COLLABORATIVE-ENHANCED MULTI-VIEW LEARNING FOR NOISY MULTIVARIATE TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time series forecasting fundamentally relies on cross-channel dependencies. However, the presence of noise may distort such dependencies, posing significant challenges to reliable relation modeling. Existing methods either model temporal patterns independently for each channel or capture cross-channel dependencies within a single shared space. The former overlooks informative channel interactions, while the latter is susceptible to noise-distorted dependencies. As a result, neither is able to fully exploit reliable and complementary relational information, which constrains their forecasting robustness under noisy conditions. To address this challenge, we propose a collaborative-enhanced multi-view framework that constructs multiple overlapping latent relation views through adaptive filtering and dynamic soft routing. Bidirectional view interaction enables mutual guidance and collaborative enhancement by leveraging complementary relational information, reducing reliance on potentially distorted dependencies. Extensive experiments demonstrate improved forecasting accuracy and stable performance under Constant, Missing, and Gaussian noise.

1 INTRODUCTION

Multivariate time series forecasting has a wide range of real-world applications, including energy systems (Zhou et al., 2022; Boussif et al., 2023), weather forecasting (Bi et al., 2023; Lam et al., 2023), and traffic management (Wang et al., 2023; Jiang et al., 2023). Given historical observations from multiple variables, the objective is to predict their future values over a specified horizon. Since these variables are often interdependent, accurate forecasting requires not only capturing the temporal dynamics (Wu et al., 2023; Wang et al., 2024a) of individual channels but also exploiting informative cross-channel dependencies. Accordingly, recent channel-dependent models have demonstrated the benefits of explicitly modeling interactions among variables through attention-based or graph-based mechanisms. However, the effectiveness of such interactions depends on the reliability of the relational information extracted from historical observations, which is frequently degraded by noise in many real-world scenarios.

Real-world multivariate time series frequently contain outliers, missing observations, and random perturbations. Beyond disrupting the temporal patterns of individual channels, these corruptions can substantially distort the estimation of cross-channel dependencies. As illustrated in Figure 1(a), we examine the Pearson correlation matrices of seven channels in ETTh1 under clean and Gaussian-corrupted observations. The results reveal that informative correlations can be considerably weakened by noise, leading to significant differences between the clean and noisy dependency estimates. For instance, a strongly correlated channel pair exhibits a substantial reduction in its observed correlation after corruption. Noise-induced weakening of cross-channel relationships can lead to unreliable dependency estimates, making it difficult for forecasting models to capture informative channel interactions.

Existing forecasting methods address cross-channel modeling in different ways, but many rely on a limited relational perspective. Channel-independent methods model each variable separately, avoiding direct interference from other channels while potentially overlooking useful cross-channel information. In contrast, many channel-dependent methods explicitly capture variable interactions

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

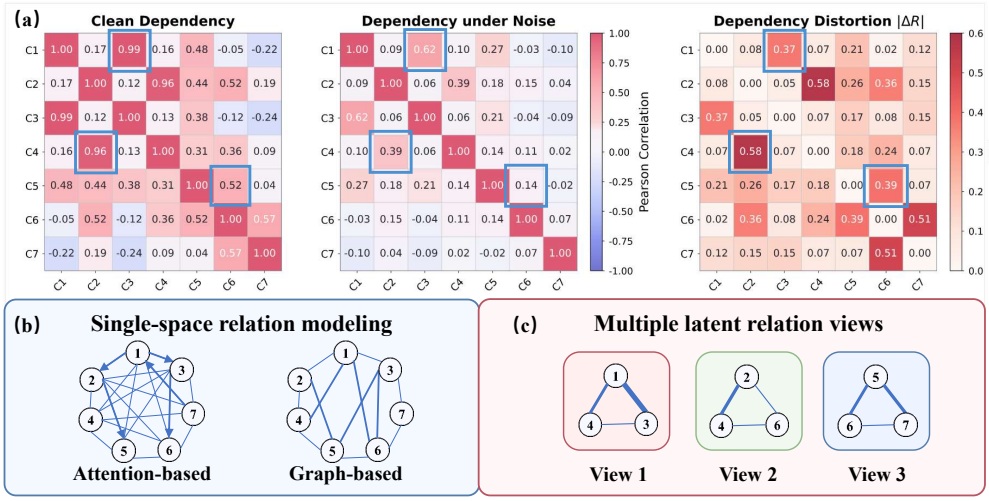

Figure 1: (a) Cross-channel dependency distortion under Gaussian noise, illustrated by clean and noisy Pearson correlation matrices and their absolute differences. (b) Single-space relation modeling using attention-based and graph-based methods, which may be vulnerable to corrupted dependencies. (c) Multiple overlapping latent relation views for capturing complementary channel relationships.

through attention, graph propagation, or shared representations. In representative single-space formulations, diverse channel interactions are organized within a shared dependency space, as illustrated in Figure 1(b). If dependency estimates are distorted by noise, single-space modeling tends to rely excessively on corrupted relational information, thereby limiting its capacity to capture reliable cross-channel interactions. The key challenge is thus to learn reliable cross-channel relationships despite the distortions introduced by noise.

To address this challenge, we explore multiple overlapping latent relation views, as illustrated in Figure 1(c). Rather than requiring all channel relationships to be accurately estimated within a single global interaction space, we construct multiple overlapping latent views that capture distinct relational contexts. This allows channels to participate with varying degrees of membership. Although noise may severely weaken a particular relationship, other views can still provide complementary relational evidence that is less affected. These views collaboratively enhance their representations by leveraging such complementary evidence, thereby reducing reliance on any single potentially distorted dependency representation. Building on this, we propose a collaborative-enhanced multi-view framework for robust multivariate time series forecasting. The main contributions of this study are summarized as follows:

- We propose a noise-aware multi-view forecasting framework that combines adaptive filtering with dynamic soft routing to construct multiple overlapping latent relation views, enabling diverse cross-channel dependencies to be modeled beyond a single shared relation space.
- We introduce a collaborative enhancement mechanism that enables bidirectional information exchange across latent relation views, allowing complementary relational information to enhance the representations learned in each view under noisy observations.
- We extend the RobustTSF evaluation setting to multivariate long-term forecasting with constant, missing, and Gaussian corruption at different severity levels, enabling systematic evaluation of forecasting robustness.

2 RELATED WORK

Multivariate time series forecasting: MTS forecasting methods commonly adopt channel-independent (CI) or channel-dependent (CD) strategies (Han et al., 2024b). CI methods model

channels separately (Zeng et al., 2023; Nie et al., 2023; Zhou et al., 2023; Jin et al., 2024; Lin et al., 2024; Dai et al., 2024), whereas CD methods capture variable interactions through attention (Zhang & Yan, 2023; Liu et al., 2024; Ilbert et al., 2024; Wang et al., 2024b;c; Yang et al., 2024; Yue et al., 2025), graphs (Wu et al., 2020; Huang et al., 2023; Yi et al., 2023; Cai et al., 2024), or feature mixing and global representations (Han et al., 2024a; Chen et al., 2023). However, CI methods overlook channel interactions, while many CD methods capture dependencies within a shared space vulnerable to noise, limiting their use of complementary relational information. Our framework enables mutual guidance across overlapping latent relation views, exploiting complementary information.

Robustness Evaluation: Existing robustness evaluations differ in data sources, corruption types, severity levels, and corruption stages. Several studies inject Gaussian noise of varying intensities into real-world datasets (Huang et al., 2023; Wu et al., 2025; Han et al., 2024a), but a single noise type provides limited evidence of robustness to diverse corruptions. Other evaluations rely on synthetic datasets (Yi et al., 2024; Tan et al., 2026); despite diverse corruption settings, such data may not fully reflect real-world cross-channel dependencies or their interactions with noise. LGA (Kim et al., 2026) introduces TSRBench to evaluate models trained on clean data under multiple test-time corruptions and severity levels, leaving training-data contamination unaddressed. RobustTSF (Cheng et al., 2024) considers contaminated training data under several anomaly types, primarily for individual time series. Building on these settings, we evaluate multivariate long-term forecasting on real-world datasets under a unified protocol covering multiple corruption types and ratios. We corrupt training inputs and targets, as well as test inputs, while retaining clean test targets to assess robustness during both training and inference.

