COMPASS: CHANNEL-WISE CALIBRATION OF PERIODIC PRIORS FOR MULTIVARIATE TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multivariate time series forecasting requires effectively leveraging both historical periodic patterns and current dynamics. Existing periodic forecasting methods typically employ a uniform strategy to incorporate historical periodic priors across channels, overlooking the channel-wise differences between historical priors and current dynamics. Such uniform fusion can introduce mismatched periodic information into channels where historical priors differ more from current dynamics. To address this issue, we propose **COMPASS**, a channel-wise periodic prior adaptation framework for multivariate time series forecasting. COMPASS focuses on the channel-wise inconsistency between historical periodic priors and current dynamics and first performs channel-wise relationship calibration to quantify this inconsistency and adaptively adjust their contributions across channels. To further exploit useful periodic information, COMPASS employs dynamic-guided prior enhancement, which aggregates periodic information across channels according to the current dynamics. The enhanced periodic prior is then adaptively fused with current dynamics for forecasting. Extensive experiments on 12 public forecasting benchmarks demonstrate the effectiveness of COMPASS, reducing MSE by an average of 2.23% over the strongest competing methods. COMPASS achieves the best performance on 6 of 8 long-term forecasting datasets and all 4 short-term forecasting datasets.

1 INTRODUCTION

Multivariate time series forecasting (MTSF) plays a critical role in a wide range of applications, including energy demand forecasting , traffic management , financial analysis, and healthcare monitoring (Wu et al., 2021; Qiu et al., 2025a; Zhou et al., 2022a; Khan et al., 2023; Zhou et al., 2022b; Chengqing et al., 2023; Chen et al., 2024; 2026; Huang et al., 2025). Accurate forecasting requires not only capturing recent variations but also identifying temporal patterns from historical observations that can persist into the future (Wu et al., 2022; Zeng et al., 2023; Wang et al., 2024b). In particular, long-term trends and periodic patterns provide structural cues beyond local observations by repeatedly occurring across time. Among them, periodicity is a prevalent form of long-term temporal structure that connects historical observations across different time scales and provides persistent signals for future prediction. Therefore, effectively extracting and leveraging historical periodic patterns is essential for accurate long-horizon forecasting.

To exploit historical periodicity, existing MTSF methods have mainly followed several representative paradigms. One line improves neural architectures, such as Transformers for modeling long-range dependencies (Liu et al., 2023), or employs seasonal-trend decomposition to separate temporal components at different scales and extract their periodic and trend information (Zhou et al., 2022b; Wang et al., 2024b; 2025b). Another line performs frequency-domain modeling, leveraging classical signal-processing tools such as the Fast Fourier Transform (FFT) (Brigham & Morrow, 1967) to represent global temporal structures and periodic patterns in the frequency domain (Xu et al., 2024). A third line explicitly retrieves or constructs historical periodic patterns and uses them as long-term priors to assist future prediction (Lin et al., 2024; 2025a; Zhang et al., 2026; Cao et al., 2026). Despite their differences, these approaches share a common principle: extracting stable temporal structures from historical observations and using them as priors for forecasting. However, they typically rely on an implicit assumption that historical periodic priors can be utilized similarly across different channels, overlooking the possibility that individual channels may exhibit substantially different compatibility with the historical patterns.

This assumption often breaks down in real-world multivariate time series, where channels exhibit heterogeneous dynamics. As illustrated in Figure 1(a), the historical periodic prior π and the current observation x exhibit substantially different relationships across channels: some channels remain consistent with the historical prior, while others show weak or even opposite correlations. Thus, the usefulness of historical periodic information is inherently channel-dependent. For channels whose current dynamics remain consistent with historical patterns, the periodic prior can provide valuable long-term structural information. In contrast, when a channel's dynamics have evolved, blindly relying on a mismatched historical prior can introduce conflicting information and interfere with modeling the current dynamics. This channel-level inconsistency between historical periodic priors and current

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.

dynamics is largely overlooked by existing periodic forecasting methods, motivating adaptive prior utilization across channels.

Motivated by this observation, we revisit periodic modeling from the perspective of channel-wise consistency and propose **COMPASS**, as illustrated in Figure 1(b). Rather than blindly trusting the historical periodic prior and applying a uniform strategy across channels, COMPASS explicitly evaluates its consistency with current dynamics at the channel level and adaptively determines how much each channel should rely on the prior. Specifically, COMPASS first performs *channel-wise relationship calibration*, which adjusts the contributions of the periodic prior and current dynamics according to their relationship differences across channels. It then introduces *dynamic-guided prior enhancement*, where the current dynamics guide cross-channel aggregation to leverage relevant periodic information from other channels, thereby enhancing the calibrated prior. Finally, an *adaptive fusion mechanism* balances the enhanced periodic prior and current dynamics to preserve stable historical patterns while reducing interference from mismatched prior information. Overall, COMPASS enables channel-wise adaptive utilization of historical periodic information, allowing periodic priors to better accommodate evolving dynamics in multivariate time series.

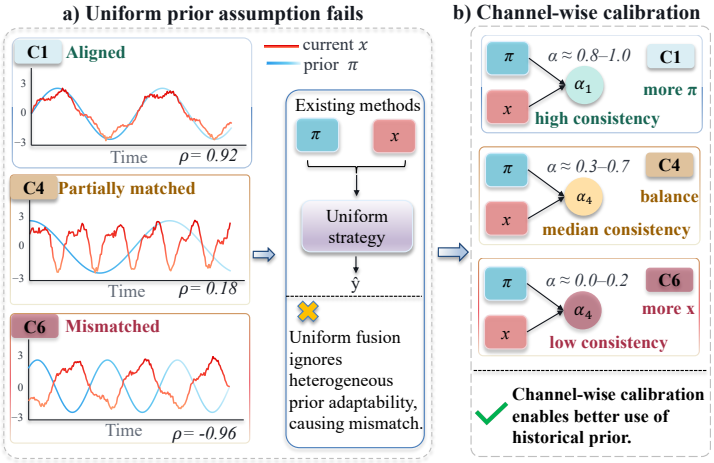

Figure 1: Motivation of COMPASS. (a) The relationship between historical periodic priors and current dynamics varies across channels, while existing methods apply the same strategy to all channels. (b) COMPASS performs channel-wise calibration to adaptively utilize the periodic prior according to its consistency with current dynamics.

Our main contributions are summarized as follows:

- We pinpoint a key limitation in existing periodic forecasting methods: uniform utilization of historical periodic priors overlooks the varying degree of mismatch between the prior and current dynamics across channels.
- We propose **COMPASS**, a channel-aware framework that calibrates historical periodic priors according to their channel-wise mismatch with current dynamics and further enhances the calibrated priors using dynamic-guided cross-channel information.
- Experiments on 12 public benchmarks show that COMPASS reduces MSE by 2.23% on average over the strongest competing methods, with the best results on 6 of 8 long-term and all 4 short-term datasets.

2 Related Work

Recent advances in deep learning have led to substantial progress in multivariate time series forecasting (MTSF), with existing studies exploring both diverse architectural backbones and modeling strategies tailored to the characteristics of multivariate temporal data. Representative architectures include: recurrent neural networks (RNNs) (Guo et al., 2023; Wang et al., 2024a; Lin et al., 2025b), convolutional neural networks (CNNs) (Liu et al., 2022; Luo & Wang, 2024), multilayer perceptrons (MLPs) (Das et al., 2023; Lin et al., 2024), Transformers (Liu et al., 2023; Wang et al., 2024c), and more recent Mamba-based architectures (Wang et al., 2025d; Ma et al., 2025). Beyond architectural design, increasing attention has been devoted to explicitly modeling the intrinsic structures of multivariate time series, particularly cross-variable dependencies and recurring temporal regularities. Specifically, channel modeling aims to capture inter-variable relationships and exploit cross-channel information to enhance forecasting representations, whereas periodicity modeling leverages recurring historical patterns to provide informative temporal priors for future prediction. Accordingly, we organize the related literature along two primary axes: channel modeling and periodicity modeling.

