **ETTm1, DLinear,** $H = 192$**.** RC reduces the persistent underprediction visible in Vanilla and Safe-linear, especially in the later part of the forecast. Selected-window MSE (Vanilla / Safe-linear / RC): 0.0803 / 0.0509 / 0.0185; RC reductions: 77.0% / 63.7%. Full-test MSE (same method order): 0.5081 / 0.3969 / 0.3799.

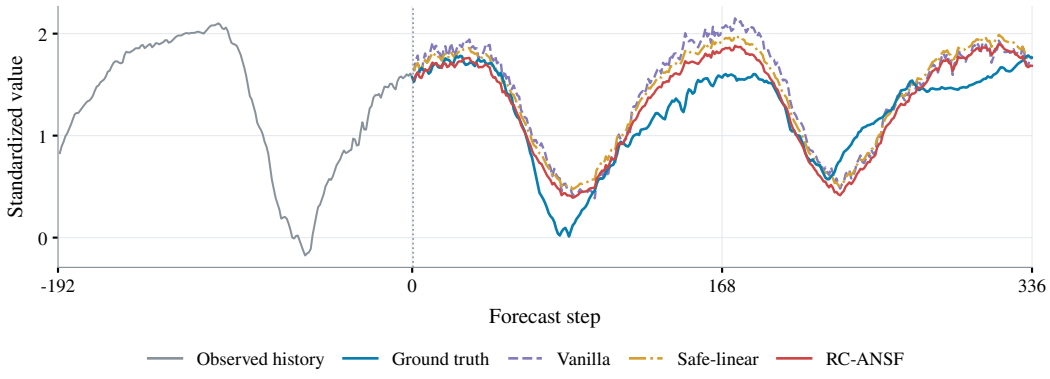

Figure 4: **Weather, DLinear,** $H = 336$**.** RC reduces the excessive level of the second predicted peak while preserving the broad periodic pattern. Selected-window MSE (Vanilla / Safe-linear / RC): 0.0847 / 0.0669 / 0.0435; RC reductions: 48.6% / 35.0%. Full-test MSE (same method order): 0.2759 / 0.2654 / 0.2602.

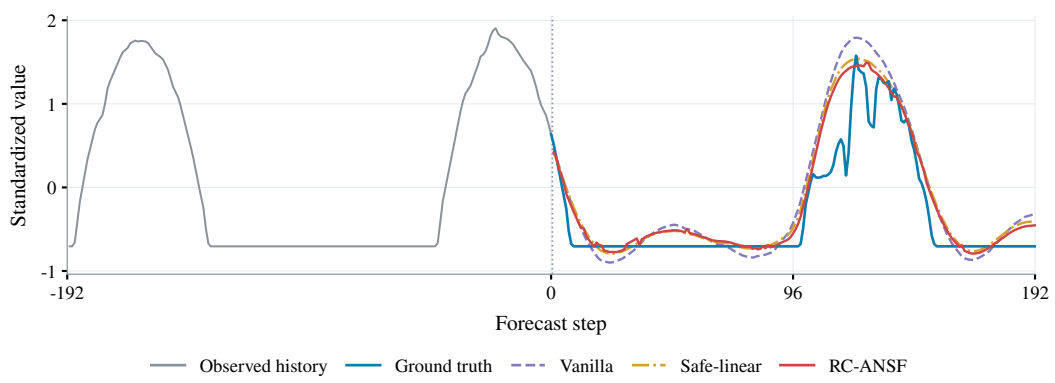

Figure 5: **Solar-Energy, TimeMixer-Core,** $H = 192$**.** RC reduces the excessive daytime peak of Vanilla and its undershoot around the low-output interval; sharp intraday fluctuations remain unresolved. Selected-window MSE (Vanilla / Safe-linear / RC): 0.1370 / 0.0967 / 0.0781; RC reductions: 43.0% / 19.2%. Full-test MSE (same method order): 0.2188 / 0.2110 / 0.2069.