Multi-view Learning: Multi-view learning aims to capture complementary information from different representations. In time series modeling, existing methods often construct predefined views with explicit semantic meanings, such as temporal-frequency, spatial-temporal, contextual, and global-local views, to exploit complementary information (Wu et al., 2024b; Xie et al., 2026; Gao et al., 2025; Yu et al., 2025; Guo et al., 2024; Wu et al., 2024a; Zhang et al., 2024; Shen et al., 2025; Zhang et al., 2026; Deng et al., 2023; Huang et al., 2026). However, such predefined views may limit the discovery of diverse and overlapping cross-channel relationships. Co-training, originally developed for semi-supervised learning, enables two complementary views to mutually guide each other (Blum & Mitchell, 1998), and has been extended to regression and deep neural networks (Zhou & Li, 2005; Qiao et al., 2018). Inspired by this mutual guidance principle, we propose a collaborative-enhanced framework that learns multiple overlapping latent relation views without predefined semantics. Through mutual guidance among these views, our framework collaboratively enhances their representations by exploiting complementary relational evidence, reducing reliance on a single potentially distorted dependency representation under noisy observations.

3 PRELIMINARIES

3.1 Multivariate Long-term Time Series Forecasting

A multivariate time series consists of multiple variables synchronously observed over time, where each variable is also referred to as a channel. Given a historical sequence of length T with C channels, we represent it as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T] \in \mathbb{R}^{C \times T}$, where $\mathbf{x}_t = [x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(C)}]^\top \in \mathbb{R}^C$ denotes the observations of all C channels at time step t, and $x_t^{(c)}$ denotes the observation of the c-th channel at time step t. Given the historical observations $\mathbf{X}$, the objective of multivariate time series forecasting is to predict future values over a horizon of $\hat{T}$ time steps through a learned mapping function $f(\cdot)$. Specifically, the prediction is represented as $\widehat{\mathbf{Y}} = f(\mathbf{X}) = [\widehat{\mathbf{y}}_{T+1}, \widehat{\mathbf{y}}_{T+2}, \ldots, \widehat{\mathbf{y}}_{T+\hat{T}}] \in \mathbb{R}^{C \times \hat{T}}$, where each $\widehat{\mathbf{y}}_{T+\hat{t}} \in \mathbb{R}^C$ denotes the predicted values at future time step $T + \hat{t}$.

3.2 Robust Forecasting under Corrupted Data

Real-world time series observations may be corrupted by sensor failures, missing data, or random perturbations. To describe such scenarios, we use $\mathbf{X}$ to denote the clean historical sequence and $\widetilde{\mathbf{X}}$ to denote its corrupted counterpart. Following the corruption setting of RobustTSF, we formulate

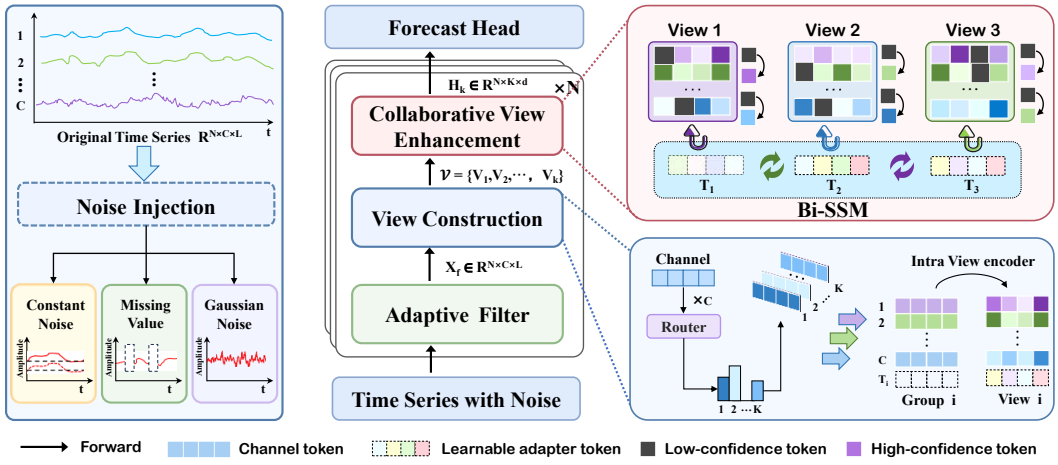

Figure 2: Overview framework. The left part illustrates the noise injection process, while the right part shows the forecasting pipeline consisting of adaptive filtering, view construction, and collaborative view enhancement.

the corruption process as $\widetilde{\mathbf{X}} = \mathcal{C}_{\tau,\eta}(\mathbf{X})$, where $\tau \in \{Constant, Missing, Gaussian\}$ denotes the corruption type and $\eta \in [0, 0.5]$ denotes the corruption ratio.

Under the corrupted-data setting considered in this work, the training data itself contains corrupted observations. Specifically, for an original training sample $(\mathbf{X}, \mathbf{Y})$, the model is trained on the corrupted pair $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$, where $\widetilde{\mathbf{X}} = \mathcal{C}_{\tau,\eta_{\text{train}}}(\mathbf{X})$ and $\widetilde{\mathbf{Y}} = \mathcal{C}_{\tau,\eta_{\text{train}}}(\mathbf{Y})$. Thus, both the historical inputs and their corresponding future targets are drawn from corrupted sequences during training. During testing, we focus on the model's robustness to corrupted historical observations. Given a test sample $(\mathbf{X}, \mathbf{Y})$, only the historical observations are corrupted, $\widetilde{\mathbf{X}} = \mathcal{C}_{\tau,\eta_{\text{test}}}(\mathbf{X})$, and the model produces the prediction $\widehat{\mathbf{Y}} = f(\widetilde{\mathbf{X}})$. The prediction $\widehat{\mathbf{Y}}$ is evaluated against the corresponding clean future sequence $\mathbf{Y}$, measuring the model's ability to recover the underlying future dynamics from corrupted inputs. Therefore, our goal is to learn a forecasting model $f(\cdot)$ from corrupted training data that can accurately predict the corresponding clean future sequence under different corruption types and ratios at test time.

4 The Proposed Method

4.1 Overall Framework

Given a corrupted historical sequence $\widetilde{\mathbf{X}} \in \mathbb{R}^{C \times T}$, our objective is to learn reliable forecasting representations while exploiting complementary relationships among channels. As illustrated in Figure 2, the proposed framework consists of three main stages: adaptive filtering, latent relation view construction, and collaborative view enhancement. The noise injection process on the left illustrates the corruption settings used in our robustness evaluation, while the right part presents the forecasting framework.

We first apply adaptive filtering to reduce noise interference before modeling cross-channel relationships. The filtered sequence is then encoded into channel tokens, which are dynamically assigned to multiple overlapping latent relation views. Within each view, a relation encoder learns cross-channel interactions and produces an adapter token that summarizes the learned relational information. However, independently learned views may still contain incomplete information under noisy observations. We therefore introduce collaborative view enhancement, which enables mutual guidance among views and uses complementary relational evidence to enhance their channel representations. Finally, the enhanced representations are integrated to predict.

4.2 ADAPTIVE FILTERING

Noise in historical observations can distort temporal patterns and interfere with subsequent cross-channel relation modeling. We therefore introduce an adaptive filter to reduce noise interference before constructing latent relation views. Rather than applying a fixed frequency threshold (Xu et al., 2024), the filter learns a smooth frequency mask that can be optimized together with the forecasting model.

We first normalize the corrupted input along the temporal dimension and transform it into the frequency domain using the real Fourier transform. Let $\mathbf{X}^n = \mathrm{Norm}(\widetilde{\mathbf{X}})$ denote the normalized input and$\mathbf{Z} = \mathcal{F}_r(\mathbf{X}^n)$ its frequency representation. A learnable cutoff f_c determines the soft low-pass mask, where the mask value at frequency index f is defined as

$$m_f = \sigma\left(\frac{f_c - f}{\tau_f}\right), \qquad \mathbf{X}^f = \mathcal{F}_r^{-1}\left(\mathbf{Z} \odot \mathbf{m}\right). \tag{1}$$

Here, $\sigma(\cdot)$ denotes the sigmoid function, τ_f controls the smoothness of the transition, and $\odot$ denotes element-wise multiplication. The cutoff is learnable and shared across channels, while the zero-frequency component is preserved. The resulting sequence $\mathbf{X}^f \in \mathbb{R}^{C\times T}$ retains the lower-frequency temporal information used by the subsequent modules.