2.1 Channel Modeling in Time Series Forecasting

Channel modeling aims to capture dependencies among variables to improve the effectiveness of multivariate representation learning. Representative methods primarily enhance cross-variable interactions from different perspectives. Crossformer (Zhang & Yan, 2023) introduces a cross-dimension attention mechanism to explicitly model correlations between variables. CSformer (Wang et al., 2025a) proposes a parameter-shared two-stage self-attention mechanism, employing a strategy of channel-independence followed by mixing to jointly capture and integrate sequential and channel features. S-D-Mamba (Wang et al., 2025d) utilizes a dual state-space branch

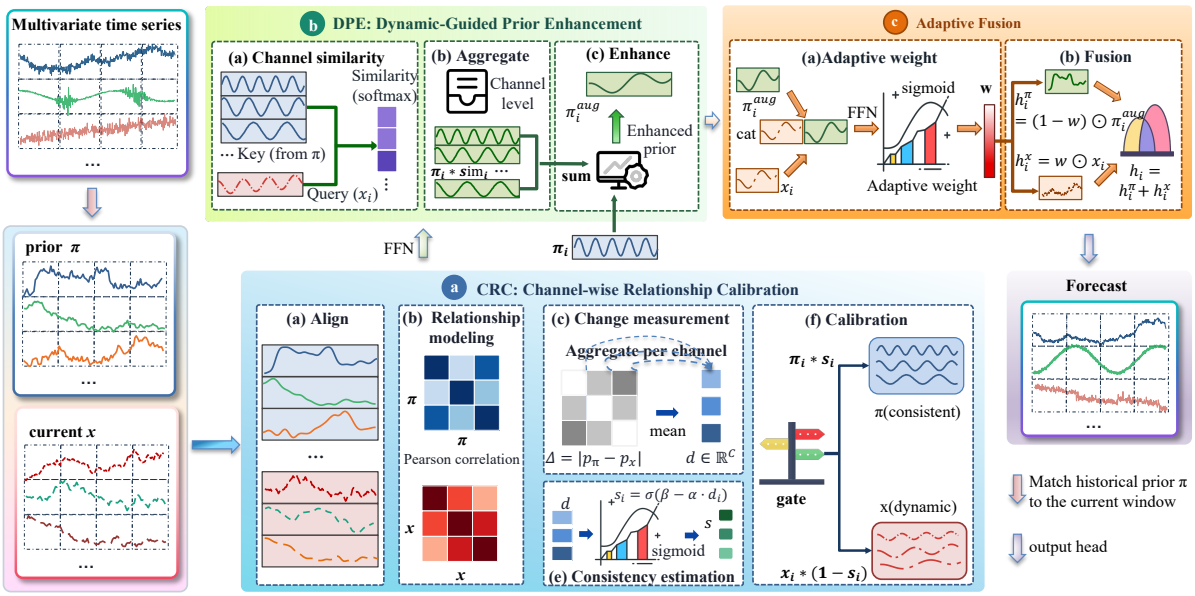

Figure 2: Overview of COMPASS, consisting of three stages: (a) **Channel-wise Relationship Calibration (CRC)**, which calibrates the periodic prior according to its channel-wise discrepancy with current dynamics; (b) **Dynamic-Guided Prior Enhancement (DPE)**, which enhances the calibrated prior through dynamics-guided cross-channel aggregation; and (c) **Adaptive Fusion**, which dynamically balances the enhanced prior and current dynamics for forecasting.

to jointly capture temporal dynamics and channel interactions, while TimeMosaic (Ding et al., 2026) uniformly injects global temporal priors into each variable by sharing prompt tokens across channels. Although these methods effectively enhance inter-variable representations, they primarily focus on variable interactions rather than temporal periodic structures.

2.2 Periodicity Modeling in Time Series Forecasting

Periodicity modeling has long been a crucial avenue for improving the accuracy of long-term forecasting. To effectively capture and utilize historical periodic patterns, existing research primarily unfolds along several distinct directions. One line of methods explicitly extracts periodic components through season-trend decomposition, such as Autoformer (Wu et al., 2021), FEDformer (Zhou et al., 2022b), and the TimeMixer series (Wang et al., 2024b; 2025b). Another line of methods integrates frequency-domain modeling into deep networks, utilizing spectral representations like FFT to capture periodic patterns that are difficult to characterize in the time domain, including FiLM (Zhou et al., 2022a), FITS (Xu et al., 2024), and FilterNet (Yi et al., 2024). In recent years, research has further shifted towards the explicit learning and retrieval of historical periodic patterns. CycleNet (Lin et al., 2024) captures stable periodic laws through learnable cyclic representations; TQNet (Lin et al., 2025a) retrieves global periodic dependencies using temporal queries; and GTR (Cao et al., 2026) maintains long-term temporal memory to dynamically retrieve relevant historical periodic segments. FreqCycle (Zhang et al., 2026) further synergizes time-domain and frequency-domain information to jointly model low-frequency periodicity and high-frequency dynamics. However, most existing methods implicitly assume that historical periodic patterns can be utilized holistically or uniformly, lacking an in-depth investigation into the adaptation discrepancies between periodic priors and current dynamics across different channels.

3 Method

In multivariate time series forecasting (MTSF), the goal is to forecast $\mathbf{Y} \in \mathbb{R}^{C \times T}$ from a historical observation window $\mathbf{X} \in \mathbb{R}^{C \times L}$, where C, L, and T denote the number of variables, look-back length, and forecasting horizon, respectively. Historical observations can provide useful prior information for forecasting the current window, motivating the use of periodic priors.

3.1 Overview

We revisit periodicity modeling in time series forecasting from a channel-wise perspective. Different channels exhibit varying degrees of consistency between historical periodic priors and current dynamics. However, existing methods typically utilize historical periodic priors uniformly across channels, which may introduce mismatched

prior information and interfere with modeling current dynamics. To address this issue, we propose COMPASS (Figure 2), a channel-aware periodic prior adaptation framework. Specifically, COMPASS first calibrates the utilization of periodic priors and current dynamics for each channel according to their degree of mismatch. It then uses current dynamics to guide cross-channel information aggregation, leveraging relevant periodic information from other channels to further enhance the calibrated prior. Finally, an adaptive fusion mechanism integrates the two representations for forecasting.

To capture global periodic patterns beyond the local receptive window, we adopt the periodic embedding mechanism of CycleNet Lin et al. (2024). Given a predefined period length P, we construct a learnable periodic embedding matrix $\mathbf{Q} \in \mathbb{R}^{C \times P}$, which is globally shared across the temporal dimension. For an input sequence starting at time index t, periodic embeddings are retrieved according to their phase indices, yielding the window-aligned periodic prior $\boldsymbol{\pi} = \text{Lookup}\left(\mathbf{Q}, [(t+\tau) \bmod P]_{\tau=0}^{L-1}\right) \in \mathbb{R}^{C \times L}$.

3.2 Channel-wise Relation Calibration (CRC)

Existing methods typically apply historical periodic priors to all channels using the same strategy, which may mix stable long-term patterns with rapidly evolving local dynamics. Our goal is to preserve the long-term stable information in the periodic prior while allowing the current input to retain more dynamic information, thereby decoupling the two sources of information. However, the degree of mismatch between the periodic prior and current dynamics can vary substantially across channels. Therefore, the periodic prior should be utilized adaptively at the channel level rather than treated uniformly across all channels. Based on this observation, we propose the **Channel-wise Relationship Calibration (CRC)** module. By comparing the cross-channel relationships of the periodic prior and current dynamics, CRC explicitly estimates their consistency for each channel and calibrates the utilization of the periodic prior accordingly, encouraging the prior to focus on stable patterns while preserving channel-specific dynamic information in the current input.

