# FISC: Time-Series Forecasting via First-Layer Statistical Calibration Constraints

**Hua Wang**
School of Computer and Artificial Intelligence
Ludong University
Yantai 264025, China
hwa229@163.com

**Xiyuan Zhang**
School of Computer and Artificial Intelligence
Ludong University
Yantai 264025, China
zhangxy@ldu.edu.cn

**Fan Zhang***
School of Computer Science and Technology
Shandong Technology and Business University
Yantai 264005, China
zhangfan@sdtbu.edu.cn

## Abstract

Multivariate time-series forecasting requires reliable cross-series interaction, yet many attention- and mixing-based models couple heterogeneous variables without explicit entry-stage alignment. Such uncalibrated early interaction can lead to poorly conditioned optimization and degraded generalization when cross-series statistics drift. We propose First-layer Statistical Calibration (FISC), a simple framework that imposes statistical calibration constraints only on the first cross-series coupling layer while leaving deeper coupling layers learnable without repeated calibration. FISC estimates training-set cross-series statistics, converts them into a statistical affinity matrix, and convexly fuses this matrix with learnable coupling parameters. A row-wise simplex projection then yields a nonnegative row-stochastic mixing operator, providing bounded and interpretable entry-layer interaction. By restricting calibration to the first layer, FISC calibrates the entrance interaction without repeatedly constraining deeper feature transformations. To reduce temporal redundancy, FISC further combines a time-domain branch with a data-adaptive orthogonal-domain branch derived from training-time temporal statistics. Experiments on standard long- and short-term forecasting benchmarks show that FISC improves forecasting accuracy while maintaining favorable efficiency. Ablations and non-stationarity analyses further show that first-layer-only calibration consistently outperforms no calibration, last-layer calibration, and all-layer calibration, supporting a simple principle for robust cross-series learning: constrain the entrance interaction with training-set statistics, while letting deeper representations remain data-adaptive.

**Code Repository:** https://github.com/Q7I7/FISC-code

## 1 Introduction

Multivariate time-series forecasting underpins applications in meteorology, transportation, energy, and finance [Hyndman and Athanasopoulos, 2018, Lim et al., 2021]. Real-world time series are typically non-stationary and multi-scale: strong temporal autocorrelation coexists with heterogeneous and drifting cross-series dependencies [Yan and Wang, 2023, Xiao and Hou, 2026]. In such settings,

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

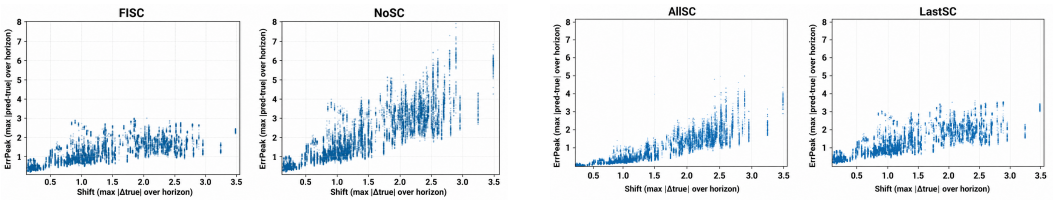



Figure 1: Motivating evidence for first-layer statistical calibration. Under increasing calibration mismatch and structural drift, FISC degrades more gradually than variants that impose calibration constraints at deeper or all layers, suggesting that calibration is most useful at the entrance while deeper layers should remain adaptive.

accurate forecasting requires not only temporal modeling, but also reliable cross-variable interaction under well-conditioned representations [Lu et al., 2025, Gao et al., 2021, Xiao et al., 2026]. However, many attention- and mixing-based models couple variables directly at early layers without explicitly calibrating the entrance interaction. When variables have mismatched scales, correlations, or noise patterns, such uncalibrated early coupling can distort optimization and amplify distribution shift [Liang et al., 2025, Xu and Xie, 2023, Xu et al., 2026]. Existing linear and frequency-domain forecasters provide efficient alternatives, but linear models can underfit high-dimensional interactions, especially when $N \gg T$ [Zeng et al., 2023, Challu et al., 2023], while fixed transform bases may mismatch dataset-specific temporal structures [Zhou et al., 2022, Cao et al., 2020, Yi et al., 2024]. These observations raise a more specific question: where should training-set statistics constrain cross-series interaction? We argue that such statistics are most reliable at the representation entrance, where variables still retain explicit series identities, but should not be repeatedly imposed on deeper representations whose structure is learned from data.

We propose FISC (First-layer Statistical Calibration), a simple framework for stable cross-series learning. Its central principle is to impose statistical calibration constraints only on the first cross-series coupling layer, while leaving deeper coupling layers learnable without repeated calibration. FISC estimates cross-series statistics from the training set, converts them into a statistical affinity matrix, and convexly fuses this matrix with learnable coupling parameters. A row-wise simplex projection then yields a nonnegative row-stochastic mixing operator, providing bounded and interpretable entry-layer interaction. This first-layer-only design calibrates the initial interaction space without forcing the same fixed calibration structure onto evolving deep representations. As illustrated in Fig. 1, this distinction becomes important when training-time calibration becomes mismatched under structural drift.

Along the temporal dimension, FISC further introduces a dual-domain representation to reduce autocorrelation-induced redundancy. A time-domain branch captures local dynamics, while a data-adaptive orthogonal-domain branch, derived from training-time temporal correlation statistics, reorganizes redundant temporal dependencies into a better-conditioned basis. The two branches are fused before cross-series coupling, allowing temporal de-correlation and first-layer cross-variable calibration to operate jointly. Experiments on long- and short-term forecasting benchmarks show that FISC improves forecasting accuracy while maintaining favorable efficiency. Ablation, sensitivity, and non-stationarity analyses further support the effectiveness of first-layer statistical calibration constraints.

**Contributions.**

- We formulate entry-stage cross-series misalignment as a layer-specific source of unstable variable interaction, motivating first-layer statistical calibration constraints that stabilize entrance coupling while preserving the data-adaptive flexibility of deeper representations.
- FISC constructs the entrance mixing operator by convexly fusing training-set statistics with learnable coupling parameters and applying row-wise simplex projection. The resulting nonnegative row-stochastic operator provides bounded and interpretable variable interaction.
- FISC combines time-domain features with an orthogonal-domain branch derived from training-time temporal statistics. Extensive accuracy, ablation, efficiency, and non-stationarity analyses support the proposed first-layer calibration design.

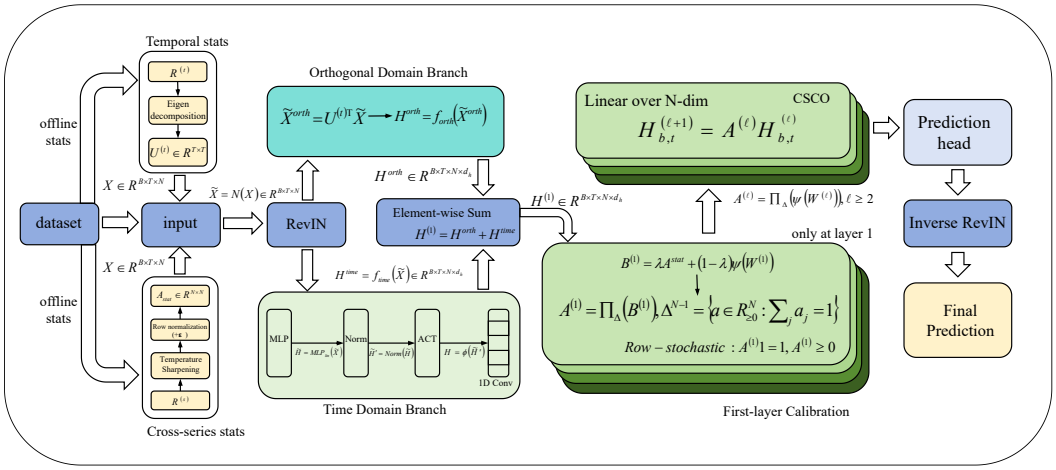



Figure 2: Overview of FISC. Statistical calibration constraints are imposed only on the first cross-series coupling layer, while deeper CSCO layers remain learnable without repeated statistical calibration.

## 2 Related work

**Statistically guided cross-series modeling.** Multivariate forecasting requires reliable cross-series interaction, yet such dependencies are often non-stationary and difficult to estimate from finite data Huang et al. [2025], Zhou et al. [2026]. Existing methods introduce cross-series structure through predefined or adaptive graphs Yu et al. [2017], Wu et al. [2019], Bai et al. [2020], Wu et al. [2020], or improve distributional consistency through normalization-based calibration such as RevIN Kim et al. [2021]. These methods show the value of structural or distributional guidance Yao et al. [2026], but usually do not specify where training-set statistics should constrain the network. FISC addresses this layer-specific question by imposing statistical calibration constraints only on the entrance cross-series coupling layer, while leaving deeper couplings learnable.

**Orthogonal-domain temporal modeling.** Transform-domain forecasters reduce temporal entanglement by modeling sequences in spectral or basis spaces Zhou et al. [2022], Yi et al. [2023], Xu et al. [2023], Piao et al. [2024]. However, fixed bases may mismatch dataset-specific temporal correlations under non-stationarity or multi-scale dynamics Zhang et al. [2026]. FISC instead estimates a data-adaptive orthogonal temporal basis from training-time statistics and fuses it with a time-domain branch before cross-series coupling.

**Cross-series dependency operators.** Cross-series interaction has been modeled by attention, graph propagation, linear mixing, and MLP-based operators Zhou et al. [2021], Wu et al. [2021], Zhang and Yan [2023], Nie et al. [2023], Liu et al. [2024]. FISC is complementary to this line: rather than increasing interaction complexity, it constrains only the entrance coupling to be bounded, row-stochastic, and statistically calibrated, while deeper representations remain data-adaptive.

## 3 Method

### 3.1 Problem setup and design principle

Given an input window $\mathbf{X} \in \mathbb{R}^{B\times T\times N}$, where $B$ is the batch size, $T$ is the look-back length, and $N$ is the number of variables, multivariate forecasting aims to predict $\mathbf{Y} \in \mathbb{R}^{B\times P\times N}$ over horizon $P$. FISC operates on normalized inputs $\tilde{\mathbf{X}} = \mathcal{N}(\mathbf{X})$, where $\mathcal{N}(\cdot)$ denotes reversible normalization along the temporal dimension of each variable. Predictions are mapped back to the original scale by $\mathcal{N}^{-1}$. This normalization reduces scale heterogeneity and makes the statistics used by FISC consistent with the model input.

FISC uses two statistics estimated only from the training split: the cross-series correlation matrix $\mathbf{R}^{(s)} \in \mathbb{R}^{N\times N}$ and the temporal correlation matrix $\mathbf{R}^{(t)} \in \mathbb{R}^{T\times T}$. Both are kept fixed during training

and testing to avoid test-time leakage. The central design principle is layer-asymmetric statistical calibration: cross-series statistics are used as calibration constraints only at the entrance coupling layer, while deeper couplings remain learnable and data-adaptive.

Fig. 2 summarizes the architecture. FISC first constructs dual-domain temporal features and then applies an $L$-layer constrained cross-series coupling stack. Only the entrance coupling matrix is statistically calibrated; all deeper coupling matrices are learned under the same row-stochastic feasible set, without repeated statistical calibration constraints.

### 3.2 Constrained cross-series coupling operator

Cross-series interaction is performed by a constrained cross-series coupling operator (CSCO). For layer $\ell$, sample $b$, and time step $t$, CSCO mixes representations along the variable dimension:

$$\mathbf{H}_{b,t}^{(\ell+1)} = \mathbf{A}^{(\ell)}\mathbf{H}_{b,t}^{(\ell)}, \qquad \mathbf{A}^{(\ell)}\mathbf{1} = \mathbf{1}, \quad \mathbf{A}^{(\ell)} \geq 0. \tag{1}$$

Here $\mathbf{H}_{b,t}^{(\ell)} \in \mathbb{R}^{N\times d_h}$ denotes the hidden representation of all variables at time $t$, and $\mathbf{A}^{(\ell)} \in \mathbb{R}^{N\times N}$ is shared across samples, time steps, and hidden channels. Equivalently, $H_{b,t,i,h}^{(\ell+1)} = \sum_{j=1}^{N} A_{ij}^{(\ell)} H_{b,t,j,h}^{(\ell)}$. Thus, each row of $\mathbf{A}^{(\ell)}$ defines a convex aggregation over source variables, yielding bounded and interpretable cross-series mixing.

To enforce this feasible set, each coupling matrix is obtained by row-wise Euclidean projection onto the probability simplex $\Delta^{N-1} = \{\mathbf{a} \in \mathbb{R}_{\geq 0}^{N} : \sum_{j=1}^{N} a_j = 1\}$. For a pre-projection score matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$, we define

$$\Pi_{\mathcal{C}}(\mathbf{S})_{i,:} = \arg\min_{\mathbf{a}\in\Delta^{N-1}} \|\mathbf{a} - \mathbf{S}_{i,:}\|_2^2, \tag{2}$$

where $\Pi_{\mathcal{C}}(\cdot)$ denotes row-wise projection onto the set of row-stochastic matrices. This projection keeps the coupling bounded and avoids unconstrained amplification across variables. In implementation, the projection is applied row by row and differentiated through its piecewise-linear mapping.

### 3.3 First-Layer statistical calibration

FISC imposes statistical calibration constraints only on the entrance CSCO layer. The calibration is computed from normalized training windows $\{\tilde{\mathbf{X}}^{(m)}\}_{m=1}^{M}$, where $\tilde{\mathbf{X}}^{(m)} \in \mathbb{R}^{T\times N}$. The empirical cross-series correlation is

$$(\mathbf{R}^{(s)})_{ij} = \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T} \tilde{x}_{t,i}^{(m)}\tilde{x}_{t,j}^{(m)}. \tag{3}$$

Because the inputs are centered and normalized per variable, this statistic corresponds to standardized covariance over the training split.

FISC converts $\mathbf{R}^{(s)}$ into a nonnegative statistical affinity by $\tilde{\mathbf{R}}^{(s)} = g_\tau(\mathbf{R}^{(s)})$, where $g_\tau(r) = \exp(r/\tau)$, followed by row normalization:

$$(\mathbf{A}_{\text{stat}})_{ij} = \frac{(\tilde{\mathbf{R}}^{(s)})_{ij}}{\sum_{k=1}^{N}(\tilde{\mathbf{R}}^{(s)})_{ik} + \varepsilon}, \qquad \mathbf{A}_{\text{stat}} \in \mathbb{R}^{N\times N}. \tag{4}$$

Here $\tau > 0$ controls the concentration of the affinity distribution and $\varepsilon > 0$ ensures numerical stability. This construction assigns larger calibration weights to stronger positive correlations and down-weights weak or negative correlations, while keeping the coupling nonnegative and row-stochastic. The nonnegative constraint applies to the coupling weights, while hidden features and downstream projections remain signed. This allows signed end-to-end responses but does not explicitly encode inhibitory edges in the coupling matrix. Strong inhibitory cross-series structure can therefore impose an expressivity cost, which we quantify in Sec. 4.2 and discuss in App. B.2.

Let $\mathbf{W}^{(\ell)} \in \mathbb{R}^{N\times N}$ be learnable coupling parameters, and let $\psi(\cdot)$ be a monotone nonnegative map such as softplus. At the first layer, FISC fuses the statistical affinity with learnable coupling scores:

$$\mathbf{A}^{(1)} = \Pi_{\mathcal{C}}\Big(\lambda\mathbf{A}_{\text{stat}} + (1-\lambda)\psi\big(\mathbf{W}^{(1)}\big)\Big), \qquad \lambda \in [0,1]. \tag{5}$$

The coefficient $\lambda$ controls calibration strength, while the learnable component allows the model to correct imperfect calibration.

For deeper layers, the statistical calibration term is removed:

$$\mathbf{A}^{(\ell)} = \Pi_{\mathcal{C}}\Big(\psi\big(\mathbf{W}^{(\ell)}\big)\Big), \qquad \ell \geq 2. \tag{6}$$

This first-layer-only design is central to FISC. At the entrance, variables still retain their original identities, so training-set cross-series statistics can constrain the initial interaction in a meaningful way. After temporal encoding and several cross-series transformations, however, hidden representations no longer correspond to raw variables in the same form. Repeatedly imposing the same calibration constraint at deeper layers may restrict the data-adaptive degrees of freedom needed to learn task-specific representations. FISC therefore uses statistics to calibrate the entrance interaction, while leaving deeper cross-series couplings learnable. A boundary analysis of when fixed statistical calibration may become unreliable is provided in App. B.

### 3.4 Dual-Domain temporal representation

To reduce autocorrelation-induced redundancy, FISC uses a dual-domain temporal representation composed of a time-domain branch and a data-adaptive orthogonal-domain branch. The temporal basis is estimated once from the training split. Let $\tilde{\mathbf{x}}_{:,n}^{(m)} \in \mathbb{R}^T$ denote the normalized window of variable $n$ in training window $m$. The temporal correlation matrix is

$$(\mathbf{R}^{(t)})_{uv} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\tilde{x}_{u,n}^{(m)}\tilde{x}_{v,n}^{(m)}, \qquad u, v \in \{1, \ldots, T\}. \tag{7}$$

The eigendecomposition $\mathbf{R}^{(t)} = \mathbf{U}^{(t)}\mathbf{\Lambda}^{(t)}\mathbf{U}^{(t)\top}$ gives an orthogonal temporal basis. Each normalized sequence is projected as $\tilde{\mathbf{x}}_{:,n}^{\mathrm{orth}} = \mathbf{U}^{(t)\top}\tilde{\mathbf{x}}_{:,n}$.

The time-domain branch models local dynamics from $\tilde{\mathbf{X}}$, while the orthogonal-domain branch models $\tilde{\mathbf{X}}^{\mathrm{orth}}$. Both use lightweight temporal encoders applied independently per variable: $\mathbf{H}^{\mathrm{time}} = f_{\mathrm{time}}(\tilde{\mathbf{X}})$ and $\widehat{\mathbf{H}}^{\mathrm{orth}} = f_{\mathrm{orth}}(\tilde{\mathbf{X}}^{\mathrm{orth}})$, with both outputs in $\mathbb{R}^{B\times T\times N\times d_h}$. The orthogonal output is mapped back to the temporal domain by $\mathbf{H}_{b,:,n,h}^{\mathrm{orth}} = \mathbf{U}^{(t)}\widehat{\mathbf{H}}_{b,:,n,h}^{\mathrm{orth}}$. Finally, the two branches are fused as $\mathbf{H}^{(1)} = \mathbf{H}^{\mathrm{time}} + \mathbf{H}^{\mathrm{orth}}$.

Thus, temporal de-correlation is performed before CSCO, while the time-domain branch preserves local dynamics not captured by the orthogonal basis.

### 3.5 Prediction head and design interpretation

After $L$ CSCO layers, FISC obtains $\mathbf{H}^{(L+1)} \in \mathbb{R}^{B\times T\times N\times d_h}$. A temporal prediction head maps the look-back dimension to the prediction horizon, and a linear decoder produces the normalized forecast $\widehat{\widetilde{\mathbf{Y}}} \in \mathbb{R}^{B\times P\times N}$. The final forecast is transformed back to the original scale:

$$\widehat{\mathbf{Y}} = \mathcal{N}^{-1}\big(\widehat{\widetilde{\mathbf{Y}}}\big). \tag{8}$$

FISC uses training-set statistics as entrance-level calibration constraints rather than repeated constraints across depth. The first-layer calibration provides an empirical constraint when variables still retain explicit semantics, while the simplex projection keeps the coupling bounded and interpretable. Deeper layers remain fully learnable to avoid enforcing potentially mismatched calibration constraints on evolving representations or shifted test distributions. Similarly, the orthogonal temporal branch organizes redundant temporal structure, whereas the time-domain branch preserves local flexibility.

## 4 Experimental setting

**Datasets.** We evaluate FISC on standard public multivariate forecasting benchmarks covering long- and short-term scenarios, including electricity load, traffic flow, weather, ETT, exchange rate, epidemiological, financial, web-traffic, and transportation datasets. MSE and MAE are used as

Table 1: Long-term forecasting results. For PEMS, horizons are $H \in \{12, 24, 48, 96\}$; for other datasets, $H \in \{96, 192, 336, 720\}$. The look-back window is $T = 96$. *Avg* denotes averaging within each horizon subset. Best and second-best results are marked in **red** and blue, respectively. Full per-horizon results are reported in Appendix Table 14.

| Model | FISC ours | | OLinear 2025 | | TimeMixer 2024a | | FilterNet 2024a | | FITS 2024 | | DLinear 2023 | | TimeMixer++ 2025a | | Leddam 2024 | | CARD 2024b | | Fredformer 2024 | | iTransformer 2024a | | PatchTST 2023 | | TimesNet 2023b | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric | 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 |
| ETT(Avg) | **0.348** | **0.370** | 0.359 | 0.376 | 0.367 | 0.388 | 0.375 | 0.394 | 0.400 | 0.403 | 0.442 | 0.444 | 0.349 | 0.378 | 0.368 | 0.388 | 0.366 | 0.380 | 0.366 | 0.385 | 0.383 | 0.399 | 0.380 | 0.396 | 0.391 | 0.404 |
| weather | **0.226** | **0.256** | 0.237 | 0.260 | 0.240 | 0.272 | 0.245 | 0.272 | 0.273 | 0.292 | 0.265 | 0.317 | **0.226** | 0.262 | 0.242 | 0.272 | 0.239 | 0.265 | 0.246 | 0.272 | 0.258 | 0.279 | 0.259 | 0.281 | 0.259 | 0.287 |
| Traffic | 0.436 | **0.244** | 0.451 | 0.247 | 0.485 | 0.298 | 0.463 | 0.310 | 0.615 | 0.370 | 0.625 | 0.383 | **0.416** | 0.264 | 0.467 | 0.294 | 0.453 | 0.282 | 0.433 | 0.291 | 0.428 | 0.282 | 0.531 | 0.343 | 0.620 | 0.336 |
| Electricity | **0.155** | **0.244** | 0.159 | 0.248 | 0.182 | 0.273 | 0.173 | 0.268 | 0.384 | 0.434 | 0.212 | 0.300 | 0.165 | 0.253 | 0.169 | 0.263 | 0.168 | 0.258 | 0.176 | 0.269 | 0.178 | 0.270 | 0.208 | 0.295 | 0.192 | 0.295 |
| Exchange | **0.332** | 0.394 | 0.355 | 0.399 | 0.387 | 0.416 | 0.388 | 0.419 | 0.365 | 0.408 | 0.354 | 0.414 | 0.357 | 0.409 | 0.354 | 0.402 | 0.362 | 0.402 | 0.333 | **0.391** | 0.360 | 0.403 | 0.367 | 0.404 | 0.416 | 0.443 |
| PEMS03 | **0.094** | **0.198** | 0.095 | 0.199 | 0.167 | 0.267 | 0.145 | 0.251 | 0.489 | 0.465 | 0.278 | 0.375 | 0.165 | 0.263 | 0.107 | 0.210 | 0.174 | 0.275 | 0.135 | 0.243 | 0.113 | 0.221 | 0.180 | 0.291 | 0.147 | 0.248 |
| PEMS04 | **0.090** | **0.189** | 0.091 | 0.190 | 0.185 | 0.287 | 0.146 | 0.258 | 0.531 | 0.489 | 0.295 | 0.388 | 0.136 | 0.251 | 0.103 | 0.210 | 0.206 | 0.299 | 0.162 | 0.261 | 0.111 | 0.221 | 0.195 | 0.307 | 0.129 | 0.241 |
| PEMS07 | **0.076** | **0.163** | 0.077 | 0.164 | 0.181 | 0.271 | 0.123 | 0.229 | 0.500 | 0.472 | 0.329 | 0.395 | 0.152 | 0.258 | 0.084 | 0.180 | 0.149 | 0.247 | 0.121 | 0.222 | 0.101 | 0.204 | 0.211 | 0.303 | 0.124 | 0.225 |
| PEMS08 | **0.111** | **0.192** | 0.113 | 0.194 | 0.226 | 0.299 | 0.172 | 0.260 | 0.534 | 0.487 | 0.379 | 0.416 | 0.200 | 0.279 | 0.122 | 0.211 | 0.201 | 0.280 | 0.161 | 0.250 | 0.150 | 0.226 | 0.280 | 0.321 | 0.193 | 0.271 |
| 1st Count | 8 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 2: Short-term forecasting results. *Avg* averages over eight horizons from two settings: S (Input-12, Predict-$\{3, 6, 9, 12\}$) and L (Input-36, Predict-$\{24, 36, 48, 60\}$). Best and second-best results are marked in **red** and blue, respectively. Full results are provided in Appendix Tables 15 and 16.

| Dataset | FISC ours | | OLinear 2025 | | TimeMixer 2024a | | FilterNet 2024a | | FITS 2024 | | DLinear 2023 | | TimeMixer++ 2025a | | Leddam 2024 | | CARD 2024b | | Fredformer 2024 | | iTransformer 2024a | | PatchTST 2023 | | TimesNet 2023b | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metric | 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 |
| ILI | **1.412** | **0.682** | 1.429 | 0.690 | 1.864 | 0.806 | 1.792 | 0.791 | 3.495 | 1.267 | 2.742 | 1.126 | 1.806 | 0.794 | 1.725 | 0.777 | 1.959 | 0.822 | 1.733 | 0.798 | 1.715 | 0.773 | 1.905 | 0.804 | 1.809 | 0.808 |
| COVID-19 | **5.020** | **1.181** | 5.187 | 1.211 | 5.919 | 1.350 | 5.607 | 1.322 | 6.910 | 1.534 | 8.279 | 1.601 | 5.973 | 1.369 | 5.252 | 1.286 | 5.536 | 1.315 | 5.279 | 1.286 | 5.301 | 1.294 | 5.836 | 1.363 | 6.106 | 1.369 |
| METR-LA | 0.582 | **0.307** | 0.586 | 0.311 | 0.608 | 0.372 | 0.603 | 0.366 | 0.636 | 0.380 | 0.580 | 0.422 | **0.567** | 0.362 | 0.602 | 0.367 | 0.639 | 0.349 | 0.617 | 0.368 | 0.627 | 0.372 | 0.614 | 0.373 | 0.617 | 0.370 |
| NASDAQ | **0.117** | **0.199** | 0.121 | 0.201 | 0.121 | 0.204 | 0.128 | 0.211 | 0.136 | 0.222 | 0.150 | 0.251 | 0.124 | 0.211 | 0.128 | 0.210 | 0.125 | 0.207 | 0.126 | 0.210 | 0.134 | 0.217 | 0.128 | 0.209 | 0.162 | 0.247 |
| Wiki | 6.391 | **0.413** | 6.395 | 0.414 | 6.444 | 0.438 | 6.457 | 0.439 | 8.164 | 0.617 | 6.420 | 0.510 | 6.431 | 0.444 | 6.416 | 0.434 | 6.419 | 0.427 | **6.318** | 0.430 | 6.422 | 0.432 | 6.368 | 0.424 | 7.633 | 0.571 |
| SP500 | **0.144** | **0.248** | 0.146 | 0.250 | 0.153 | 0.265 | 0.163 | 0.279 | 0.183 | 0.307 | 0.179 | 0.298 | 0.156 | 0.270 | 0.163 | 0.282 | 0.147 | 0.252 | 0.167 | 0.286 | 0.162 | 0.279 | 0.159 | 0.277 | 0.150 | 0.262 |
| DowJones | **7.664** | **0.615** | 7.685 | 0.619 | 8.498 | 0.634 | 8.282 | 0.633 | 8.458 | 0.647 | 7.894 | 0.626 | 8.895 | 0.643 | 8.257 | 0.634 | 7.699 | 0.619 | 8.041 | 0.625 | 8.176 | 0.630 | 7.991 | 0.626 | 10.960 | 0.737 |
| CarSales | 0.327 | **0.302** | 0.330 | 0.305 | 0.333 | 0.323 | 0.329 | 0.319 | 0.396 | 0.371 | 0.387 | 0.376 | 0.337 | 0.321 | 0.334 | 0.322 | 0.348 | 0.324 | 0.336 | 0.326 | **0.311** | 0.308 | 0.327 | 0.318 | 0.334 | 0.328 |
| Power | **1.234** | **0.831** | 1.248 | 0.834 | **1.234** | 0.841 | 1.309 | 0.870 | 1.405 | 0.908 | 1.278 | 0.870 | 1.235 | 0.841 | 1.296 | 0.868 | 1.288 | 0.847 | 1.303 | 0.870 | 1.325 | 0.874 | 1.311 | 0.874 | 1.318 | 0.870 |
| Website | **0.178** | **0.294** | 0.225 | 0.312 | 0.279 | 0.359 | 0.297 | 0.367 | 0.441 | 0.484 | 0.303 | 0.389 | 0.260 | 0.344 | 0.264 | 0.348 | 0.303 | 0.366 | 0.266 | 0.350 | 0.179 | 0.297 | 0.284 | 0.361 | 0.251 | 0.341 |
| Unemp | 0.615 | **0.438** | 0.729 | 0.462 | 1.581 | 0.708 | 1.286 | 0.628 | 2.884 | 1.122 | **0.564** | 0.510 | 1.506 | 0.678 | 1.502 | 0.689 | 1.163 | 0.596 | 2.048 | 0.789 | 1.408 | 0.666 | 1.237 | 0.624 | 2.328 | 0.852 |
| **1st Count** | 7 | 11 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

primary metrics, while $R^2$, Pearson $r$, and MASE are additionally reported for scale-free evaluation. Dataset statistics, splits, preprocessing details, and full per-horizon results are provided in App. C.