4.3 VIEW CONSTRUCTION

A single global relation space may be overly dependent on distorted cross-channel relationships when the input is corrupted. To capture diverse relational evidence, we construct multiple overlapping latent relation views without assigning predefined semantics. As shown in the view construction module of Figure 2, this process involves dynamic soft routing followed by relation modeling within each view.

Dynamic Soft Routing. We first embed the filtered sequence into channel tokens, $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_C]^\top \in \mathbb{R}^{C\times D}$, where each token represents the historical information of one channel. Instead of assigning channels to fixed groups, we use a dynamic router to determine their participation in different latent views. For each channel c, the router combines its statistical characteristics, a learnable channel embedding, and its channel token. The resulting channel representation is

$$\mathbf{r}_c = \mathrm{LN}\left(g_s(\mathbf{s}_c) + \mathbf{u}_c + g_e(\mathbf{e}_c)\right), \tag{2}$$

where $\mathbf{s}_c$ denotes the channel statistics, $\mathbf{u}_c$ is the learnable channel embedding, and g_s and g_e are projection functions. To construct K latent views, we introduce K learnable view prototypes $\{\mathbf{p}_k\}_{k=1}^K$. The assignment of channel c to view k is determined by its similarity to the corresponding prototype:

$$a_{c,k} = \frac{\exp\left(\mathrm{sim}(\mathbf{r}_c, \mathbf{p}_k)/\gamma\right)}{\sum_{j=1}^K \exp\left(\mathrm{sim}(\mathbf{r}_c, \mathbf{p}_j)/\gamma\right)}, \tag{3}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes normalized similarity and γ is the routing temperature.

The resulting assignment matrix $\mathbf{A} \in \mathbb{R}^{C\times K}$ specifies the contribution of each channel to different views. Since the assignments are soft and depend on the current input, a channel can participate in several views with different weights. This allows the model to learn different channel combinations and overlapping relational contexts without imposing a fixed partition.

Intra-view Relation Modeling. After obtaining the routing assignments, we construct each latent view by weighting the channel tokens according to their corresponding assignment scores:

$$\mathbf{V}_k = \mathrm{Diag}(\mathbf{a}_{:,k})\mathbf{E}, \tag{4}$$

where $\mathbf{a}_{:,k}$ contains the assignment weights for view k. Each view retains all channel tokens, but their contributions differ according to the routing weights. To learn cross-channel relationships within these different contexts, we employ an independent relation encoder for each view. We further incorporate routing information into the attention mechanism by adding a routing-derived bias to the attention scores. This encourages the encoder to place greater emphasis on channels that are more relevant to the current view. As illustrated in the lower-right part of Figure 2, a learnable

adapter token is appended to the channel tokens of each view. The adapter token participates in the same attention process and aggregates relational information from the corresponding view. The encoder therefore produces two outputs:

$$[\mathbf{H}_k, \mathbf{t}_k] = \mathrm{ViewEnc}_k\left(\mathbf{V}_k, \mathbf{t}_k^{(0)}, \mathbf{a}_{:,k}\right), \tag{5}$$

where $\mathbf{H}_k \in \mathbb{R}^{C\times D}$ denotes the learned representation of the k-th view, and $\mathbf{t}_k \in \mathbb{R}^D$ is the corresponding adapter token. This design enables mutual guidance among views.

4.4 Collaborative View Enhancement

The constructed views capture cross-channel relationships from different relational contexts. Nevertheless, each view is learned from the same noisy observations, and its representation may remain incomplete when some relationships are severely distorted. Importantly, the corruption of a particular relationship does not necessarily imply that all complementary relational views are equally affected. We therefore introduce collaborative view enhancement to enable mutual guidance among latent views and strengthen their representations using complementary information.

Bidirectional View Guidance. We first collect the adapter tokens from all views into $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_K] \in \mathbb{R}^{K\times D}$. Since each adapter token summarizes a different relational context, interactions among these tokens allow the views to draw on complementary information learned by one another.

To establish such interactions, we employ a bidirectional state-space module along the view dimension. The forward and backward state-space models process the adapter tokens in opposite orders, and their outputs are combined through a residual projection:

$$\begin{aligned}
\mathbf{T}^{\rightarrow} &= \mathrm{SSM}_f(\mathbf{T}),\\
\mathbf{T}^{\leftarrow} &= \mathrm{Rev}\left(\mathrm{SSM}_b(\mathrm{Rev}(\mathbf{T}))\right),\\
\mathbf{T}^{+} &= \mathrm{LN}\left(\mathbf{T} + \mathrm{Linear}([\mathbf{T}^{\rightarrow}; \mathbf{T}^{\leftarrow}])\right).
\end{aligned} \tag{6}$$

Here, $\mathrm{Rev}(\cdot)$ reverses the order of the view tokens, and $\mathbf{T}^{+}$ denotes the representations after bidirectional interaction. Through bidirectional interaction, each adapter token incorporates complementary information from other views while retaining its original information through the residual connection. The updated adapter tokens then provide guidance for enhancing the corresponding view representations.

Collaborative Representation Enhancement. The enhanced adapter tokens incorporate complementary relational information from other views. To leverage this information, we use the difference between the updated and original adapter tokens to adaptively enhance each view:

$$\overline{\mathbf{H}}_k = \mathbf{H}_k \odot \mathcal{G}\left(\mathbf{t}_k^{+} - \mathbf{t}_k\right), \tag{7}$$

where $\mathcal{G}(\cdot)$ denotes a learnable gating function and $\overline{\mathbf{H}}_k$ is the enhanced representation of the k-th view. By using information exchanged across views, this mechanism adaptively refines each view with complementary relational information.

Representation Fusion and Forecasting. Finally, the enhanced view representations are concatenated and fused through layer normalization and linear projection. The fused representation is then passed to a forecasting head to generate future predictions.

5 Experiments

5.1 Experimental Setup

Datasets and Data Splitting. We conduct experiments on multiple public time-series forecasting datasets (Zhou et al., 2021) to evaluate the robustness of our method under noisy conditions. All datasets are split into training and test sets with a ratio of $7:3$, and different levels of noise are injected into both the training and test sets. It is worth noting that we do not create an additional

validation set for two reasons: (1) under noisy environments, the training data contain noise perturbations, making model selection based on a noisy validation set unreliable; (2) using a portion of the test set for model selection would introduce test information into the training procedure, resulting in information leakage.

Noise Settings and Training Setup. Following the experimental protocol of RobustTSF, we consider three representative types of time-series corruption: Constant Noise, Missing Noise, and Gaussian Noise. For each noise type, the noise ratios of the training and test sets are configured as (0,0), (0.5,0.01), (0.5,0.05), (0.5,0.10), (0.5,0.20), (0.5,0.30), where each pair denotes the noise ratio applied to the training set and test set, respectively. The setting $(0, 0)$ represents the clean scenario without noise perturbations. For the remaining settings, the training noise ratio is fixed at 0.5, while the test noise ratio is gradually increased to evaluate model robustness under different levels of test-time corruption.

All methods are trained using the Adam optimizer for 30 epochs. We adopt a two-stage learning-rate strategy, where the learning rate remains fixed during the first 10 epochs and is adjusted to another fixed value for the subsequent 20 epochs. For forecasting tasks, the input window length is fixed to 96, the prediction horizons are set to $\{96,192,336,720\}$ for most long-term forecasting datasets, while the prediction lengths for the Illness dataset are $\{24,36,48,60\}$. All experiments are conducted with a fixed random seed of 2021 to ensure reproducibility.All experiments are conducted on a Linux server equipped with seven NVIDIA Tesla V100-SXM2 GPUs (16GB memory per GPU), an Intel Xeon Gold 6240 CPU @ 2.60GHz, and 251GB RAM. The implementation is based on PyTorch 2.0.1 with Python 3.9.25 and CUDA 11.8. All experiments share the same hardware and software environment for fair comparison.

Comparing Methods and Evaluation Protocol. We compare our method with PatchTST(Nie et al., 2023) , TimeXer(Wang et al., 2024c) and Dynamic-TMoE(Zhu et al., 2026). We use Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the evaluation metrics. Following the evaluation protocol adopted by RobustTSF (Cheng et al., 2024),we further extend it to noisy test scenarios. Specifically, the model performance is evaluated on the corresponding noisy test set at the end of each training epoch. After the training process is completed, we report both the best epoch and last epoch test performances for each method. This allows us to evaluate both the best performance achieved during training and the final performance after the complete training process. The best result in each comparison is highlighted in **bold**.