Temporal Alignment. To eliminate the temporal deviation between the periodic prior and the real-time dynamics, we first utilize a linear layer mapping to align π and x in the time domain:

$$\pi^{align} = \text{Linear}_{\pi}(\pi), \quad x^{align} = \text{Linear}_{x}(x). \tag{1}$$

Channel Relation Modeling. The correlation structure among different channels reflects the dependencies between variables. We first model the channel-wise relationships of the periodic prior and current dynamics by computing their Pearson correlation matrices:

$$R_{\pi} = \text{Pearson}(\pi^{align}), \quad R_{x} = \text{Pearson}(x^{align}). \tag{2}$$

Relation Difference Measurement. We then measure the discrepancy between the historical periodic structure and the current dynamics by taking the element-wise absolute difference between the two correlation matrices:

$$D = |R_{\pi} - R_{x}|. \tag{3}$$

Channel-wise Difference Aggregation. To quantify the overall relation change for each channel, we average the relation differences with all other channels, excluding the self-relation, to obtain the channel-wise discrepancy d_i:

$$d_i = \frac{1}{C-1} \sum_{j \neq i} D_{ij}. \tag{4}$$

Consistency Score Estimation. We map the channel-wise discrepancy d_i to a learnable consistency score s_i using a sigmoid function:

$$s_i = \sigma(\beta - \alpha \cdot d_i), \tag{5}$$

where σ denotes the sigmoid activation function, and α and β are learnable scaling and bias parameters, respectively.

Channel-wise Calibration. Finally, we perform complementary weighting on the periodic prior and the real-time dynamics based on the consistency score s_i. Consequently, channels with higher consistency retain more of the periodic prior, whereas channels with lower consistency rely more heavily on the current dynamics, thereby achieving a channel-wise adaptive calibration of periodic information:

$$\pi_i^c = \pi_i \cdot s_i, \quad x_i^c = x_i \cdot (1 - s_i), \tag{6}$$

where π^c and x^c represent the calibrated periodic prior and dynamic representation, respectively.

3.3 Intermediate Feature Refinement

The calibrated periodic and dynamic representations are further refined through independent Feed-Forward Networks (FFNs) with residual connections:

$$\pi^r = \text{FFN}_{\pi}(\pi^c) + \pi^c, \quad x^r = \text{FFN}_{x}(x^c) + x^c. \tag{7}$$

The resulting π^r and x^r serve as refined periodic and dynamic representations for the subsequent cross-channel enhancement.

3.4 Dynamic-guided Prior Enhancement (DPE)

While CRC calibrates the periodic prior according to its channel-wise consistency with current dynamics, it does not explicitly exploit latent dependencies among periodic patterns across channels. Such cross-channel dependencies may provide complementary periodic information relevant to the current dynamics. Therefore, we introduce **Dynamic-guided Prior Enhancement (DPE)** to enrich the calibrated periodic representation through dynamics-guided cross-channel aggregation.

Specifically, we use the refined dynamic representation x^r as the Query and the refined periodic representation π^r as the Key and Value. The attention weights are computed as

$$A = \mathrm{softmax}\left(\frac{Q_{x^r}(K_{\pi^r})^T}{\sqrt{d}}\right), \tag{8}$$

where A measures the relevance of periodic information across channels to the current dynamics. Based on these attention weights, we aggregate the periodic representations to obtain a dynamics-relevant periodic context:

$$h_\pi^{\mathrm{ctx}} = AV_{\pi^r}. \tag{9}$$

Finally, the aggregated context is added to the refined periodic representation to obtain the enhanced periodic representation:

$$\pi^e = \pi^r + h_\pi^{\mathrm{ctx}}. \tag{10}$$

3.5 Adaptive Fusion

After the channel-wise calibration and dynamic-guided prior enhancement, we obtain the enhanced periodic representation π^e and the refined dynamic representation x^r, which contain complementary information. To fully exploit both types of representations, we employ an adaptive fusion mechanism to dynamically generate fusion weights based on their joint features:

$$w = \sigma\left(\mathrm{MLP}\left([\pi^e; x^r]\right)\right) \tag{11}$$

where $[\cdot;\cdot]$ denotes feature concatenation, and $\sigma(\cdot)$ represents the Sigmoid activation function. Subsequently, we utilize these weights to perform a weighted fusion of the two representations:

$$h = w \odot x^r + (1 - w) \odot \pi^e \tag{12}$$

Consequently, the model can adaptively regulate the contributions of the periodic and dynamic information according to the current representation, thereby achieving an effective complementarity between the two to form the final unified representation h.

3.6 Final Output Projection

Finally, the unified representation h is linearly projected to generate the final prediction sequence of length T:

$$\hat{Y} = \mathrm{Linear}_{\mathrm{out}}(h) \quad \in \mathbb{R}^{C\times T} \tag{13}$$

4 Experiments

4.1 Experimental Details

Datasets. To rigorously evaluate the performance and generalizability of our proposed method, we conduct extensive experiments across 12 widely recognized multivariate time series forecasting benchmarks. These datasets span diverse application domains and exhibit a broad spectrum of statistical properties, providing a comprehensive and challenging testbed. We categorize the benchmarks into two primary tasks:

- **Long-term Forecasting (8 Datasets):** This includes the ETT series (ETTh1, ETTh2, ETTm1, ETTm2) (Zhou et al., 2021), alongside Electricity, Weather, Traffic (Wu et al., 2022), and Solar-Energy (Lin et al., 2024). These benchmarks evaluate the model's capacity to capture long-range dependencies and complex, overlapping periodic patterns.
- **Short-term Forecasting (4 Datasets):** This comprises the PEMS traffic network datasets (PEMS03, PEMS04, PEMS07, and PEMS08) (Liu et al., 2022). These datasets are utilized to assess the model's responsiveness to high-frequency dynamics and localized temporal shifts.

Baselines. We compare our method with a diverse set of representative state-of-the-art baselines, including Transformer-based models (iTransformer (Liu et al., 2023)), periodic modeling methods (CycleNet (Lin et al., 2024), TQNet (Lin et al., 2025a), GTR (Cao et al., 2026), and FreqCycle (Zhang et al., 2026)), channel modeling

approaches (S-Mamba (Wang et al., 2025d) and TimeMosaic (Ding et al., 2026)), distribution-aware frameworks (DUET (Qiu et al., 2025b)), frequency modeling methods (FilterTS (Wang et al., 2025c)), and hybrid architectures (SOFTS (Han et al., 2024)).

Setup. All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 3090 GPU with 24GB memory. To ensure a fair comparison, we follow the experimental setup of iTransformer Liu et al. (2023). We report mean absolute error (MAE) and mean squared error (MSE) as evaluation metrics. The input length is fixed to $L = 96$. For the PEMS datasets, the prediction horizon is set to $H \in \{12, 24, 48, 96\}$, while for the remaining datasets, $H \in \{96, 192, 336, 720\}$.

4.2 Main Results

Extensive experiments on 12 real-world datasets show that COMPASS achieves the best performance on 10 datasets, reducing MSE by 2.23% on average compared with the best non-COMPASS methods.

Long-term Forecasting. Table 1 reports the long-term forecasting results with an input length of 96, averaged over $H \in 96, 192, 336, 720$. COMPASS achieves 9 first-place results across the eight datasets, with the best overall performance on ETTh1, ETTm1, Weather, Electricity, and Solar, and remains competitive on the remaining datasets. FreqCycle and FilterTS reports results on only seven datasets and is therefore compared within this subset.