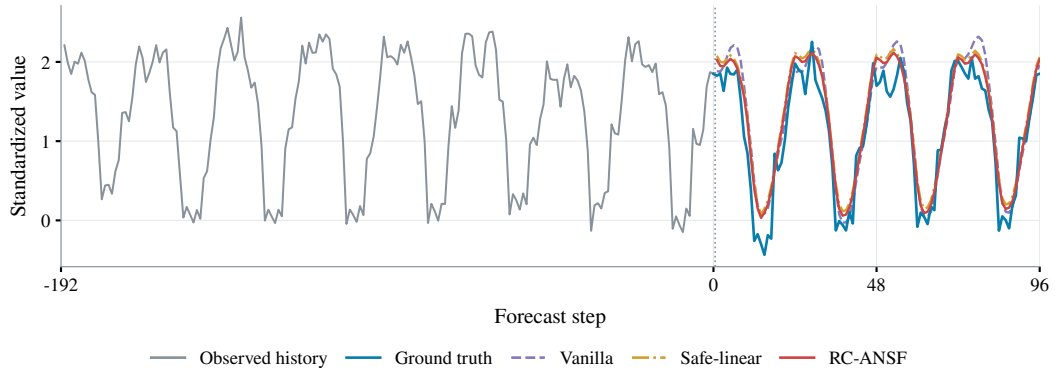

Figure 6: **Electricity, iTransformer,** $H = 96$**.** RC brings the repeated troughs closer to the target than Safe-linear, while all three forecasts capture the recurring load pattern. Selected-window MSE (Vanilla / Safe-linear / RC): 0.0920 / 0.0878 / 0.0688; RC reductions: 25.2% / 21.6%. Full-test MSE (same method order): 0.1836 / 0.1763 / 0.1732.

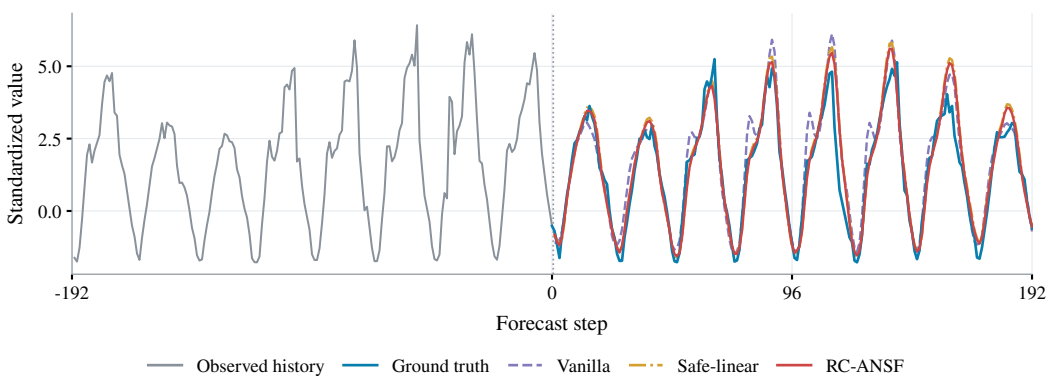

Figure 7: **Traffic, iTransformer,** $H = 192$**.** RC reduces the extra shoulders in Vanilla and moderates excessive peak amplitudes, retaining the recurring traffic pattern. Selected-window MSE (Vanilla / Safe-linear / RC): 0.4205 / 0.3507 / 0.2970; RC reductions: 29.4% / 15.3%. Full-test MSE (same method order): 0.4557 / 0.4577 / 0.4538.