**Baselines.** We compare with OLinear Yue et al. [2025], TimeMixer Wang et al. [2024a], FilterNet Yi et al. [2024], FITS Xu et al. [2023], DLinear Zeng et al. [2023], TimeMixer++ Wang et al. [2025], Leddam Yu et al. [2024], CARD Wang et al. [2024b], Fredformer Piao et al. [2024], iTransformer Liu et al. [2024], PatchTST Nie et al. [2023], and TimesNet Wu et al. [2023]. All comparisons use matched chronological splits, input lengths, horizons, and metrics. Baselines follow their public implementations and recommended settings. Published results are used only under matching evaluation protocols; otherwise, the implementations are rerun.

**Training and model selection.** FISC uses Adam and CARD's weighted $L_1$ loss, with at most 50 epochs and validation-based early stopping with patience 10. The search ranges are learning rate $\{3, 4, 5\} \times 10^{-4}$, $D \in \{128, 256, 512\}$, $L \in \{2, 3, 4, 5\}$, $\lambda \in \{0.2, 0.4, 0.6, 0.8\}$, and $\tau \in \{0.1, 0.2, 0.5, 1.0\}$. We fix $d = 16$ and choose batch sizes from $\{4, 8, 16, 32, 64\}$ according to memory availability. All configuration and checkpoint selection uses validation data, with $\lambda$ and $\tau$ shared across horizons within each dataset. Test data are reserved for final evaluation. Both correlation matrices are estimated from training data only and remain fixed during validation and testing.

### 4.1 Overall forecasting performance

**Long-term forecasting.** Table 1 summarizes the long-term forecasting results. FISC achieves the best or second-best performance on most datasets and horizons, with particularly consistent performance at longer horizons. These results support the usefulness of constraining the entrance cross-series coupling, where accumulated temporal and cross-series errors become more pronounced.

**Short-term forecasting.** Table 2 reports the short-term forecasting results. FISC remains competitive in short-horizon settings, including cases where heavier attention-based models or single-domain representations are strong baselines. This suggests that first-layer statistical calibration constraints and dual-domain temporal modeling are not limited to long-horizon regimes.

Table 3: Forecasting performance under scale-free metrics: coefficient of determination ($R^2$), Pearson correlation ($r$), and MASE. ↑ / ↓ indicates higher/lower is better. Best results are marked in red bold and second-best results in blue underlined. The look-back length is fixed to $T = 96$.

| Dataset | H | FISC (Ours) | | | OLinear (2025) | | | CARD (2024b) | | | iTrans. (2024a) | | | TimeMixer (2024a) | | | FITS (2024) | | | DLinear (2023) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ | $R^2$ ↑ | $r$ ↑ | MASE↓ |
| ECL | 96 | **0.671** | **0.960** | **0.896** | 0.640 | 0.916 | 0.902 | 0.533 | 0.905 | 0.975 | 0.550 | 0.907 | 0.990 | 0.456 | 0.897 | 1.022 | 0.352 | 0.883 | 1.144 | 0.120 | 0.870 | 1.300 |
| | 192 | **0.660** | **0.936** | **0.976** | 0.638 | 0.908 | 0.983 | 0.547 | 0.899 | 1.028 | 0.590 | 0.900 | 1.057 | 0.575 | 0.893 | 1.076 | 0.446 | 0.880 | 1.153 | 0.215 | 0.867 | 1.312 |
| | 336 | **0.732** | **0.943** | **1.048** | 0.710 | 0.901 | 1.061 | 0.678 | 0.891 | 1.113 | 0.678 | 0.893 | 1.133 | 0.640 | 0.886 | 1.155 | 0.588 | 0.873 | 1.229 | 0.247 | 0.858 | 1.387 |
| | 720 | **0.735** | **0.928** | **1.158** | 0.701 | 0.889 | 1.174 | 0.653 | 0.877 | 1.273 | 0.638 | 0.872 | 1.315 | 0.634 | 0.871 | 1.314 | 0.593 | 0.858 | 1.376 | 0.148 | 0.844 | 1.536 |
| | Avg | **0.699** | **0.938** | **1.019** | 0.672 | 0.903 | 1.030 | 0.603 | 0.893 | 1.097 | 0.614 | 0.893 | 1.124 | 0.576 | 0.887 | 1.142 | 0.495 | 0.873 | 1.225 | 0.182 | 0.860 | 1.384 |
| Traffic | 96 | **0.761** | **0.937** | **0.718** | 0.737 | 0.899 | 0.729 | 0.690 | 0.882 | 0.855 | 0.690 | 0.889 | 0.860 | 0.640 | 0.888 | 0.979 | 0.458 | 0.811 | 1.284 | 0.441 | 0.811 | 1.346 |
| | 192 | **0.751** | **0.916** | **0.746** | 0.719 | 0.884 | 0.756 | 0.683 | 0.869 | 0.856 | 0.689 | 0.877 | 0.879 | 0.660 | 0.862 | 0.959 | 0.542 | 0.814 | 1.151 | 0.551 | 0.815 | 1.189 |
| | 336 | **0.756** | **0.910** | **0.747** | 0.730 | 0.875 | 0.758 | 0.702 | 0.862 | 0.843 | 0.713 | 0.867 | 0.883 | 0.685 | 0.852 | 0.960 | 0.582 | 0.810 | 1.119 | 0.582 | 0.810 | 1.167 |
| | 720 | **0.738** | **0.903** | **0.806** | 0.703 | 0.860 | 0.815 | 0.679 | 0.848 | 0.894 | 0.693 | 0.857 | 0.913 | 0.672 | 0.842 | 0.986 | 0.568 | 0.794 | 1.169 | 0.564 | 0.794 | 1.224 |
| | Avg | **0.746** | **0.911** | **0.753** | 0.722 | 0.879 | 0.764 | 0.689 | 0.865 | 0.862 | 0.696 | 0.873 | 0.884 | 0.664 | 0.861 | 0.971 | 0.537 | 0.807 | 1.181 | 0.535 | 0.808 | 1.232 |

**Scale-free evaluation.** Table 3 reports coefficient of determination ($R^2$), Pearson correlation ($r$), and MASE. These metrics complement MSE/MAE by assessing goodness-of-fit, linear association, and relative error across datasets with different scales, providing a metric-robust view of forecasting quality.

### 4.2 Model analysis

We analyze the contributions of FISC through targeted ablations over its core design choices.

**Placement of statistical calibration constraints.**

FISC imposes statistical calibration constraints only on the first cross-series coupling layer. To isolate this design choice, we vary the placement of the calibration constraints while keeping the remaining architecture and training protocol unchanged. The variants include NoSC, which removes statistical calibration and keeps only learnable simplex-constrained coupling; LastSC, which applies calibration only at the last layer; and AllSC, which applies calibration at every coupling layer.

Table 4: Study of different placements for statistical calibration constraints. Best and second-best results are marked in **red** and blue, respectively. Full results are provided in Appendix Table 17.

| Model | FISC | | NoSC | | LastSC | | AllSC | |
|---|---|---|---|---|---|---|---|---|
| Dataset | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ETTh1 | **0.416** | **0.416** | 0.425 | 0.425 | 0.421 | 0.423 | 0.423 | 0.424 |
| Weather | **0.226** | **0.256** | 0.240 | 0.261 | 0.235 | 0.258 | 0.238 | 0.259 |
| Electricity | **0.155** | **0.244** | 0.161 | 0.250 | 0.158 | 0.248 | 0.159 | 0.249 |
| PEMS04 | **0.090** | **0.189** | 0.093 | 0.192 | 0.092 | 0.191 | 0.092 | 0.191 |
| PEMS08 | **0.111** | **0.192** | 0.115 | 0.196 | 0.114 | 0.196 | 0.114 | 0.195 |
| ILI | **1.412** | **0.682** | 1.483 | 0.699 | 1.469 | 0.693 | 1.482 | 0.698 |
| NASDAQ | **0.117** | **0.199** | 0.122 | 0.203 | 0.121 | 0.202 | 0.121 | 0.202 |

As shown in Table 4, removing first-layer calibration degrades performance, while moving calibration to the last layer or applying it to all layers brings no consistent gain. These results support FISC's central principle: training-set statistics are most useful as entrance-layer calibration constraints, whereas deeper couplings should remain data-adaptive. Plug-in results in Table 12 and multi-seed diagnostics in App. E.4 further support the transferability and stability of this design.

**Dual-domain Representation: Time vs. Ortho.**

Beyond cross-variable coupling, FISC uses a dual-domain temporal representation to condition features before variable mixing. The time-domain branch preserves local dynamics, while the orthogonal branch reorganizes autocorrelation-induced redundancy using a data-adaptive basis estimated from training statistics. We compare FISC with OnlyTime and OnlyOrth to test branch complementarity, and replace the learned basis with fixed Fourier, Wavelet, and Chebyshev transforms to examine whether the gain comes from data-adaptive temporal de-correlation.

Table 5 shows that removing either branch degrades performance, supporting their complementary roles. Fixed transforms also underperform the learned basis, indicating that the gains are not merely due to using an alternative transform domain, but to a data-matched temporal representation before cross-series coupling.

**CSCO constraints and statistical calibration.**

To isolate gains from variable coupling, we ablate the feasible-set constraints on the coupling matrix and the first-layer statistical calibration constraints. We compare unconstrained mixing, nonnegativity

Table 5: Average MSE and MAE over seven datasets for assessing branch complementarity and transform-basis choices. Best and second-best results are marked in **red** and blue, respectively. Full results are provided in Appendix Table 18.

| Model | FISC | | OnlyTime | | OnlyOrth | | Fourier | | Wavelet | | Chebyshev | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ETTh1 | **0.416** | **0.416** | 0.441 | 0.433 | 0.429 | 0.427 | 0.426 | 0.429 | 0.431 | 0.423 | 0.436 | 0.432 |
| Weather | **0.226** | **0.256** | 0.236 | 0.258 | 0.238 | 0.261 | 0.235 | 0.266 | 0.239 | 0.264 | 0.243 | 0.263 |
| Electricity | **0.155** | **0.244** | 0.163 | 0.251 | 0.159 | 0.248 | 0.160 | 0.248 | 0.159 | 0.252 | 0.162 | 0.256 |
| PEMS04 | **0.090** | **0.189** | 0.141 | 0.245 | 0.096 | 0.269 | 0.091 | 0.193 | 0.093 | 0.193 | 0.092 | 0.197 |
| PEMS08 | **0.111** | **0.192** | 0.117 | 0.195 | 0.455 | 0.427 | 0.113 | 0.199 | 0.112 | 0.195 | 0.116 | 0.199 |
| ILI | **1.412** | **0.682** | 1.473 | 0.698 | 1.476 | 0.697 | 1.443 | 0.702 | 1.436 | 0.705 | 1.482 | 0.714 |
| NASDAQ | **0.117** | **0.199** | 0.179 | 0.202 | 0.121 | 0.202 | 0.122 | 0.206 | 0.122 | 0.203 | 0.122 | 0.207 |

only, row normalization, row-sum constraint, and NoFusion, which removes the first-layer calibration fusion while keeping the simplex projection.

Table 6: Average MSE and MAE over seven representative datasets for quantifying the contributions of feasible-set constraints and first-layer statistical calibration constraints. Best and second-best results are marked in **red** and blue, respectively. Full results are provided in Appendix Table 19.

| Model | FISC | | Unconstrained | | Nonneg | | Rownorm | | Rowsum | | NoFusion | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| ETTh1 | **0.416** | **0.416** | 0.429 | 0.427 | 0.436 | 0.430 | 0.430 | 0.427 | 0.428 | 0.427 | 0.431 | 0.428 |
| Weather | **0.226** | **0.256** | 0.239 | 0.262 | 0.251 | 0.271 | 0.244 | 0.266 | 0.249 | 0.269 | 0.239 | 0.260 |
| Electricity | **0.155** | **0.244** | 0.191 | 0.269 | 0.215 | 0.286 | 0.194 | 0.272 | 0.190 | 0.269 | 0.159 | 0.249 |
| PEMS04 | **0.090** | **0.189** | 0.092 | 0.265 | 0.511 | 0.461 | 0.525 | 0.478 | 0.546 | 0.483 | 0.535 | 0.477 |
| PEMS08 | **0.111** | **0.192** | 0.113 | 0.195 | 0.436 | 0.411 | 0.452 | 0.422 | 0.447 | 0.418 | 0.447 | 0.427 |
| ILI | **1.412** | **0.682** | 1.633 | 0.731 | 1.964 | 0.831 | 1.670 | 0.741 | 1.570 | 0.714 | 1.489 | 0.703 |
| NASDAQ | **0.117** | **0.199** | 0.121 | 0.202 | 0.126 | 0.208 | 0.121 | 0.201 | 0.121 | 0.201 | 0.125 | 0.208 |

As shown in Table 6, feasible-set constraints improve stability and consistency, while removing the first-layer calibration fusion (NoFusion) degrades performance. These results indicate that simplex-constrained coupling and entrance-layer statistical calibration constraints provide complementary benefits: the former bounds cross-series mixing, whereas the latter calibrates the initial variable interaction. **Signed-coupling stress test.**

We evaluate controlled signed-VAR systems with negative-edge ratios of 0%, 25%, and 50%. SignedRes permits bounded negative coupling with unit row sums (App. B.2), whereas Unconstrained removes the feasible-set constraints. Temporal encoding, coupling depth, prediction head, data, and training settings are matched. Table 7 reports mean±std MSE ratios relative to FISC over five graph seeds, normalized within each setting and seed.

Table 7: Controlled signed-VAR stress test. Entries are mean±std of same-seed MSE ratios over five graph seeds; lower is better. FISC equals $1.000 \pm 0.000$ by normalization. $\beta$ and $\rho_A^-$ denote the validation-selected SignedRes coefficient and its negative-entry proportion.

| Neg. edges | FISC | SignedRes | Unconstrained | NoSC | $\beta$ | $\rho_A^-$ |
|---|---|---|---|---|---|---|
| 0% | $1.000 \pm 0.000$ | $1.007 \pm 0.003$ | $1.053 \pm 0.002$ | $1.064 \pm 0.003$ | 0.02 | 0.027 |
| 25% | $1.000 \pm 0.000$ | $0.989 \pm 0.006$ | $1.013 \pm 0.015$ | $1.047 \pm 0.020$ | 0.05 | 0.084 |
| 50% | $1.000 \pm 0.000$ | $0.955 \pm 0.010$ | $0.994 \pm 0.016$ | $1.027 \pm 0.033$ | 0.10 | 0.277 |

FISC outperforms NoSC at all three ratios, supporting entrance calibration under signed dependencies. Relative to FISC, SignedRes increases MSE by 0.7% at 0% negative edges but reduces it by 1.1% and 4.5% at 25% and 50%, respectively. Unconstrained remains weaker than SignedRes throughout and improves over FISC by only 0.6% at 50%. These results reveal an empirical stability–expressivity trade-off: nonnegative coupling benefits the positive system, while controlled signed coupling becomes more effective as inhibitory structure increases in the tested systems.

### 4.3 Robustness and interpretability of statistical calibration

**Regime shifts and non-stationarity.**

We further examine FISC under abrupt non-stationarity, where test-time dynamics deviate from training statistics. This setting stresses a natural boundary of statistical calibration: a calibration matrix estimated from the training split may become less reliable after a regime change. Fig. 3 visualizes the Electricity event window centered at the detected change point. FISC shows a smaller post-change error spike and faster recovery toward a stable low-error region. This behavior suggests that entrance-layer calibration can stabilize the initial cross-series interaction without repeatedly constraining deeper representations with mismatched statistics. The result complements the structural-drift stress test in the appendix and supports the first-layer-only calibration design under non-stationary conditions.

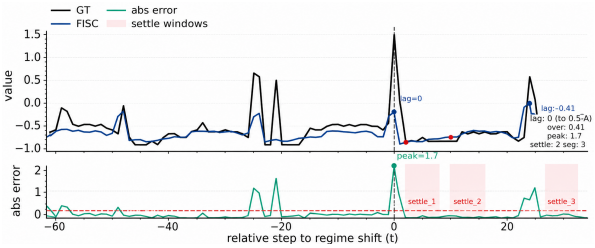



Figure 3: Event-window trajectories centered at the detected regime shift. Lower post-change error spikes and faster recovery are consistent with stronger robustness under non-stationarity.

**Interpretable Decomposition via Statistical Calibration.**

First-layer statistical calibration makes the learned entrance coupling inspectable relative to the training-set calibration matrix. Rather than fixing the coupling to $A_{\text{stat}}$, FISC learns $A_{\text{full}}$ through calibration fusion and simplex projection, and the deviation $\Delta = A_{\text{full}} - A_{\text{stat}}$ highlights data-adaptive adjustments beyond the calibration matrix. On Electricity (Fig. 4), $A_{\text{full}}$ shows a structured coupling pattern, while $\Delta$ concentrates on localized corrections. The row sparsity curves $k_{90/95/99}$ indicate that most variables allocate coupling mass to a small subset of sources, suggesting compact cross-series aggregation. The energy spectrum is also concentrated on a few dominant basis directions, providing further evidence that FISC learns compact and interpretable interaction structure. Results on other datasets and random-seed stability diagnostics are provided in the appendix.

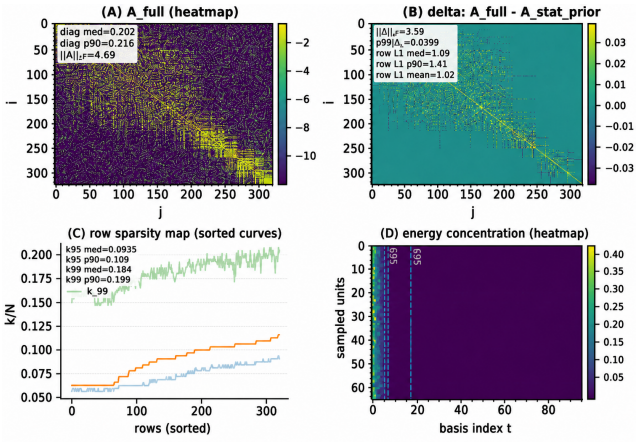



Figure 4: On Electricity, (A) learned coupling $A_{\text{full}}$; (B) deviation $\Delta = A_{\text{full}} - A_{\text{stat}}$; (C) row sparsity $k_{90/95/99}$; (D) energy concentration over basis directions.

### 4.4 Practicality and hyperparameter sensitivity

**Efficiency and practical configuration.**

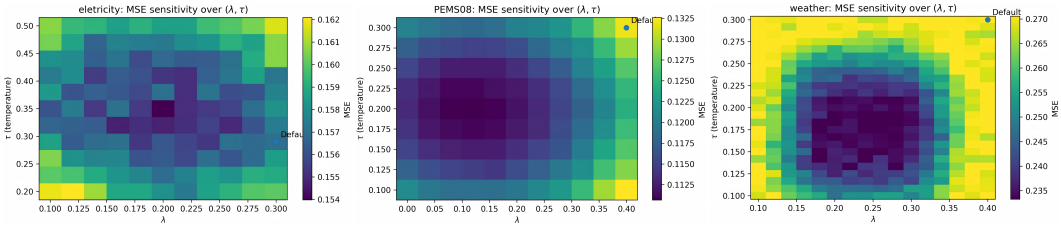
To assess practical deployability, Table 8 reports parameters, FLOPs, and measured training/inference time and memory under a unified ETT setting. FISC adds first-layer statistical calibration constraints and dual-domain temporal modeling with modest overhead. It achieves lower inference time than the attention-based baselines in this comparison and maintains memory usage comparable to lightweight models, supporting a favorable accuracy–cost trade-off.

**Sensitivity to calibration hyperparameters.**

We further examine the sensitivity to calibration strength $\lambda$ and temperature $\tau$, which control the contribution of the statistical affinity and the sharpness of the induced coupling distribution, respectively. As shown in Fig. 5, performance changes smoothly across a broad range of values rather than depending on a narrow optimum. This pattern suggests that FISC is not overly sensitive to delicate hyperparameter tuning, and that the first-layer calibration constraint remains effective under moderate variations in fusion strength and affinity concentration.

Table 8: Computational cost under the ETT setting (input 96, predict 96, batch 16) on an RTX 4060 16GB GPU. T.T./T.M.: training time / peak training memory; I.T./I.M.: inference time / inference memory.

| Model | Params M | FLOPs M | T.T. ms | T.M. GB | I.T. ms | I.M. MB |
|---|---|---|---|---|---|---|
| DLinear | 0.02 | 0.13 | 1.42 | 0.02 | 0.21 | 8.48 |
| TimesNet | 299.94 | 289708 | 501.98 | 5.80 | 155.97 | 1396.24 |
| TimeMixer++ | 1.19 | 1396.90 | 74.74 | 0.27 | 28.15 | 149.58 |
| PatchTST | 3.76 | 272.20 | 8.37 | 0.14 | 1.38 | 51.83 |
| iTransformer | 4.83 | 33.99 | 10.11 | 0.21 | 1.72 | 156.00 |
| TimeMixer | 0.08 | 10.72 | 27.99 | 0.05 | 3.86 | 18.06 |
| OLinear | 4.52 | 33.74 | 8.09 | 0.20 | 1.19 | 156.41 |
| **FISC** | 4.27 | 31.06 | 7.53 | 0.14 | 1.02 | 137.20 |




(a) Electricity (b) PEMS08 (c) Weather

Figure 5: **Sensitivity to SC hyperparameters.** Forecasting performance as a function of fusion strength $\lambda$ and temperature $\tau$, showing smooth trends over a broad range of values.

## 5 Conclusion

We proposed FISC for multivariate forecasting, calibrating only the first cross-series coupling layer with training-set statistics while keeping deeper couplings data-adaptive. Convex fusion of statistical affinity and learnable coupling parameters, followed by row-wise simplex projection, yields bounded and interpretable entrance interactions. A dual-domain temporal representation complements this design by organizing temporal dependencies before cross-variable mixing. Long- and short-term forecasting experiments demonstrate competitive accuracy and efficiency, while ablation, sensitivity, and non-stationarity analyses support first-layer-only calibration. Fixed training statistics may become stale under persistent distribution shift, and standard FISC does not include online recalibration. The signed-VAR stress test also reveals an expressivity cost under strong inhibitory interactions. These limitations motivate controlled signed extensions and causal online recalibration as future work.

## 6 Acknowledgments

This work was supported in part by the following: the National Natural Science Foundation of China under Grant Nos. U24A20328, U24A20219, 62272281, the Natural Science Foundation of Shandong Province under Grant No. ZR2026MS1059, the Youth Innovation Technology Project of Higher School in Shandong Province under Grant No. 2023KJ212, and the Yantai Natural Science Foundation under Grant No. 2024JCYJ034.

## 7 Competing interests.

The authors declare no competing financial or non-financial interests related to this work.

## References


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

Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Congrui Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, et al. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems*, 33:17766–17778, 2020.

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, pages 6989–6997, 2023.

Ruobin Gao, Liang Du, Okan Duru, and Kum Fai Yuen. Time series forecasting based on echo state network and empirical wavelet transformation. *Applied Soft Computing*, 102:107111, 2021.

Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. Timebase: The power of minimalism in efficient long-term time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025.

Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International conference on learning representations*, 2021.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pages 95–104, 2018.

Mengfan Liang, Shixiang Jia, Yepeng Liu, Xiaofeng Zhang, Hua Wang, and Yujuan Sun. Tritracknet: A dual-channel time series forecasting model with multi-path interaction and perturbation optimization. *Neurocomputing*, page 132519, 2025.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International journal of forecasting*, 37(4): 1748–1764, 2021.

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 *The Twelfth International Conference on Learning Representations*, 2024.

Jinghao Lu, Fan Zhang, Xiaofeng Zhang, Yujuan Sun, and Hua Wang. Mcnr: Multiscale feature-based latent data component extraction linear regression model. *Expert Systems with Applications*, page 128634, 2025.

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

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2400–2410, 2024.

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. *arXiv preprint arXiv:2405.14616*, 2024a.

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 *ICLR*, 2025.

Xue Wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. Card: Channel aligned robust blend transformer for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 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 *The Eleventh International Conference on Learning Representations*, 2023.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. Graph wavenet for deep spatial-temporal graph modeling. *arXiv preprint arXiv:1906.00121*, 2019.

Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 753–763, 2020.

Canran Xiao and Liwei Hou. Prototype-aligned federated soft-prompts for continual web personalization. In *Proceedings of the ACM Web Conference 2026*, pages 6743–6754, 2026.

Canran Xiao, Tianxiang Xu, Siyuan Ma, Yiyang Jiang, Haoyu Gao, and Yuhan Wu. Reversible primitive–composition alignment for continual vision–language learning. In *International Conference on Learning Representations*, volume 2026, pages 88700–88722, 2026.