5.2 Main Results

As shown in Table 1, our method achieves robust forecasting performance across four benchmark datasets under three different types of noise. The complete results are provided in the Appendix. The reported results are averaged over six noise settings and four forecasting horizons. Overall, our method consistently achieves competitive performance. Taking the ILI dataset as an example, our method improves the MSE by 35.3%, 12.9%, and 13.6% over PatchTST, TimeXer, and Dynamic-TMoE under Missing noise, respectively. Moreover, our method maintains strong performance in both the best and final checkpoints, indicating stable optimization and reliable generalization during training. These results demonstrate the effectiveness of the proposed method for robust time series forecasting in noisy environments.

5.3 Ablation Study

We conduct ablation studies to investigate the contribution of each key component in our framework. Specifically, we construct three variants: (1) w/o AF removes the adaptive filtering module; (2) w/o VC replaces the Dynamic Soft Routing strategy with random channel grouping; (3) w/o Bi-SSM removes the bidirectional state-space view interaction module. The results are summarized in Table 2. As can be seen, all model components are necessary. Specifically, w/o AF indicates that adaptive filtering effectively mitigates noise interference and preserves reliable temporal patterns. The performance degradation of w/o VC is more significant, especially on the ETTh1 and ILI datasets, suggesting that replacing Dynamic Soft Routing with random channel grouping limits the ability to capture noise-aware channel dependencies and construct meaningful views. Furthermore, removing Bi-SSM degrades performance, indicating that view-guided collaboration effectively exploits complementary information to improve forecasting accuracy.

Table 1: Main forecasting results under different noise types. The reported results are averaged over six noise configurations and four forecasting horizons. For each setting, we report the MSE/MAE of both the best and final checkpoints. The best results are highlighted in **bold**.

Noise Type	Dataset	ETTh1				ETTh2			
	Model	Ours	PatchTST	TimeXer	Dynamic-TMoE	Ours	PatchTST	TimeXer	Dynamic-TMoE
Gaussian	Best	**0.614/0.567**	0.764/0.631	0.664/0.591	0.626/0.576	**0.248/0.348**	0.256/0.355	0.249/0.350	0.266/0.367
	Last	**0.630/0.574**	0.775/0.635	0.669/0.593	0.655/0.593	**0.250/0.351**	0.259/0.357	0.252/0.352	0.290/0.385
Constant	Best	**0.661/0.599**	0.683/0.606	0.708/0.616	0.700/0.622	0.307/0.405	0.315/0.409	**0.296/0.396**	0.365/0.444
	Last	**0.674/0.606**	0.730/0.630	0.720/0.621	0.744/0.646	**0.315/0.412**	0.326/0.420	0.318/0.414	0.528/0.502
Missing	Best	0.615/0.563	0.756/0.623	0.641/0.575	**0.604/0.559**	**0.241/0.346**	0.248/0.351	0.242/0.347	0.259/0.359
	Last	**0.625/0.568**	0.760/0.625	0.644/0.576	0.646/0.572	**0.244/0.348**	0.250/0.353	**0.244/0.348**	0.519/0.401
Noise Type	Dataset	Exchange				ILI			
	Model	Ours	PatchTST	TimeXer	Dynamic-TMoE	Ours	PatchTST	TimeXer	Dynamic-TMoE
Gaussian	Best	**0.262/0.357**	0.269/0.367	**0.262**/0.361	0.285/0.380	**2.519/1.103**	3.784/1.427	2.679/1.151	2.785/1.153
	Last	**0.273/0.364**	0.293/0.381	0.285/0.374	0.321/0.397	**2.608/1.122**	3.807/1.423	2.700/1.154	2.830/1.163
Constant	Best	0.450/0.494	0.490/0.526	**0.436/0.485**	0.528/0.541	**2.268/1.055**	3.742/1.441	2.411/1.099	2.565/1.093
	Last	**0.476/0.512**	0.588/0.593	0.507/0.529	0.834/0.649	**2.361/1.077**	3.757/1.448	2.446/1.106	2.607/1.104
Missing	Best	**0.234/0.351**	**0.234**/0.356	**0.234**/0.355	0.269/0.376	**2.301/1.049**	3.555/1.349	2.643/1.128	2.664/1.135
	Last	**0.242/0.357**	0.254/0.368	0.260/0.371	0.346/0.406	**2.345/1.060**	3.599/1.349	2.690/1.132	2.711/1.146

Table 2: Ablation results of different components of the proposed framework. The reported results are averaged over six noise configurations and four forecasting horizons. All results are reported as MSE/MAE, and the best performance is highlighted in **bold**.

Model	Dataset	ETTh1			Exchange			ILI		
	Noise type	Gaussian	Constant	Missing	Gaussian	Constant	Missing	Gaussian	Constant	Missing
Full	Best	**0.614/0.567**	**0.661/0.599**	**0.615/0.563**	**0.262/0.357**	0.450/0.494	**0.234/0.351**	**2.519/1.103**	**2.268/1.055**	**2.301/1.049**
	Last	0.630/0.574	**0.674/0.606**	**0.625/0.568**	**0.273/0.364**	**0.476/0.512**	**0.242/0.357**	**2.608/1.122**	**2.361/1.077**	**2.345/1.060**
w/o AF	Best	0.619/0.570	0.663/0.602	0.621/0.567	**0.262/0.357**	**0.448**/0.496	0.238/0.353	2.549/1.115	2.340/1.069	2.349/1.058
	Last	0.635/0.577	0.680/0.610	0.630/0.571	0.295/0.378	0.507/0.534	0.296/0.389	2.659/1.134	2.448/1.097	2.428/1.072
w/o VC	Best	0.732/0.630	0.769/0.649	0.699/0.612	0.265/0.358	**0.448/0.491**	0.240/0.355	2.624/1.112	2.334/1.062	2.523/1.088
	Last	0.734/0.630	0.779/0.653	0.706/0.615	0.298/0.377	0.491/0.515	0.275/0.374	2.707/1.141	2.413/1.082	2.627/1.111
w/o Bi-SSM	Best	**0.614/0.567**	0.666/0.601	0.619/0.565	0.266/0.359	0.457/0.499	0.234/0.352	2.587/1.119	2.338/1.068	2.432/1.077
	Last	**0.621/0.570**	0.677/0.608	0.626/0.567	0.294/0.379	0.504/0.527	0.279/0.375	2.644/1.129	2.406/1.088	2.479/1.085

5.4 Hyperparameter Analysis and Visualization

Varying Look-back Window. We investigate the impact of different look-back windows on forecasting performance. In principle, a longer look-back window provides more historical information and a larger temporal receptive field. As observed in Fig. 3, our method consistently achieves lower MSE across different look-back windows and forecasting horizons. In particular, as the look-back window increases, our model maintains stable performance improvements, demonstrating its capability to effectively exploit long-range temporal dependencies. In contrast, existing baselines exhibit larger fluctuations when increasing the historical window.

Visualization under Different Noisy Conditions. To qualitatively analyze the forecasting behavior of different models under noisy conditions, we visualize prediction results on the Illness dataset. Specifically, we set the prediction length to 60 and adopt the most severe noise configuration. Fig. 4 presents representative forecasting results of our method, PatchTST, TimeXer, and Dynamic-TMOE under three different types of noise. different models exhibit varying degrees of performance degradation. PatchTST tends to produce overly smooth or even nearly constant predictions in some cases, making it difficult to adequately capture the dynamic variations in the ground-truth sequence. Although TimeXer can capture the overall trend to some extent, noticeable discrepancies remain in modeling local peaks and variation amplitudes. Dynamic-TMoE, in contrast, exhibits pronounced prediction fluctuations, with evident high-frequency oscillations in its forecasting trajectories. Compared with these baselines, our method consistently produces more stable predictions across differ-

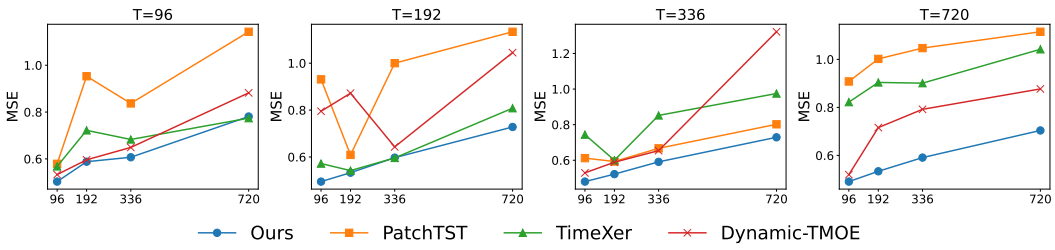

Figure 3: Forecasting performance (MSE at the best checkpoint) with varying look-back windows on ETTh1. Each subplot corresponds to a look-back window $T \in \{96, 192, 336, 720\}$, as indicated by its title, while the horizontal axis represents the prediction horizons $\hat{T} \in \{96, 192, 336, 720\}$.