Table 1: Long-term forecasting results. The look-back length is fixed at 96. All results are averaged across four different forecasting horizons $H \in \{96, 192, 336, 720\}$. The best results are highlighted in **bold**, while the second-best results are underlined. Lower values indicate better performance. See Table 5 in Appendix B for the full results.

model	COMPASS (Ours)		GTR (2026)		TimeMosatic (2026)		FreqCycle (2026)		DUET (2025)		FilterTS (2025)		TQNet (2025)		S-Mamba (2025)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)	
metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	**0.406**	**0.424**	0.439	0.434	0.425	**0.424**	0.428	0.427	0.443	0.436	0.433	0.430	0.441	0.434	0.455	0.450	0.457	0.441	0.449	0.443	0.454	0.448
ETTh2	**0.362**	0.395	0.372	0.400	0.363	**0.388**	0.371	0.399	0.372	0.397	0.372	0.396	0.378	0.402	0.381	0.405	0.388	0.409	0.373	0.400	0.383	0.407
ETTm1	**0.362**	0.387	0.367	0.389	0.381	**0.381**	0.372	0.389	0.390	0.393	0.385	0.396	0.377	0.393	0.398	0.405	0.379	0.396	0.393	0.402	0.407	0.410
ETTm2	0.264	0.314	0.268	0.315	0.273	0.314	**0.263**	**0.311**	0.280	0.324	0.276	0.321	0.277	0.323	0.288	0.332	0.266	0.314	0.287	0.330	0.288	0.332
Weather	**0.238**	**0.266**	0.239	0.268	0.251	0.267	0.243	0.270	0.251	0.273	0.244	0.274	0.242	0.269	0.251	0.276	0.243	0.271	0.255	0.278	0.258	0.278
Electricity	**0.160**	**0.258**	0.166	0.260	0.187	0.279	0.168	0.259	0.172	**0.258**	0.180	0.271	0.164	0.259	0.170	0.265	0.168	0.259	0.174	0.264	0.178	0.270
Traffic	0.455	0.275	0.470	0.280	0.458	0.283	0.448	**0.261**	0.451	0.269	0.471	0.315	0.445	0.276	0.414	0.276	0.472	0.301	**0.409**	0.267	0.428	0.282
Solar	**0.194**	0.243	**0.194**	0.245	0.240	0.270	-	-	0.237	**0.233**	-	-	0.198	0.256	0.240	0.273	0.210	0.261	0.229	0.256	0.233	0.262
1st Cnt	9		1		3		3		1		0		0		0		0		1		0	

Compared with existing periodic modeling methods, including GTR, FreqCycle, TQNet, and CycleNet, COMPASS provides a more adaptive way to utilize historical periodic information. These methods primarily exploit periodic patterns to construct or enhance forecasting representations, but generally apply the resulting periodic information without explicitly calibrating its utilization according to channel-wise mismatch with current dynamics. In contrast, COMPASS first calibrates the periodic prior at the channel level and then enhances the calibrated prior with dynamics-guided cross-channel information, allowing each channel to adapt its use of historical periodic information to its current dynamics. The consistent improvements over these periodic baselines, particularly on ETTh1, ETTm1, Weather, and Electricity, suggest that explicitly accounting for channel-wise differences can further improve periodic forecasting.

Short-term Forecasting. Table 2 reports the short-term forecasting results with input length 96, averaged over $H \in \{12, 24, 48, 96\}$. FSR (Ours) achieves the best performance on all four benchmarks, obtaining 8 first-place entries. It ranks first on PEMS03, PEMS04, PEMS07, and PEMS08 in both MSE and MAE. GTR is the closest competitor on PEMS03/PEMS04, while TQNet and SOFTS achieve some second-best results on PEMS07 and PEMS08, respectively. TimeMosatic, FreqCycle, FilterTS and DUET are not included due to the lack of publicly available results or training code.

4.3 Ablation Study

We conduct ablation studies to evaluate the contributions of each module (Table 3). The results demonstrate that removing any of the CRC, DPE, or Fusion modules generally degrades overall performance, although results remain stable on certain datasets. Furthermore, we conduct replacement experiments on the CRC and DPE modules.

COMPASS Ablation Variants. To evaluate the effectiveness of individual components, we construct three ablation variants: w/o CRC, w/o DPE, and w/o AF. To further investigate the design choices within CRC, we introduce several replacement variants, denoted as R-C1–R-C6. To evaluate the effectiveness of dynamically guided prior adaptation, we further design R-D variants that modify the design of DPE. The specific variants are described below: **w/o CRC** removes CRC, i.e., $\pi^c = \pi^{align}, x^c = x^{align}$. **w/o DPE** removes DPE, i.e., $\pi^e = \pi^r$. **w/o AF** removes AF and directly adds the two representations, i.e., $h = \pi^e + x^r$. **R-C1** replaces complementary weighting in CRC

Table 2: Short-term forecasting results. The look-back length is fixed at 96. All results are averaged across four different forecasting horizons $H \in \{12, 24, 48, 96\}$. The best results are highlighted in **bold**, while the second-best results are underlined. Lower values indicate better performance. See Table 6 in Appendix B for the full results.

model	COMPASS (Ours)		GTR (2026)		TQNet (2025)		S-Mamba (2025)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)	
metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS03	**0.086**	**0.185**	0.087	0.189	0.097	0.203	0.122	0.228	0.118	0.226	0.104	0.210	0.113	0.222
PEMS04	**0.085**	**0.188**	0.087	0.189	0.091	0.197	0.103	0.211	0.119	0.232	0.102	0.208	0.111	0.221
PEMS07	**0.074**	**0.165**	0.076	0.169	0.075	0.171	0.089	0.188	0.113	0.214	0.086	0.184	0.101	0.204
PEMS08	**0.104**	**0.189**	0.142	0.222	0.142	0.230	0.148	0.224	0.150	0.246	0.138	0.220	0.150	0.226
1st Cnt	8		0		0		0		0		0		0	

with direct subtraction, i.e., $x_i^c = x_i - \pi_i$. **R-C2** averages the consistency score across channels and broadcasts the resulting scalar, i.e., $s_i = \text{mean}(s_i, \text{dim} = 1)$. **R-C3** uses the maximum rather than the average relation change, i.e., $d_i = \max_{j \neq i} D_{ij}$. **R-C4** removes the learnable α, β and sets $s_i = \sigma(-d_i)$. **R-C5** replaces correlation-based consistency with an MLP, i.e., $s_i = \sigma(\text{MLP}([\pi^{align}; x^{align}]))$. **R-C6** uses cosine similarity instead of correlation differences, i.e., $s_i = \text{CosineSimilarity}(\pi^{align}, x^{align})$. **R-D1** reverses the Query–Key roles in DPE, using π^r as Query and x^r as Key, i.e., $A = \text{softmax}(Q_{\pi^r}(K_{x^r})^\top / \sqrt{d})$.

Table 3: Ablation analysis results. The look-back length is fixed at 96. All results are averaged across four different forecasting horizons, where $T \in \{12, 24, 48, 96\}$ for the PEMS datasets and $H \in \{96, 192, 336, 720\}$ for all other datasets. T he best results are highlighted in **bold**. Lower values indicate better performance. See Table 7 in Appendix C for the full results.