Chen Xu and Yao Xie. Sequential predictive conformal inference for time series. In *International Conference on Machine Learning*, pages 38707–38727. PMLR, 2023.

Mengzhu Xu, Hanzhi Liu, Ningkang Peng, Qianyu Chen, and Canran Xiao. Affordance-first decomposition for continual learning in video-language understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3908–3919, June 2026.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. *arXiv preprint arXiv:2307.03756*, 2023.

Jingquan Yan and Hao Wang. Self-interpretable time series prediction with counterfactual explanations. In *International Conference on Machine Learning*, pages 39110–39125. PMLR, 2023.

Zhen Yao, Bo Lang, Zhihao Zheng, Haoyan Xu, Zhengtao Yao, Zhikang Dong, and Mooi Choo Chuah. From semantic to reasoing: A survey of language-guided segmentation in the multimodal era. 2026.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. *Advances in Neural Information Processing Systems*, 37:55115–55140, 2024.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.

Guoqi Yu, Jing Zou, Xiaowei Hu, Angelica I Aviles-Rivero, Jing Qin, and Shujun Wang. Revitalizing multivariate time series forecasting: Learnable decomposition with inter-series dependencies and intra-series variations modeling. In *International Conference on Machine Learning*, pages 57818–57841. PMLR, 2024.

Wenzhen Yue, Yong Liu, Haoxuan Li, Hao Wang, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. Olinear: A linear model for time series forecasting in orthogonally transformed domain. *arXiv preprint arXiv:2505.08550*, 2025.

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, pages 11121–11128, 2023.

Jiayu Zhang, Chuangxin Zhao, Canran Xiao, Ruibo Duan, Wenyi Mo, Haoyu Gao, and Wenshuo Wang. Pi-cca: Prompt-invariant cca certificates for replay-free continual multimodal learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*, 2023.

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, pages 11106–11115, 2021.

Heng Zhou, Jing Tang, Jusheng Zhang, Yanshu Li, Canran Xiao, Liwei Hou, Zong Ke, and Jiawei Yao. Comem: Compositional concept-graph memory for vision–language adaptation. In *International Conference on Learning Representations*, volume 2026, pages 20009–20032, 2026.

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*, pages 27268–27286. PMLR, 2022.

# A Theoretical interpretation of First-layer Statistical Calibration constraints

This appendix provides a simplified optimization analysis of the first-layer statistical calibration constraints in FISC. The goal is not to prove global convergence of the full nonlinear forecasting model, but to isolate their local effect on the entrance cross-series coupling layer. Under a quadratic surrogate aligned with CSCO, we show that first-layer calibration can reduce the initial distance between the effective entrance coupling and a useful constrained coupling, thereby tightening the prefactor in the projected-gradient convergence bound.

**Surrogate setup.** Let $\mathbf{z} \in \mathbb{R}^N$ denote the correlation-aligned input representation used by the first CSCO layer, and consider the linear surrogate

$$\mathbf{y} = \mathbf{A}^\star \mathbf{z} + \boldsymbol{\xi}, \qquad \mathbb{E}[\boldsymbol{\xi} \mid \mathbf{z}] = 0. \tag{9}$$

Let $\boldsymbol{\Sigma}_z = \mathbb{E}[\mathbf{z}\mathbf{z}^\top]$ satisfy

$$0 < \mu \le \lambda_{\min}(\boldsymbol{\Sigma}_z) \le \lambda_{\max}(\boldsymbol{\Sigma}_z) \le L < \infty.$$

We study the population loss

$$\mathcal{L}(\mathbf{A}) = \frac{1}{2}\mathbb{E}\left[\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2\right], \tag{10}$$

over the row-stochastic feasible set $\mathcal{C} = \{\mathbf{A} \in \mathbb{R}^{N\times N} : \mathbf{A}\mathbf{1} = \mathbf{1}, \mathbf{A} \ge 0\}$. Let $\mathbf{A}^{\mathcal{C}} = \arg\min_{\mathbf{A}\in\mathcal{C}} \mathcal{L}(\mathbf{A})$ be the constrained optimum. Projected gradient descent updates the effective coupling by

$$\mathbf{A}_{m+1} = \Pi_{\mathcal{C}}\left(\mathbf{A}_m - \eta \nabla \mathcal{L}(\mathbf{A}_m)\right), \tag{11}$$

where $\Pi_{\mathcal{C}}$ denotes row-wise Euclidean projection onto the probability simplex.

**Statistical calibration constraint.** Let $\mathbf{A}_{\mathrm{stat}}$ be the population calibration matrix induced by training-time cross-series statistics, and let $\widehat{\mathbf{A}}_{\mathrm{stat}}$ be its empirical estimate from training windows. Assume that the statistics-to-coupling map is stable enough that, with high probability,

$$\|\widehat{\mathbf{A}}_{\mathrm{stat}} - \mathbf{A}_{\mathrm{stat}}\|_F \le \epsilon_{\mathrm{stat}}. \tag{12}$$

Define the calibration mismatch to the constrained entrance optimum as

$$\delta_{\mathrm{mis}} = \|\mathbf{A}_{\mathrm{stat}} - \mathbf{A}^{\mathcal{C}}\|_F. \tag{13}$$

FISC forms the effective first-layer coupling as

$$\mathbf{A}_0^{\mathrm{SC}} = \Pi_{\mathcal{C}}\left(\lambda \widehat{\mathbf{A}}_{\mathrm{stat}} + (1-\lambda)\mathbf{Q}_0\right), \qquad \lambda \in [0,1], \tag{14}$$

where $\mathbf{Q}_0 = \psi(\mathbf{W}_0^{(1)})$ is the nonnegative learnable component before projection.

**Theorem A.1** (Entrance calibration reduces the optimization prefactor)**.** *Consider the projected gradient updates above with step size $0 < \eta \le 2/(L+\mu)$. Let $\rho = \max\{|1-\eta\mu|, |1-\eta L|\} < 1$. Then the projected iterates satisfy*

$$\|\mathbf{A}_m - \mathbf{A}^{\mathcal{C}}\|_F \le \rho^m \|\mathbf{A}_0 - \mathbf{A}^{\mathcal{C}}\|_F. \tag{15}$$

*Moreover, the first-layer calibration in Eq.* (14) *satisfies*

$$\|\mathbf{A}_0^{\mathrm{SC}} - \mathbf{A}^{\mathcal{C}}\|_F \le \lambda\left(\epsilon_{\mathrm{stat}} + \delta_{\mathrm{mis}}\right) + (1-\lambda)\|\mathbf{Q}_0 - \mathbf{A}^{\mathcal{C}}\|_F. \tag{16}$$

*Thus, when the empirical statistic is reliable and the calibration mismatch is moderate, first-layer calibration can reduce the convergence prefactor relative to an uncalibrated learnable coupling, while the contraction rate $\rho$ remains governed by the conditioning of the surrogate objective.*

*Proof.* The loss is quadratic in $\mathbf{A}$. Since

$$\nabla \mathcal{L}(\mathbf{A}) = (\mathbf{A} - \mathbf{A}^\star)\boldsymbol{\Sigma}_z,$$

the gradient is $L$-Lipschitz and $\mu$-strongly monotone in the Frobenius norm over the linear space of coupling matrices. The projection $\Pi_{\mathcal{C}}$ is non-expansive because $\mathcal{C}$ is closed and convex. Since $\mathbf{A}^{\mathcal{C}}$ is the constrained minimizer, it is a fixed point of the projected gradient mapping:

$$\mathbf{A}^{\mathcal{C}} = \Pi_{\mathcal{C}}\left(\mathbf{A}^{\mathcal{C}} - \eta \nabla \mathcal{L}(\mathbf{A}^{\mathcal{C}})\right).$$

Therefore,

$$\|\mathbf{A}_{m+1} - \mathbf{A}^{\mathcal{C}}\|_F \le \left\|(\mathbf{A}_m - \mathbf{A}^{\mathcal{C}}) - \eta\left(\nabla\mathcal{L}(\mathbf{A}_m) - \nabla\mathcal{L}(\mathbf{A}^{\mathcal{C}})\right)\right\|_F.$$

Using the quadratic form of the gradient,

$$\nabla\mathcal{L}(\mathbf{A}_m) - \nabla\mathcal{L}(\mathbf{A}^{\mathcal{C}}) = (\mathbf{A}_m - \mathbf{A}^{\mathcal{C}})\mathbf{\Sigma}_z.$$

Since the eigenvalues of $\mathbf{\Sigma}_z$ lie in $[\mu, L]$, the linear operator $\mathbf{M} \mapsto \mathbf{M}(\mathbf{I} - \eta\mathbf{\Sigma}_z)$ has operator norm at most $\rho = \max\{|1-\eta\mu|,\ |1-\eta L|\}$. Thus,

$$\|\mathbf{A}_{m+1} - \mathbf{A}^{\mathcal{C}}\|_F \le \rho\|\mathbf{A}_m - \mathbf{A}^{\mathcal{C}}\|_F,$$

and iterating gives Eq. (15).

It remains to bound the first-layer calibration error. Since $\mathbf{A}^{\mathcal{C}} \in \mathcal{C}$ and Euclidean projection is non-expansive,

$$\begin{aligned}\|\mathbf{A}_0^{\mathrm{SC}} - \mathbf{A}^{\mathcal{C}}\|_F &= \left\|\Pi_{\mathcal{C}}\left(\lambda\widehat{\mathbf{A}}_{\mathrm{stat}} + (1-\lambda)\mathbf{Q}_0\right) - \Pi_{\mathcal{C}}(\mathbf{A}^{\mathcal{C}})\right\|_F \\ &\le \lambda\|\widehat{\mathbf{A}}_{\mathrm{stat}} - \mathbf{A}^{\mathcal{C}}\|_F + (1-\lambda)\|\mathbf{Q}_0 - \mathbf{A}^{\mathcal{C}}\|_F.\end{aligned}$$

By the triangle inequality,

$$\|\widehat{\mathbf{A}}_{\mathrm{stat}} - \mathbf{A}^{\mathcal{C}}\|_F \le \|\widehat{\mathbf{A}}_{\mathrm{stat}} - \mathbf{A}_{\mathrm{stat}}\|_F + \|\mathbf{A}_{\mathrm{stat}} - \mathbf{A}^{\mathcal{C}}\|_F \le \epsilon_{\mathrm{stat}} + \delta_{\mathrm{mis}}.$$

Substituting this into the previous inequality proves Eq. (16). Combining Eq. (15) and Eq. (16) completes the proof. □

**Implication.** Theorem A.1 explains the local role of first-layer statistical calibration constraints in FISC. These constraints do not change the conditioning-controlled contraction rate $\rho$ of the surrogate problem. Instead, they can reduce the initial distance to a useful constrained coupling, which tightens the multiplicative prefactor in Eq. (15). This interpretation is consistent with the empirical observation that first-layer calibration improves optimization stability. The same bound also clarifies the mechanism boundary: the benefit weakens when the empirical statistic is unreliable or mismatched, motivating moderate calibration strength and avoiding repeated calibration constraints in deeper layers.

# B Failure modes and boundary analysis of FISC

This appendix complements the theoretical analysis in App. A. While Theorem A.1 explains why first-layer calibration can improve the entrance coupling under a simplified surrogate, this section clarifies when the assumptions behind that benefit may weaken. The goal is not to claim that training-set statistics are universally reliable, but to specify when FISC is expected to help, when it may become less effective, and which design choices mitigate these risks.

## B.1 Boundary of First-layer Statistical Calibration

Theorem A.1 shows that the advantage of first-layer calibration depends on two quantities: the estimation error of the empirical calibration matrix and its mismatch to a useful constrained entrance coupling. Therefore, calibration is most helpful when the training statistic is reliably estimated and remains close to the cross-series structure needed at the model entrance.

This condition can fail in two ways. First, if the number of training windows is small or the series are highly noisy, $\widehat{\mathbf{A}}_{\mathrm{stat}}$ may be unstable. Second, if deployment-time dependencies differ substantially from training-time dependencies, the calibration matrix may become systematically mismatched. In either case, a large calibration strength $\lambda$ may over-constrain the first layer.

FISC addresses this boundary by using calibration as a soft entrance-stage constraint rather than a hard coupling rule. The first coupling layer combines the calibration matrix with a learnable correction, and all deeper CSCO layers are learned without repeated calibration constraints. Thus, training statistics constrain the interaction near the input, where they are most meaningful, while deeper representations remain data-adaptive.

### B.2 Cross-series failure modes

**Persistent calibration mismatch.** A direct failure mode occurs when training-time cross-series statistics no longer represent test-time dependencies. This may happen under abrupt regime changes, sensor replacement, market reconfiguration, policy changes, or long-term concept drift. In such cases, the calibration matrix may no longer provide a useful entrance constraint.

FISC mitigates this risk by restricting calibration to the first layer and by controlling its strength through $\lambda$. Since deeper CSCO layers do not reuse the same calibration matrix, mismatch is not repeatedly imposed through depth. However, under severe and persistent drift, fixed calibration may still be insufficient. A natural extension is online or uncertainty-aware first-layer recalibration based only on past observations.

**Repeated calibration of transformed representations.** Applying the same training-set statistic to multiple layers may appear to strengthen calibration, but it can reduce adaptability. Deeper layers operate on transformed representations whose geometry may no longer match the raw cross-series correlation matrix. Reusing the same calibration matrix at these layers can restrict the data-adaptive degrees of freedom needed for task-specific representations.

This boundary explains the NoSC, LastSC, and AllSC ablations. NoSC removes the entrance calibration, LastSC applies calibration after representations have already evolved, and AllSC repeatedly applies the same calibration matrix across depth. Their weaker performance supports the design principle that statistical calibration constraints are most useful at the entrance, while deeper cross-series interactions should remain data-adaptive.

**Limits of simplex-constrained coupling.** CSCO constrains each coupling row to the probability simplex, producing bounded nonnegative aggregation. Although hidden features and output projections remain signed, the coupling weights cannot explicitly represent inhibitory edges.

To measure this restriction, we construct a controlled signed variant, SignedRes. For each layer, let

$$V_{ij}^{(\ell)} = \tanh\big(W_{ij}^{(\ell)}\big) - \frac{1}{N}\sum_{k=1}^{N}\tanh\big(W_{ik}^{(\ell)}\big), \qquad E_{ij}^{(\ell)} = \frac{V_{ij}^{(\ell)}}{\sum_{k=1}^{N}|V_{ik}^{(\ell)}| + \varepsilon}.$$

SignedRes uses

$$\mathbf{A}_{\mathrm{SR}}^{(\ell)} = \mathbf{A}_{\mathrm{FISC}}^{(\ell)} + \beta\mathbf{E}^{(\ell)}.$$

Since each row of $\mathbf{E}^{(\ell)}$ sums to zero, the resulting coupling retains unit row sums while allowing bounded negative entries. The construction reuses $\mathbf{W}^{(\ell)}$ and introduces no additional $N \times N$ learnable matrix. The coefficient $\beta$ is selected on validation data from $\{0.02, 0.05, 0.10, 0.20\}$.

The signed-VAR results in Table 7 show that the nonnegative constraint has a measurable expressivity cost when inhibitory interactions become strong. Its benefit should therefore be understood as a bounded aggregation bias with a specific scope, rather than universal robustness to signed dependencies. Applications known to contain substantial inhibitory cross-series structure may benefit from relaxing nonnegativity through a controlled signed extension. SignedRes is a diagnostic variant and is not part of the standard FISC formulation.

### B.3 Boundary of Dual-domain temporal modeling

The orthogonal-domain branch is designed for time series with redundant autocorrelated temporal structure. Let $\mathbf{R}^{(t)} = \mathbf{U}^{(t)}\mathbf{\Lambda}^{(t)}\mathbf{U}^{(t)\top}$ be the temporal correlation matrix estimated from the training split. When the spectrum is concentrated, the effective rank $r_{\mathrm{eff}} = \mathrm{Tr}(\mathbf{R}^{(t)})/\|\mathbf{R}^{(t)}\|_2$ is much smaller than the input length $T$, and the orthogonal branch can reorganize dominant temporal regularities into a compact basis.

A standard complexity view illustrates this boundary. For an $\ell_2$-bounded single-path temporal map,

$$\mathfrak{R}_n(\mathcal{G}_{\mathrm{base}}) \lesssim \frac{B}{\sqrt{n}}\sqrt{\mathrm{Tr}(\mathbf{\Lambda}^{(t)})}.$$

For a dual-path map that allocates capacity to the leading $k$ temporal modes and a residual subspace,

$$\mathfrak{R}_n(\mathcal{G}_{\mathrm{dual}}) \lesssim \frac{1}{\sqrt{n}}\left(B_{\mathrm{orth}}\sqrt{\mathrm{Tr}(\mathbf{\Lambda}_k^{(t)})} + B_{\mathrm{res}}\sqrt{\mathrm{Tr}(\mathbf{\Lambda}_\perp^{(t)})}\right).$$

Thus, the dual-domain design is most useful when the dominant temporal energy is concentrated in a small number of modes, i.e., $\mathrm{Tr}(\mathbf{\Lambda}_{\perp}^{(t)}) \ll \mathrm{Tr}(\mathbf{\Lambda}_{k}^{(t)})$. If the temporal spectrum is nearly flat, highly irregular, or rapidly changing, the fixed orthogonal basis provides less benefit.

This boundary also explains why FISC keeps the time-domain branch. The orthogonal branch captures global correlated temporal structure, while the time-domain branch preserves local and residual patterns that may not align with the dominant temporal basis. Removing either branch weakens the model's coverage of one temporal regime.

### B.4 Temporal basis mismatch

The temporal basis $\mathbf{U}^{(t)}$ is estimated only from the training split and kept fixed to avoid test-time leakage. This is appropriate when temporal autocorrelation patterns are relatively stable, but it can become limiting under abrupt seasonality changes, intervention-driven dynamics, or irregular event bursts. In these cases, the dominant temporal modes at test time may differ from those learned from training data.

FISC partially mitigates this through the time-domain branch, which remains directly tied to the observed sequence and can model local deviations. Nevertheless, under persistent temporal drift, adaptive basis updates may be beneficial. Such updates should rely only on past observations and should be controlled by a drift detector to avoid leakage and instability.

### B.5 Practical diagnostics

The above analysis suggests several diagnostics for applying FISC.

**Calibration reliability.** Estimate $\widehat{\mathbf{A}}_{\mathrm{stat}}$ on different training subsets. Large variability suggests that the calibration matrix is unstable and that a smaller $\lambda$ may be preferable.

**Calibration mismatch.** Inspect the deviation map $\Delta = \mathbf{A}_{\mathrm{full}} - \mathbf{A}_{\mathrm{stat}}$. Localized deviations suggest that the learnable component performs targeted correction around a useful calibration matrix. Large diffuse deviations suggest poor calibration match.

**Sensitivity to $\lambda$ and $\tau$.** A broad low-error region in the $\lambda$–$\tau$ sensitivity map indicates robust calibration. A sharp optimum suggests stronger dependence on calibration hyperparameters and should be reported carefully.

**Structural-drift stress tests.** Regime-shift or structural-drift experiments directly evaluate whether fixed calibration remains useful under non-stationarity. Graceful degradation supports the first-layer-only design, while severe degradation indicates the need for weaker calibration or online recalibration.

**Temporal spectrum concentration.** The spectrum of $\mathbf{R}^{(t)}$ diagnoses the usefulness of the orthogonal branch. Strong energy concentration supports dual-domain temporal modeling, whereas a flat spectrum suggests greater reliance on the time-domain branch.

### B.6 Scope of the analysis

This appendix provides a boundary analysis rather than a full theory of the nonlinear FISC architecture. The optimization result in App. A isolates the entrance coupling layer and explains why first-layer calibration can reduce the effective initialization error under simplex constraints. The temporal-complexity discussion isolates the benefit of the orthogonal branch under low-effective-rank autocorrelation. These simplified views do not prove global convergence or universal generalization.

Overall, FISC is expected to be most effective when training statistics provide a useful but imperfect entrance calibration, temporal dependencies contain nontrivial autocorrelation structure, and distribution shifts are moderate enough that deeper learnable layers can adapt after the calibrated entrance interaction. Its main limitations arise under severe persistent drift, unreliable statistical estimates, coupling patterns dominated by direct signed interactions, or temporal regimes whose dominant basis changes rapidly.

## C Datasets and evaluation protocol

This appendix details the datasets and evaluation protocols used in the long- and short-term forecasting experiments. The benchmarks cover electricity, traffic, weather, health, finance, web traffic, energy, and socioeconomic time series, enabling evaluation across different dimensionalities, sampling frequencies, and degrees of non-stationarity. Dataset statistics, splits, and forecasting horizons are summarized in Table 9. All statistics used by FISC, including cross-series and temporal correlation matrices, are estimated only from the training split and kept fixed during validation and testing to avoid leakage.

**Long-term forecasting benchmarks.** The long-term setting follows the standard protocol with look-back length $T = 96$ and forecasting horizons $H \in \{96, 192, 336, 720\}$, except for PEMS datasets where $H \in \{12, 24, 48, 96\}$. The ETT family Zhou et al. [2021] contains power transformer monitoring data with 7 variables, including hourly ETTh1/ETTh2 and 15-minute ETTm1/ETTm2 subsets. Weather Wu et al. [2021] contains 21 meteorological variables recorded every 10 minutes. Electricity (ECL) Zhou et al. [2021] records hourly electricity consumption from 321 customers. Traffic Wu et al. [2021] contains hourly road occupancy from 862 Bay Area sensors. Exchange Zhou et al. [2021] provides daily exchange rates from 8 countries. Solar-Energy Lai et al. [2018] contains 10-minute solar power generation from 137 photovoltaic plants. The PEMS family contains 5-minute California traffic sensor data, including PEMS03, PEMS04, PEMS07, and PEMS08.

**Short-term forecasting benchmarks.** The short-term setting uses two horizon groups: S with input length 12 and prediction horizons $\{3, 6, 9, 12\}$, and L with input length 36 and prediction horizons $\{24, 36, 48, 60\}$. ILI Wu et al. [2021] contains weekly influenza-like illness statistics released by the U.S. CDC. COVID-19 contains daily hospitalization records in California. METR-LA contains 5-minute traffic speed measurements from 207 loop detectors in Los Angeles. NASDAQ contains daily financial and macroeconomic indicators. Wiki contains daily page-view series of selected Wikipedia articles. SP500 and DowJones contain daily financial market series. CarSales, Power, Website, and Unemp cover market demand, renewable generation, website traffic, and unemployment-rate forecasting, respectively. These datasets test whether first-layer statistical calibration constraints transfer beyond standard long-horizon benchmarks to smaller, noisier, and more heterogeneous regimes. For nonstandard datasets, data sources, preprocessing scripts, and split files are included in the released code.

**Preprocessing and metrics.** For each dataset, train/validation/test splits follow Table 9. All models are evaluated under the same splits, input lengths, forecasting horizons, and metrics. MSE and MAE are the primary error metrics. We additionally report coefficient of determination ($R^2$), Pearson correlation ($r$), and MASE for scale-free evaluation across datasets with different units and variability. Normalization statistics and FISC calibration matrices are computed from the training split only.

## D Implementation details

All experiments are implemented in PyTorch and run on a single NVIDIA RTX 4060 GPU with 16GB memory. FISC is trained with Adam. The initial learning rate is selected from $\{3 \times 10^{-4}, 4 \times 10^{-4}, 5 \times 10^{-4}\}$ according to validation performance. The hidden dimension $D$ is chosen from $\{128, 256, 512\}$, and the embedding dimension used in the input-side statistical calibration and CSCO module is fixed to $d = 16$. The batch size is selected from $\{4, 8, 16, 32, 64\}$ depending on dataset scale and memory usage. The number of stacked CSCO blocks is denoted by $L$ and selected from $\{2, 3, 4, 5\}$.

For statistical calibration, the cross-series and temporal correlation matrices are computed only from the training split and kept fixed during validation and testing. The fusion strength $\lambda$ and temperature $\tau$ are selected using validation performance from $\lambda \in \{0.2, 0.4, 0.6, 0.8\}$ and $\tau \in \{0.1, 0.2, 0.5, 1.0\}$, respectively, where larger $\tau$ produces a smoother and more uniform calibration matrix. Once selected, the same values are fixed across all horizons of a dataset.

Models are trained for at most 50 epochs with early stopping. Training is terminated if the validation metric does not improve for 10 consecutive epochs. We use the same weighted $L_1$ training loss as CARD Wang et al. [2024b], while reporting MSE and MAE as the primary evaluation metrics. For

Table 9: **Dataset summary and forecasting horizons.** $N$ denotes the number of variables; Len. denotes the number of time steps. S denotes short-horizon evaluation with input length 12 and horizons $\{3, 6, 9, 12\}$; L denotes longer short-term evaluation with input length 36 and horizons $\{24, 36, 48, 60\}$.

| Dataset | $N$ | Freq. | Len. | Split | Horizons | Domain |
|---|---|---|---|---|---|---|
| ETTh1/ETTh2 | 7 | 1h | 17420 | 6:2:2 | $\{96, 192, 336, 720\}$ | Electricity |
| ETTm1/ETTm2 | 7 | 15min | 69680 | 6:2:2 | $\{96, 192, 336, 720\}$ | Electricity |
| Weather | 21 | 10min | 52696 | 7:1:2 | $\{96, 192, 336, 720\}$ | Weather |
| Electricity | 321 | 1h | 26304 | 7:1:2 | $\{96, 192, 336, 720\}$ | Electricity |
| Traffic | 862 | 1h | 17544 | 7:1:2 | $\{96, 192, 336, 720\}$ | Transportation |
| Exchange | 8 | 1d | 7588 | 7:1:2 | $\{96, 192, 336, 720\}$ | Economy |
| Solar-Energy | 137 | 10min | 52560 | 7:1:2 | $\{96, 192, 336, 720\}$ | Energy |
| PEMS03 | 358 | 5min | 26209 | 6:2:2 | $\{12, 24, 48, 96\}$ | Transportation |
| PEMS04 | 307 | 5min | 16992 | 6:2:2 | $\{12, 24, 48, 96\}$ | Transportation |
| PEMS07 | 883 | 5min | 28224 | 6:2:2 | $\{12, 24, 48, 96\}$ | Transportation |
| PEMS08 | 170 | 5min | 17856 | 6:2:2 | $\{12, 24, 48, 96\}$ | Transportation |
| ILI | 7 | 1w | 966 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Health |
| COVID-19 | 55 | 1d | 335 | 7:1:2 / 6:2:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Health |
| METR-LA | 207 | 5min | 34272 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Transportation |
| NASDAQ | 12 | 1d | 3914 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Finance |
| Wiki | 99 | 1d | 730 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Web |
| SP500 | 5 | 1d | 8077 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Finance |
| DowJones | 27 | 1d | 6577 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Finance |
| CarSales | 10 | 1d | 6728 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Market |
| Power | 2 | 1d | 1186 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Energy |
| Website | 4 | 1d | 2167 | 7:1:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Web |
| Unemp | 53 | 1mo | 531 | 6:2:2 | S:$\{3, 6, 9, 12\}$; L:$\{24, 36, 48, 60\}$ | Society |

baselines, we use public implementations whenever available and evaluate them under the same data splits, input lengths, horizons, and metrics. When reported numbers are taken from prior papers, we only use results that follow the same benchmark protocol; otherwise, we rerun the official code under our evaluation setting.