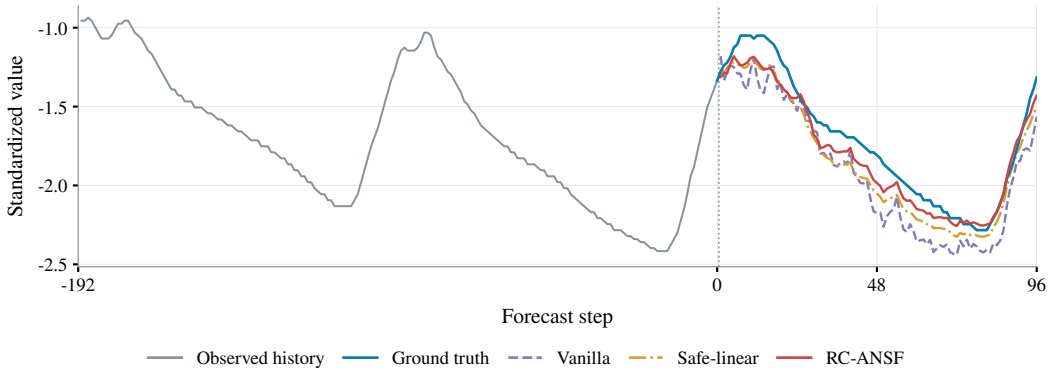

Figure 8: **ETTm2, DLinear,** $H = 96$**.** RC reduces the overly deep trough in Vanilla and the remaining downward bias in Safe-linear. Selected-window MSE (Vanilla / Safe-linear / RC): 0.0468 / 0.0218 / 0.0112; RC reductions: 76.1% / 48.6%. Full-test MSE (same method order): 0.2132 / 0.1771 / 0.1774. The local RC gain over Safe-linear does not hold for the full-test mean.

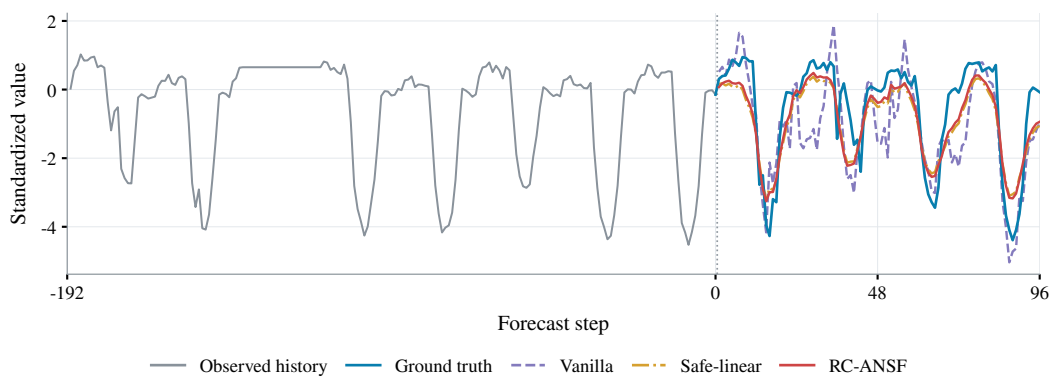

Figure 9: **ETTh1, DLinear,** $H = 96$**.** RC reduces the excessive excursions of Vanilla. Its gain over Safe-linear is smaller and occurs mainly around the troughs and intermediate levels. Selected-window MSE (Vanilla / Safe-linear / RC): 1.2488 / 0.6553 / 0.5553; RC reductions: 55.5% / 15.2%. Full-test MSE (same method order): 0.5960 / 0.4324 / 0.4191.

H.2 ALTERNATIVE FUSION METHODS

The seven methods in Appendix E share frozen predictors. Base is their finest-scale path, distinct from the separately trained Vanilla in H.1. Backbones occupy columns and methods occupy rows, with matched test inputs and channels. RC has the lowest error in these selected cases; they illustrate possible gains, not their frequency. Dashed curves in the RC row repeat Safe-linear for direct comparison.