Dataset	Metric	w/o CRC	w/o DPE	w/o AF	R-C1	R-C2	R-C3	R-C4	R-C5	R-C6	R-D1	**COMPASS**
ETTh1	MSE	0.417	0.408	0.411	0.422	0.407	0.407	0.415	0.412	**0.406**	0.416	**0.406**
	MAE	0.429	0.425	0.429	0.432	0.425	0.425	0.430	0.428	**0.424**	0.429	**0.424**
Weather	MSE	0.241	**0.238**	0.240	0.241	0.242	0.241	0.242	0.242	0.240	0.239	**0.238**
	MAE	0.268	**0.266**	0.268	0.268	0.268	0.268	0.268	0.268	**0.266**	**0.266**	**0.266**
Solar	MSE	**0.194**	0.196	0.198	0.196	**0.194**	**0.194**	0.197	0.197	**0.194**	**0.194**	**0.194**
	MAE	0.244	0.247	0.255	**0.243**	**0.243**	**0.243**	0.244	0.251	**0.243**	**0.243**	**0.243**
PEMS03	MSE	0.088	**0.086**	0.094	0.090	**0.086**	0.092	0.090	0.114	0.092	0.088	**0.086**
	MAE	0.187	**0.185**	0.198	0.187	0.186	0.190	0.188	0.220	0.187	0.187	**0.185**

Table 3 shows that removing CRC, DPE, or AF generally causes slight performance degradation or comparable results, confirming their complementary contributions. The CRC variants (R-C1–R-C6) can match the full model on individual datasets, but exhibit less consistent performance across datasets and metrics. In particular, R-C2, which averages the consistency scores across channels and applies a shared score to all channels, performs worse than COMPASS on most datasets, indicating that different channels benefit from different strategies for utilizing the periodic prior. In contrast, R-C6 uses cosine similarity to measure the consistency between the periodic prior and the current input and achieves performance comparable to COMPASS on most datasets, suggesting that the effectiveness of channel-wise consistency modeling is not limited to a specific similarity formulation. R-C5 shows a notable degradation on PEMS08. Similarly, reversing the Query–Key roles in DPE (R-D1) leads to small but consistent changes. Overall, the full COMPASS provides the most stable performance across datasets, supporting the effectiveness of its channel-wise calibration, dynamic prior enhancement, and adaptive fusion.

4.4 Model Analysis

Robustness Analysis. We evaluate COMPASS across four random seeds (2024–2027) on ETTh1, Solar-Energy, and PEMS04, reporting the mean $\pm$ standard deviation of the error metrics in Figure 3, with detailed results provided in Table 9 in Appendix C. Compared to other baseline methods, COMPASS achieves the best forecasting accuracy, consistently yielding the lowest average MSE. In terms of stability, COMPASS demonstrates minimal variance across most datasets, with particularly small standard deviations on ETTh1 and PEMS04. On Solar-Energy, its standard deviation is slightly larger than that of CycleNet, while its mean MSE remains lower than all competing methods. Notably, COMPASS achieves standard deviations of only 0.0031 and 0.0004 on ETTh1 and PEMS04, respectively, indicating highly consistent performance across different random seeds. Overall, COMPASS maintains consistently strong accuracy with low variability across different random seeds, demonstrating a favorable balance between predictive precision and robustness.

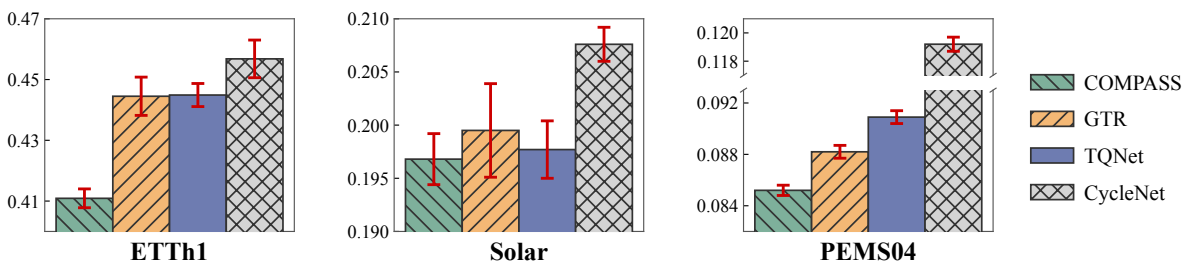

Figure 3: Robustness across random seeds. Results are averaged over four forecasting horizons per seed and reported as mean ± standard deviation across four seeds (2024–2027).

Impact of hyperparameter d_{model}. We investigate the sensitivity of different methods to the model dimension d_{model} on three representative datasets. As shown in Fig. 4, COMPASS consistently maintains a lower MSE than GTR, TQNet, and CycleNet across the tested dimensions, while exhibiting relatively small performance variations. On ETTh1, COMPASS achieves its best MSE of 0.406 at $d_{\text{model}} = 256$ and maintains a clear margin over the three competing methods across all dimensions. On Weather, COMPASS is particularly stable, with MSEs ranging only from 0.238 to 0.240, whereas the competing methods show more noticeable variations with changes in d_{model}. On PEMS03, COMPASS improves from 0.090 to 0.086 as the dimension increases and largely saturates beyond $d_{\text{model}} = 256$, while GTR, TQNet, and CycleNet exhibit more gradual performance changes over the same range. Overall, COMPASS demonstrates both strong performance and low sensitivity to d_{model}, indicating that its advantage over existing periodic forecasting methods is not dependent on a narrowly tuned model dimension.

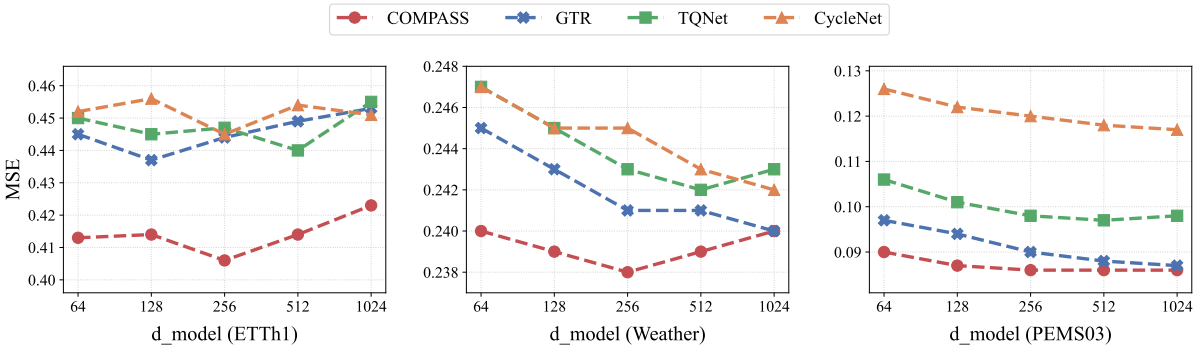

Figure 4: Performance sensitivity to the model dimension d_{model} on three representative datasets. COMPASS is compared with GTR, TQNet, and CycleNet across different model dimensions.

Model Complexity. We compare COMPASS with representative forecasting methods in terms of parameter count, training time, and forecasting accuracy, as shown in Figure 5. We focus on the high-dimensional Electricity dataset with 321 channels, where the computational cost of channel-wise modeling is more evident. COMPASS achieves the lowest MSE of 0.200, with 1.52M parameters and a training time of 17.3s per epoch. Compared with periodic modeling methods, including CycleNet and TQNet, and the frequency-based FreqCycle, COMPASS achieves better forecasting accuracy while maintaining a moderate computational cost. Compared with Transformer-based iTransformer and the Mamba-based S-Mamba, COMPASS uses fewer parameters and requires less training time while achieving lower MSE. Compared with the distribution-aware GTR, COMPASS also achieves lower MSE with comparable computational cost. Although lightweight methods such as CycleNet and FreqCycle require fewer parameters, their forecasting performance is inferior to COMPASS. These comparisons show that COMPASS does not rely on excessive model capacity to achieve its performance gains, but provides a favorable balance between forecasting accuracy, parameter efficiency, and training cost.