# E Additional qualitative and robustness diagnostics

This appendix provides visualization-based diagnostics that complement the averaged forecasting results in the main text. The goal is to examine how FISC organizes cross-variable coupling and how it behaves under abrupt non-stationarity or calibration mismatch. These analyses are not used for model selection; they provide diagnostic evidence for the first-layer statistical calibration constraints and the dual-domain temporal design.

## E.1 Calibration mismatch and structural drift

A natural boundary of FISC arises when the training-set statistics used for first-layer calibration become mismatched at test time. This may happen not only under marginal distribution shift, but also when the cross-variable dependency structure itself changes abruptly. To probe this boundary, we construct a controlled stress test that varies (i) the mismatch intensity of the calibration matrix and (ii) the magnitude of structural drift. We compare FISC with calibration-placement variants (NoSC, LastSC, and AllSC) to isolate the effect of applying statistical calibration only at the first layer.

**Controlled calibration mismatch.** Let $A_{\mathrm{stat}}$ denote the clean calibration matrix estimated from the training split. To create a controlled mismatch, we perturb it before model evaluation:

$$A_{\mathrm{stat}}^{(\gamma)} = \Pi_{\mathcal{C}}\big((1-\gamma)A_{\mathrm{stat}} + \gamma A_{\mathrm{mis}}\big), \qquad \gamma \in [0, 1], \tag{17}$$

where $A_{\mathrm{mis}}$ is a mismatched calibration matrix obtained by disrupting the original cross-series correspondence, and $\Pi_{\mathcal{C}}$ denotes row-wise simplex projection. The mismatch level $\gamma$ controls the deviation from the training-time calibration: $\gamma = 0$ recovers the original matrix, while larger $\gamma$ applies a more mismatched calibration constraint. The same perturbed matrix is used for all calibration-placement variants. NoSC does not use this calibration term.

**Structural drift construction and measurement.** We evaluate models around event windows centered at drift onsets. A candidate onset $t_0$ is detected by a label-free change score computed from adjacent temporal windows, capturing abrupt changes in level, scale, or local dynamics. For each prediction origin, we measure the realized structural drift magnitude from the normalized ground-truth trajectory:

$$\mathrm{Shift}(t) = \max_{1\le h\le H} \max_{1\le i\le N} \left| y_{t+h,i} - y_{t+h-1,i} \right|, \tag{18}$$

where $H$ is the forecasting horizon and all values are measured after training-split normalization. Thus, Shift is not learned; it is the maximum realized ground-truth change over the forecast horizon and is used to order test cases by drift severity.

**Robustness metric.** For the same prediction origin, we compute the worst post-event forecasting error:

$$\mathrm{ErrPeak}(t) = \max_{1\le h\le H} \max_{1\le i\le N} \left| \hat{y}_{t+h,i} - y_{t+h,i} \right|. \tag{19}$$

The error window is fixed for all methods within each dataset and horizon. ErrPeak is always computed over the same post-event forecasting horizon $H$, and the surrounding event-window length is used only for visualization and sample selection. We do not tune the window length or error range separately for different models.

We then plot ErrPeak against Shift. This gives a compact robustness diagnostic: a flatter ErrPeak–Shift curve indicates more graceful degradation as calibration mismatch and structural drift increase. This test is not intended to simulate a full deployment protocol; rather, it isolates whether the placement of statistical calibration constraints affects boundary behavior under controlled mismatch.

Figs. 6–7 show that FISC generally maintains lower ErrPeak and a flatter ErrPeak–Shift trend across datasets. This suggests that first-layer statistical calibration constraints help stabilize entrance coupling even when training-time statistics are imperfectly matched at test time. By contrast, removing calibration entirely (NoSC) weakens entrance conditioning, while applying calibration at deeper layers or throughout the network (LastSC/AllSC) tends to produce steeper degradation, consistent with reduced adaptability after representations have evolved away from raw-variable statistics.

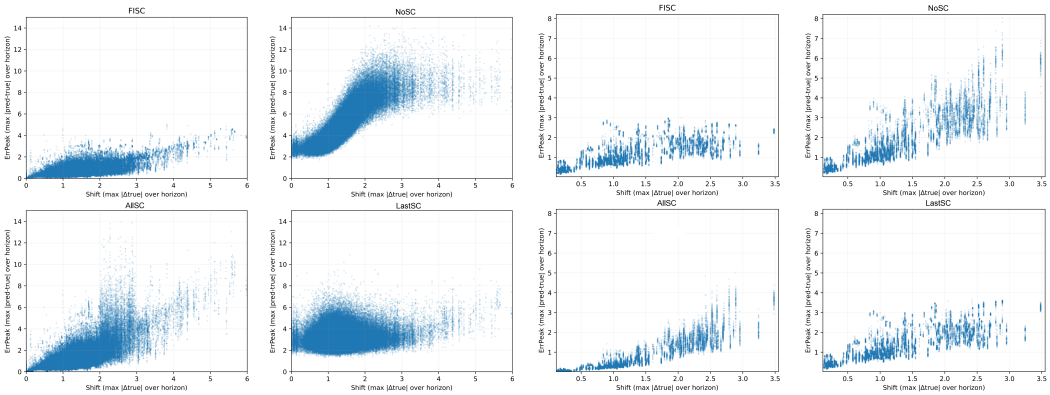
### E.2 Qualitative decomposition and interpretability

FISC makes the entrance coupling inspectable relative to the training-set calibration matrix, while reorganizing temporal dependencies through a data-adaptive orthogonal basis. Figs. 8–9 extend the main-text Electricity visualization to additional datasets.

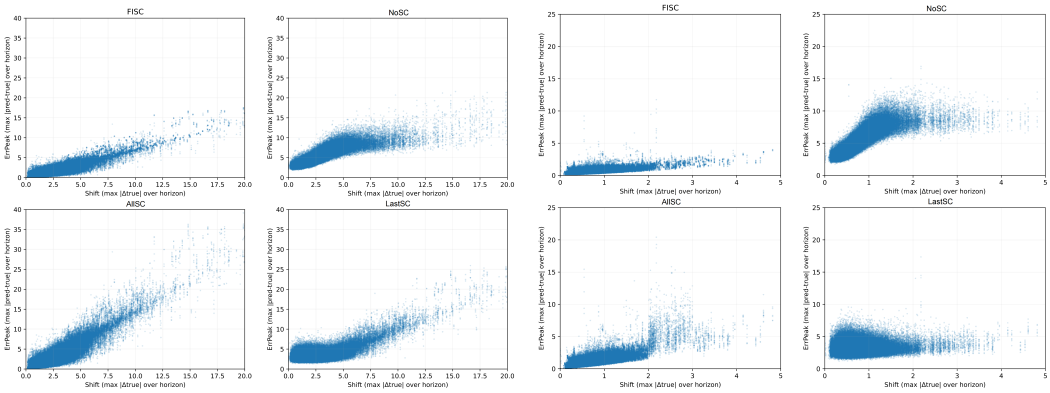
In each internal panel, (A) shows the learned row-stochastic coupling matrix $A_{\mathrm{full}}$, where block- or band-like patterns indicate non-uniform dependency allocation rather than diffuse variable mixing. Panel (B) visualizes the signed deviation $\Delta A = A_{\mathrm{full}} - A_{\mathrm{stat}}$; its localized structure suggests that FISC performs targeted data-adaptive correction around the calibration matrix. Panel (C) reports row-wise sparsity via $k$-at-mass, i.e., the minimum number of source variables needed to accumulate a fixed proportion of each row's probability mass. Smaller values indicate more compact cross-series aggregation. Panel (D) shows the energy distribution over orthogonal temporal bases, where concentration on a limited number of dominant bases suggests that the adaptive basis captures major temporal regularities while leaving residual components for local fluctuations. Together, these visualizations support the interpretation that FISC learns compact cross-series coupling and organized temporal structure across diverse datasets.

### E.3 Regime shifts and non-stationarity

The main text highlights Electricity as a representative regime-shift case. Here we provide additional event-window visualizations on ETTh1, Weather, Traffic, and Electricity. These visualizations stress

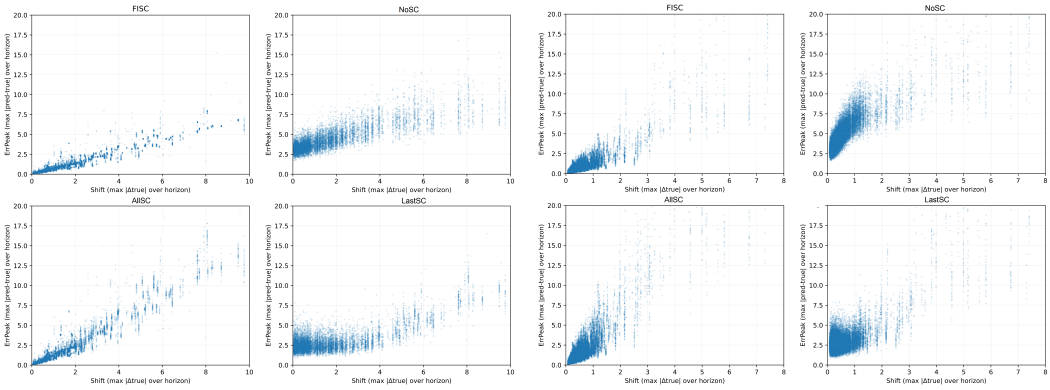



(a) Electricity (b) ETTh1



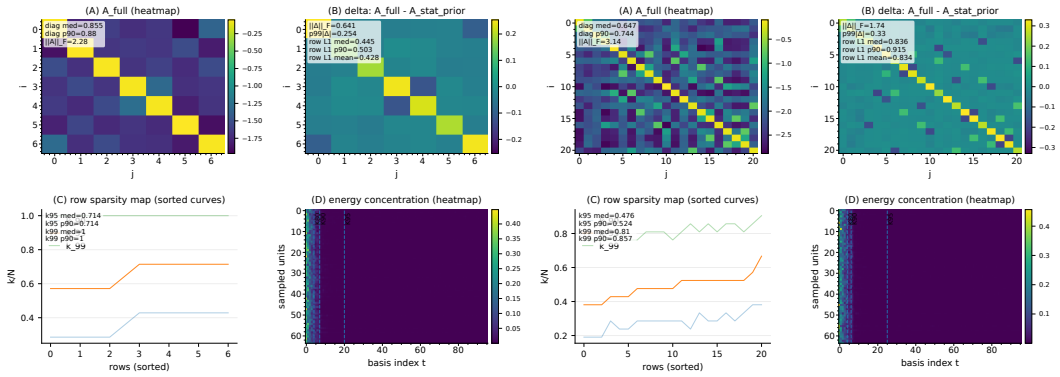

(c) Traffic (d) PEMS03

Figure 6: **Calibration mismatch and structural drift stress test, part I.** Each subfigure plots drift magnitude (Shift) against the worst post-event error peak (ErrPeak) for FISC and its calibration-placement variants (NoSC, LastSC, AllSC). A flatter ErrPeak–Shift curve indicates more graceful degradation when training-time calibration becomes mismatched under structural drift.




(e) Wiki (f) DowJones

Figure 7: **Calibration mismatch and structural drift stress test, part II.** Additional stress-test results on Wiki and DowJones. The same ErrPeak–Shift diagnostic is used to assess whether performance degradation remains controlled as calibration mismatch and structural drift increase.




(a) ETTh1 (b) Weather

Figure 8: **Qualitative decomposition of FISC, part I.** Each internal $2 \times 2$ panel shows: (A) learned row-stochastic coupling $A_{\text{full}}$; (B) deviation from the calibration matrix $\Delta A = A_{\text{full}} - A_{\text{stat}}$; (C) row-wise sparsity measured by $k$-at-mass; and (D) energy concentration over the adaptive orthogonal temporal bases.

a central boundary of FISC: the calibration matrix is estimated from training data and may become mismatched when test-time dynamics change abruptly.

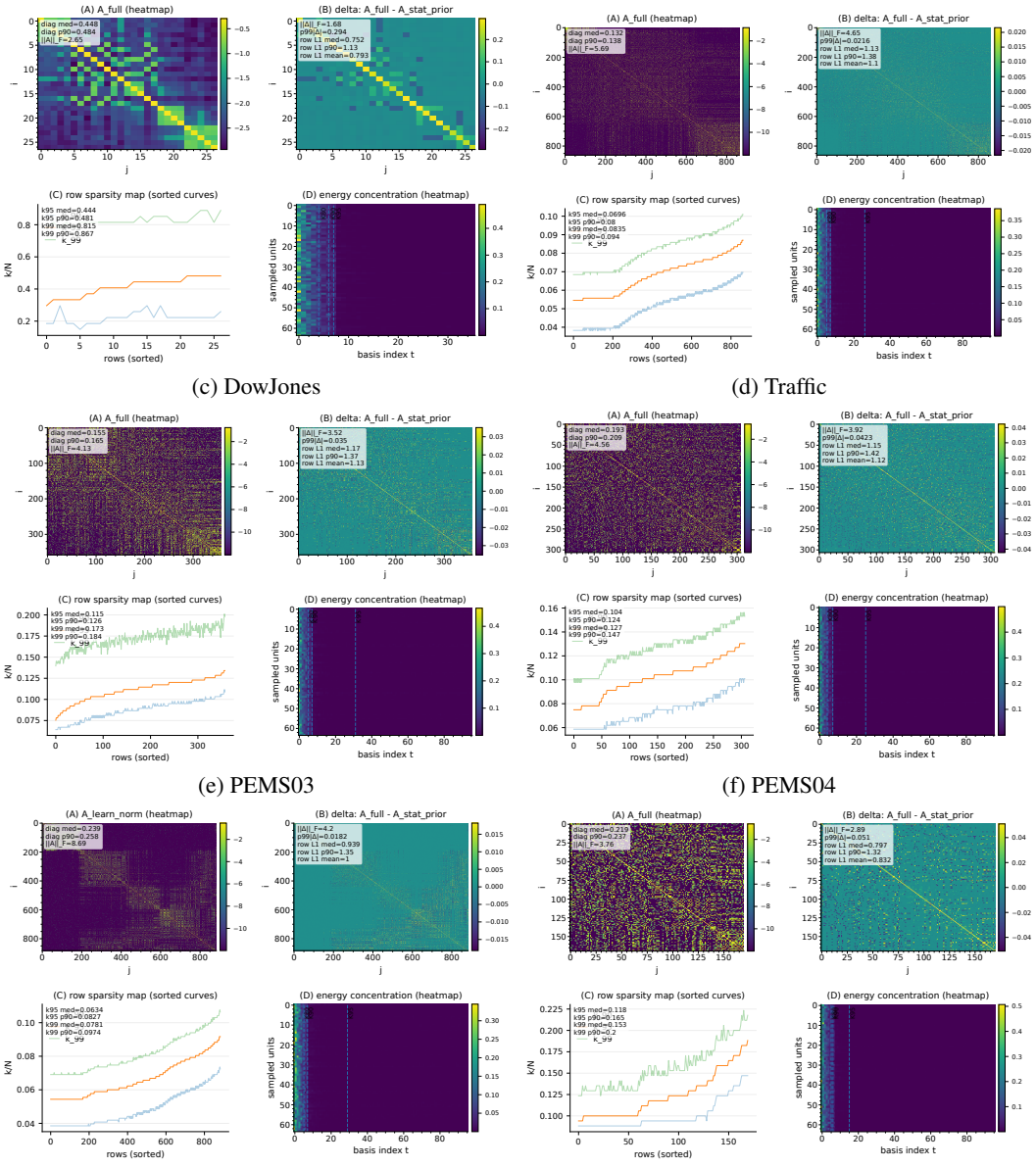
We detect a candidate regime-change timestamp $t_0$ using a lightweight, label-free change-point heuristic based on discrepancies between adjacent temporal windows. This captures abrupt changes in level, scale, or local dynamics without using future labels for model selection. We then form an event window $W = [t_0 - L_{\text{pre}},\, t_0 + L_{\text{post}}]$ and re-center the visualization at $t = 0$. Within the post-event segment, we summarize robustness using three quantities: ErrPeak, the maximum post-event absolute error; Osc, the degree of post-event ringing or oscillation; and Settle, the first sustained return to a low-error band defined from pre-event error statistics. Lower ErrPeak/Osc and smaller Settle indicate smoother adaptation and faster recovery after non-stationary changes.

As shown in Fig. 10, FISC generally exhibits smaller post-change error spikes and smoother recovery around detected regime shifts. This behavior is consistent with the intended role of first-layer statistical calibration constraints: training-set statistics calibrate the entrance interaction, while deeper coupling layers remain learnable rather than being repeatedly constrained by a potentially mismatched calibration matrix. The orthogonal temporal branch further helps reduce multi-scale temporal entanglement near change points. These observations complement the quantitative ablations in the main text and support the boundary-aware interpretation of FISC: calibration is useful at the entrance, whereas deeper representations should remain data-adaptive under non-stationarity.

### E.4 Random-seed stability and paired significance test

Forecasting performance can be affected by random initialization, mini-batch ordering, and stochastic optimization. To verify that the reported gains are not due to favorable seeds, we conduct a multi-seed stability study. For each compared method, we repeat training with $K = 10$ random seeds under the same split, input length, horizon, hyperparameter selection protocol, and early-stopping criterion. The same seed set is shared across methods, so each seed defines a matched comparison between FISC and a baseline or ablation variant.

This evaluation serves three purposes. The mean test error measures average forecasting accuracy, the standard deviation measures seed sensitivity, and the paired significance test evaluates whether the improvement of FISC is consistent under matched stochastic conditions. For each dataset and method, we report the mean and standard deviation of MSE and MAE over valid runs. Let $e^{(k)}_{\text{FISC}}$ and $e^{(k)}_{\text{base}}$ denote the test error of FISC and a compared method under seed $k$. We compute the paired difference as $d^{(k)} = e^{(k)}_{\text{base}} - e^{(k)}_{\text{FISC}}$, where $k = 1, \ldots, K$. A positive mean indicates lower error for FISC on average. We conduct paired tests on these seed-wise differences, using a paired $t$-test when the differences are approximately symmetric and additionally checking the Wilcoxon signed-rank

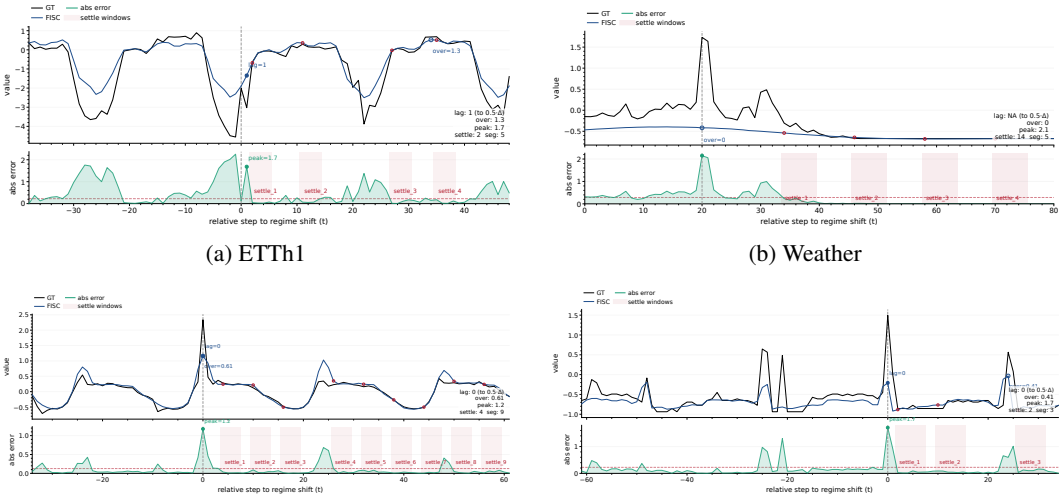



(c) DowJones (d) Traffic

(e) PEMS03 (f) PEMS04

(g) PEMS07 (h) PEMS08

Figure 9: **Qualitative decomposition of FISC, part II.** Additional decomposition results on financial and traffic datasets. The panels illustrate how FISC combines entrance-layer calibration, localized correction, compact row-wise aggregation, and adaptive temporal-basis organization.



(a) ETTh1 (b) Weather

(c) Traffic (d) Electricity

Figure 10: **Robustness under regime shifts.** Event-window trajectories are centered at the detected regime shift ($t = 0$). Lower post-event error spikes, weaker oscillation, and earlier sustained recovery are consistent with stronger robustness under non-stationarity.

test otherwise. The reported $p$-value is computed against FISC, so the FISC row is marked as the reference method.

If a run terminates with numerical failure, such as NaN loss, divergence, out-of-memory error, or abnormal early stopping, it is counted in the failure-rate diagnostic. Error means and paired tests are computed on valid matched runs, while failure rates are reported separately to avoid conflating numerical breakdown with ordinary forecasting error.

Table 10: **Random-seed stability over $K = 10$ runs.** Each entry reports mean $\pm$ standard deviation over random seeds. The standard deviation measures seed sensitivity, while paired $p$-values test whether the improvement over each compared method is consistent under matched seeds. Lower MSE/MAE and smaller standard deviation indicate better accuracy and seed stability.

| Dataset | Model | MSE (↓) | MAE (↓) | $p$-value vs. FISC (MSE) | $p$-value vs. FISC (MAE) |
|---|---|---|---|---|---|
| ETTh1 | FISC | $\mathbf{0.416 \pm 0.004}$ | $\mathbf{0.416 \pm 0.003}$ | Ref. | Ref. |
| | NoSC | $0.425 \pm 0.008$ | $0.425 \pm 0.006$ | 0.003 | 0.006 |
| | LastSC | $0.421 \pm 0.006$ | $0.423 \pm 0.005$ | 0.018 | 0.032 |
| | AllSC | $0.423 \pm 0.007$ | $0.424 \pm 0.005$ | 0.009 | 0.021 |
| Weather | FISC | $\mathbf{0.226 \pm 0.002}$ | $\mathbf{0.256 \pm 0.002}$ | Ref. | Ref. |
| | NoSC | $0.240 \pm 0.005$ | $0.261 \pm 0.004$ | 0.001 | 0.004 |
| | LastSC | $0.235 \pm 0.004$ | $0.258 \pm 0.003$ | 0.004 | 0.041 |
| | AllSC | $0.238 \pm 0.004$ | $0.259 \pm 0.003$ | 0.002 | 0.018 |

Beyond final forecasting errors, we further analyze optimization stability across seeds. This diagnostic measures whether a method reaches its final performance through smooth training, rather than through unstable trajectories that occasionally produce good checkpoints. For each trajectory, we record the log-gradient norm $g(t) = \log(1 + \|\nabla_t\|)$, where $\|\nabla_t\|$ is the gradient norm at step $t$. We define the gradient spike rate as

$$\text{SpikeRate} = \frac{1}{T_{\text{train}}} \sum_{t=1}^{T_{\text{train}}} \mathbf{1}[g(t) > \text{median}(g) + c \cdot \text{MAD}(g)] \,, \tag{20}$$

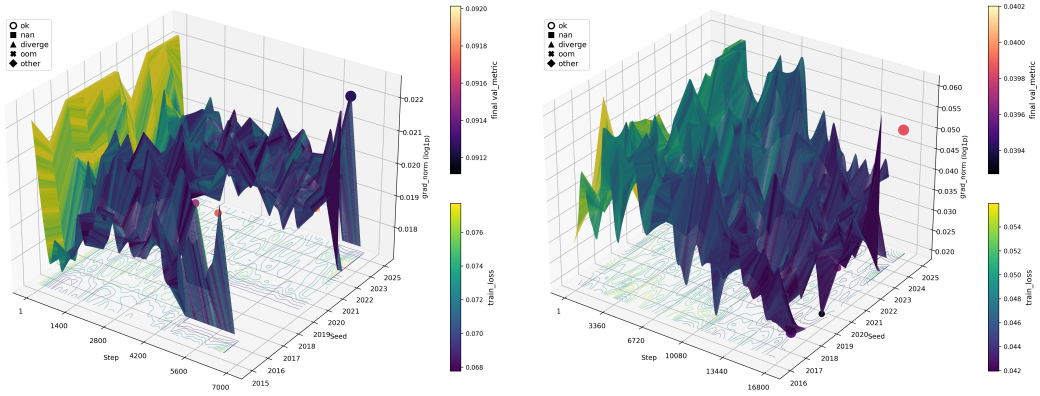
where $\text{MAD}(\cdot)$ is the median absolute deviation and $c$ is a fixed threshold shared by all methods. This robust threshold detects abrupt gradient bursts relative to each run's own trajectory. A lower spike rate indicates smoother optimization and fewer unstable gradient excursions.

We also report the failure rate, defined as the proportion of runs that terminate with numerical instability, NaN loss, divergence, out-of-memory error, or abnormal early termination. Finally, we report the standard deviation of the final validation metric across seeds, which measures checkpoint stability under stochastic training. Together, these metrics characterize optimization robustness from complementary perspectives: spike rate measures local training smoothness, failure rate measures numerical reliability, and validation-metric standard deviation measures cross-seed consistency.

Table 11: **Optimization stability over $K = 10$ seeds.** Avg. Spike Rate measures abrupt gradient bursts; Failure Rate measures numerical breakdowns; Final Val. Metric Std. measures cross-seed variability of selected checkpoints. Lower values indicate smoother and more reliable optimization.

| Dataset | Model | Avg. Spike Rate (↓) | Failure Rate (↓) | Final Val. Metric Std. (↓) |
|---|---|---|---|---|
| ETTh1 | FISC | **0.031** | **0/10** | **0.004** |
| | NoSC | 0.078 | 1/10 | 0.009 |
| | LastSC | 0.056 | 0/10 | 0.006 |
| | AllSC | 0.064 | 0/10 | 0.007 |
| Weather | FISC | **0.024** | **0/10** | **0.002** |
| | NoSC | 0.061 | 0/10 | 0.005 |
| | LastSC | 0.043 | 0/10 | 0.004 |
| | AllSC | 0.049 | 0/10 | 0.004 |

Fig. 11 visualizes the same multi-seed behavior. The surface height represents the log-gradient norm $g(t)$, the surface color represents training loss, and the terminal marker summarizes the final validation metric and run status. Compared with variants without entrance calibration or with repeated calibration constraints, FISC shows more consistent trajectories across seeds, fewer abrupt gradient spikes, and more stable terminal validation behavior. This supports the interpretation that first-layer statistical calibration constraints improve the entrance coupling, while leaving deeper layers adaptive avoids repeatedly applying potentially noisy or mismatched calibration matrices.




(a) ETTh1 (b) Weather

Figure 11: **Multi-channel 3D seed summary over $K = 10$ runs.** Surface height denotes the log-gradient norm $g(t) = \log(1 + \|\nabla_t\|)$, and surface color denotes training loss. The terminal marker encodes final validation metric, run status, and spike behavior. More stable surfaces and fewer abnormal terminal markers indicate stronger optimization robustness across random seeds.

Overall, the multi-seed results provide two complementary forms of evidence. The mean–standard deviation statistics verify that FISC's forecasting accuracy is stable across random seeds, while the paired tests assess whether the improvements are consistent under matched stochastic conditions. The optimization diagnostics further show that the gains are accompanied by smoother training behavior and fewer unstable runs. These results strengthen the empirical support for FISC's main design principle: training-set statistics are most useful as entrance-layer calibration constraints, whereas deeper cross-series representations should remain learnable and data-adaptive.