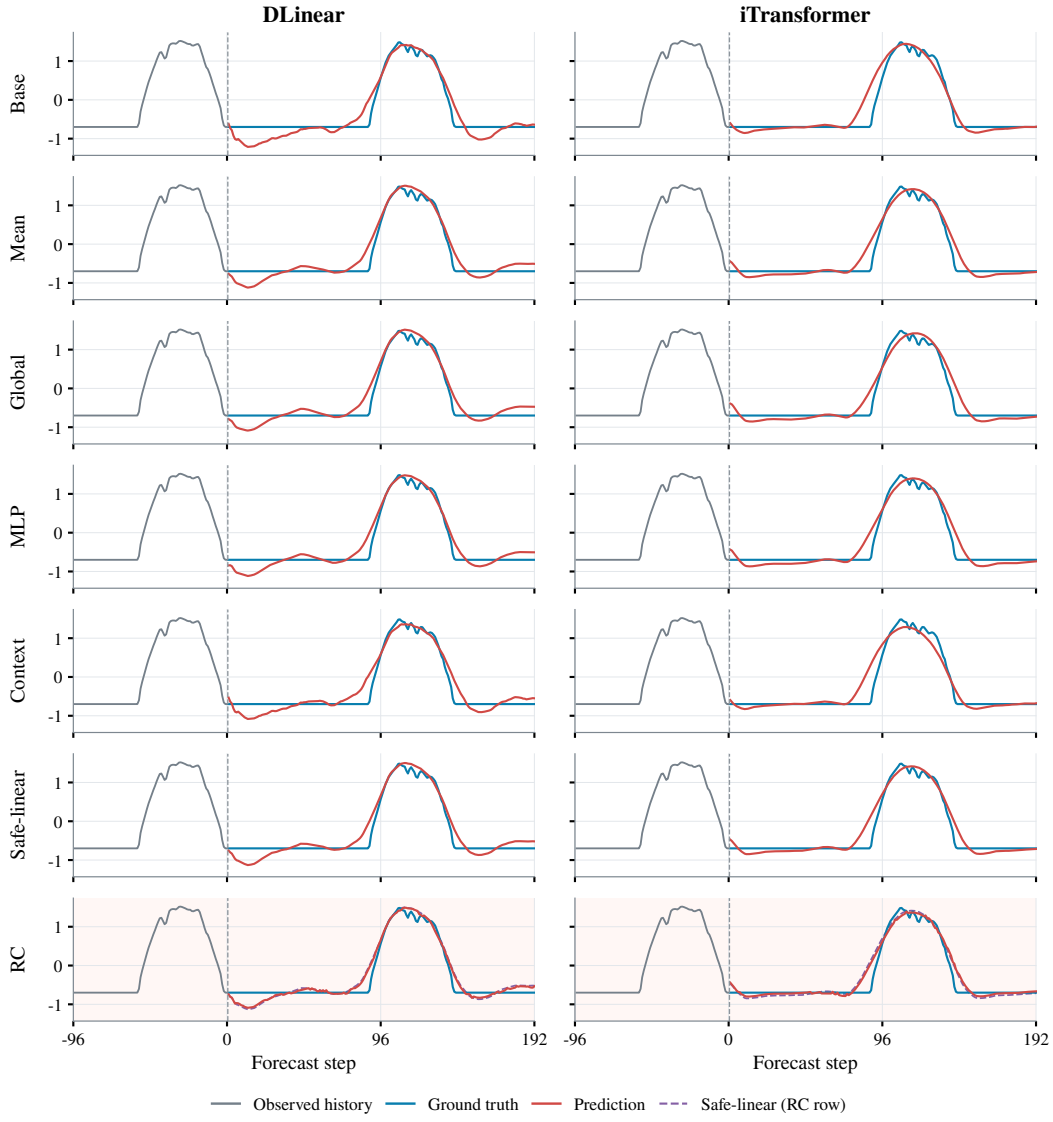

Figure 10: **Solar-Energy, $H = 192$: seven-method fusion comparison.** RC reduces spurious excursions during the low-output intervals while retaining the daytime pattern. Selected-window MSE falls by 45.4% and 42.6% relative to Base for DLinear and iTransformer, respectively; the reductions relative to the best other displayed method are 19.7% and 11.9%. Full-horizon results remain in Appendix E.