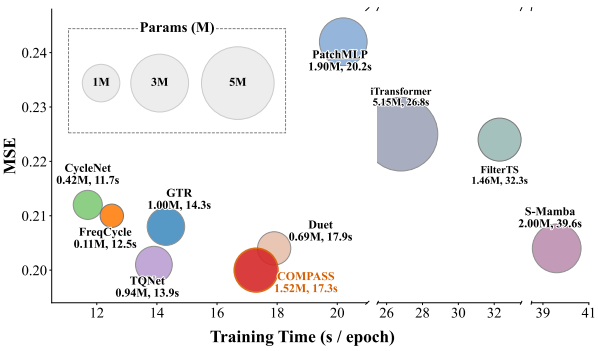

Figure 5: Comprehensive performance comparison of different models on Electricity (321 channels) datasets. The horizontal axis represents training time per epoch, the vertical axis represents Mean Squared Error (MSE), and circle size indicates the number of parameters (Params) of the models.

4.5 MECHANISM ANALYSIS OF COMPASS

We analyze the mechanism of COMPASS from two perspectives: **Channel-wise Relationship Calibration** and **Dynamic-guided Prior Enhancement**. CRC aims to disentangle the periodic prior from the current input, while DPE further enriches the calibrated prior with dynamic information.

Channel-wise Relationship Calibration. We first examine whether CRC effectively decouples the periodic prior from the current input. CRC measures the mismatch between their cross-channel relationship structures and derives a channel-specific consistency score to calibrate their contributions. As shown in Fig. 6, the Pearson correlation between the original prior π and input x is substantially reduced after CRC calibration. This reduction indicates that CRC suppresses shared information while preserving their characteristics. Consequently, the calibrated prior is encouraged to focus more on stable long-range patterns, whereas the calibrated input retains more current dynamics. These results demonstrate that channel-wise relationship-based calibration adapts prior utilization across channels and promotes separation between stable prior information and dynamic input information.

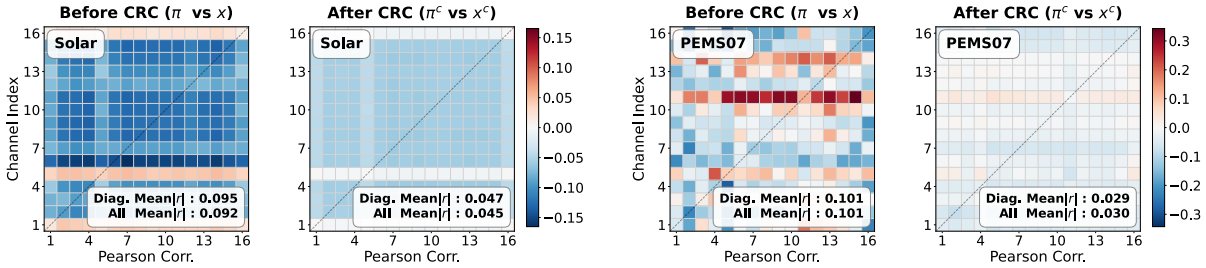

Figure 6: **Pearson Correlation Before and After CRC Calibration.** Comparison of the correlation between the original prior π and input x before and after CRC calibration on Solar and PEMS07. CRC reduces the mean absolute correlation (Mean $|r|$), indicating improved decoupling between the calibrated prior and input.

Dynamic-guided Prior Enhancement. We further investigate whether DPE improves the representation of the calibrated prior. As shown in Fig. 7, the Pearson correlation matrices of the enhanced prior π^e exhibit patterns that more closely resemble those of the ground truth Y on both Solar and PEMS07. This observation is quantitatively supported by the global mean Pearson correlation: on Solar, the correlation increases from 0.680 (π^r) to 0.842 (π^e), approaching the ground-truth value of 0.902; on PEMS07, it increases from 0.248 to 0.393, compared with 0.847 for the ground truth. These results indicate that DPE effectively injects dynamic information into the calibrated prior, aligning its correlation structure more closely with the underlying target.

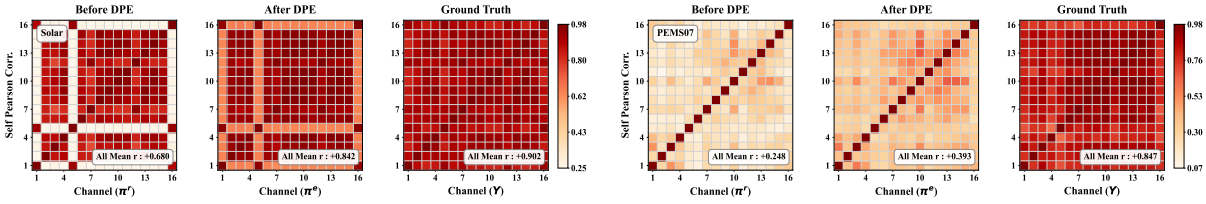

Figure 7: **Mechanism Analysis of the DPE Module.** Pearson self-correlation matrices of the prior π^r, the enhanced prior π^e, and the ground truth Y on Solar and PEMS07. DPE makes the correlation pattern of π^e more closely align with Y, while substantially increasing the overall mean correlation toward the ground-truth level.

5 CONCLUSION

In this work, we identify the channel-wise inconsistency between historical periodic priors and current dynamics as an important yet underexplored challenge in multivariate time series forecasting. To address this issue, we propose **COMPASS**, a channel-wise periodic prior adaptation framework that enables periodic information to be utilized adaptively across channels. COMPASS first employs Channel-wise Relationship Calibration (CRC) to quantify the discrepancy between historical periodic patterns and current dynamics and accordingly calibrate their contributions at the channel level. It then introduces Dynamic-guided Prior Enhancement (DPE), which uses current dynamics to guide cross-channel aggregation and enrich the calibrated periodic prior with relevant information. Finally, an adaptive fusion mechanism balances the enhanced periodic prior and current dynamics for forecasting. Extensive experiments on 12 public benchmarks demonstrate the effectiveness of COMPASS. It achieves an average MSE reduction of 2.23% over the strongest competing methods, with the best performance on 6 of 8 long-term forecasting datasets and all 4 short-term forecasting datasets. These results demonstrate that explicitly accounting for channel-wise prior–dynamics consistency and dynamically enhancing periodic information can provide a more effective way to exploit historical periodic patterns for multivariate time series forecasting.

AI USE STATEMENT

We used Large Language Models (LLMs) solely as auxiliary tools for language editing, text polishing, and LaTeX formatting. They had no influence on the research design, experiments, analysis, or scientific claims. All research contributions and conclusions were developed and verified independently by the authors.

ETHICS STATEMENT

This work uses only publicly available datasets and does not involve human subjects or personally identifiable information. All datasets were used in accordance with their publicly stated terms of use.

REPRODUCIBILITY STATEMENT

The model architecture, training objectives, evaluation metrics, and datasets are detailed in the paper and supplementary materials. The complete source code will be released upon acceptance.

REFERENCES

E Oran Brigham and Richard E Morrow. The fast fourier transform. *IEEE spectrum*, 4(12):63–70, 1967.

Fanpu Cao, Lu Dai, Jindong Han, and Hui Xiong. Enhancing multivariate time series forecasting with global temporal retrieval. *arXiv preprint arXiv:2602.10847*, 2026.

Hui Chen, Xiangyang Wang, Yang Xiao, Beixian Wu, Zhuo Wang, Yao Liu, Peiyi Wang, Chunjie Chen, and Xinyu Wu. ip3t: an interpretable multimodal time-series model for enhanced gait phase prediction in wearable exoskeletons. *Frontiers in Neuroscience*, 18:1457623, 2024.

Hui Chen, Zhuo Wang, Fangliang Yang, Xiangyang Wang, Chunjie Chen, and Xinyu Wu. Enhancing terrain recognition with a transformer-based model: Integrating imus for motion intent detection. *IEEE Transactions on Human-Machine Systems*, 56(1):22–31, 2026. doi: 10.1109/THMS.2025.3631855.

Yu Chengqing, Yan Guangxi, Yu Chengming, Zhang Yu, and Mi Xiwei. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy*, 263: 126034, 2023.

Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with tide: Time-series dense encoder. *arXiv preprint arXiv:2304.08424*, 2023.