### E.5 Plug-in first-layer statistical calibration

To examine whether first-layer statistical calibration is specific to the full FISC architecture or transferable across forecasting backbones, we conduct plug-in experiments on representative models. For each backbone, we construct a calibrated variant, denoted as +SC, by applying the first-layer SC module at the earliest compatible cross-variable interaction point. The original temporal encoder, prediction head, optimizer, training protocol, and backbone-specific hyperparameters are kept unchanged.

**Plug-in SC module.** The plug-in module follows the same entrance-layer calibration design as FISC. We estimate the cross-series statistic $\widehat{\mathbf{R}}^{(s)}$ from the training split only, transform it into a nonnegative row-normalized calibration matrix $\widehat{\mathbf{A}}_{\mathrm{stat}}$, and fuse it with a learnable correction:

$$\mathbf{A}_{\mathrm{SC}} = \Pi_{\Delta}\Big(\lambda\widehat{\mathbf{A}}_{\mathrm{stat}} + (1-\lambda)\psi(\mathbf{W}_{\mathrm{SC}})\Big), \tag{21}$$

where $\Pi_{\Delta}$ is the row-wise simplex projection, $\psi(\cdot)$ is a monotone nonnegative map, and $\mathbf{W}_{\mathrm{SC}} \in \mathbb{R}^{N\times N}$ is the only additional learnable parameter introduced by the plug-in module.

For input-level backbones, SC is applied as entrance cross-series calibration, $\bar{\mathbf{X}}_{b,t,i} = \sum_{j=1}^{N}(\mathbf{A}_{\mathrm{SC}})_{ij}\tilde{\mathbf{X}}_{b,t,j}$. For token- or representation-level backbones, SC is applied to the earliest variable-level representation, $\bar{\mathbf{E}}_{b,i,:} = \sum_{j=1}^{N}(\mathbf{A}_{\mathrm{SC}})_{ij}\mathbf{E}_{b,j,:}$. In all cases, SC is applied only once at the entrance stage; all subsequent layers remain backbone-specific and fully learnable.

**Backbone-specific integration.** For DLinear/RLinear, SC is placed before the temporal linear projection, providing one calibrated cross-series interaction before the original channel-wise forecasting head. For PatchTST, SC is applied before patch construction, so the Transformer encoder receives calibrated multivariate inputs. For iTransformer, SC is placed after variate-token embedding and before the first encoder block. For OLinear, SC replaces the purely learnable first cross-series mixing matrix, while the orthogonal transformation, intra-series learner, and prediction head remain unchanged. For the TCN-based backbone, represented by TimesNet, SC is applied before the first temporal convolutional or periodic modeling block.

**Evaluation protocol.** Each vanilla backbone and its +SC variant are evaluated under the same data split, input length, prediction horizons, optimizer, early-stopping criterion, and validation protocol. The calibration hyperparameters $\lambda$ and $\tau$ are selected on the validation set using the same search space as FISC and then fixed across all horizons of each dataset. We report horizon-averaged MSE

Table 12: **Plug-in first-layer statistical calibration on representative forecasting backbones.** Each +SC variant applies only the first-layer SC module to the corresponding backbone, while keeping the original temporal encoder, prediction head, optimizer, and training protocol unchanged. Results are averaged over prediction horizons. For each backbone, bold values indicate improvement over its vanilla counterpart.

| Dataset | Metric | DLinear | | PatchTST | | iTransformer | | OLinear | | TimesNet | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Van. | +SC | Van. | +SC | Van. | +SC | Van. | +SC | Van. | +SC |
| ETTm1 | MSE | 0.403 | **0.389** | 0.387 | **0.378** | 0.407 | **0.392** | 0.374 | **0.370** | 0.400 | **0.389** |
| | MAE | 0.407 | **0.394** | 0.400 | **0.392** | 0.410 | **0.398** | 0.377 | **0.374** | 0.406 | **0.397** |
| ECL | MSE | 0.212 | **0.195** | 0.208 | **0.190** | 0.178 | **0.169** | 0.159 | **0.157** | 0.192 | **0.181** |
| | MAE | 0.300 | **0.283** | 0.295 | **0.280** | 0.270 | **0.262** | 0.248 | **0.246** | 0.295 | **0.279** |
| Traffic | MSE | 0.625 | **0.574** | 0.531 | **0.493** | 0.428 | **0.421** | 0.451 | **0.444** | 0.620 | **0.576** |
| | MAE | 0.383 | **0.350** | 0.343 | **0.318** | 0.282 | **0.275** | 0.247 | **0.245** | 0.336 | **0.313** |
| Avg. Imp. | MSE | 6.6% | | 6.0% | | 3.5% | | 1.3% | | 5.2% | |
| | MAE | 5.8% | | 4.8% | | 2.8% | | 0.8% | | 4.8% | |

and MAE. The relative improvement is computed as

$$\text{Imp.}(\%) = \frac{\text{Metric}_{\text{Van.}} - \text{Metric}_{+\text{SC}}}{\text{Metric}_{\text{Van.}}} \times 100, \tag{22}$$

where the metric is MSE or MAE.

Table 12 shows that first-layer SC improves all tested backbones across the representative datasets. The gains are larger for backbones with weaker explicit cross-series calibration, such as DLinear, PatchTST, and TimesNet, and smaller for OLinear, which already contains normalized cross-series modeling. These results support the view that SC is not merely an architecture-specific component of FISC, but a transferable entrance-layer calibration design. The plug-in variants are not intended to replace full FISC, since they do not include the dual-domain temporal representation or the stacked simplex-constrained CSCO layers.

# F Complete forecasting and ablation results

This appendix reports the complete per-horizon forecasting and ablation results corresponding to the summarized tables in the main text. The main paper uses compact average tables to show overall trends, while this appendix provides horizon-wise results for reproducibility and detailed comparison. Table 13 maps each main-text table to its complete appendix version.

Table 13: Mapping between summarized tables in the main text and complete tables in the appendix.

| Main-text table | Appendix table | Content |
|---|---|---|
| Table 1 | Table 14 | Long-term forecasting |
| Table 2 | Tables 15–16 | Short-term forecasting |
| Table 4 | Table 17 | Placement of statistical calibration constraints |
| Table 5 | Table 18 | Dual-domain representation and basis selection |
| Table 6 | Table 19 | CSCO constraints and first-layer calibration |

# G Existing assets and supplementary code

## G.1 Existing datasets, baselines, and license documentation

All datasets used in this paper are public time-series forecasting benchmarks or publicly accessible datasets commonly used in forecasting research. We do not collect private data, conduct user studies, or use human-subject data. Dataset statistics, horizons, splits, and preprocessing protocols are reported in App. C. All normalization statistics, cross-series statistics, temporal correlation matrices, and calibration matrices used by FISC are estimated only from the training split and kept fixed during validation and testing to avoid leakage.

For existing assets, including datasets and baseline implementations, we follow the terms of use and licenses specified by the original providers when available. Dataset and baseline sources are cited in the main text and appendix. For datasets requiring additional preprocessing beyond standard benchmark usage, the supplementary material includes preprocessing scripts or configuration information. For baselines, we use public or official implementations whenever available and document the evaluation protocol in App. D. When previously reported results do not match our split, input length, horizon, or metric, we rerun the official implementation under the unified protocol used in this paper.

The public datasets used in this work are employed only for forecasting benchmark evaluation. We do not attempt to infer sensitive attributes, identify individuals, or link records across external sources. When applying FISC beyond these benchmarks, users should verify the license, privacy requirements, and domain-specific restrictions of the data before training or deployment.

## G.2 Supplementary code documentation

We provide FISC model code in the supplementary material to support inspection and reproducibility. The code includes the core implementation of first-layer statistical calibration constraints, the con-

Table 14: Complete long-term forecasting results. The look-back window is fixed to $T = 96$ for all methods. For ETT, Weather, ECL, Traffic, Exchange, and Solar-Energy, horizons are $H \in \{96, 192, 336, 720\}$; for PEMS datasets, horizons are $H \in \{12, 24, 48, 96\}$. The Avg row reports the average over four horizons. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 1.

| **Category** | | Linear-Based | | | | | | | | | | | | | | Transformer-Based | | | | | | | | | | | | TCN-Based | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Models** | | FISC ours | | OLinear 2025 | | TimeMixer 2024a | | FilterNet 2024a | | FITS 2024 | | DLinear 2023 | | TimeMixer++ 2025a | | Leddam 2024 | | CARD 2024b | | Fredformer 2024 | | iTransformer 2024a | | PatchTST 2023 | | TimesNet 2023b | |
| **Metric** | | 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 | 96 | **0.357** | **0.379** | 0.360 | 0.382 | 0.375 | 0.400 | 0.382 | 0.402 | 0.385 | 0.394 | 0.386 | 0.400 | 0.361 | 0.403 | 0.377 | 0.394 | 0.383 | 0.391 | 0.373 | 0.392 | 0.386 | 0.405 | 0.414 | 0.419 | 0.384 | 0.402 |
| | 192 | **0.410** | **0.412** | 0.416 | 0.414 | 0.429 | 0.421 | 0.430 | 0.429 | 0.434 | 0.422 | 0.437 | 0.432 | 0.416 | 0.441 | 0.424 | 0.422 | 0.435 | 0.420 | 0.433 | 0.420 | 0.441 | 0.436 | 0.460 | 0.445 | 0.436 | 0.429 |
| | 336 | 0.439 | **0.433** | 0.457 | 0.438 | 0.484 | 0.458 | 0.472 | 0.451 | 0.476 | 0.444 | 0.481 | 0.459 | **0.430** | 0.434 | 0.459 | 0.442 | 0.479 | 0.442 | 0.470 | 0.437 | 0.487 | 0.458 | 0.501 | 0.466 | 0.491 | 0.469 |
| | 720 | **0.457** | **0.438** | 0.463 | 0.462 | 0.498 | 0.482 | 0.481 | 0.473 | 0.465 | 0.462 | 0.519 | 0.516 | 0.467 | 0.451 | 0.463 | 0.459 | 0.471 | 0.461 | 0.467 | 0.456 | 0.503 | 0.491 | 0.500 | 0.488 | 0.521 | 0.500 |
| | Avg | **0.416** | **0.416** | 0.424 | 0.424 | 0.447 | 0.440 | 0.441 | 0.439 | 0.440 | 0.431 | 0.456 | 0.452 | 0.419 | 0.432 | 0.431 | 0.429 | 0.442 | 0.429 | 0.435 | 0.426 | 0.454 | 0.447 | 0.469 | 0.454 | 0.458 | 0.450 |
| ETTh2 | 96 | **0.274** | **0.327** | 0.284 | 0.329 | 0.289 | 0.341 | 0.293 | 0.343 | 0.292 | 0.340 | 0.333 | 0.387 | 0.276 | 0.328 | 0.292 | 0.343 | 0.281 | 0.330 | 0.293 | 0.342 | 0.297 | 0.349 | 0.292 | 0.342 | 0.340 | 0.374 |
| | 192 | 0.354 | **0.375** | 0.360 | 0.379 | 0.372 | 0.392 | 0.374 | 0.396 | 0.377 | 0.391 | 0.477 | 0.476 | **0.342** | 0.379 | 0.367 | 0.389 | 0.363 | 0.381 | 0.371 | 0.389 | 0.380 | 0.400 | 0.387 | 0.400 | 0.402 | 0.414 |
| | 336 | 0.381 | 0.407 | 0.409 | 0.415 | 0.386 | 0.414 | 0.417 | 0.430 | 0.416 | 0.425 | 0.594 | 0.541 | **0.346** | **0.398** | 0.412 | 0.424 | 0.411 | 0.418 | 0.411 | 0.418 | 0.382 | 0.409 | 0.428 | 0.432 | 0.426 | 0.433 |
| | 720 | **0.390** | **0.413** | 0.415 | 0.431 | 0.412 | 0.434 | 0.449 | 0.460 | 0.418 | 0.437 | 0.831 | 0.657 | 0.392 | 0.415 | 0.419 | 0.438 | 0.416 | 0.431 | 0.415 | 0.434 | 0.427 | 0.445 | 0.431 | 0.446 | 0.462 | 0.468 |
| | Avg | 0.350 | 0.381 | 0.367 | 0.388 | 0.365 | 0.395 | 0.383 | 0.407 | 0.376 | 0.398 | 0.559 | 0.515 | **0.339** | **0.380** | 0.373 | 0.399 | 0.368 | 0.390 | 0.365 | 0.393 | 0.383 | 0.407 | 0.384 | 0.405 | 0.414 | 0.427 |
| ETTm1 | 96 | **0.300** | **0.332** | 0.302 | 0.334 | 0.320 | 0.357 | 0.321 | 0.361 | 0.353 | 0.375 | 0.345 | 0.372 | 0.310 | 0.334 | 0.319 | 0.359 | 0.316 | 0.347 | 0.326 | 0.361 | 0.334 | 0.368 | 0.329 | 0.367 | 0.338 | 0.375 |
| | 192 | **0.346** | **0.362** | 0.357 | 0.363 | 0.361 | 0.381 | 0.367 | 0.387 | 0.486 | 0.445 | 0.380 | 0.389 | 0.348 | **0.362** | 0.369 | 0.383 | 0.363 | 0.370 | 0.363 | 0.380 | 0.377 | 0.391 | 0.367 | 0.385 | 0.374 | 0.387 |
| | 336 | **0.375** | **0.383** | 0.387 | 0.385 | 0.390 | 0.404 | 0.401 | 0.409 | 0.531 | 0.475 | 0.413 | 0.413 | 0.376 | 0.391 | 0.394 | 0.402 | 0.392 | 0.390 | 0.395 | 0.403 | 0.426 | 0.420 | 0.399 | 0.410 | 0.410 | 0.411 |
| | 720 | **0.437** | **0.422** | 0.452 | 0.426 | 0.454 | 0.441 | 0.477 | 0.448 | 0.600 | 0.513 | 0.474 | 0.453 | 0.440 | 0.423 | 0.460 | 0.442 | 0.458 | 0.425 | 0.458 | 0.425 | 0.453 | 0.438 | 0.491 | 0.459 | 0.454 | 0.439 |
| | Avg | **0.365** | **0.375** | 0.374 | 0.377 | 0.381 | 0.395 | 0.392 | 0.401 | 0.493 | 0.452 | 0.403 | 0.407 | 0.369 | 0.378 | 0.386 | 0.397 | 0.383 | 0.384 | 0.384 | 0.395 | 0.407 | 0.410 | 0.387 | 0.400 | 0.400 | 0.406 |
| ETTm2 | 96 | **0.166** | **0.244** | 0.169 | 0.249 | 0.175 | 0.258 | 0.175 | 0.258 | 0.182 | 0.266 | 0.193 | 0.292 | 0.170 | 0.245 | 0.176 | 0.257 | 0.169 | 0.248 | 0.177 | 0.259 | 0.180 | 0.264 | 0.175 | 0.259 | 0.187 | 0.267 |
| | 192 | **0.227** | **0.290** | 0.232 | **0.290** | 0.237 | 0.299 | 0.240 | 0.301 | 0.253 | 0.312 | 0.284 | 0.362 | 0.229 | 0.291 | 0.243 | 0.303 | 0.234 | 0.292 | 0.243 | 0.301 | 0.250 | 0.309 | 0.241 | 0.302 | 0.249 | 0.309 |
| | 336 | **0.287** | **0.326** | 0.291 | 0.328 | 0.298 | 0.340 | 0.311 | 0.347 | 0.313 | 0.349 | 0.369 | 0.427 | 0.303 | 0.343 | 0.303 | 0.341 | 0.294 | 0.339 | 0.302 | 0.340 | 0.311 | 0.348 | 0.305 | 0.343 | 0.321 | 0.351 |
| | 720 | 0.381 | **0.385** | 0.389 | 0.387 | 0.391 | 0.396 | 0.414 | 0.405 | 0.416 | 0.406 | 0.554 | 0.522 | **0.373** | 0.399 | 0.400 | 0.398 | 0.390 | 0.388 | 0.397 | 0.396 | 0.412 | 0.407 | 0.402 | 0.400 | 0.408 | 0.403 |
| | Avg | **0.265** | **0.311** | 0.270 | 0.313 | 0.275 | 0.323 | 0.285 | 0.328 | 0.291 | 0.333 | 0.350 | 0.401 | 0.269 | 0.320 | 0.281 | 0.325 | 0.272 | 0.317 | 0.279 | 0.324 | 0.288 | 0.332 | 0.281 | 0.326 | 0.291 | 0.333 |
| Weather | 96 | **0.148** | **0.185** | 0.153 | 0.190 | 0.163 | 0.209 | 0.162 | 0.207 | 0.196 | 0.236 | 0.196 | 0.255 | 0.155 | 0.205 | 0.156 | 0.202 | 0.150 | 0.188 | 0.163 | 0.207 | 0.174 | 0.214 | 0.177 | 0.218 | 0.172 | 0.220 |
| | 192 | **0.194** | **0.233** | 0.200 | 0.235 | 0.208 | 0.250 | 0.210 | 0.250 | 0.240 | 0.271 | 0.237 | 0.296 | 0.201 | 0.245 | 0.207 | 0.250 | 0.202 | 0.238 | 0.211 | 0.251 | 0.221 | 0.254 | 0.225 | 0.259 | 0.219 | 0.261 |
| | 336 | 0.241 | 0.276 | 0.258 | 0.280 | 0.251 | 0.287 | 0.265 | 0.290 | 0.292 | 0.307 | 0.283 | 0.335 | **0.237** | **0.265** | 0.262 | 0.291 | 0.260 | 0.282 | 0.267 | 0.292 | 0.278 | 0.296 | 0.278 | 0.297 | 0.280 | 0.306 |
| | 720 | 0.322 | **0.329** | 0.337 | 0.333 | 0.339 | 0.341 | 0.342 | 0.340 | 0.365 | 0.354 | 0.345 | 0.381 | **0.312** | 0.334 | 0.343 | 0.343 | 0.343 | 0.353 | 0.343 | 0.341 | 0.358 | 0.349 | 0.354 | 0.348 | 0.365 | 0.359 |
| | Avg | **0.226** | **0.256** | 0.237 | 0.260 | 0.240 | 0.272 | 0.245 | 0.272 | 0.273 | 0.292 | 0.265 | 0.317 | **0.226** | 0.262 | 0.242 | 0.272 | 0.239 | 0.265 | 0.246 | 0.272 | 0.258 | 0.279 | 0.259 | 0.281 | 0.259 | 0.287 |
| Traffic | 96 | **0.390** | **0.224** | 0.398 | 0.226 | 0.462 | 0.285 | 0.430 | 0.294 | 0.601 | 0.361 | 0.650 | 0.396 | 0.392 | 0.253 | 0.426 | 0.276 | 0.419 | 0.269 | 0.406 | 0.277 | 0.395 | 0.268 | 0.446 | 0.283 | 0.593 | 0.321 |
| | 192 | 0.428 | **0.240** | 0.439 | 0.241 | 0.473 | 0.296 | 0.452 | 0.307 | 0.603 | 0.365 | 0.598 | 0.370 | **0.402** | 0.258 | 0.458 | 0.289 | 0.443 | 0.276 | 0.426 | 0.290 | 0.417 | 0.276 | 0.540 | 0.354 | 0.617 | 0.336 |
| | 336 | 0.451 | **0.245** | 0.464 | 0.250 | 0.498 | 0.296 | 0.470 | 0.316 | 0.609 | 0.366 | 0.605 | 0.373 | **0.428** | 0.263 | 0.486 | 0.297 | 0.460 | 0.283 | 0.437 | 0.292 | 0.433 | 0.283 | 0.551 | 0.358 | 0.629 | 0.336 |
| | 720 | 0.475 | **0.268** | 0.502 | 0.270 | 0.506 | 0.313 | 0.498 | 0.323 | 0.648 | 0.387 | 0.645 | 0.394 | **0.441** | 0.282 | 0.498 | 0.313 | 0.490 | 0.299 | 0.462 | 0.305 | 0.467 | 0.302 | 0.586 | 0.375 | 0.640 | 0.350 |
| | Avg | 0.436 | **0.244** | 0.451 | 0.247 | 0.485 | 0.298 | 0.463 | 0.310 | 0.615 | 0.370 | 0.625 | 0.383 | **0.416** | 0.264 | 0.467 | 0.294 | 0.453 | 0.282 | 0.433 | 0.291 | 0.428 | 0.282 | 0.531 | 0.343 | 0.620 | 0.336 |
| Electricity | 96 | **0.130** | **0.220** | 0.131 | 0.221 | 0.153 | 0.247 | 0.147 | 0.245 | 0.198 | 0.274 | 0.197 | 0.282 | 0.135 | 0.222 | 0.141 | 0.235 | 0.141 | 0.233 | 0.147 | 0.241 | 0.148 | 0.240 | 0.161 | 0.250 | 0.168 | 0.272 |
| | 192 | **0.146** | **0.235** | 0.150 | 0.238 | 0.166 | 0.256 | 0.160 | 0.250 | 0.363 | 0.422 | 0.196 | 0.285 | 0.147 | **0.235** | 0.159 | 0.252 | 0.160 | 0.250 | 0.165 | 0.258 | 0.162 | 0.253 | 0.199 | 0.289 | 0.184 | 0.289 |
| | 336 | **0.163** | 0.248 | 0.165 | 0.254 | 0.185 | 0.277 | 0.173 | 0.267 | 0.444 | 0.490 | 0.209 | 0.301 | 0.164 | **0.245** | 0.173 | 0.268 | 0.173 | 0.263 | 0.177 | 0.273 | 0.178 | 0.269 | 0.215 | 0.305 | 0.198 | 0.300 |
| | 720 | **0.181** | **0.273** | 0.191 | 0.279 | 0.225 | 0.310 | 0.210 | 0.309 | 0.532 | 0.551 | 0.245 | 0.333 | 0.212 | 0.310 | 0.201 | 0.295 | 0.197 | 0.284 | 0.213 | 0.304 | 0.225 | 0.317 | 0.256 | 0.337 | 0.220 | 0.320 |
| | Avg | **0.155** | **0.244** | 0.159 | 0.248 | 0.182 | 0.273 | 0.173 | 0.268 | 0.384 | 0.434 | 0.212 | 0.300 | 0.165 | 0.253 | 0.169 | 0.263 | 0.168 | 0.258 | 0.176 | 0.269 | 0.178 | 0.270 | 0.208 | 0.295 | 0.192 | 0.295 |
| Exchange | 96 | **0.081** | **0.199** | 0.082 | 0.200 | 0.086 | 0.205 | 0.091 | 0.211 | 0.087 | 0.208 | 0.088 | 0.218 | 0.085 | 0.214 | 0.086 | 0.207 | 0.084 | 0.202 | 0.084 | 0.202 | 0.086 | 0.206 | 0.088 | 0.205 | 0.107 | 0.234 |
| | 192 | **0.169** | **0.292** | 0.171 | 0.293 | 0.193 | 0.312 | 0.186 | 0.305 | 0.185 | 0.306 | 0.176 | 0.315 | 0.175 | 0.313 | 0.175 | 0.301 | 0.179 | 0.298 | 0.178 | 0.302 | 0.177 | 0.299 | 0.176 | 0.299 | 0.226 | 0.344 |
| | 336 | 0.317 | 0.406 | 0.331 | 0.414 | 0.356 | 0.433 | 0.380 | 0.449 | 0.342 | 0.425 | 0.313 | 0.427 | 0.316 | 0.420 | 0.325 | 0.415 | 0.333 | 0.418 | 0.319 | 0.408 | 0.331 | 0.417 | **0.301** | **0.397** | 0.367 | 0.448 |
| | 720 | 0.762 | 0.677 | 0.837 | 0.688 | 0.912 | 0.712 | 0.896 | 0.712 | 0.846 | 0.694 | 0.839 | 0.695 | 0.851 | 0.689 | 0.831 | 0.686 | 0.851 | 0.691 | **0.749** | **0.651** | 0.847 | 0.691 | 0.901 | 0.714 | 0.964 | 0.746 |
| | Avg | **0.333** | 0.394 | 0.355 | 0.399 | 0.387 | 0.416 | 0.388 | 0.419 | 0.365 | 0.408 | 0.354 | 0.414 | 0.357 | 0.409 | 0.354 | 0.402 | 0.362 | 0.402 | **0.333** | **0.391** | 0.360 | 0.403 | 0.367 | 0.404 | 0.416 | 0.443 |
| PEMS03 | 12 | **0.059** | **0.158** | 0.060 | 0.159 | 0.076 | 0.188 | 0.071 | 0.177 | 0.117 | 0.226 | 0.122 | 0.243 | 0.097 | 0.208 | 0.063 | 0.164 | 0.072 | 0.177 | 0.068 | 0.174 | 0.071 | 0.174 | 0.099 | 0.216 | 0.085 | 0.192 |
| | 24 | **0.076** | **0.177** | 0.078 | 0.179 | 0.113 | 0.226 | 0.102 | 0.213 | 0.235 | 0.324 | 0.201 | 0.317 | 0.120 | 0.230 | 0.080 | 0.185 | 0.107 | 0.217 | 0.093 | 0.202 | 0.093 | 0.201 | 0.142 | 0.259 | 0.118 | 0.223 |
| | 48 | **0.102** | **0.209** | 0.104 | 0.210 | 0.191 | 0.292 | 0.162 | 0.272 | 0.541 | 0.521 | 0.333 | 0.425 | 0.170 | 0.272 | 0.124 | 0.226 | 0.194 | 0.302 | 0.146 | 0.258 | 0.125 | 0.236 | 0.211 | 0.319 | 0.155 | 0.260 |
| | 96 | **0.138** | **0.246** | 0.140 | 0.247 | 0.288 | 0.363 | 0.244 | 0.340 | 1.062 | 0.790 | 0.457 | 0.515 | 0.274 | 0.342 | 0.160 | 0.266 | 0.323 | 0.402 | 0.228 | 0.330 | 0.164 | 0.275 | 0.269 | 0.370 | 0.228 | 0.317 |
| | Avg | **0.094** | **0.198** | 0.095 | 0.199 | 0.167 | 0.267 | 0.145 | 0.251 | 0.489 | 0.465 | 0.278 | 0.375 | 0.165 | 0.263 | 0.107 | 0.210 | 0.174 | 0.275 | 0.135 | 0.243 | 0.113 | 0.221 | 0.180 | 0.291 | 0.147 | 0.248 |
| PEMS04 | 12 | **0.067** | **0.161** | 0.068 | 0.163 | 0.092 | 0.204 | 0.082 | 0.190 | 0.129 | 0.239 | 0.148 | 0.272 | 0.099 | 0.214 | 0.071 | 0.172 | 0.089 | 0.194 | 0.085 | 0.189 | 0.078 | 0.183 | 0.105 | 0.224 | 0.087 | 0.195 |
| | 24 | **0.078** | **0.174** | 0.079 | 0.176 | 0.128 | 0.243 | 0.110 | 0.224 | 0.246 | 0.337 | 0.224 | 0.340 | 0.115 | 0.231 | 0.087 | 0.193 | 0.128 | 0.234 | 0.117 | 0.224 | 0.095 | 0.205 | 0.153 | 0.275 | 0.103 | 0.215 |
| | 48 | **0.093** | **0.196** | 0.095 | 0.197 | 0.213 | 0.315 | 0.160 | 0.276 | 0.568 | 0.539 | 0.355 | 0.437 | 0.144 | 0.261 | 0.113 | 0.222 | 0.224 | 0.321 | 0.174 | 0.276 | 0.120 | 0.233 | 0.229 | 0.339 | 0.136 | 0.250 |
| | 96 | **0.121** | **0.224** | 0.122 | 0.226 | 0.307 | 0.384 | 0.234 | 0.343 | 1.181 | 0.843 | 0.452 | 0.504 | 0.185 | 0.297 | 0.141 | 0.252 | 0.382 | 0.445 | 0.273 | 0.354 | 0.150 | 0.262 | 0.291 | 0.389 | 0.190 | 0.303 |
| | Avg | **0.090** | **0.189** | 0.091 | 0.190 | 0.185 | 0.287 | 0.146 | 0.258 | 0.531 | 0.489 | 0.295 | 0.388 | 0.136 | 0.251 | 0.103 | 0.210 | 0.206 | 0.299 | 0.162 | 0.261 | 0.111 | 0.221 | 0.195 | 0.307 | 0.129 | 0.241 |
| PEMS07 | 12 | **0.051** | **0.136** | 0.052 | 0.138 | 0.073 | 0.184 | 0.064 | 0.163 | 0.109 | 0.222 | 0.115 | 0.242 | 0.090 | 0.197 | 0.055 | 0.145 | 0.068 | 0.166 | 0.063 | 0.158 | 0.067 | 0.165 | 0.095 | 0.207 | 0.082 | 0.181 |
| | 24 | **0.064** | **0.151** | 0.065 | **0.151** | 0.111 | 0.219 | 0.093 | 0.200 | 0.230 | 0.327 | 0.210 | 0.329 | 0.110 | 0.219 | 0.070 | 0.164 | 0.103 | 0.206 | 0.089 | 0.192 | 0.088 | 0.190 | 0.150 | 0.262 | 0.101 | 0.204 |
| | 48 | **0.082** | **0.168** | 0.084 | 0.171 | 0.237 | 0.328 | 0.137 | 0.248 | 0.551 | 0.531 | 0.398 | 0.458 | 0.149 | 0.256 | 0.094 | 0.192 | 0.165 | 0.268 | 0.136 | 0.241 | 0.110 | 0.215 | 0.253 | 0.340 | 0.134 | 0.238 |
| | 96 | **0.106** | 0.198 | 0.108 | **0.196** | 0.303 | 0.354 | 0.198 | 0.306 | 1.112 | 0.809 | 0.594 | 0.553 | 0.258 | 0.359 | 0.117 | 0.217 | 0.258 | 0.346 | 0.197 | 0.298 | 0.139 | 0.245 | 0.346 | 0.404 | 0.181 | 0.279 |
| | Avg | **0.076** | **0.163** | 0.077 | 0.164 | 0.181 | 0.271 | 0.123 | 0.229 | 0.500 | 0.472 | 0.329 | 0.395 | 0.152 | 0.258 | 0.084 | 0.180 | 0.149 | 0.247 | 0.121 | 0.222 | 0.101 | 0.204 | 0.211 | 0.303 | 0.124 | 0.225 |
| PEMS08 | 12 | **0.065** | **0.157** | 0.068 | 0.159 | 0.091 | 0.201 | 0.080 | 0.182 | 0.122 | 0.233 | 0.154 | 0.276 | 0.119 | 0.222 | 0.071 | 0.171 | 0.080 | 0.181 | 0.081 | 0.185 | 0.079 | 0.182 | 0.168 | 0.232 | 0.112 | 0.212 |
| | 24 | **0.086** | **0.177** | 0.089 | 0.178 | 0.137 | 0.246 | 0.114 | 0.219 | 0.236 | 0.330 | 0.248 | 0.353 | 0.149 | 0.249 | 0.091 | 0.189 | 0.118 | 0.220 | 0.112 | 0.214 | 0.115 | 0.219 | 0.224 | 0.281 | 0.141 | 0.238 |
| | 48 | **0.121** | **0.201** | 0.123 | 0.204 | 0.265 | 0.343 | 0.184 | 0.284 | 0.562 | 0.540 | 0.440 | 0.470 | 0.206 | 0.292 | 0.128 | 0.219 | 0.199 | 0.289 | 0.174 | 0.267 | 0.186 | 0.235 | 0.321 | 0.354 | 0.198 | 0.283 |
| | 96 | **0.170** | **0.233** | 0.173 | 0.236 | 0.410 | 0.407 | 0.309 | 0.356 | 1.216 | 0.846 | 0.674 | 0.565 | 0.329 | 0.355 | 0.198 | 0.266 | 0.405 | 0.431 | 0.277 | 0.335 | 0.221 | 0.267 | 0.408 | 0.417 | 0.320 | 0.351 |
| | Avg | **0.111** | **0.192** | 0.113 | 0.194 | 0.226 | 0.299 | 0.172 | 0.260 | 0.534 | 0.487 | 0.379 | 0.416 | 0.200 | 0.279 | 0.122 | 0.211 | 0.201 | 0.280 | 0.161 | 0.250 | 0.150 | 0.226 | 0.280 | 0.321 | 0.193 | 0.271 |
| 1st Count | | 47 | 50 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 6 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 0 |