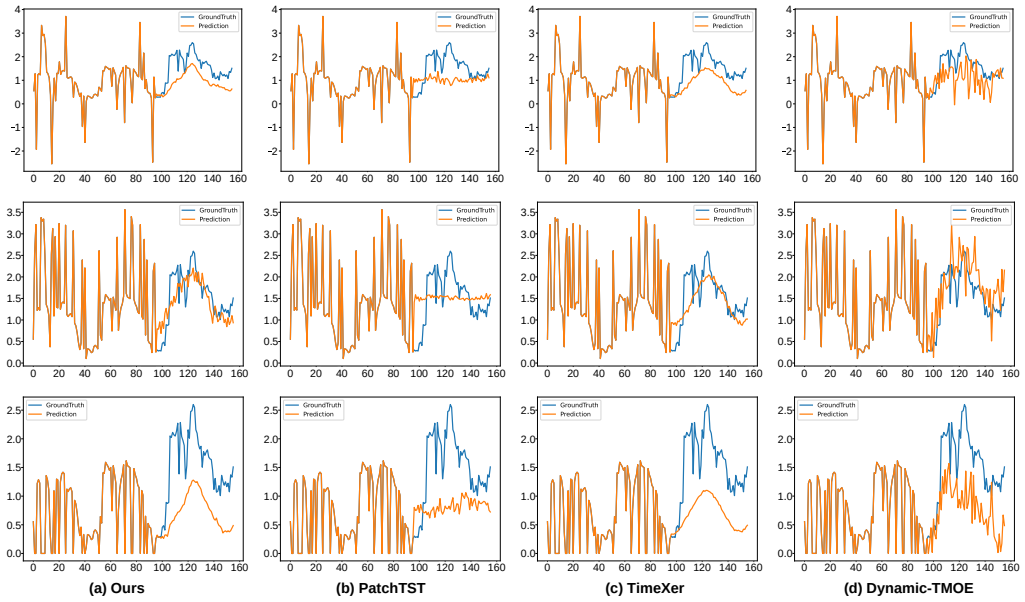

Figure 4: Qualitative forecasting results on the Illness dataset with a prediction length of 60 ($\eta_{\text{train}} = 0.5$, $\eta_{\text{test}} = 0.3$). The rows from top to bottom correspond to Gaussian, Constant, and Missing noise.

ent noise types and better follows the overall trends and major dynamic patterns of the ground-truth sequence.

6 Conclusions

To improve forecasting robustness under noisy conditions by fully exploiting reliable and complementary relational information, we propose a collaborative-enhanced multi-view framework that constructs multiple overlapping latent relation views through adaptive filtering and dynamic soft routing. This enables diverse cross-channel dependencies to be modeled beyond a single shared relation space. We further introduce a collaborative enhancement mechanism that enables bidirectional information exchange across latent relation views, allowing complementary relational information to enhance the representations learned in each view under noisy observations. Extensive experiments demonstrated improved forecasting accuracy and stable performance under constant, missing, and Gaussian noise. However, the current method is limited to a single modality, and future work will consider multimodal information such as text to further improve robustness.

AI USE STATEMENT

In this work, we have not used generative AI tools for research ideation or execution, drafting sections of the paper, generating synthetic datasets, proving mathematical claims, or any other purposes beyond those described below. In particular, AI tools were not used to develop theoretical or conceptual frameworks, formulate hypotheses or mathematical claims, design the methodology or experiments, implement methods, translate text, process datasets, conduct qualitative analysis, or interpret results. Any remaining tasks requiring disclosure were not applicable to this work. Additionally, we used generative AI tools to polish author-written text for grammar, clarity, and fluency, and to search for potentially relevant related work. We reviewed all AI-assisted work. The authors read the original publications, verified the relevance and details of the cited references, and reviewed all suggested edits to ensure that the text accurately reflects our work. The research ideas, scientific conclusions, and interpretation of results were developed and verified by the authors. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work follows the ICLR Code of Ethics. It develops and evaluates a time-series forecasting method using publicly available benchmark datasets and does not involve human participants or animal experiments. The datasets were used in accordance with their respective usage terms. We made every effort to avoid bias or discriminatory outcomes throughout the research process. No personally identifiable information was used, and no experiments that could raise privacy or security concerns were conducted. We evaluated forecasting performance under different noise conditions on benchmark datasets. Before applying the method in a real-world setting, its stability and the potential consequences of prediction errors should be assessed using data from that setting. We are committed to transparency and research integrity: we report our experimental procedures and results and provide the code and configurations in the anonymous supplementary materials to support scrutiny and reproducibility.

REPRODUCIBILITY STATEMENT

We have taken several steps to facilitate the reproduction of our results. Section 5.1 describes the datasets, evaluation protocol, and experimental environment, while the appendix reports the complete experimental results. The anonymous supplementary materials contain the source code, configuration files, and instructions needed to reproduce the main results. Together, these materials specify the data preparation, model training, and evaluation procedures used in this study. The experiments use publicly available datasets, which can be obtained from their original sources. We believe that the details provided in the paper, appendix, and supplementary materials enable independent verification and reproduction of our findings.

REFERENCES

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, July 2023. doi: 10.1038/s41586-023-06185-3. URL https://ideas.repec.org/a/nat/nature/v619y2023i7970d10.1038_s41586-023-06185-3.html.

Avrim Blum and Tom Mitchell. Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98, pp. 92–100, New York, NY, USA, 1998. Association for Computing Machinery. ISBN 1581130570. doi: 10.1145/279943.279962. URL https://doi.org/10.1145/279943.279962.

Oussama Boussif, Ghait Boukachab, Dan Assouline, Stefano Massaroli, Tianle Yuan, Loubna Benabbou, and Yoshua Bengio. Improving day-ahead solar irradiance time series forecasting by leveraging spatio-temporal context. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 2342–2367. Curran Associates, Inc., 2023. doi: 10.52202/075280-0109. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/070a57c5ef1e58cc90201b11d369b3c2-Paper-Conference.pdf.

Wanlin Cai, Yuxuan Liang, Xianggen Liu, Jianshuai Feng, and Yuankai Wu. MSGNet: Learning Multi-Scale Inter-Series Correlations for Multivariate Time Series Forecasting. In *AAAI Conference on Artificial Intelligence*, pp. 11141–11149, 2024. doi: 10.1609/AAAI.V38I10.28991. URL https://mlanthology.org/aaai/2024/cai2024aaai-msgnet/.

Si-An Chen, Chun-Liang Li, Sercan O Arik, Nathanael Christian Yoder, and Tomas Pfister. TSMixer: An All-MLP Architecture for Time Series Forecast-Ing. *Transactions on Machine Learning Research*, 2023. URL https://mlanthology.org/tmlr/2023/chen2023tmlr-tsmixer/.

Hao Cheng, Qingsong Wen, Yang Liu, and Liang Sun. Robusttsf: Towards theory and design of robust time series forecasting with anomalies. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 5787–5813, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/179f5dcdeedc149443ebd3ba70811dbd-Paper-Conference.pdf.

Tao Dai, Beiliang Wu, Peiyuan Liu, Naiqi Li, Jigang Bao, Yong Jiang, and Shu-Tao Xia. Periodicity decoupling framework for long-term series forecasting. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 48067–48078, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/d298cf34e4539f9134db7f38b42f69fe-Paper-Conference.pdf.

Jinliang Deng, Xiusi Chen, Renhe Jiang, Xuan Song, and Ivor W. Tsang. A multi-view multi-task learning framework for multi-variate time series forecasting. *IEEE Trans. on Knowl. and Data Eng.*, 35(8):7665–7680, August 2023. ISSN 1041-4347. doi: 10.1109/TKDE.2022.3218803. URL https://doi.org/10.1109/TKDE.2022.3218803.