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

Qiang Guo, Lexin Fang, Ren Wang, and Caiming Zhang. Multivariate time series forecasting using multiscale recurrent networks with scale attention and cross-scale guidance. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1):540–554, 2023.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. Softs: Efficient multivariate time series forecasting with series-core fusion. *Advances in Neural Information Processing Systems*, 37:64145–64175, 2024.

Qihe Huang, Zhengyang Zhou, Yangze Li, Kuo Yang, Binwu Wang, and Yang Wang. Many minds, one goal: Time series forecasting via sub-task specialization and inter-agent cooperation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Anwar Khan, Mostafa M Fouda, Dinh-Thuan Do, Abdulaziz Almaleh, and Atiq Ur Rahman. Short-term traffic prediction using deep learning long short-term memory: Taxonomy, applications, challenges, and future trends. *IEEE access*, 11:94371–94391, 2023.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37:106315–106345, 2024.

Shengsheng Lin, Haojun Chen, Haijie Wu, Chunyun Qiu, and Weiwei Lin. Temporal query network for efficient multivariate time series forecasting. *arXiv preprint arXiv:2505.12917*, 2025a.

Shengsheng Lin, Weiwei Lin, Wentai Wu, Feiyu Zhao, Ruichao Mo, and Haotong Zhang. Segrnn: Segment recurrent neural network for long-term time series forecasting. *IEEE Internet of Things Journal*, 2025b.

Minhao Liu, Ailing Zeng, Muxi Chen, Zhijian Xu, Qiuxia Lai, Lingna Ma, and Qiang Xu. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in neural information processing systems*, 35:5816–5828, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625*, 2023.

Donghao Luo and Xue Wang. Moderntcn: A modern pure convolution structure for general time series analysis. In *ICLR*, 2024.

Xiaowen Ma, Zhenliang Ni, Shuai Xiao, and Xinghao Chen. Timepro: Efficient multivariate long-term time series forecasting with variable-and time-aware hyper-state. *arXiv preprint arXiv:2505.20774*, 2025.

Xiangfei Qiu, Xingjian Wu, Hanyin Cheng, Xvyuan Liu, Chenjuan Guo, Jilin Hu, and Bin Yang. Dbloss: Decomposition-based loss function for time series forecasting. *arXiv preprint arXiv:2510.23672*, 2025a.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1185–1196, 2025b.

Said E Said and David A Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

Haoxin Wang, Yipeng Mo, Kunlan Xiang, Nan Yin, Honghe Dai, Bixiong Li, Songhai Fan, and Site Mo. Csformer: Combining channel independence and mixing for robust multivariate time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 21090–21098, 2025a.

Shiyu Wang, Zhixuan Chu, Yinbo Sun, Yu Liu, Yuliang Guo, Yang Chen, Huiyang Jian, Lintao Ma, Xingyu Lu, and Jun Zhou. Multiscale representation enhanced temporal flow fusion model for long-term workload forecasting. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 4948–4956, 2024a.

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*, 2024b.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin. Timemixer++: A general time series pattern machine for universal predictive analysis. In *International Conference on Learning Representations*, volume 2025, pp. 16980–17016, 2025b.

Yulong Wang, Yushuo Liu, Xiaoyi Duan, and Kai Wang. Filterts: Comprehensive frequency filtering for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21375–21383, 2025c.

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

Zihan Wang, Fanheng Kong, Shi Feng, Ming Wang, Xiaocui Yang, Han Zhao, Daling Wang, and Yifei Zhang. Is mamba effective for time series forecasting? *Neurocomputing*, 619:129178, 2025d.

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. *arXiv preprint arXiv:2210.02186*, 2022.

Zhijian Xu, Ailing Zeng, and Qiang Xu. Fits: Modeling time series with $10k$ parameters. In *International Conference on Learning Representations*, volume 2024, pp. 26295–26318, 2024.

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.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11121–11128, 2023.

Boya Zhang, Shuaijie Yin, Huiwen Zhu, and Xing He. Freqcycle: A multi-scale time-frequency analysis method for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28159–28167, 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, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, Rong Jin, et al. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in neural information processing systems*, 35:12677–12690, 2022a.

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

A Implementation Details

A.1 Dataset Details

We conduct experiments on 12 widely used time series datasets, including benchmarks for long-term forecasting and PEMS datasets for short-term forecasting. Dataset statistics are summarized in Table 4. ADF denotes the stationarity distribution metrics, which are calculated using the Augmented Dickey-Fuller (ADF) (Said & Dickey, 1984) test. Detailed descriptions are provided below.

- **ETT** (Zhou et al., 2021) records transformer oil temperature and load data from China (2016–2018), provided in hourly (ETTh1, ETTh2) and 15-minute (ETTm1, ETTm2) resolutions.
- **Weather** (Wu et al., 2022) contains 21 meteorological indicators (e.g., temperature, visibility) collected in Germany throughout 2020 at 10-minute intervals.
- **Electricity** (Wu et al., 2022) comprises hourly electricity consumption data (in kWh) from 321 clients spanning 2012 to 2014.
- **Traffic** (Wu et al., 2022) includes hourly road occupancy rates from 862 highway sensors in the San Francisco Bay Area between 2015 and 2016.
- **Solar-Energy** (Lin et al., 2024) contains solar power generation records from 137 photovoltaic plants in Alabama, USA, sampled every 10 minutes in 2006.
- **PEMS** (Liu et al., 2022) comprises four public highway traffic datasets (PEMS03, PEMS04, PEMS07, PEMS08) from the Caltrans Performance Measurement System, aggregated at 5-minute intervals.

Task	Pred. Length	Dataset	Ch.	Period	Split	Freq.	ADF
Long-term Forecast	{96, 192, 336, 720}	ETTh1	7	24	6:2:2	1h	-5.91
		ETTh2	7	24	6:2:2	1h	-4.13
		ETTm1	7	96	6:2:2	15m	-14.98
		ETTm2	7	96	6:2:2	15m	-5.66
		Weather	21	144	7:1:2	10m	-26.68
		Electricity	321	168	7:1:2	1h	-8.44
		Traffic	862	168	7:1:2	1h	-15.02
		Solar	137	144	7:1:2	10m	-37.23
Short-term Forecast	{12, 24, 48, 96}	PEMS03	358	288	3:1:1	5m	-19.05
		PEMS04	307	288	3:1:1	5m	-15.66
		PEMS07	883	288	3:1:1	5m	-20.6
		PEMS08	170	288	3:1:1	5m	-16.04

Table 4: Detailed statistics of the datasets. Ch. denotes the number of variables, and Period denotes the length of a complete cycle. Split indicates the ratio of training/validation/test sets. ADF denotes the stationarity distribution metrics.

A.2 Evaluation Metrics

To evaluate the performance of our model on both long-term and short-term forecasting tasks, we employ two standard metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics are defined as follows:

$$\text{MSE} = \frac{1}{H \times C} \sum_{i=1}^{H} \sum_{j=1}^{C} (y_{i,j} - \hat{y}_{i,j})^2, \tag{14}$$

$$\text{MAE} = \frac{1}{H \times C} \sum_{i=1}^{H} \sum_{j=1}^{C} |y_{i,j} - \hat{y}_{i,j}|, \tag{15}$$

where $y_{i,j}$ and $\hat{y}_{i,j}$ denote the ground truth and the predicted values for the j-th channel at the i-th time step, respectively. Here, H represents the number of future time steps (forecasting horizon) and C denotes the number of channels (variables). For both metrics, a lower value indicates superior forecasting performance.