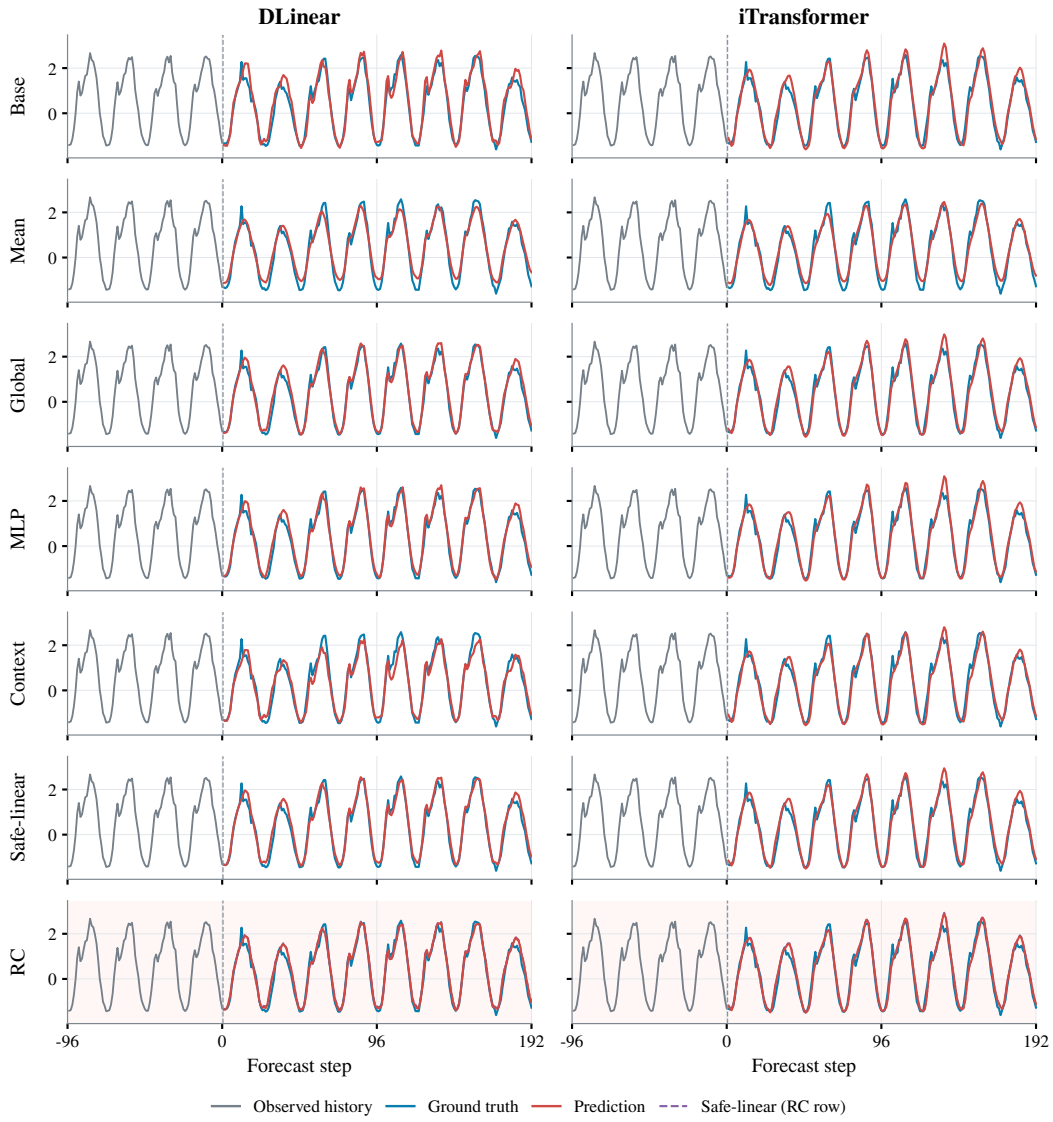

Figure 11: **Traffic, $H = 192$: seven-method fusion comparison.** RC reduces the extra within-cycle oscillations in DLinear and improves the trough levels in iTransformer while retaining the recurring pattern. Selected-window MSE falls by 53.4% and 31.0% relative to Base for DLinear and iTransformer, respectively, and by 14.9% and 10.5% relative to the best other displayed method.

H.3 NATIVE-SCALE FORECASTS AND THE FUSED PREDICTION

Figure 12 compares the finest-scale forecast, two coarse forecasts, and the final prediction using the same Solar-Energy window and channel for all backbones. Coarse forecasts are linearly interpolated to the fine temporal grid for display; the RC row is the actual output after adaptive alignment and fusion. The displayed native predictions are inputs to fusion, not the adapted candidates C_s. The bottom row repeats Base as a dashed reference. In this selected case, RC has lower fine-grid error than each displayed individual forecast; this is not a guarantee that fusion always outperforms every branch.