Wentao Gao, Ziqi Xu, Jiuyong Li, Lin Liu, Jixue Liu, Thuc Duy Le, Debo Cheng, Yanchang Zhao, and Yun Chen. Tsi: A multi-view representation learning approach for time series forecasting. In Mingming Gong, Yiliao Song, Yun Sing Koh, Wei Xiang, and Derui Wang (eds.), *AI 2024: Advances in Artificial Intelligence*, pp. 291–302, Singapore, 2025. Springer Nature Singapore. ISBN 978-981-96-0348-0.

Canyang Guo, Chi-Hua Chen, Feng-Jang Hwang, Ching-Chun Chang, and Chin-Chen Chang. Multi-view spatiotemporal learning for traffic forecasting. *Information Sciences*, 657:119868, 2024. ISSN 0020-0255. doi: https://doi.org/10.1016/j.ins.2023.119868. URL https://www.sciencedirect.com/science/article/pii/S0020025523014536.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. Softs: Efficient multivariate time series forecasting with series-core fusion. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 64145–64175. Curran Associates, Inc., 2024a. doi: 10.52202/079017-2046. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/754612bde73a8b65ad8743f1f6d8ddf6-Paper-Conference.pdf.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. volume 36, pp. 7129–7142, 2024b. doi: 10.1109/TKDE.2024.3400008.

Qihe Huang, Lei Shen, Ruixin Zhang, Shouhong Ding, Binwu Wang, Zhengyang Zhou, and Yang Wang. CrossGNN: Confronting noisy multivariate time series via cross interaction refinement. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=xOzlW2vUYc.

Yaohui Huang, Runmin Zou, Yun Wang, Laeeq Aslam, and Ruipeng Dong. M2fmoe: multi-resolution multi-view frequency mixture-of-experts for extreme-adaptive time series forecasting. In *Proceedings of the Fortieth AAAI Conference on Artificial Intelligence and Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence and Sixteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'26/IAAI'26/EAAI'26. AAAI Press, 2026. ISBN 978-1-57735-906-7. doi: 10.1609/aaai.v40i26.39362. URL https://doi.org/10.1609/aaai.v40i26.39362.

Romain Ilbert, Ambroise Odonnat, Vasilii Feofanov, Aladin Virmaux, Giuseppe Paolo, Themis Palpanas, and Ievgen Redko. Samformer: unlocking the potential of transformers in time series forecasting with sharpness-aware minimization and channel-wise attention. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Jiawei Jiang, Chengkai Han, Wayne Zhao, and Jingyuan Wang. Pdformer: Propagation delay-aware dynamic long-range transformer for traffic flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37:4365–4373, 06 2023. doi: 10.1609/aaai.v37i4.25556.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y. Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Unb5CVPtae.

Dongbin Kim, Youngjoo Park, Woojin Jeong, and Jaewook Lee. Local geometry attention for time series forecasting under realistic corruptions. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 43205–43244, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/476ab8f369e489c04187ba84f68cfa68-Paper-Conference.pdf.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336. URL https://www.science.org/doi/abs/10.1126/science.adi2336.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 106315–106345. Curran Associates, Inc., 2024. doi: 10.52202/079017-3373. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/bfe7998398779dde03cad7a73b1f81b6-Paper-Conference.pdf.

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 B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 11116–11140, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/2ea18fdc667e0ef2ad82b2b4d65147ad-Paper-Conference.pdf.

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. URL https://openreview.net/forum?id=Jbdc0vTOcol.

Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. Deep co-training for semi-supervised image recognition. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision – ECCV 2018*, pp. 142–159, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01267-0.

ChengAo Shen, Wenchao Yu, Ziming Zhao, Dongjin Song, Wei Cheng, Haifeng Chen, and Jingchao Ni. Multi-modal view enhanced large vision models for long-term time series forecasting. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 93269–93296. Curran Associates, Inc., 2025. doi: 10.52202/085713-3119. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/860c1c657deafe09f64c013c2888bd7b-Paper-Conference.pdf.

Qitai Tan, Yiyun Chen, Mo Li, Ruiwen Gu, Yilin Su, and Xiao-Ping Zhang. SynTSBench: Rethinking temporal pattern learning in deep learning models for time series. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2026. URL https://openreview.net/forum?id=rNw128KrYU.

Binwu Wang, Yudong Zhang, Xu Wang, Pengkun Wang, Zhengyang Zhou, Lei Bai, and Yang Wang. Pattern expansion and consolidation on evolving graphs for continual traffic prediction. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023. URL https://api.semanticscholar.org/CorpusID:260500002.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024a.

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. URL https://openreview.net/forum?id=MJksrOhurE.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. In *Advances in Neural Information Processing Systems*, 2024c.

Chenrui Wu, Haishuai Wang, Xiang Zhang, Zhen Fang, and Jiajun Bu. Spatio-temporal heterogeneous federated learning for time series classification with multi-view orthogonal training. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, pp. 2613–2622, New York, NY, USA, 2024a. Association for Computing Machinery. ISBN 9798400706868. doi: 10.1145/3664647.3680733. URL https://doi.org/10.1145/3664647.3680733.

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. URL https://openreview.net/forum?id=ju_Uqw384Oq.

Yuhan Wu, Xiyu Meng, Yang He, Junru Zhang, Haowen Zhang, Yabo Dong, and Dongming Lu. Multi-view self-supervised contrastive learning for multivariate time series. In *ACM Multimedia 2024*, 2024b. URL https://openreview.net/forum?id=BgjLy3chju.

Yuhan Wu, Xiyu Meng, Huajin Hu, Junru Zhang, Yabo Dong, and Dongming Lu. Affirm: interactive mamba with adaptive fourier filters for long-term time series forecasting. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025. ISBN 978-1-57735-897-8. doi: 10.1609/aaai.v39i20.35463. URL https://doi.org/10.1609/aaai.v39i20.35463.

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*, KDD '20, pp. 753–763, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3403118. URL https://doi.org/10.1145/3394486.3403118.

Tao Xie, Zexi Tan, Haoyi Xiao, Mengke Li, Yiqun Zhang, Yang Lu, Cuie Yang, and Yiu-ming Cheung. Anchormoe: Interpretable time series classification via anchor-routed moe. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '26, pp. 5720–5731, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400722592. doi: 10.1145/3770855.3817786. URL https://doi.org/10.1145/3770855.3817786.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with 10k parameters. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.),

International Conference on Learning Representations, volume 2024, pp. 26295–26318, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/701251e1db4a2e4dd2ef23f5265d5936-Paper-Conference.pdf.

Yingnan Yang, Qingling Zhu, and Jianyong Chen. Vcformer: Variable correlation transformer with inherent lagged correlation for multivariate time series forecasting. In Kate Larson (ed.), *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 5335–5343. International Joint Conferences on Artificial Intelligence Organization, 8 2024. doi: 10.24963/ijcai.2024/590. URL https://doi.org/10.24963/ijcai.2024/590. Main Track.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. Fouriergnn: Rethinking multivariate time series forecasting from a pure graph perspective. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 69638–69660. Curran Associates, Inc., 2023. doi: 10.52202/075280-3050. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/dc1e32dd3eb381dbc71482f6a96cbf86-Paper-Conference.pdf.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. 11 2024. doi: 10.48550/arXiv.2411.01623.

Chengqing Yu, Fei Wang, Chuanguang Yang, Zezhi Shao, Tao Sun, Tangwen Qian, Wei Wei, Zhulin An, and Yongjun Xu. Merlin: Multi-view representation learning for robust multivariate time series forecasting with unfixed missing rates. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, pp. 3633–3644, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714542. doi: 10.1145/3711896.3737046. URL https://doi.org/10.1145/3711896.3737046.

Wenzhen Yue, Yong Liu, Xianghua Ying, Bowei Xing, Ruohao Guo, and Ji Shi. Freeformer: Frequency enhanced transformer for multivariate time series forecasting. In James Kwok (ed.), *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pp. 3606–3614. International Joint Conferences on Artificial Intelligence Organization, 8 2025. doi: 10.24963/ijcai.2025/401. URL https://doi.org/10.24963/ijcai.2025/401. Main Track.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i9.26317. URL https://doi.org/10.1609/aaai.v37i9.26317.