Table 15: Complete short-term forecasting results on the first group of datasets. For the short setting, the look-back window is 12 and horizons are $H \in \{3, 6, 9, 12\}$; for the long setting, the look-back window is 36 and horizons are $H \in \{24, 36, 48, 96\}$. The Avg row reports the average over four horizons. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 2.

| Category | | Linear-Based | | | | | | | | | | | | Transformer-Based | | | | | | | | | | | | TCN-Based | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Models | | FISC ours | | OLinear 2025 | | TimeMixer 2024a | | FilterNet 2024a | | FITS 2024 | | DLinear 2023 | | TimeMixer++ 2025a | | Leddam 2024 | | CARD 2024b | | Fredformer 2024 | | iTransformer 2024a | | PatchTST 2023 | | TimesNet 2023b | |
| Metric | | 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 |
| ILI | 3 | **0.440** | **0.338** | 0.468 | 0.349 | 0.659 | 0.435 | 0.660 | 0.437 | 1.461 | 0.743 | 1.280 | 0.747 | 0.658 | 0.430 | 0.551 | 0.388 | 0.597 | 0.392 | 0.528 | 0.403 | 0.555 | 0.395 | 0.646 | 0.417 | 0.627 | 0.420 |
| | 6 | **0.891** | **0.510** | 0.923 | 0.516 | 1.306 | 0.643 | 1.140 | 0.606 | 2.337 | 0.974 | 2.054 | 0.967 | 1.273 | 0.645 | 1.021 | 0.569 | 1.246 | 0.610 | 1.128 | 0.604 | 1.124 | 0.586 | 1.269 | 0.629 | 1.147 | 0.610 |
| | 9 | **1.287** | **0.651** | 1.289 | 0.655 | 2.070 | 0.842 | 1.815 | 0.798 | 3.397 | 1.197 | 2.771 | 1.138 | 2.009 | 0.840 | 1.881 | 0.795 | 2.041 | 0.829 | 1.804 | 0.786 | 1.794 | 0.772 | 2.021 | 0.830 | 1.796 | 0.785 |
| | 12 | **1.650** | **0.769** | 1.698 | 0.791 | 2.792 | 1.018 | 2.435 | 0.953 | 4.244 | 1.361 | 3.497 | 1.284 | 2.683 | 1.004 | 2.421 | 0.964 | 2.746 | 0.996 | 2.610 | 0.989 | 2.273 | 0.884 | 2.788 | 1.016 | 2.349 | 0.920 |
| | Avg | **1.067** | **0.567** | 1.094 | 0.578 | 1.707 | 0.734 | 1.512 | 0.698 | 2.860 | 1.069 | 2.400 | 1.034 | 1.656 | 0.730 | 1.468 | 0.679 | 1.658 | 0.707 | 1.518 | 0.696 | 1.437 | 0.659 | 1.681 | 0.723 | 1.480 | 0.684 |
| COVID-19 | 24 | **1.729** | **0.786** | 1.737 | 0.800 | 2.110 | 0.879 | 2.190 | 0.870 | 4.265 | 1.523 | 3.158 | 1.243 | 1.877 | 0.826 | 2.085 | 0.883 | 2.407 | 0.970 | 2.098 | 0.894 | 2.004 | 0.860 | 2.046 | 0.849 | 2.317 | 0.934 |
| | 36 | **1.710** | **0.794** | 1.714 | 0.795 | 2.084 | 0.890 | 1.902 | 0.862 | 3.718 | 1.363 | 3.009 | 1.200 | 2.276 | 0.912 | 2.017 | 0.892 | 2.324 | 0.948 | 1.712 | 0.867 | 1.910 | 0.880 | 2.344 | 0.912 | 1.972 | 0.920 |
| | 48 | **1.813** | **0.801** | 1.821 | 0.804 | 1.961 | 0.866 | 2.051 | 0.882 | 3.994 | 1.422 | 2.994 | 1.194 | 1.921 | 0.850 | 1.860 | 0.847 | 2.133 | 0.911 | 2.054 | 0.922 | 2.036 | 0.891 | 2.123 | 0.883 | 2.238 | 0.940 |
| | 60 | 1.774 | **0.807** | 1.785 | 0.810 | 1.926 | 0.878 | 2.151 | 0.925 | 4.543 | 1.554 | 3.172 | 1.232 | **1.745** | 0.838 | 1.967 | 0.879 | 2.177 | 0.921 | 1.925 | 0.913 | 2.022 | 0.919 | 2.001 | 0.895 | 2.027 | 0.928 |
| | Avg | **1.757** | **0.797** | 1.764 | 0.802 | 2.020 | 0.878 | 2.073 | 0.885 | 4.130 | 1.465 | 3.083 | 1.217 | 1.955 | 0.857 | 1.982 | 0.875 | 2.260 | 0.938 | 1.947 | 0.899 | 1.993 | 0.887 | 2.128 | 0.885 | 2.139 | 0.931 |
| | 3 | **1.089** | **0.482** | 1.100 | 0.487 | 1.237 | 0.547 | 1.195 | 0.555 | 2.039 | 0.790 | 2.386 | 0.909 | 1.298 | 0.584 | 1.216 | 0.570 | 1.103 | 0.521 | 1.165 | 0.548 | 1.193 | 0.561 | 1.220 | 0.573 | 2.021 | 0.704 |
| | 6 | 1.725 | **0.612** | 1.750 | 0.619 | 2.003 | 0.739 | 1.839 | 0.711 | 2.683 | 0.919 | 3.220 | 1.053 | 1.833 | 0.682 | 1.782 | 0.689 | 1.919 | 0.735 | **1.465** | 0.685 | 1.933 | 0.755 | 1.982 | 0.762 | 2.405 | 0.808 |
| | 9 | 2.217 | **0.730** | 2.239 | 0.734 | 2.594 | 0.860 | 2.537 | 0.897 | 3.147 | 1.050 | 3.803 | 1.160 | 2.472 | 0.822 | 2.407 | 0.866 | 2.358 | 0.841 | **2.145** | 0.845 | 2.441 | 0.879 | 2.633 | 0.916 | 2.858 | 0.969 |
| | 12 | **2.524** | **0.821** | 2.538 | 0.831 | 3.103 | 0.981 | 2.782 | 0.956 | 3.630 | 1.156 | 4.524 | 1.288 | 3.273 | 1.084 | 2.851 | 0.991 | 2.857 | 0.971 | 2.833 | 0.984 | 2.819 | 0.984 | 3.050 | 1.030 | 2.993 | 0.964 |
| | Avg | **1.889** | **0.661** | 1.907 | 0.668 | 2.234 | 0.782 | 2.088 | 0.780 | 2.875 | 0.979 | 3.483 | 1.102 | 2.219 | 0.793 | 2.064 | 0.779 | 2.059 | 0.767 | 1.902 | 0.765 | 2.096 | 0.795 | 2.221 | 0.820 | 2.569 | 0.861 |
| METR-LA | 24 | **4.468** | **1.174** | 4.474 | 1.180 | 6.335 | 1.554 | 5.926 | 1.517 | 8.248 | 1.829 | 9.780 | 1.851 | 6.539 | 1.618 | 4.860 | 1.342 | 5.133 | 1.394 | 4.799 | 1.347 | 4.715 | 1.321 | 5.528 | 1.450 | 5.634 | 1.442 |
| | 36 | **7.232** | **1.663** | 7.241 | 1.670 | 8.222 | 1.787 | 7.696 | 1.733 | 10.345 | 2.050 | 12.804 | 2.083 | 7.986 | 1.770 | 7.378 | 1.708 | 7.377 | 1.725 | 7.536 | 1.727 | 7.299 | 1.681 | 8.351 | 1.830 | 9.114 | 1.848 |
| | 48 | **9.621** | **1.943** | 10.076 | 1.985 | 11.669 | 2.157 | 11.572 | 2.141 | 11.999 | 2.197 | 14.244 | 2.189 | 11.655 | 2.156 | 10.051 | 1.999 | 11.013 | 2.103 | 9.833 | 1.951 | 10.141 | 2.012 | 11.259 | 2.114 | 10.940 | 2.033 |
| | 60 | **11.287** | **2.019** | 12.079 | 2.182 | 12.188 | 2.173 | 11.311 | 2.066 | 13.185 | 2.282 | 15.472 | 2.275 | 12.734 | 2.235 | 11.467 | 2.119 | 12.528 | 2.227 | 12.455 | 2.209 | 11.871 | 2.156 | 12.666 | 2.225 | 12.888 | 2.186 |
| | Avg | **8.152** | **1.700** | 8.467 | 1.754 | 9.604 | 1.918 | 9.126 | 1.864 | 10.944 | 2.089 | 13.075 | 2.099 | 9.728 | 1.945 | 8.439 | 1.792 | 9.013 | 1.862 | 8.656 | 1.808 | 8.506 | 1.792 | 9.451 | 1.905 | 9.644 | 1.877 |
| | 3 | **0.203** | **0.171** | 0.207 | **0.171** | 0.205 | 0.192 | 0.204 | 0.189 | 0.211 | 0.198 | 0.218 | 0.231 | 0.204 | 0.189 | 0.204 | 0.191 | 0.210 | 0.180 | 0.205 | 0.188 | 0.205 | 0.188 | 0.204 | 0.190 | 0.221 | 0.204 |
| | 6 | **0.293** | **0.206** | 0.301 | 0.207 | 0.297 | 0.230 | 0.296 | 0.228 | 0.306 | 0.237 | 0.307 | 0.278 | 0.298 | 0.230 | **0.293** | 0.227 | 0.311 | 0.219 | 0.298 | 0.227 | 0.300 | 0.229 | 0.298 | 0.227 | 0.308 | 0.238 |
| | 9 | 0.376 | **0.235** | 0.382 | 0.238 | 0.381 | 0.264 | 0.377 | 0.261 | 0.392 | 0.271 | 0.386 | 0.316 | 0.384 | 0.266 | **0.369** | 0.264 | 0.401 | 0.253 | 0.385 | 0.263 | 0.386 | 0.265 | 0.382 | 0.263 | 0.387 | 0.273 |
| | 12 | 0.447 | **0.260** | 0.452 | 0.263 | 0.455 | 0.295 | 0.449 | 0.290 | 0.467 | 0.300 | 0.452 | 0.353 | 0.456 | 0.294 | **0.442** | 0.292 | 0.474 | 0.281 | 0.457 | 0.292 | 0.460 | 0.295 | 0.456 | 0.292 | 0.462 | 0.298 |
| | Avg | 0.330 | **0.218** | 0.335 | 0.220 | 0.334 | 0.245 | 0.331 | 0.242 | 0.344 | 0.251 | 0.341 | 0.294 | 0.335 | 0.245 | **0.327** | 0.243 | 0.349 | 0.233 | 0.336 | 0.242 | 0.338 | 0.244 | 0.335 | 0.243 | 0.344 | 0.253 |
| NASDAQ | 24 | 0.647 | **0.334** | 0.650 | 0.337 | 0.671 | 0.413 | 0.670 | 0.402 | 0.698 | 0.416 | 0.645 | 0.458 | **0.617** | 0.394 | 0.680 | 0.405 | 0.700 | 0.378 | 0.676 | 0.408 | 0.700 | 0.413 | 0.679 | 0.410 | 0.698 | 0.415 |
| | 36 | 0.799 | **0.382** | 0.800 | 0.388 | 0.841 | 0.480 | 0.824 | 0.471 | 0.874 | 0.490 | 0.785 | 0.533 | **0.781** | 0.457 | 0.841 | 0.471 | 0.874 | 0.448 | 0.852 | 0.477 | 0.867 | 0.480 | 0.845 | 0.484 | 0.856 | 0.475 |
| | 48 | 0.901 | **0.424** | 0.905 | 0.427 | 0.964 | 0.531 | 0.955 | 0.521 | 1.013 | 0.546 | 0.885 | 0.585 | **0.842** | 0.520 | 0.963 | 0.528 | 1.017 | 0.498 | 0.982 | 0.526 | 1.017 | 0.539 | 0.972 | 0.536 | 0.972 | 0.518 |
| | 60 | 0.993 | **0.447** | 0.999 | 0.457 | 1.047 | 0.573 | 1.050 | 0.563 | 1.122 | 0.589 | 0.959 | 0.623 | **0.958** | 0.551 | 1.029 | 0.556 | 1.126 | 0.541 | 1.084 | 0.569 | 1.079 | 0.572 | 1.077 | 0.578 | 1.033 | 0.543 |
| | Avg | 0.835 | **0.397** | 0.838 | 0.402 | 0.881 | 0.499 | 0.875 | 0.489 | 0.927 | 0.510 | 0.819 | 0.550 | **0.799** | 0.480 | 0.878 | 0.490 | 0.929 | 0.466 | 0.898 | 0.495 | 0.916 | 0.501 | 0.893 | 0.502 | 0.890 | 0.488 |
| | 3 | **0.033** | **0.089** | 0.036 | 0.092 | 0.035 | 0.093 | 0.038 | 0.100 | 0.042 | 0.111 | 0.044 | 0.123 | 0.036 | 0.098 | 0.040 | 0.103 | 0.037 | 0.096 | 0.039 | 0.102 | 0.040 | 0.105 | 0.038 | 0.099 | 0.049 | 0.123 |
| | 6 | **0.047** | **0.114** | 0.049 | 0.117 | 0.049 | 0.118 | 0.052 | 0.126 | 0.056 | 0.134 | 0.062 | 0.155 | 0.050 | 0.121 | 0.054 | 0.128 | 0.052 | 0.123 | 0.053 | 0.126 | 0.054 | 0.129 | 0.053 | 0.124 | 0.061 | 0.142 |
| | 9 | **0.060** | 0.138 | 0.062 | **0.137** | 0.062 | 0.139 | 0.065 | 0.145 | 0.069 | 0.154 | 0.082 | 0.189 | 0.063 | 0.144 | 0.066 | 0.147 | 0.063 | 0.141 | 0.067 | 0.147 | 0.068 | 0.150 | 0.065 | 0.145 | 0.073 | 0.161 |
| | 12 | **0.071** | **0.152** | 0.073 | 0.154 | 0.073 | 0.156 | 0.076 | 0.162 | 0.081 | 0.172 | 0.100 | 0.215 | 0.075 | 0.161 | 0.078 | 0.164 | 0.075 | 0.158 | 0.079 | 0.165 | 0.078 | 0.165 | 0.077 | 0.161 | 0.088 | 0.179 |
| | Avg | **0.053** | **0.124** | 0.055 | 0.125 | 0.055 | 0.126 | 0.058 | 0.133 | 0.062 | 0.142 | 0.072 | 0.170 | 0.056 | 0.131 | 0.059 | 0.135 | 0.057 | 0.130 | 0.059 | 0.135 | 0.060 | 0.137 | 0.058 | 0.132 | 0.068 | 0.151 |
| | 24 | **0.117** | **0.214** | 0.121 | 0.216 | 0.122 | 0.221 | 0.130 | 0.230 | 0.140 | 0.244 | 0.155 | 0.274 | 0.132 | 0.233 | 0.125 | 0.222 | 0.124 | 0.220 | 0.128 | 0.226 | 0.137 | 0.237 | 0.127 | 0.224 | 0.198 | 0.299 |
| | 36 | **0.157** | **0.259** | 0.163 | 0.261 | 0.183 | 0.279 | 0.175 | 0.273 | 0.184 | 0.284 | 0.196 | 0.306 | 0.177 | 0.278 | 0.174 | 0.271 | 0.167 | 0.266 | 0.170 | 0.268 | 0.184 | 0.280 | 0.174 | 0.269 | 0.229 | 0.326 |
| | 48 | **0.198** | **0.294** | 0.205 | 0.296 | 0.200 | 0.298 | 0.224 | 0.314 | 0.234 | 0.324 | 0.244 | 0.344 | 0.216 | 0.311 | 0.222 | 0.312 | 0.218 | 0.307 | 0.218 | 0.306 | 0.229 | 0.318 | 0.225 | 0.314 | 0.267 | 0.352 |
| | 60 | 0.255 | 0.332 | 0.259 | 0.336 | **0.238** | **0.328** | 0.259 | 0.340 | 0.282 | 0.357 | 0.318 | 0.401 | 0.249 | 0.337 | 0.264 | 0.341 | 0.264 | 0.341 | 0.262 | 0.339 | 0.279 | 0.352 | 0.265 | 0.339 | 0.327 | 0.394 |
| | Avg | **0.185** | **0.275** | 0.187 | 0.277 | 0.186 | 0.281 | 0.197 | 0.289 | 0.210 | 0.302 | 0.228 | 0.331 | 0.193 | 0.290 | 0.196 | 0.286 | 0.193 | 0.284 | 0.194 | 0.285 | 0.207 | 0.297 | 0.198 | 0.286 | 0.255 | 0.343 |
| Wiki | 3 | 6.153 | **0.366** | 6.161 | 0.368 | 6.209 | 0.392 | 6.234 | 0.402 | 7.470 | 0.496 | 6.254 | 0.438 | 6.149 | 0.389 | 6.148 | 0.383 | 6.183 | 0.378 | 6.190 | 0.387 | 6.237 | 0.393 | **6.112** | 0.380 | 7.597 | 0.510 |
| | 6 | 6.445 | **0.381** | 6.453 | 0.385 | 6.475 | 0.402 | 6.460 | 0.401 | 8.326 | 0.544 | 6.579 | 0.467 | 6.436 | 0.401 | 6.455 | 0.397 | 6.465 | 0.393 | 6.696 | 0.404 | 6.484 | 0.400 | **6.425** | 0.395 | 7.962 | 0.515 |
| | 9 | **6.657** | **0.398** | 6.666 | **0.398** | 6.702 | 0.418 | 6.697 | 0.416 | 8.869 | 0.564 | 6.776 | 0.508 | 6.714 | 0.420 | 6.687 | 0.412 | 6.714 | 0.415 | 6.768 | 0.411 | 6.689 | 0.411 | 6.743 | 0.426 | 8.150 | 0.524 |
| | 12 | 6.827 | **0.403** | 6.834 | 0.406 | 6.902 | 0.426 | 6.899 | 0.426 | 9.394 | 0.608 | 6.927 | 0.513 | 6.852 | 0.421 | 6.899 | 0.424 | 6.852 | 0.415 | 7.168 | 0.424 | 6.868 | 0.419 | **6.814** | 0.414 | 8.117 | 0.533 |
| | Avg | **6.521** | **0.389** | 6.528 | **0.389** | 6.572 | 0.409 | 6.572 | 0.411 | 8.515 | 0.553 | 6.634 | 0.481 | 6.538 | 0.408 | 6.547 | 0.404 | 6.553 | 0.400 | 6.705 | 0.406 | 6.569 | 0.405 | 6.523 | 0.404 | 7.956 | 0.520 |
| | 24 | 6.889 | **0.422** | 6.894 | 0.423 | 6.900 | 0.446 | 6.907 | 0.443 | 7.887 | 0.613 | 6.883 | 0.520 | 6.902 | 0.460 | 6.919 | 0.450 | 6.925 | 0.440 | **6.531** | 0.432 | 6.886 | 0.437 | 6.858 | 0.430 | 8.023 | 0.612 |
| | 36 | 6.444 | **0.437** | 6.446 | 0.439 | 6.520 | 0.467 | 6.514 | 0.467 | 7.774 | 0.660 | 6.393 | 0.538 | 6.539 | 0.473 | 6.456 | 0.457 | 6.463 | 0.451 | **5.935** | 0.453 | 6.431 | 0.452 | 6.400 | 0.445 | 7.229 | 0.595 |
| | 48 | 6.013 | **0.443** | 6.004 | 0.446 | 6.108 | 0.484 | 6.135 | 0.478 | 7.737 | 0.706 | 5.940 | 0.547 | 6.115 | 0.487 | 6.031 | 0.468 | 6.031 | 0.460 | **5.871** | 0.464 | 6.101 | 0.483 | 5.959 | 0.449 | 7.184 | 0.641 |
| | 60 | 5.702 | 0.453 | 5.705 | 0.454 | 5.732 | 0.476 | 5.811 | 0.482 | 7.855 | 0.746 | 5.605 | 0.552 | 5.736 | 0.497 | 5.740 | 0.478 | 5.723 | 0.463 | **5.389** | 0.463 | 5.681 | 0.462 | 5.633 | **0.452** | 6.805 | 0.645 |
| | Avg | 6.262 | **0.439** | 6.262 | 0.440 | 6.315 | 0.468 | 6.342 | 0.467 | 7.813 | 0.681 | 6.205 | 0.539 | 6.323 | 0.479 | 6.286 | 0.463 | 6.285 | 0.453 | **5.931** | 0.453 | 6.275 | 0.458 | 6.212 | 0.444 | 7.310 | 0.623 |
| 1st Count | | 30 | 47 | 0 | 4 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 4 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 3 | 1 | 0 | 0 |

strained cross-series coupling operator, the dual-domain temporal representation, and the prediction head. It is intended for research use on public multivariate time-series forecasting benchmarks.

The experimental protocol needed to interpret and reproduce the reported results is documented in the paper and appendix. App. C describes the datasets, splits, horizons, preprocessing, and metrics, while App. D describes the optimizer, learning-rate search, model dimensions, number of CSCO layers, calibration hyperparameters, early stopping, and baseline evaluation procedure. The supplementary material further provides implementation files and configuration information corresponding to the reported settings.

The supplementary code does not include private datasets, pretrained foundation models, user data, or high-risk generative components. Its purpose is to make the proposed forecasting architecture easier to inspect and reproduce. Users who extend the code to non-public or sensitive datasets should ensure that data usage, storage, and deployment comply with relevant institutional, legal, and ethical requirements.