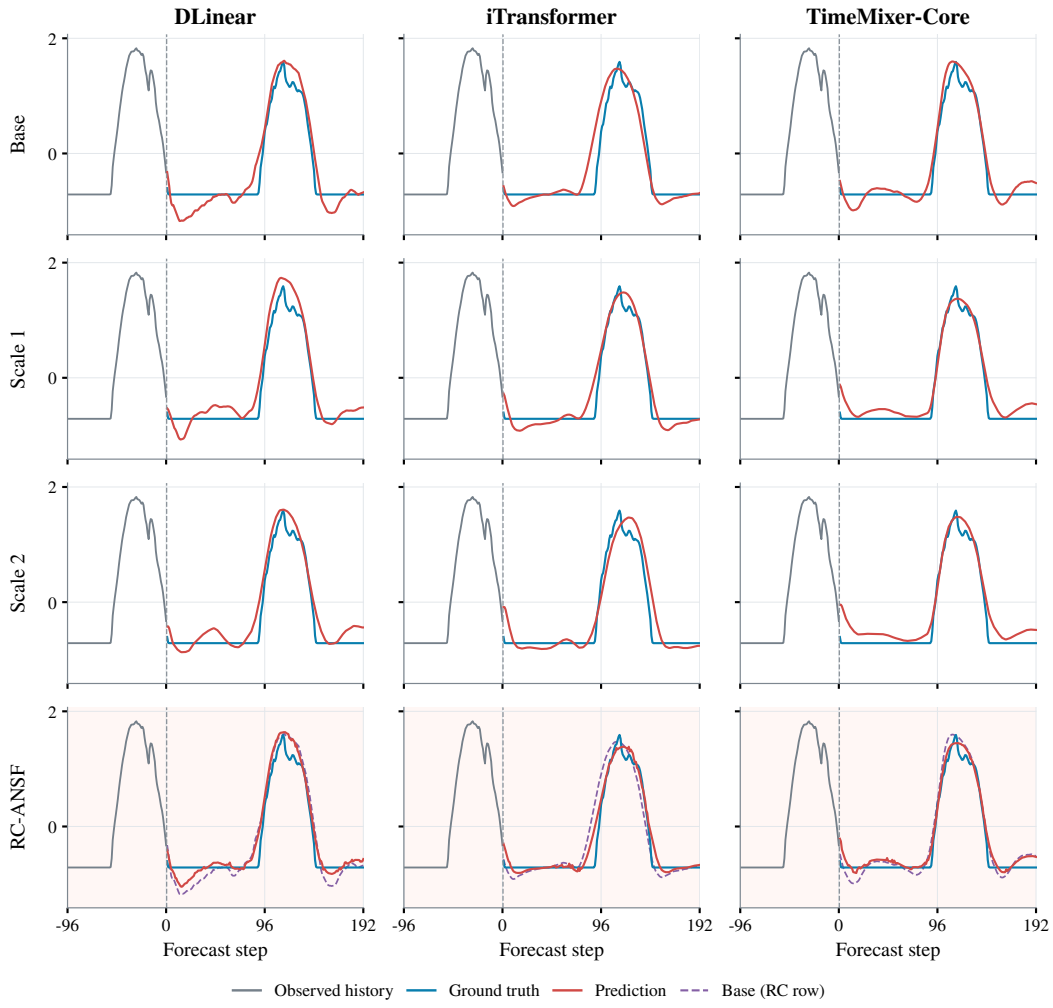

Figure 12: **Solar-Energy, $H = 192$: native-scale predictions and RC.** Columns correspond to the three backbones; rows show Base, Scale 1, Scale 2, and RC-ANSF. The native prediction lengths are 192, 96, and 48. The coarse forecasts preserve the broad cycle but differ in peak amplitude and low-output excursions. RC reduces these excursions, with selected-window MSE reductions of 42.6%, 67.3%, and 36.3% relative to Base for DLinear, iTransformer, and TimeMixer-Core, respectively. Relative to the most accurate individual forecast in each column, the reductions are 24.4%, 25.1%, and 36.3%.

H.4 GATE-INDUCED SCALE PARTICIPATION

This diagnostic asks whether the learned mixture retains appreciable coarse-scale participation rather than collapsing onto the finest-scale reference. We visualize the effective mixing coefficients induced by the participation gates, not error-attribution scores. For channel c,

$$w_{s,c} = \frac{\alpha}{S} g_{s,c}, \qquad w_{B,c} = 1 - \sum_{s=1}^{S} w_{s,c}, \qquad \widehat{Y}^F = w_B \odot B + \sum_{s=1}^{S} w_s \odot C_s. \tag{41}$$

Figure 13 shows channel-averaged coefficients from one fixed checkpoint per backbone, with $S = 2$ and $\alpha = 1$. The horizontal axis follows 64 evenly spaced inputs in chronological test order, not the steps within a forecast: each gate is broadcast over the prediction horizon. All sampled inputs and all channels are retained. The plotted coefficients are read from the fitted model, without retraining or branch removal; the displayed areas are w_s, not the raw gates g_s.