Liang Zhang, Jianping Zhu, Bo Jin, and Xiaopeng Wei. Multiview spatial-temporal meta-learning for multivariate time series forecasting. volume 24, 2024. doi: 10.3390/s24144473. URL https://www.mdpi.com/1424-8220/24/14/4473.

Wenhao Zhang, Chun Zhang, Wei Bai, Ning Zhang, Changxia Gao, Yuxin Jia, Chenhao Shi, and Shaoxiong Pang. Unifying channel independence and mixing: Multi-scale patch recursion for global–local representation synergy in multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28427–28436, 2026.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *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 *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*, volume 35, pp. 11106–11115. AAAI Press, 2021.

Jingbo Zhou, Xinjiang Lu, Yixiong Xiao, Jian Tang, Jian Su, Yu Li, Ji Liu, Junfu Lyu, Yanjun Ma, and De-Jing Dou. Sdwpf: A dataset for spatial dynamic wind power forecasting over a large turbine array. *Scientific Data*, 11, 2022. URL https://api.semanticscholar.org/CorpusID:251442543.

Tian Zhou, Peisong Niu, Xue Wang, Liang Sun, and Rong Jin. One fits all: Power general time series analysis by pretrained lm. *Advances in Neural Information Processing Systems 36*, 2023. URL https://api.semanticscholar.org/CorpusID:258741419.

Zhi-Hua Zhou and Ming Li. Semi-supervised regression with co-training. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, IJCAI'05, pp. 908–913, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc.

Jiawen Zhu, Shuhan Liu, Di Weng, and Yingcai Wu. Dynamic TMoe: A drift-aware dynamic mixture of experts framework for non-stationary time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=JabkBcaoa9.

A APPENDIX

The complete forecasting results are presented in Tables 3 and 4. We report the performance of our method and three baselines on four benchmark datasets under Gaussian, Constant, and Missing noise. For each noise type, six noise configurations (S0–S5) are evaluated, with results averaged over four prediction horizons. Both the best-checkpoint (Best) and final-checkpoint (Last) results are provided in terms of MSE and MAE. These detailed results complement the aggregated results in the main text and enable a more comprehensive examination of forecasting performance and training stability under different noise conditions.

Table 3: Forecasting results under different noise conditions. S_0–S_5 correspond to $(0, 0)$, $(0.5, 0.01)$, $(0.5, 0.05)$, $(0.5, 0.10)$, $(0.5, 0.20)$, and $(0.5, 0.30)$, respectively. All values are reported as MSE/MAE.

Noise Type	Severity	Dataset	ETTh1				ETTh2			
		Model	Ours	PatchTST	TimeXer	Dynamic-TMoE	Ours	PatchTST	TimeXer	Dynamic-TMoE
Gaussian	S_0	Best	**0.557/0.530**	0.560/0.534	0.571/0.537	**0.557/0.531**	0.242/0.341	**0.230/0.329**	0.232/0.334	0.246/0.345
		Last	**0.562/0.531**	0.564/0.537	0.572/0.538	0.585/0.548	0.244/0.343	**0.232/0.332**	0.237/0.337	0.256/0.353
	S_1	Best	**0.636/0.575**	0.872/0.672	0.717/0.614	**0.629/0.575**	**0.246/0.344**	0.256/0.352	0.248/0.346	**0.243/0.344**
		Last	**0.657/0.587**	0.882/0.675	0.731/0.620	**0.657/0.589**	**0.248/0.345**	0.256/0.352	0.249/0.347	0.268/0.362
	S_2	Best	**0.630/0.575**	0.898/0.688	0.678/0.598	0.632/0.579	**0.247/0.346**	0.257/0.354	0.250/0.348	0.264/0.365
		Last	**0.646/0.583**	0.904/0.689	0.681/0.600	0.657/0.593	**0.248/0.347**	0.259/0.355	0.251/0.350	0.283/0.378
	S_3	Best	**0.624/0.574**	0.726/0.618	0.656/0.589	0.637/0.583	**0.246/0.348**	0.258/0.357	0.251/0.353	0.266/0.369
		Last	**0.642/0.581**	0.747/0.626	0.657/0.591	0.661/0.598	**0.251/0.352**	0.261/0.360	0.254/0.355	0.297/0.392
	S_4	Best	**0.619/0.572**	0.652/0.592	0.675/0.600	0.638/0.585	**0.249/0.352**	0.264/0.364	0.253/0.356	0.284/0.384
		Last	**0.630/0.577**	0.668/0.599	0.682/0.603	0.668/0.604	**0.252/0.356**	0.266/0.366	0.255/0.357	0.310/0.406
	S_5	Best	**0.620/0.576**	0.878/0.682	0.686/0.607	0.665/0.602	**0.256/0.359**	0.270/0.371	0.259/0.361	0.294/0.395
		Last	**0.645/0.585**	0.886/0.686	0.688/0.608	0.702/0.623	**0.258/0.362**	0.279/0.379	0.263/0.364	0.327/0.421
Constant	S_0	Best	0.564/0.534	0.561/0.534	0.569/0.537	**0.552/0.529**	0.243/0.342	0.230/0.329	0.233/0.333	**0.227/0.328**
		Last	**0.568/0.536**	**0.566/0.539**	0.576/0.541	0.579/0.545	0.245/0.344	**0.230/0.330**	0.237/0.336	0.238/0.337
	S_1	Best	**0.592/0.554**	0.600/0.559	0.621/0.569	0.605/0.560	**0.242/0.342**	0.246/0.344	**0.242/0.342**	0.274/0.359
		Last	**0.596/0.557**	0.615/0.568	0.629/0.571	0.642/0.579	**0.245/0.344**	0.252/0.348	0.248/0.345	0.321/0.385
	S_2	Best	**0.595/0.558**	0.669/0.590	0.628/0.574	0.626/0.574	**0.246/0.346**	0.254/0.352	**0.246/0.348**	0.309/0.398
		Last	**0.601/0.560**	0.696/0.602	0.633/0.576	0.662/0.596	**0.247/0.348**	0.255/0.354	**0.247/0.348**	0.381/0.432
	S_3	Best	**0.616/0.570**	0.669/0.598	0.674/0.594	0.660/0.600	**0.259/0.364**	0.278/0.381	**0.259/0.364**	0.349/0.435
		Last	**0.629/0.577**	0.676/0.601	0.686/0.598	0.705/0.628	**0.268/0.373**	0.284/0.388	0.272/0.378	0.418/0.478
	S_4	Best	**0.722/0.639**	0.745/0.645	0.809/0.666	0.775/0.675	0.349/0.457	0.356/0.458	**0.333/0.439**	0.445/0.522
		Last	**0.743/0.651**	0.772/0.662	0.834/0.676	0.831/0.708	**0.358/0.463**	0.396/0.495	0.373/0.477	1.139/0.698
	S_5	Best	0.875/0.736	**0.851/0.710**	0.947/0.753	0.982/0.791	0.504/0.581	0.527/0.592	**0.461/0.547**	0.585/0.624
		Last	**0.909/0.754**	1.057/0.810	0.964/0.763	1.042/0.817	**0.529/0.598**	0.541/0.607	0.530/0.601	0.669/0.681
Missing	S_0	Best	**0.559/0.530**	0.670/0.585	0.566/0.535	**0.559/0.533**	0.241/0.341	0.228/0.329	0.232/0.333	**0.226/0.327**
		Last	**0.565/0.532**	0.687/0.595	0.572/0.539	0.581/0.546	0.241/0.341	**0.233/0.333**	0.234/0.335	0.238/0.337
	S_1	Best	**0.610/0.562**	0.713/0.603	0.656/0.584	**0.608/0.562**	**0.232/0.334**	0.250/0.347	0.241/0.340	0.245/0.344
		Last	**0.618/0.567**	0.717/0.607	0.659/0.586	0.680/0.582	**0.241/0.339**	0.250/0.347	0.242/0.341	0.248/0.347
	S_2	Best	**0.602/0.557**	0.801/0.639	0.663/0.586	0.605/0.559	**0.234/0.337**	0.245/0.346	0.239/0.341	0.278/0.369
		Last	**0.609/0.561**	0.801/0.639	0.667/0.587	0.666/0.577	**0.236/0.339**	0.246/0.347	0.241/0.342	0.306/0.379
	S_3	Best	0.642/0.578	0.798/0.641	0.630/0.568	**0.609/0.561**	**0.232/0.338**	0.242/0.347	0.233/0.339	0.243/0.348
		Last	0.657/0.584	0.799/0.642	**0.631/0.568**	0.651/0.572	0.236/0.341	0.243/0.347	**0.234/0.339**	0.249/0.352
	S_4	Best	0.649/0.581	0.747/0.621	0.665/0.586	**0.611/0.563**	**0.241/0.351**	0.250/0.359	0.244/0.353	0.283/0.383
		Last	0.667/0.589	0.748/0.621	0.665/0.586	**0.646/0.573**	**0.243/0.353**	0.250/0.359	0.246/0.354	1.785/0.597
	S_5	Best	**0.627/0.571**	0.809/0.646	0.666/0.588	0.630/0.573	**0.265/0.374**	0.273/0.380	**0.265/0.374**	0.278/0.385
		Last	**0.633/0.574**	0.809/0.646	0.671/0.590	0.651/0.580	**0.267/0.376**	0.276/0.383	**0.267/0.376**	0.288/0.391