## H Broader impact and responsible use

FISC is a general-purpose multivariate time-series forecasting framework. Its potential positive impact lies in improving forecasting accuracy and efficiency for applications such as energy-load

Table 16: Complete short-term forecasting results on the second group of datasets. For the short setting, the look-back window is 12 and horizons are $H \in \{3, 6, 9, 12\}$; for the long setting, the look-back window is 36 and horizons are $H \in \{24, 36, 48, 60\}$. The Avg row reports the average over four horizons. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 2.

| **Category** | | Linear-Based | | | | | | | | | | | | | | Transformer-Based | | | | | | | | | | TCN-Based | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Models** | | FISC ours | | OLinear 2025 | | TimeMixer 2024a | | FilterNet 2024a | | FITS 2024 | | DLinear 2023 | | TimeMixer++ 2025a | | Leddam 2024 | | CARD 2024b | | Fredformer 2024 | | iTransformer 2024a | | PatchTST 2023 | | TimesNet 2023b | |
| **Metric** | | 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 |
| SP500 | 3 | **0.034** | **0.125** | 0.035 | 0.126 | 0.038 | 0.137 | 0.046 | 0.154 | 0.046 | 0.157 | 0.047 | 0.152 | 0.040 | 0.141 | 0.047 | 0.155 | 0.036 | 0.130 | 0.045 | 0.153 | 0.046 | 0.155 | 0.042 | 0.146 | 0.037 | 0.133 |
| | 6 | **0.050** | **0.157** | 0.053 | 0.158 | 0.056 | 0.167 | 0.071 | 0.193 | 0.067 | 0.190 | 0.070 | 0.197 | 0.057 | 0.168 | 0.064 | 0.181 | 0.054 | 0.162 | 0.066 | 0.187 | 0.066 | 0.187 | 0.063 | 0.183 | 0.055 | 0.165 |
| | 9 | **0.067** | **0.178** | 0.070 | 0.181 | 0.074 | 0.192 | 0.080 | 0.203 | 0.087 | 0.218 | 0.091 | 0.220 | 0.076 | 0.196 | 0.081 | 0.206 | 0.072 | 0.184 | 0.080 | 0.205 | 0.080 | 0.203 | 0.081 | 0.207 | 0.071 | 0.187 |
| | 12 | **0.086** | **0.201** | 0.088 | 0.204 | 0.092 | 0.213 | 0.097 | 0.224 | 0.106 | 0.242 | 0.113 | 0.251 | 0.093 | 0.213 | 0.104 | 0.236 | 0.089 | 0.206 | 0.103 | 0.234 | 0.099 | 0.228 | 0.101 | 0.233 | 0.089 | 0.210 |
| | Avg | **0.059** | **0.165** | 0.061 | 0.167 | 0.065 | 0.177 | 0.073 | 0.193 | 0.076 | 0.202 | 0.080 | 0.205 | 0.066 | 0.179 | 0.074 | 0.194 | 0.063 | 0.170 | 0.073 | 0.195 | 0.073 | 0.193 | 0.072 | 0.192 | 0.063 | 0.174 |
| | 24 | **0.153** | **0.270** | 0.155 | 0.271 | 0.159 | 0.288 | 0.181 | 0.317 | 0.193 | 0.334 | 0.189 | 0.330 | 0.172 | 0.305 | 0.175 | 0.308 | 0.156 | 0.276 | 0.181 | 0.315 | 0.180 | 0.309 | 0.164 | 0.298 | 0.162 | 0.291 |
| | 36 | **0.206** | **0.315** | 0.209 | 0.317 | 0.218 | 0.343 | 0.224 | 0.341 | 0.259 | 0.389 | 0.250 | 0.363 | 0.227 | 0.344 | 0.232 | 0.358 | **0.206** | 0.319 | 0.239 | 0.365 | 0.225 | 0.346 | 0.221 | 0.341 | 0.219 | 0.344 |
| | 48 | **0.256** | 0.356 | 0.258 | 0.358 | 0.264 | 0.367 | 0.280 | 0.384 | 0.324 | 0.439 | 0.291 | 0.398 | 0.272 | 0.383 | 0.276 | 0.388 | 0.258 | **0.354** | 0.283 | 0.394 | 0.275 | 0.383 | 0.278 | 0.397 | 0.262 | 0.371 |
| | 60 | **0.302** | **0.383** | 0.305 | 0.387 | 0.322 | 0.416 | 0.332 | 0.416 | 0.391 | 0.486 | 0.377 | 0.475 | 0.319 | 0.413 | 0.325 | 0.423 | 0.303 | 0.385 | 0.341 | 0.438 | 0.322 | 0.418 | 0.321 | 0.409 | 0.305 | 0.399 |
| | Avg | **0.229** | **0.331** | 0.231 | 0.333 | 0.241 | 0.353 | 0.254 | 0.365 | 0.291 | 0.412 | 0.277 | 0.391 | 0.247 | 0.361 | 0.252 | 0.369 | 0.231 | 0.333 | 0.261 | 0.378 | 0.250 | 0.364 | 0.246 | 0.361 | 0.237 | 0.351 |
| DowJones | 3 | 1.540 | **0.275** | 1.550 | 0.276 | 1.568 | 0.276 | 1.587 | 0.278 | 1.611 | 0.283 | 1.574 | 0.280 | 1.555 | 0.278 | **1.538** | 0.277 | 1.563 | 0.276 | 1.563 | 0.277 | 1.551 | 0.276 | 1.556 | 0.276 | 2.269 | 0.325 |
| | 6 | **2.568** | **0.360** | 2.569 | 0.361 | 2.593 | 0.362 | 2.634 | 0.364 | 2.661 | 0.370 | 2.613 | 0.367 | 2.591 | 0.364 | 2.588 | 0.364 | 2.579 | 0.362 | 2.619 | 0.364 | 2.589 | 0.363 | 2.588 | 0.362 | 3.934 | 0.430 |
| | 9 | 3.565 | **0.427** | 3.565 | 0.430 | **3.564** | 0.431 | 3.651 | 0.434 | 3.704 | 0.443 | 3.621 | 0.436 | 3.572 | 0.433 | 3.644 | 0.440 | 3.569 | 0.430 | 3.575 | 0.432 | 3.599 | 0.433 | 3.585 | 0.432 | 4.715 | 0.492 |
| | 12 | 4.501 | **0.488** | 4.517 | 0.490 | **4.472** | 0.490 | 4.634 | 0.496 | 4.666 | 0.503 | 4.586 | 0.497 | 4.630 | 0.496 | 4.622 | 0.498 | 4.528 | 0.490 | 4.605 | 0.493 | 4.579 | 0.493 | 4.533 | 0.492 | 5.999 | 0.561 |
| | Avg | **3.044** | **0.388** | 3.050 | 0.389 | 3.049 | 0.390 | 3.126 | 0.393 | 3.160 | 0.400 | 3.099 | 0.395 | 3.087 | 0.393 | 3.098 | 0.395 | 3.060 | 0.389 | 3.090 | 0.391 | 3.079 | 0.391 | 3.065 | 0.390 | 4.229 | 0.452 |
| | 24 | 7.427 | **0.662** | 7.432 | 0.664 | 8.327 | 0.683 | 8.000 | 0.683 | 7.974 | 0.690 | 7.590 | 0.670 | 8.283 | 0.689 | 8.029 | 0.679 | **7.416** | 0.665 | 7.758 | 0.672 | 7.925 | 0.677 | 7.641 | 0.670 | 11.535 | 0.834 |
| | 36 | 10.832 | **0.784** | 10.848 | 0.799 | 11.192 | 0.813 | 12.011 | 0.823 | 11.907 | 0.837 | 10.986 | 0.803 | 14.754 | 0.856 | 11.962 | 0.828 | **10.799** | 0.798 | 11.456 | 0.808 | 12.087 | 0.827 | 11.210 | 0.807 | 16.922 | 0.982 |
| | 48 | 14.024 | **0.910** | 14.045 | 0.914 | 15.278 | 0.945 | 14.814 | 0.933 | 15.821 | 0.969 | 14.157 | 0.922 | 16.893 | 0.970 | 15.266 | 0.937 | **13.881** | 0.912 | 14.696 | 0.921 | 14.787 | 0.930 | 14.866 | 0.935 | 19.501 | 1.093 |
| | 60 | **16.853** | **1.012** | 16.959 | 1.017 | 20.997 | 1.067 | 18.932 | 1.054 | 19.320 | 1.077 | 18.018 | 1.035 | 18.881 | 1.059 | 18.407 | 1.045 | 17.257 | 1.021 | 18.058 | 1.032 | 18.298 | 1.041 | 17.947 | 1.036 | 22.804 | 1.177 |
| | Avg | **12.285** | **0.842** | 12.321 | 0.848 | 13.948 | 0.877 | 13.439 | 0.873 | 13.755 | 0.893 | 12.688 | 0.857 | 14.703 | 0.893 | 13.416 | 0.872 | 12.338 | 0.849 | 12.992 | 0.858 | 13.274 | 0.869 | 12.916 | 0.862 | 17.690 | 1.021 |
| CarSales | 3 | 0.293 | **0.274** | 0.303 | 0.277 | 0.307 | 0.296 | 0.304 | 0.291 | 0.411 | 0.375 | 0.396 | 0.383 | 0.309 | 0.289 | 0.316 | 0.300 | 0.318 | 0.295 | 0.311 | 0.300 | **0.281** | 0.276 | 0.300 | 0.290 | 0.315 | 0.303 |
| | 6 | 0.312 | **0.283** | 0.315 | 0.285 | 0.327 | 0.307 | 0.317 | 0.300 | 0.401 | 0.367 | 0.390 | 0.376 | 0.324 | 0.300 | 0.320 | 0.303 | 0.335 | 0.307 | 0.326 | 0.309 | **0.294** | 0.285 | 0.311 | 0.295 | 0.319 | 0.306 |
| | 9 | 0.325 | **0.292** | 0.327 | 0.293 | 0.336 | 0.315 | 0.329 | 0.308 | 0.418 | 0.380 | 0.406 | 0.390 | 0.331 | 0.306 | 0.332 | 0.311 | 0.351 | 0.319 | 0.339 | 0.318 | **0.305** | **0.292** | 0.324 | 0.306 | 0.338 | 0.318 |
| | 12 | 0.334 | **0.294** | 0.336 | 0.301 | 0.343 | 0.319 | 0.335 | 0.312 | 0.424 | 0.385 | 0.410 | 0.390 | 0.343 | 0.314 | 0.336 | 0.313 | 0.357 | 0.324 | 0.348 | 0.324 | **0.314** | 0.298 | 0.331 | 0.310 | 0.338 | 0.320 |
| | Avg | 0.316 | **0.286** | 0.320 | 0.289 | 0.328 | 0.309 | 0.321 | 0.303 | 0.413 | 0.377 | 0.400 | 0.385 | 0.327 | 0.302 | 0.326 | 0.306 | 0.340 | 0.311 | 0.331 | 0.313 | **0.298** | 0.288 | 0.316 | 0.300 | 0.327 | 0.312 |
| | 24 | 0.317 | **0.301** | 0.320 | 0.302 | 0.320 | 0.318 | 0.318 | 0.319 | 0.359 | 0.347 | 0.354 | 0.350 | 0.323 | 0.320 | 0.325 | 0.322 | 0.337 | 0.321 | 0.319 | 0.326 | **0.303** | 0.312 | 0.319 | 0.319 | 0.316 | 0.328 |
| | 36 | 0.332 | **0.313** | 0.334 | 0.315 | 0.332 | 0.331 | 0.331 | 0.330 | 0.373 | 0.360 | 0.368 | 0.365 | 0.351 | 0.348 | 0.337 | 0.333 | 0.348 | 0.333 | 0.333 | 0.335 | **0.318** | 0.323 | 0.332 | 0.330 | 0.338 | 0.343 |
| | 48 | 0.346 | **0.325** | 0.347 | 0.327 | 0.345 | 0.343 | 0.342 | 0.341 | 0.385 | 0.370 | 0.382 | 0.379 | 0.351 | 0.342 | 0.351 | 0.346 | 0.362 | 0.345 | 0.349 | 0.344 | **0.331** | 0.332 | 0.347 | 0.344 | 0.349 | 0.351 |
| | 60 | 0.355 | **0.336** | 0.358 | 0.337 | 0.355 | 0.351 | 0.352 | 0.349 | 0.399 | 0.385 | 0.388 | 0.380 | 0.363 | 0.352 | 0.361 | 0.353 | 0.372 | 0.353 | 0.359 | 0.349 | **0.344** | 0.342 | 0.355 | 0.348 | 0.359 | 0.359 |
| | Avg | 0.338 | **0.319** | 0.340 | 0.320 | 0.338 | 0.336 | 0.336 | 0.335 | 0.379 | 0.365 | 0.373 | 0.368 | 0.347 | 0.340 | 0.343 | 0.338 | 0.355 | 0.338 | 0.340 | 0.338 | **0.324** | 0.327 | 0.338 | 0.335 | 0.340 | 0.345 |
| Power | 3 | 0.848 | **0.681** | 0.864 | 0.688 | 0.850 | 0.683 | 0.842 | 0.685 | 0.899 | 0.712 | 0.876 | 0.708 | 0.843 | 0.685 | 0.859 | 0.696 | 0.896 | 0.691 | 0.865 | 0.697 | 0.863 | 0.696 | **0.830** | 0.687 | 0.929 | 0.716 |
| | 6 | 0.987 | **0.740** | 0.991 | 0.742 | 0.971 | 0.743 | 0.988 | 0.747 | 1.079 | 0.785 | 0.991 | 0.761 | 0.997 | 0.748 | 1.023 | 0.761 | 1.013 | 0.750 | 0.994 | 0.751 | 0.990 | 0.748 | **0.948** | 0.741 | 1.070 | 0.773 |
| | 9 | **1.017** | 0.769 | 1.062 | 0.770 | 1.024 | **0.763** | 1.063 | 0.778 | 1.150 | 0.804 | 1.051 | 0.790 | 1.050 | 0.767 | 1.071 | 0.783 | 1.085 | 0.781 | 1.073 | 0.784 | 1.065 | 0.780 | 1.027 | 0.773 | 1.104 | 0.788 |
| | 12 | 1.114 | **0.787** | 1.119 | 0.789 | 1.087 | 0.788 | 1.125 | 0.803 | 1.266 | 0.851 | 1.110 | 0.814 | 1.107 | 0.792 | 1.130 | 0.804 | 1.124 | 0.792 | 1.135 | 0.807 | 1.142 | 0.808 | **1.083** | 0.797 | 1.212 | 0.825 |
| | Avg | 0.992 | **0.744** | 1.009 | 0.747 | 0.983 | **0.744** | 1.004 | 0.753 | 1.098 | 0.788 | 1.007 | 0.768 | 0.999 | 0.748 | 1.021 | 0.761 | 1.029 | 0.753 | 1.017 | 0.760 | 1.015 | 0.758 | **0.972** | 0.749 | 1.079 | 0.775 |
| | 24 | **1.337** | **0.866** | 1.343 | 0.870 | 1.341 | 0.881 | 1.410 | 0.916 | 1.491 | 0.944 | 1.390 | 0.916 | 1.340 | 0.877 | 1.397 | 0.909 | 1.406 | 0.886 | 1.410 | 0.913 | 1.462 | 0.924 | 1.468 | 0.935 | 1.494 | 0.924 |
| | 36 | 1.437 | **0.901** | 1.445 | 0.903 | **1.420** | 0.914 | 1.590 | 0.968 | 1.621 | 0.994 | 1.518 | 0.957 | 1.446 | 0.920 | 1.509 | 0.951 | 1.506 | 0.921 | 1.538 | 0.953 | 1.582 | 0.964 | 1.593 | 0.972 | 1.526 | 0.950 |
| | 48 | 1.538 | 0.940 | 1.559 | 0.946 | 1.567 | 0.963 | 1.680 | 1.009 | 1.775 | 1.052 | 1.610 | 0.995 | **1.467** | **0.933** | 1.646 | 0.999 | 1.583 | 0.957 | 1.652 | 1.008 | 1.696 | 1.011 | 1.710 | 1.020 | 1.581 | 0.981 |
| | 60 | **1.597** | **0.967** | 1.602 | 0.971 | 1.609 | 0.988 | 1.776 | 1.053 | 1.958 | 1.122 | 1.679 | 1.020 | 1.626 | 1.006 | 1.727 | 1.043 | 1.693 | 1.003 | 1.752 | 1.049 | 1.796 | 1.061 | 1.829 | 1.064 | 1.625 | 1.010 |
| | Avg | 1.477 | **0.919** | 1.487 | 0.922 | 1.484 | 0.937 | 1.614 | 0.986 | 1.711 | 1.028 | 1.549 | 0.972 | **1.470** | 0.934 | 1.570 | 0.975 | 1.547 | 0.942 | 1.588 | 0.981 | 1.634 | 0.990 | 1.650 | 0.998 | 1.556 | 0.966 |
| Website | 3 | **0.070** | **0.194** | 0.077 | 0.199 | 0.086 | 0.215 | 0.084 | 0.213 | 0.191 | 0.320 | 0.159 | 0.288 | 0.083 | 0.208 | 0.082 | 0.210 | 0.099 | 0.225 | 0.080 | 0.207 | 0.072 | 0.198 | 0.089 | 0.217 | 0.075 | 0.200 |
| | 6 | **0.092** | **0.224** | 0.103 | 0.227 | 0.124 | 0.248 | 0.116 | 0.242 | 0.235 | 0.356 | 0.182 | 0.302 | 0.115 | 0.237 | 0.114 | 0.239 | 0.135 | 0.255 | 0.116 | 0.241 | 0.098 | 0.227 | 0.121 | 0.246 | 0.105 | 0.232 |
| | 9 | **0.121** | **0.249** | 0.135 | 0.251 | 0.159 | 0.275 | 0.151 | 0.269 | 0.276 | 0.372 | 0.220 | 0.330 | 0.156 | 0.273 | 0.147 | 0.266 | 0.171 | 0.282 | 0.150 | 0.268 | 0.124 | 0.252 | 0.157 | 0.273 | 0.138 | 0.262 |
| | 12 | 0.149 | **0.269** | 0.176 | 0.281 | 0.204 | 0.306 | 0.194 | 0.297 | 0.409 | 0.484 | 0.255 | 0.355 | 0.193 | 0.296 | 0.196 | 0.298 | 0.213 | 0.307 | 0.196 | 0.301 | **0.144** | **0.269** | 0.200 | 0.302 | 0.169 | 0.285 |
| | Avg | **0.108** | **0.234** | 0.123 | 0.240 | 0.143 | 0.261 | 0.136 | 0.255 | 0.278 | 0.383 | 0.204 | 0.319 | 0.137 | 0.253 | 0.135 | 0.253 | 0.154 | 0.267 | 0.135 | 0.254 | 0.109 | 0.237 | 0.141 | 0.259 | 0.122 | 0.244 |
| | 24 | **0.179** | **0.302** | 0.186 | 0.306 | 0.229 | 0.335 | 0.273 | 0.357 | 0.431 | 0.469 | 0.315 | 0.393 | 0.231 | 0.349 | 0.240 | 0.345 | 0.325 | 0.370 | 0.216 | 0.335 | 0.181 | 0.305 | 0.245 | 0.350 | 0.276 | 0.358 |
| | 36 | **0.223** | **0.340** | 0.272 | 0.356 | 0.361 | 0.420 | 0.401 | 0.441 | 0.554 | 0.552 | 0.385 | 0.447 | 0.328 | 0.397 | 0.327 | 0.405 | 0.428 | 0.442 | 0.331 | 0.411 | 0.226 | 0.342 | 0.370 | 0.429 | 0.377 | 0.436 |
| | 48 | **0.259** | 0.670 | 0.365 | 0.391 | 0.501 | 0.507 | 0.530 | 0.522 | 0.694 | 0.647 | 0.436 | 0.486 | 0.450 | 0.473 | 0.446 | 0.475 | 0.457 | 0.478 | 0.483 | 0.496 | 0.263 | **0.370** | 0.504 | 0.513 | 0.389 | 0.446 |
| | 60 | 0.330 | **0.409** | 0.486 | 0.481 | 0.571 | 0.562 | 0.630 | 0.592 | 0.736 | 0.673 | 0.468 | 0.510 | 0.525 | 0.517 | 0.561 | 0.549 | 0.596 | 0.566 | 0.556 | 0.547 | **0.323** | 0.410 | 0.587 | 0.565 | 0.478 | 0.507 |
| | Avg | **0.248** | **0.355** | 0.327 | 0.383 | 0.415 | 0.456 | 0.458 | 0.478 | 0.604 | 0.585 | 0.401 | 0.459 | 0.384 | 0.434 | 0.393 | 0.443 | 0.451 | 0.464 | 0.396 | 0.447 | **0.248** | 0.357 | 0.426 | 0.464 | 0.380 | 0.437 |
| Unemp | 3 | **0.011** | **0.046** | 0.012 | 0.047 | 0.015 | 0.074 | 0.012 | 0.062 | 0.161 | 0.289 | 0.072 | 0.200 | 0.014 | 0.069 | 0.013 | 0.061 | 0.014 | 0.070 | 0.013 | 0.068 | 0.012 | 0.061 | 0.012 | 0.060 | 0.062 | 0.171 |
| | 6 | **0.040** | **0.106** | 0.041 | 0.108 | 0.057 | 0.154 | 0.043 | 0.130 | 0.229 | 0.345 | 0.115 | 0.255 | 0.079 | 0.194 | **0.040** | 0.122 | 0.042 | 0.127 | 0.046 | 0.139 | 0.047 | 0.136 | 0.043 | 0.127 | 0.111 | 0.227 |
| | 9 | **0.083** | **0.168** | 0.084 | 0.170 | 0.109 | 0.213 | 0.107 | 0.216 | 0.369 | 0.443 | 0.191 | 0.329 | 0.101 | 0.200 | 0.119 | 0.226 | 0.093 | 0.194 | 0.095 | 0.198 | 0.098 | 0.199 | 0.093 | 0.190 | 0.176 | 0.281 |
| | 12 | **0.128** | **0.219** | 0.131 | 0.220 | 0.195 | 0.293 | 0.155 | 0.255 | 0.475 | 0.500 | 0.240 | 0.386 | 0.196 | 0.291 | 0.141 | 0.239 | 0.155 | 0.256 | 0.148 | 0.250 | 0.158 | 0.256 | 0.164 | 0.261 | 0.240 | 0.326 |
| | Avg | **0.066** | **0.135** | 0.067 | 0.136 | 0.094 | 0.183 | 0.079 | 0.166 | 0.308 | 0.394 | 0.154 | 0.292 | 0.097 | 0.188 | 0.078 | 0.162 | 0.076 | 0.162 | 0.075 | 0.163 | 0.079 | 0.163 | 0.078 | 0.160 | 0.147 | 0.251 |
| | 24 | **0.445** | **0.436** | 0.458 | 0.447 | 0.668 | 0.549 | 0.655 | 0.548 | 2.520 | 1.255 | 0.951 | 0.645 | 0.779 | 0.578 | 0.861 | 0.645 | 0.583 | 0.522 | 0.545 | 0.496 | 0.719 | 0.595 | 0.620 | 0.536 | 1.798 | 0.967 |
| | 36 | **0.867** | **0.614** | 0.870 | 0.619 | 1.580 | 0.866 | 1.247 | 0.769 | 4.478 | 1.703 | 0.906 | 0.717 | 1.318 | 0.777 | 1.741 | 0.938 | 1.205 | 0.762 | 1.107 | 0.725 | 2.212 | 1.103 | 2.245 | 1.120 | 2.820 | 1.171 |
| | 48 | 1.633 | 0.891 | 1.643 | 0.900 | 3.702 | 1.483 | 2.700 | 1.209 | 6.385 | 2.052 | **0.974** | **0.747** | 2.977 | 1.258 | 3.526 | 1.427 | 2.737 | 1.218 | 6.192 | 2.055 | 2.577 | 1.178 | 2.644 | 1.191 | 5.909 | 1.620 |
| | 60 | 1.716 | 1.024 | 2.593 | 1.182 | 6.320 | 2.030 | 5.370 | 1.830 | 8.460 | 2.392 | **1.071** | **0.798** | 6.582 | 2.060 | 5.574 | 1.857 | 4.475 | 1.624 | 8.236 | 2.384 | 5.441 | 1.799 | 4.074 | 1.511 | 7.511 | 2.053 |
| | Avg | 1.165 | 0.741 | 1.391 | 0.787 | 3.068 | 1.232 | 2.493 | 1.089 | 5.461 | 1.850 | **0.975** | **0.727** | 2.914 | 1.168 | 2.925 | 1.217 | 2.250 | 1.031 | 4.020 | 1.415 | 2.737 | 1.169 | 2.396 | 1.089 | 4.510 | 1.452 |
| **1st Count** | | 32 | 53 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 3 | 3 | 2 | 1 | 2 | 0 | 4 | 1 | 0 | 0 | 13 | 3 | 4 | 0 | 0 | 0 |

Table 17: Complete ablation results for statistical calibration injection strategies. We compare FISC with three variants: NoSC removes statistical calibration, LastSC injects statistical calibration only in the last layer, and AllSC injects statistical calibration in all coupling layers. The table reports MSE and MAE for four forecast horizons and their averages. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 4.