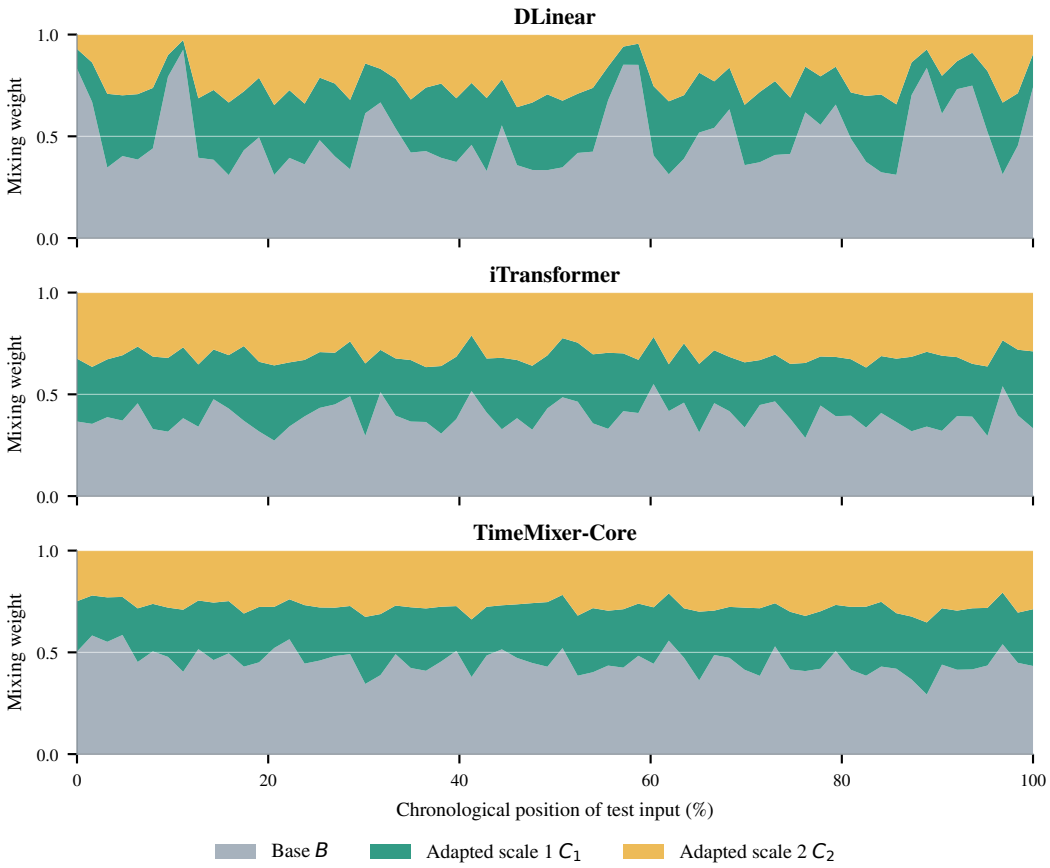

Figure 13: **Solar-Energy, $H = 192$: gate-induced mixing weights.** Stacked areas represent the channel-averaged coefficients of Base and the two adapted coarse candidates, summing to one at each sampled input. Samples are connected for display without additional smoothing. Both coarse scales receive appreciable weights across the three backbones, while the finest-scale reference retains a substantial share.

Mean coarse-scale coefficients range from 0.24 to 0.31, corresponding to mean gate values of 0.47–0.62; Base retains 0.39–0.50 of the total weight. The coefficients vary with the input, most visibly for DLinear. These patterns show non-negligible, input-dependent participation in this setting. A coefficient is not a percentage of accuracy gain: the actual correction is $w_s \odot (C_s - B)$, which also depends on the candidate's departure from Base. Thus, the plot does not establish that every scale improves every forecast or that a larger weight implies a larger benefit.