Table 4: Forecasting results under different noise conditions. S_0–S_5 correspond to $(0, 0)$, $(0.5, 0.01)$, $(0.5, 0.05)$, $(0.5, 0.10)$, $(0.5, 0.20)$, and $(0.5, 0.30)$, respectively. All values are reported as MSE/MAE.

Noise Type	Severity	Dataset	Exchange				ILI			
		Model	Ours	PatchTST	TimeXer	Dynamic-TMoE	Ours	PatchTST	TimeXer	Dynamic-TMoE
Gaussian	S_0	Best	0.241/0.337	**0.222/0.322**	0.235/0.332	0.264/0.345	**1.892**/0.951	2.216/1.003	2.186/0.967	1.971/**0.911**
		Last	0.254/0.347	0.261/0.344	**0.249/0.343**	0.365/0.392	2.010/0.978	2.276/1.002	2.219/0.971	**1.980/0.913**
	S_1	Best	**0.260/0.352**	0.265/0.362	0.267/0.363	0.265/0.358	2.651/**1.137**	4.086/1.514	**2.529**/1.210	3.068/1.215
		Last	**0.262/0.354**	0.276/0.366	0.279/0.369	0.276/0.363	2.709/**1.149**	4.104/1.511	**2.555**/1.211	3.121/1.219
	S_2	Best	0.264/**0.355**	0.275/0.369	**0.263**/0.361	0.280/0.374	**2.673/1.146**	4.110/1.517	2.882/1.201	2.938/1.198
		Last	**0.273/0.361**	0.295/0.380	0.284/0.373	0.298/0.383	**2.702/1.155**	4.131/1.511	2.894/1.205	2.979/1.207
	S_3	Best	**0.267/0.359**	0.283/0.383	0.271/0.369	0.288/0.385	**2.708/1.159**	4.129/1.524	2.893/1.202	2.861/1.187
		Last	**0.278/0.366**	0.301/0.391	0.293/0.380	0.310/0.397	**2.758/1.164**	4.143/1.517	2.901/1.203	2.894/1.195
	S_4	Best	0.269/**0.364**	0.289/0.385	**0.268**/0.368	0.299/0.400	**2.628/1.124**	4.071/1.499	2.815/1.171	2.955/1.204
		Last	**0.287/0.377**	0.304/0.394	0.299/0.385	0.327/0.415	**2.734/1.151**	4.085/1.495	2.844/1.174	2.982/1.213
	S_5	Best	0.270/**0.373**	0.278/0.383	**0.268**/0.374	0.316/0.416	**2.563/1.100**	4.094/1.502	2.767/1.152	2.916/1.205
		Last	**0.284/0.380**	0.323/0.411	0.308/0.394	0.348/0.434	**2.736/1.134**	4.101/1.500	2.786/1.157	3.022/1.229
Constant	S_0	Best	0.246/0.337	**0.225/0.322**	0.233/0.332	0.308/0.381	**1.859**/0.943	2.766/1.150	2.131/0.960	1.992/**0.920**
		Last	0.265/0.354	0.261/0.344	**0.252/0.342**	0.464/0.419	**1.972**/0.970	2.804/1.160	2.168/0.958	2.051/**0.939**
	S_1	Best	**0.247/0.343**	0.258/0.361	0.258/0.349	0.287/0.372	2.444/**1.071**	4.064/1.506	**2.382**/1.148	2.841/1.170
		Last	**0.264/0.353**	0.298/0.386	0.316/0.380	0.592/0.419	2.573/**1.112**	4.067/1.504	**2.422**/1.150	2.860/1.177
	S_2	Best	0.311/**0.387**	0.339/0.428	**0.309**/0.389	0.363/0.441	**2.412/1.086**	4.023/1.507	2.594/1.130	2.586/1.100
		Last	**0.317/0.392**	0.410/0.487	0.345/0.414	0.827/0.554	**2.468/1.098**	4.032/1.507	2.620/1.134	2.632/1.113
	S_3	Best	**0.391**/0.461	0.445/0.505	0.395/**0.460**	0.476/0.520	**2.348/1.074**	3.978/1.511	2.543/1.127	2.520/1.087
		Last	**0.404/0.472**	0.586/0.630	0.445/0.497	0.815/0.658	**2.398/1.088**	3.983/1.513	2.577/1.136	2.564/1.095
	S_4	Best	0.615/0.629	0.692/0.681	**0.589/0.610**	0.736/0.689	**2.275/1.070**	3.749/1.459	2.437/1.114	2.796/1.147
		Last	**0.650/0.657**	0.839/0.765	0.703/0.685	1.018/0.845	**2.332/1.081**	3.755/1.479	2.469/1.128	2.870/1.162
	S_5	Best	0.892/0.806	0.979/0.860	**0.834/0.769**	0.996/0.845	**2.272/1.084**	3.872/1.512	2.378/1.113	2.657/1.134
		Last	**0.957/0.845**	1.134/0.946	0.982/0.854	1.288/0.997	2.425/**1.112**	3.901/1.524	**2.419**/1.132	2.666/1.138
Missing	S_0	Best	0.247/0.341	0.228/**0.326**	**0.226/0.326**	0.309/0.379	**1.861**/0.950	2.312/1.033	2.166/0.955	2.034/**0.929**
		Last	0.256/0.346	**0.255/0.341**	0.263/0.347	0.609/0.473	**1.985**/0.975	2.355/1.041	2.198/0.962	2.055/**0.938**
	S_1	Best	**0.242/0.340**	0.258/0.360	0.246/0.348	0.251/0.346	**2.143/1.003**	3.806/1.421	2.325/1.117	2.567/1.118
		Last	**0.245/0.342**	0.276/0.370	0.261/0.355	0.290/0.363	**2.153/1.010**	3.820/1.414	2.359/1.121	2.653/1.137
	S_2	Best	0.229/0.338	**0.217/0.337**	0.226/0.339	0.259/0.357	**2.291/1.043**	3.739/1.403	2.665/1.133	2.724/1.157
		Last	**0.231/0.338**	0.256/0.360	0.249/0.350	0.295/0.376	**2.323/1.052**	3.776/1.403	2.728/1.136	2.771/1.163
	S_3	Best	**0.215/0.335**	0.216/0.345	0.220/0.344	0.241/0.356	**2.324/1.050**	3.447/1.313	2.705/1.135	2.850/1.194
		Last	**0.229/0.344**	0.236/0.350	0.243/0.356	0.273/0.373	**2.366/1.064**	3.511/1.321	2.751/1.142	2.895/1.207
	S_4	Best	**0.221/0.358**	0.230/0.366	0.226/0.365	0.255/0.384	**2.492/1.095**	3.847/1.428	2.876/1.189	2.745/1.170
		Last	**0.231/0.366**	0.239/0.375	0.252/0.385	0.284/0.402	**2.519/1.098**	3.894/1.421	2.928/1.189	2.787/1.178
	S_5	Best	**0.248/0.394**	0.257/0.404	0.262/0.406	0.299/0.431	**2.697/1.150**	4.176/1.498	3.118/1.237	3.062/1.244
		Last	**0.259/0.405**	0.264/0.411	0.291/0.432	0.326/0.446	**2.723/1.159**	4.237/1.496	3.177/1.241	3.105/1.252