| Model | Metric | ETTh1 | | Weather | | Electricity | | PEMS04 | | PEMS08 | | ILI (S) | | ILI (L) | | NASDAQ (S) | | NASDAQ (L) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| FISC | H1 | **0.357** | **0.379** | **0.148** | **0.185** | **0.130** | **0.220** | **0.067** | **0.161** | **0.065** | **0.157** | **0.440** | **0.338** | **1.729** | **0.786** | **0.033** | **0.089** | **0.117** | **0.214** |
| | H2 | **0.410** | **0.412** | **0.194** | **0.233** | **0.146** | **0.235** | **0.078** | **0.174** | **0.086** | **0.177** | **0.891** | **0.510** | **1.710** | **0.794** | **0.047** | **0.114** | **0.157** | **0.259** |
| | H3 | **0.439** | **0.433** | **0.241** | **0.276** | **0.163** | **0.248** | **0.093** | **0.196** | **0.121** | **0.201** | **1.287** | **0.651** | **1.813** | **0.801** | **0.060** | **0.138** | **0.198** | **0.294** |
| | H4 | **0.457** | **0.438** | **0.322** | **0.329** | **0.181** | **0.273** | **0.121** | **0.224** | **0.170** | **0.233** | **1.650** | **0.769** | **1.774** | **0.807** | **0.071** | **0.152** | **0.255** | **0.332** |
| | Avg | **0.416** | **0.416** | **0.226** | **0.256** | **0.155** | **0.244** | **0.090** | **0.189** | **0.111** | **0.192** | **1.067** | **0.567** | **1.757** | **0.797** | **0.053** | **0.124** | **0.185** | **0.275** |
| NoSC | H1 | 0.362 | 0.381 | 0.153 | 0.191 | 0.135 | 0.226 | 0.069 | 0.164 | 0.067 | 0.159 | 0.457 | 0.343 | 1.905 | 0.819 | 0.036 | 0.094 | 0.124 | 0.216 |
| | H2 | 0.420 | 0.423 | 0.208 | 0.240 | 0.151 | 0.239 | 0.081 | 0.177 | 0.088 | 0.179 | 0.915 | 0.513 | 1.820 | 0.818 | 0.050 | 0.118 | 0.167 | 0.263 |
| | H3 | 0.455 | 0.437 | 0.259 | 0.279 | 0.168 | 0.256 | 0.097 | 0.199 | 0.124 | 0.206 | 1.335 | 0.663 | 1.870 | 0.819 | 0.062 | 0.140 | 0.209 | 0.300 |
| | H4 | 0.462 | 0.460 | 0.339 | 0.332 | 0.190 | 0.278 | 0.125 | 0.227 | 0.179 | 0.239 | 1.761 | 0.800 | 1.800 | 0.813 | 0.074 | 0.156 | 0.257 | 0.334 |
| | Avg | 0.425 | 0.425 | 0.240 | 0.261 | 0.161 | 0.250 | 0.093 | 0.192 | 0.115 | 0.196 | 1.117 | 0.580 | 1.849 | 0.817 | 0.056 | 0.127 | 0.189 | 0.278 |
| LastSC | H1 | 0.361 | 0.381 | 0.149 | 0.188 | 0.132 | 0.221 | 0.068 | 0.164 | 0.066 | 0.158 | 0.447 | 0.341 | 1.897 | 0.810 | 0.035 | 0.094 | 0.121 | 0.216 |
| | H2 | 0.412 | 0.415 | 0.201 | 0.235 | 0.148 | 0.238 | 0.079 | 0.177 | 0.088 | 0.178 | 0.897 | 0.512 | 1.820 | 0.818 | 0.050 | 0.117 | 0.165 | 0.263 |
| | H3 | 0.453 | 0.435 | 0.256 | 0.277 | 0.166 | 0.255 | 0.096 | 0.197 | 0.124 | 0.204 | 1.334 | 0.661 | 1.863 | 0.816 | 0.061 | **0.138** | 0.206 | 0.296 |
| | H4 | 0.459 | 0.460 | 0.335 | 0.331 | 0.185 | 0.276 | 0.123 | 0.227 | 0.176 | 0.236 | 1.680 | 0.776 | 1.812 | 0.813 | 0.072 | 0.154 | 0.256 | 0.334 |
| | Avg | 0.421 | 0.423 | 0.235 | 0.258 | 0.158 | 0.248 | 0.092 | 0.191 | 0.114 | 0.194 | 1.090 | 0.573 | 1.848 | 0.814 | 0.055 | 0.126 | 0.187 | 0.277 |
| AllSC | H1 | 0.361 | 0.381 | 0.151 | 0.190 | 0.132 | 0.222 | 0.068 | 0.164 | 0.067 | 0.159 | 0.455 | 0.342 | 1.903 | 0.819 | 0.034 | 0.093 | 0.122 | 0.216 |
| | H2 | 0.417 | 0.418 | 0.204 | 0.238 | 0.150 | 0.239 | 0.079 | 0.177 | 0.088 | **0.177** | 0.912 | 0.512 | 1.820 | 0.817 | 0.050 | 0.118 | 0.165 | 0.262 |
| | H3 | 0.454 | 0.437 | 0.258 | 0.277 | 0.168 | 0.256 | 0.096 | 0.197 | 0.124 | 0.206 | 1.335 | 0.663 | 1.874 | 0.819 | 0.061 | 0.139 | 0.207 | 0.296 |
| | H4 | 0.460 | 0.461 | 0.337 | 0.331 | 0.185 | 0.279 | 0.123 | 0.227 | 0.178 | 0.236 | 1.761 | 0.801 | 1.798 | 0.813 | 0.072 | 0.154 | 0.256 | 0.334 |
| | Avg | 0.423 | 0.424 | 0.238 | 0.259 | 0.159 | 0.249 | 0.092 | 0.191 | 0.114 | 0.195 | 1.116 | 0.580 | 1.849 | 0.817 | 0.054 | 0.126 | 0.188 | 0.277 |

Table 18: Complete ablation results for the dual-domain representation and orthogonal basis selection. Rows compare FISC with single-branch variants and fixed transform bases. For S and L, the prediction horizons are $H \in \{3, 6, 9, 12\}$ and $H \in \{24, 36, 48, 60\}$, respectively. The table reports MSE and MAE. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 5.

| Model | Metric | ETTh1 | | Weather | | Electricity | | PEMS04 | | PEMS08 | | ILI (S) | | ILI (L) | | NASDAQ (S) | | NASDAQ (L) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| FISC | H1 | **0.357** | **0.379** | 0.148 | **0.185** | **0.130** | **0.220** | **0.067** | **0.161** | **0.065** | **0.157** | **0.440** | **0.338** | **1.729** | **0.786** | **0.033** | **0.089** | **0.117** | **0.214** |
| | H2 | **0.410** | **0.412** | **0.194** | **0.233** | **0.146** | **0.235** | **0.078** | **0.174** | **0.086** | **0.177** | 0.891 | **0.510** | 1.710 | **0.794** | **0.047** | **0.114** | **0.157** | **0.259** |
| | H3 | **0.439** | **0.433** | **0.241** | 0.276 | **0.163** | **0.248** | **0.093** | **0.196** | **0.121** | **0.201** | 1.287 | 0.651 | **1.813** | **0.801** | **0.060** | **0.138** | **0.198** | **0.294** |
| | H4 | **0.457** | **0.438** | **0.322** | **0.329** | **0.181** | **0.273** | **0.121** | **0.224** | **0.170** | **0.233** | **1.650** | **0.769** | 1.774 | **0.807** | **0.071** | **0.152** | 0.255 | **0.332** |
| | Avg | **0.416** | **0.416** | **0.226** | **0.256** | **0.155** | **0.244** | **0.090** | **0.189** | **0.111** | **0.192** | **1.067** | **0.567** | **1.757** | **0.797** | **0.053** | **0.124** | **0.185** | **0.275** |
| OnlyTime | H1 | 0.384 | 0.394 | **0.147** | 0.186 | 0.132 | 0.222 | 0.263 | 0.376 | 0.068 | 0.159 | 0.505 | 0.362 | 2.029 | 0.832 | 0.036 | 0.093 | 0.122 | 0.216 |
| | H2 | 0.435 | 0.424 | 0.198 | 0.235 | 0.153 | 0.241 | 0.079 | 0.177 | 0.089 | **0.177** | **0.885** | 0.511 | **1.688** | 0.797 | 0.050 | 0.119 | 0.165 | 0.262 |
| | H3 | 0.474 | 0.445 | 0.256 | 0.278 | 0.165 | 0.254 | 0.096 | 0.197 | 0.124 | 0.206 | **1.218** | **0.643** | 1.943 | 0.827 | 0.062 | **0.138** | 0.208 | 0.299 |
| | H4 | 0.469 | 0.468 | 0.337 | 0.331 | 0.200 | 0.285 | 0.124 | 0.228 | 0.187 | 0.239 | 1.750 | 0.803 | **1.768** | 0.809 | 0.073 | 0.155 | 0.254 | 0.334 |
| | Avg | 0.441 | 0.433 | 0.235 | 0.258 | 0.163 | 0.251 | 0.141 | 0.245 | 0.117 | 0.195 | 1.090 | 0.580 | 1.857 | 0.816 | 0.055 | 0.126 | 0.187 | 0.278 |
| OnlyOrth | H1 | 0.366 | 0.397 | 0.149 | 0.193 | 0.132 | 0.223 | 0.068 | 0.168 | 0.067 | 0.164 | 0.449 | 0.350 | **1.729** | 0.822 | 0.034 | 0.095 | 0.124 | 0.224 |
| | H2 | 0.413 | 0.421 | 0.201 | 0.239 | 0.152 | 0.238 | 0.079 | 0.176 | 0.088 | 0.180 | 0.900 | 0.515 | 1.790 | 0.818 | 0.051 | 0.118 | 0.169 | 0.269 |
| | H3 | 0.460 | 0.443 | 0.262 | 0.288 | 0.171 | 0.257 | 0.094 | 0.198 | **0.121** | 0.210 | 1.315 | 0.653 | 1.925 | 0.843 | 0.062 | 0.145 | 0.204 | 0.302 |
| | H4 | 0.466 | 0.454 | 0.330 | 0.343 | 0.184 | 0.275 | 0.122 | 0.232 | 0.175 | 0.241 | 1.663 | 0.781 | 1.771 | 0.835 | **0.071** | 0.156 | 0.258 | 0.340 |
| | Avg | 0.426 | 0.429 | 0.235 | 0.266 | 0.160 | 0.248 | 0.091 | 0.193 | 0.113 | 0.199 | 1.082 | 0.575 | 1.804 | 0.830 | 0.055 | 0.129 | 0.189 | 0.284 |
| Fourier | H1 | 0.371 | 0.385 | 0.154 | 0.190 | 0.134 | 0.230 | 0.070 | 0.164 | **0.065** | 0.162 | 0.450 | 0.353 | 1.745 | 0.787 | 0.034 | 0.094 | 0.126 | 0.221 |
| | H2 | 0.429 | 0.426 | 0.200 | 0.246 | 0.153 | 0.241 | 0.082 | 0.177 | **0.086** | **0.177** | 0.902 | 0.533 | 1.730 | 0.819 | 0.050 | 0.119 | 0.169 | 0.270 |
| | H3 | 0.459 | 0.439 | 0.259 | 0.276 | 0.166 | 0.260 | 0.095 | 0.205 | 0.123 | 0.203 | 1.311 | 0.679 | 1.847 | 0.845 | 0.061 | 0.144 | 0.208 | 0.302 |
| | H4 | 0.466 | 0.443 | 0.343 | 0.342 | **0.182** | 0.277 | 0.124 | 0.226 | 0.175 | 0.239 | 1.693 | 0.800 | 1.813 | 0.824 | **0.071** | 0.158 | **0.253** | 0.319 |
| | Avg | 0.431 | 0.423 | 0.239 | 0.264 | 0.159 | 0.252 | 0.093 | 0.193 | 0.112 | 0.195 | 1.089 | 0.591 | 1.784 | 0.819 | 0.054 | 0.129 | 0.189 | 0.278 |
| Wavelet | H1 | 0.361 | 0.381 | 0.149 | 0.188 | 0.132 | 0.221 | 0.068 | 0.164 | 0.066 | 0.158 | 0.447 | 0.341 | 1.897 | 0.810 | 0.035 | 0.094 | 0.121 | 0.216 |
| | H2 | 0.412 | 0.415 | 0.201 | 0.235 | 0.148 | 0.238 | 0.079 | 0.177 | 0.088 | 0.178 | 0.897 | 0.512 | 1.820 | 0.818 | 0.050 | 0.117 | 0.165 | 0.263 |
| | H3 | 0.453 | 0.435 | 0.256 | 0.277 | 0.166 | 0.255 | 0.096 | 0.197 | 0.124 | 0.204 | 1.334 | 0.661 | 1.863 | 0.816 | 0.061 | **0.138** | 0.206 | 0.296 |
| | H4 | 0.459 | 0.460 | 0.335 | 0.331 | 0.185 | 0.276 | 0.123 | 0.227 | 0.176 | 0.236 | 1.680 | 0.776 | 1.812 | 0.813 | 0.072 | 0.154 | 0.256 | 0.334 |
| | Avg | 0.421 | 0.423 | 0.235 | 0.258 | 0.158 | 0.248 | 0.092 | 0.191 | 0.114 | 0.194 | 1.090 | 0.573 | 1.848 | 0.814 | 0.055 | 0.126 | 0.187 | 0.277 |
| Chebyshev | H1 | 0.365 | 0.404 | 0.158 | 0.193 | 0.138 | 0.229 | 0.070 | 0.168 | 0.066 | 0.158 | 0.457 | 0.362 | 1.819 | 0.825 | 0.035 | 0.098 | 0.129 | 0.224 |
| | H2 | 0.422 | 0.432 | 0.200 | 0.243 | 0.155 | 0.250 | **0.078** | 0.180 | 0.091 | 0.186 | 0.947 | 0.530 | 1.808 | 0.819 | 0.050 | 0.122 | 0.174 | 0.277 |
| | H3 | 0.469 | 0.442 | 0.272 | 0.276 | 0.164 | 0.266 | 0.097 | 0.205 | 0.130 | 0.209 | 1.365 | 0.676 | 1.979 | 0.872 | 0.065 | 0.142 | 0.216 | 0.300 |
| | H4 | 0.487 | 0.448 | 0.340 | 0.340 | 0.189 | 0.280 | 0.124 | 0.235 | 0.176 | 0.245 | 1.668 | 0.791 | 1.810 | 0.834 | 0.075 | 0.160 | 0.232 | 0.330 |
| | Avg | 0.436 | 0.432 | 0.243 | 0.263 | 0.162 | 0.256 | 0.092 | 0.197 | 0.116 | 0.199 | 1.109 | 0.590 | 1.854 | 0.838 | 0.056 | 0.131 | 0.188 | 0.283 |

Table 19: Complete ablation results for CSCO constraints and first-layer statistical calibration. We compare FISC with unconstrained linear mixing, nonnegativity-only coupling (Nonneg), row normalization only (Rownorm), row-sum constraint only (Rowsum), and NoFusion, which removes the statistical calibration term while preserving the projection constraint. The table reports MSE and MAE for four forecast horizons and their averages. Best and second-best results are marked in **red** and blue, respectively. This table is the full version of Table 6.

| Model | Metric | ETTh1 | | Weather | | Electricity | | PEMS04 | | PEMS08 | | ILI (S) | | ILI (L) | | NASDAQ (S) | | NASDAQ (L) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE | MSE | MAE |
| FISC | H1 | **0.357** | **0.379** | **0.148** | **0.185** | **0.130** | **0.220** | **0.067** | **0.161** | **0.065** | **0.157** | **0.440** | **0.338** | **1.729** | **0.786** | **0.033** | **0.089** | **0.117** | **0.214** |
| | H2 | **0.410** | **0.412** | **0.194** | **0.233** | **0.146** | **0.235** | **0.078** | **0.174** | **0.086** | **0.177** | 0.891 | 0.510 | 1.710 | **0.794** | **0.047** | **0.114** | **0.157** | **0.259** |
| | H3 | **0.439** | **0.433** | **0.241** | **0.276** | **0.163** | **0.248** | **0.093** | **0.196** | **0.121** | **0.201** | 1.287 | **0.651** | **1.813** | **0.801** | **0.060** | 0.138 | **0.198** | **0.294** |
| | H4 | **0.457** | **0.438** | **0.322** | **0.329** | **0.181** | **0.273** | **0.121** | **0.224** | **0.170** | **0.233** | **1.650** | **0.769** | 1.774 | **0.807** | **0.071** | **0.152** | 0.255 | **0.332** |
| | Avg | **0.416** | **0.416** | **0.226** | **0.256** | **0.155** | **0.244** | **0.090** | **0.189** | **0.111** | **0.192** | **1.067** | **0.567** | **1.757** | **0.797** | **0.053** | **0.124** | **0.185** | **0.275** |
| Unconstrained | H1 | 0.364 | 0.383 | 0.152 | 0.191 | 0.155 | 0.250 | 0.068 | 0.461 | 0.069 | 0.158 | 0.552 | 0.374 | 1.800 | 0.816 | 0.036 | 0.093 | 0.120 | 0.215 |
| | H2 | 0.416 | 0.414 | 0.205 | 0.242 | 0.205 | 0.242 | 0.080 | 0.176 | 0.087 | 0.179 | 1.085 | 0.552 | 1.860 | 0.814 | 0.050 | 0.118 | 0.165 | 0.260 |
| | H3 | 0.460 | 0.440 | 0.256 | 0.279 | 0.190 | 0.282 | 0.096 | 0.198 | 0.126 | 0.205 | 1.721 | 0.741 | 1.858 | 0.816 | 0.061 | 0.139 | 0.209 | 0.300 |
| | H4 | 0.476 | 0.471 | 0.342 | 0.334 | 0.213 | 0.301 | 0.123 | 0.226 | 0.171 | 0.237 | 2.410 | 0.922 | 1.774 | 0.812 | 0.073 | 0.154 | 0.254 | 0.334 |
| | Avg | 0.429 | 0.427 | 0.239 | 0.262 | 0.191 | 0.269 | 0.092 | 0.265 | 0.113 | 0.195 | 1.442 | 0.647 | 1.823 | 0.815 | 0.055 | 0.126 | 0.187 | 0.277 |
| Nonneg | H1 | 0.373 | 0.387 | 0.173 | 0.211 | 0.196 | 0.265 | 0.128 | 0.234 | 0.119 | 0.226 | 0.704 | 0.461 | 2.695 | 0.950 | 0.037 | 0.097 | 0.127 | 0.224 |
| | H2 | 0.429 | 0.421 | 0.218 | 0.248 | 0.198 | 0.271 | 0.238 | 0.323 | 0.235 | 0.319 | 1.248 | 0.637 | 1.982 | 0.896 | 0.053 | 0.124 | 0.170 | 0.268 |
| | H3 | 0.464 | 0.439 | 0.268 | 0.286 | 0.212 | 0.287 | 0.560 | 0.513 | 0.247 | 0.324 | 1.947 | 0.825 | 2.190 | 0.915 | 0.066 | 0.144 | 0.218 | 0.307 |
| | H4 | 0.479 | 0.473 | 0.344 | 0.337 | 0.255 | 0.320 | 1.118 | 0.775 | 1.141 | 0.773 | 2.992 | 1.081 | 1.955 | 0.886 | 0.077 | 0.161 | 0.262 | 0.340 |
| | Avg | 0.436 | 0.430 | 0.251 | 0.271 | 0.215 | 0.286 | 0.511 | 0.461 | 0.436 | 0.411 | 1.723 | 0.751 | 2.206 | 0.912 | 0.058 | 0.132 | 0.194 | 0.285 |
| Rownorm | H1 | 0.365 | 0.383 | 0.163 | 0.197 | 0.157 | 0.253 | 0.135 | 0.244 | 0.123 | 0.232 | 0.554 | 0.374 | 1.797 | 0.814 | 0.035 | 0.091 | 0.119 | **0.214** |
| | H2 | 0.416 | 0.414 | 0.209 | 0.249 | 0.208 | 0.245 | 0.242 | 0.338 | 0.240 | 0.331 | 1.079 | 0.552 | 1.860 | 0.817 | 0.048 | 0.115 | 0.165 | 0.263 |
| | H3 | 0.459 | 0.439 | 0.262 | 0.284 | 0.197 | 0.286 | 0.592 | 0.530 | 0.251 | 0.331 | 1.709 | 0.738 | 1.859 | 0.819 | 0.061 | **0.136** | 0.211 | 0.302 |
| | H4 | 0.479 | 0.472 | 0.341 | 0.335 | 0.215 | 0.304 | 1.129 | 0.799 | 1.192 | 0.793 | 2.729 | 0.998 | **1.772** | 0.812 | 0.072 | 0.153 | 0.254 | 0.336 |
| | Avg | 0.430 | 0.427 | 0.244 | 0.266 | 0.194 | 0.272 | 0.525 | 0.478 | 0.452 | 0.422 | 1.518 | 0.666 | 1.822 | 0.816 | 0.054 | **0.124** | 0.187 | 0.279 |
| Rowsum | H1 | 0.364 | 0.383 | 0.170 | 0.208 | 0.147 | 0.246 | 0.135 | 0.243 | 0.125 | 0.229 | 0.521 | 0.362 | 1.750 | 0.791 | 0.034 | 0.091 | 0.119 | **0.214** |
| | H2 | 0.415 | 0.414 | 0.214 | 0.246 | 0.203 | 0.239 | 0.256 | 0.326 | 0.237 | 0.335 | 1.077 | 0.551 | 1.807 | 0.810 | 0.048 | 0.116 | 0.163 | 0.262 |
| | H3 | 0.457 | 0.438 | 0.265 | 0.283 | 0.196 | 0.286 | 0.603 | 0.540 | 0.248 | 0.328 | 1.636 | 0.717 | 1.840 | 0.815 | **0.060** | **0.136** | 0.207 | 0.300 |
| | H4 | 0.475 | 0.471 | 0.345 | 0.338 | 0.214 | 0.305 | 1.190 | 0.823 | 1.178 | 0.778 | 2.148 | 0.852 | 1.779 | 0.816 | 0.072 | 0.153 | **0.250** | 0.334 |
| | Avg | 0.428 | 0.427 | 0.249 | 0.269 | 0.190 | 0.269 | 0.546 | 0.483 | 0.447 | 0.418 | 1.346 | 0.621 | 1.794 | 0.808 | 0.054 | **0.124** | **0.185** | 0.278 |
| NoFusion | H1 | 0.363 | 0.384 | 0.153 | 0.191 | 0.133 | 0.229 | 0.132 | 0.248 | 0.121 | 0.236 | 0.482 | 0.350 | 2.114 | 0.855 | 0.036 | 0.094 | 0.123 | 0.218 |
| | H2 | 0.422 | 0.418 | 0.207 | 0.240 | 0.148 | 0.236 | 0.240 | 0.326 | 0.246 | 0.324 | **0.854** | **0.503** | **1.707** | 0.801 | 0.049 | 0.120 | 0.172 | 0.272 |
| | H3 | 0.459 | 0.438 | 0.256 | 0.277 | 0.166 | 0.254 | 0.579 | 0.546 | 0.257 | 0.339 | **1.276** | 0.661 | 1.904 | 0.824 | 0.063 | 0.138 | 0.217 | 0.316 |
| | H4 | 0.478 | 0.471 | 0.341 | 0.332 | 0.189 | 0.277 | 1.187 | 0.786 | 1.163 | 0.810 | 1.771 | 0.804 | 1.807 | 0.822 | 0.075 | 0.157 | 0.267 | 0.350 |
| | Avg | 0.431 | 0.428 | 0.239 | 0.260 | 0.159 | 0.249 | 0.535 | 0.477 | 0.447 | 0.427 | 1.096 | 0.580 | 1.883 | 0.826 | 0.056 | 0.127 | 0.195 | 0.289 |

planning, traffic management, weather-related monitoring, financial time-series analysis, and public-health surveillance. By imposing statistical calibration constraints only on the entrance cross-series interaction while keeping deeper representations data-adaptive, FISC provides a lightweight forecasting mechanism with moderate computational cost. The row-stochastic coupling constraint also makes cross-series aggregation more inspectable than unconstrained mixing, which may help practitioners diagnose stable dependencies and data-adaptive deviations.

Forecasting models can also have negative consequences when their outputs are used without validation or domain oversight. In high-stakes settings such as public-health response, financial decision-making, power-grid operation, or transportation control, inaccurate forecasts may lead to inappropriate resource allocation, delayed intervention, or amplified operational risk. FISC should therefore not be used as the sole basis for critical decisions. Its predictions should be combined with domain knowledge, uncertainty estimation, stress testing, and monitoring of distribution shift. This is especially important because FISC estimates calibration matrices from the training split and keeps them fixed during deployment; under severe or persistent regime changes, these matrices may become mismatched and require reassessment or recalibration.

The method is evaluated on public benchmark datasets and does not require collecting private user data or conducting human-subject studies. Nevertheless, users should respect the licenses and terms of use of all datasets and baseline implementations. When applying FISC to real-world data, practitioners should check whether the data contain sensitive attributes, whether forecasting errors may disproportionately affect specific groups or regions, and whether downstream decisions include adequate human review. We intend FISC to be used as a research tool and forecasting component, not as an autonomous decision-making system.

Table 20: Documentation provided for existing assets and supplementary code. Detailed experimental settings are described in the appendix, and implementation files and configurations are provided in the supplementary material.

| Asset type | Source and license documentation | Documentation location |
|---|---|---|
| Forecasting datasets | Source/citation and license or terms are documented when available | Paper appendix and supplementary material |
| Baseline implementations | Source/citation, version or access information, and license are documented when available | Paper appendix and supplementary material |
| Preprocessing scripts | Provided for datasets requiring additional preprocessing | Supplementary material |
| FISC model code | Included with the submission for research inspection and reproducibility | Supplementary material |
| Configuration files | Included with the submission for the reported experimental settings | Supplementary material |

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the main scope of FISC as a multivariate forecasting framework based on first-layer statistical calibration constraints, simplex-constrained cross-series coupling, and dual-domain temporal modeling. The claims are supported by the method description, ablations, non-stationarity analysis, and benchmark results, and the paper does not claim global optimality or universal robustness.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses limitations in the conclusion and in the appendix on failure modes and boundary analysis. In particular, it notes that calibration statistics are estimated from the training split and kept fixed, which may become unreliable under severe or persistent distribution shifts.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The theoretical result is presented as a simplified local optimization interpretation rather than a full proof of the nonlinear model. The appendix states the surrogate assumptions, defines the constrained objective and projected-gradient update, and provides a proof of the convergence-prefactor bound.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper describes the architecture, training-split statistics, normalization, CSCO layers, calibration hyperparameters, datasets, horizons, splits, evaluation metrics, and baseline protocols. Full per-horizon forecasting and ablation results are provided in the appendix, and supplementary code/configuration files are included for reproducibility.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The datasets are public forecasting benchmarks or publicly accessible datasets, and the supplementary material includes the FISC model code, preprocessing/configuration information, and implementation files for the reported settings. Dataset sources and baseline implementations are cited or documented when available.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies train/validation/test splits, look-back lengths, forecasting horizons, metrics, optimizer, learning-rate search, hidden dimensions, CSCO depth, calibration strength, temperature, early stopping, batch-size choices, and baseline evaluation protocol. Additional dataset and implementation details are provided in the appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: The appendix reports multi-seed stability results over matched random seeds, including mean and standard deviation, paired significance tests, and optimization-stability diagnostics. These analyses are used to support the main first-layer calibration claim beyond single-run accuracy tables.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The implementation details state that experiments are run on a single NVIDIA RTX 4060 GPU with 16GB memory. The paper also reports parameters, FLOPs, training/inference time, and training/inference memory under a unified ETT setting to characterize computational cost.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The work uses public forecasting benchmarks or publicly accessible datasets, does not collect private user data, does not conduct human-subject studies, and does not release high-risk generative or dual-use models. The paper discusses responsible use, dataset licensing, and deployment cautions.

   Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The appendix discusses potential positive impacts in energy-load planning, traffic management, weather-related monitoring, financial time-series analysis, and public-health surveillance. It also discusses risks from inaccurate forecasts in high-stakes settings and recommends domain oversight, uncertainty estimation, stress testing, and distribution-shift monitoring.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release high-risk generative models, pretrained foundation models, scraped sensitive datasets, or user data. The released supplementary code is intended for research use on public multivariate forecasting benchmarks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Existing datasets and baseline implementations are cited in the main text and appendix. The appendix states that the terms of use and licenses specified by original providers are followed when available, and the supplementary material documents source, license, or access information when available.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The paper introduces FISC model code and configuration files as supplementary material. The appendix documents the included components, including first-layer statistical calibration constraints, CSCO, the dual-domain temporal representation, the prediction head, preprocessing/configuration information, and reported experimental settings.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing, user studies, or research with human subjects. All experiments are conducted on public or publicly accessible time-series forecasting datasets.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The work does not involve human-subject experiments, crowdsourcing, private user data collection, or interventions involving participants. Therefore, IRB approval or an equivalent human-subject review is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method, experiments, statistical calibration constraints, CSCO operator, temporal representation, and empirical evaluation do not use LLMs as an important, original, or non-standard component. Any ordinary writing or formatting assistance, if used, does not affect the scientific method or results.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.