A.3 ALGORITHM DETAILS

The complete procedure of the proposed framework is summarized in **Algorithm 1**. Given the current input $\mathbf{x}$ and its starting time index t, COMPASS first constructs a periodic prior $\boldsymbol{\pi}$ by retrieving the corresponding temporal patterns from the periodic embedding matrix. **Part I** then performs Channel-wise Relationship Calibration (CRC), where the cross-channel relationships of the periodic prior and current input are compared to estimate channel-specific consistency scores. These scores adaptively regulate the contributions of the two representations, encouraging the periodic prior to retain stable long-range patterns while preserving current dynamics in the input representation. **Part II** further enhances the calibrated periodic representation through Dynamic-guided Prior Enhancement (DPE), where the current dynamic representation provides guidance to aggregate relevant information from the periodic representation across channels. The resulting enhanced prior is therefore better aligned with the current dynamics. Finally, in **Part III**, an adaptive fusion module determines the contribution of the enhanced periodic representation and the dynamic representation for each channel and feature dimension, producing the fused representation for final prediction. The detailed operations of each part are described below.

Algorithm 1: COMPASS: Channel-wise Calibration of Periodic Priors

Input: Current input $\mathbf{x} \in \mathbb{R}^{C\times L}$, starting time index t, periodic embedding matrix $\mathbf{Q} \in \mathbb{R}^{C\times P}$, learnable factors α and β

Output: Prediction $\hat{\mathbf{Y}} \in \mathbb{R}^{C\times T}$

Step 1: Periodic Prior Construction

$\boldsymbol{\pi} \leftarrow \text{Lookup}\left(\mathbf{Q}, [(t+\tau) \bmod P]_{\tau=0}^{L-1}\right), \in \mathbb{R}^{C\times L}$;

Step 2: Channel-wise Relation Calibration (CRC)

$\boldsymbol{\pi}^{align} \leftarrow \text{Linear}_{\pi}(\boldsymbol{\pi}), \in \mathbb{R}^{C\times D}$;

$\mathbf{x}^{align} \leftarrow \text{Linear}_{x}(\mathbf{x}), \in \mathbb{R}^{C\times D}$;

$\mathbf{R}_{\pi} \leftarrow \text{Pearson}(\boldsymbol{\pi}^{align}), \in \mathbb{R}^{C\times C}$;

$\mathbf{R}_{x} \leftarrow \text{Pearson}(\mathbf{x}^{align}), \in \mathbb{R}^{C\times C}$;

$\mathbf{D} \leftarrow |\mathbf{R}_{\pi} - \mathbf{R}_{x}|, \in \mathbb{R}^{C\times C}$;

for $i = 1, \ldots, C$ **do**

  $d_i \leftarrow \frac{1}{C-1}\sum_{j\neq i} D_{ij}$;

  $s_i \leftarrow \sigma(\beta - \alpha d_i)$;

end

$\boldsymbol{\pi}^{c} \leftarrow \mathbf{s} \odot \boldsymbol{\pi}, \in \mathbb{R}^{C\times D}$;

$\mathbf{x}^{c} \leftarrow (\mathbf{1} - \mathbf{s}) \odot \mathbf{x}, \in \mathbb{R}^{C\times D}$;

Step 3: Intermediate Feature Refinement

$\boldsymbol{\pi}^{r} \leftarrow \text{FFN}_{\pi}(\boldsymbol{\pi}^{c}) + \boldsymbol{\pi}^{c}, \in \mathbb{R}^{C\times D}$;

$\mathbf{x}^{r} \leftarrow \text{FFN}_{x}(\mathbf{x}^{c}) + \mathbf{x}^{c}, \in \mathbb{R}^{C\times D}$;

Step 4: Dynamic-guided Prior Enhancement (DPE)

$\mathbf{Q}_{x} \leftarrow Q(\mathbf{x}^{r}), \in \mathbb{R}^{C\times D}$;

$\mathbf{K}_{\pi} \leftarrow K(\boldsymbol{\pi}^{r}), \in \mathbb{R}^{C\times D}$;

$\mathbf{V}_{\pi} \leftarrow V(\boldsymbol{\pi}^{c}), \in \mathbb{R}^{C\times D}$;

$\mathbf{S} \leftarrow \text{softmax}\left(\frac{\mathbf{Q}_{x}\mathbf{K}_{\pi}^{\top}}{\sqrt{d}}\right), \in \mathbb{R}^{C\times C}$;

$\mathbf{h}_{\pi}^{ctx} \leftarrow \mathbf{S}\mathbf{V}_{\pi}, \in \mathbb{R}^{C\times D}$;

$\boldsymbol{\pi}^{e} \leftarrow \boldsymbol{\pi}^{r} + \mathbf{h}_{\pi}^{ctx}, \in \mathbb{R}^{C\times D}$;

Step 5: Adaptive Fusion

$\mathbf{w} \leftarrow \sigma\left(\text{MLP}([\boldsymbol{\pi}^{e}; \mathbf{x}^{r}])\right), \in \mathbb{R}^{C\times D}$;

$\mathbf{h} \leftarrow \mathbf{w} \odot \boldsymbol{\pi}^{e} + (\mathbf{1} - \mathbf{w}) \odot \mathbf{x}^{r}, \in \mathbb{R}^{C\times D}$;

Step 6: Final Output Projection

$\hat{\mathbf{Y}} \leftarrow \text{Linear}_{out}(\mathbf{h}), \in \mathbb{R}^{C\times T}$;

return $\hat{\mathbf{Y}}$

B COMPARISON OF FULL RESULTS

B.1 LONG-TERM FORECASTING RESULTS

As shown in Table 5, the COMPASS model demonstrates a highly significant leading advantage across all forecasting horizons ($H \in \{96, 192, 336, 720\}$) and their comprehensive average performance (Avg), establishing itself as the best-performing forecasting model overall. A detailed analysis is as follows:

- **Dominant Overall Performance:** COMPASS ranks first 29 times across all settings, far exceeding TimeMosaic (12). It achieves the best average MSE in 6 out of 8 datasets, and the lowest average MAE in ETTh1, Weather, and Electricity.
- **Robustness Across Horizons:** The performance degradation remains smooth from short (96) to extremely long (720) forecasting spans. Notably, it consistently maintains the best metrics at almost every individual horizon for Weather and Electricity.
- **Competitive on Specific Datasets:** For ETTm2 and Solar, COMPASS stably ranks second in average metrics, surpassed only by FreqCycle.

Table 5: Full results on 8 long-term forecasting datasets. Input length is fixed to 96. Avg denotes the average over four forecasting horizons: $\{96, 192, 336, 720\}$. **Bold** and underlined indicate the best and second-best results, respectively. Lower is better.

Dataset	T	COMPASS (Ours)		GTR (2026)		TimeMosaic (2026)		FreqCycle (2026)		DUET (2025)		FilterTS (2025)		TQNet (2025)		S-Mamba (2025)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)	
		MSE↓	MAE↓	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.364**	0.394	0.367	0.391	0.369	0.389	0.369	**0.390**	0.377	0.393	0.374	0.391	0.371	0.393	0.386	0.405	0.375	0.395	0.381	0.399	0.386	0.405
	192	**0.408**	0.420	0.420	0.421	0.416	**0.415**	0.417	0.418	0.429	0.425	0.424	0.421	0.428	0.426	0.443	0.437	0.436	0.428	0.435	0.431	0.441	0.436
	336	**0.434**	**0.435**	0.476	0.447	0.454	**0.435**	0.459	0.439	0.471	0.446	0.464	0.441	0.476	0.446	0.489	0.468	0.496	0.455	0.480	0.452	0.487	0.458
	720	**0.420**	**0.448**	0.493	0.476	0.461	0.458	0.456	0.454	0.496	0.480	0.470	0.466	0.487	0.470	0.502	0.489	0.520	0.484	0.499	0.488	0.503	0.491
	Avg	**0.406**	**0.424**	0.439	0.434	0.425	**0.424**	0.425	0.425	0.443	0.436	0.433	0.430	0.441	0.434	0.455	0.450	0.457	0.441	0.449	0.443	0.454	0.448
ETTh2	96	**0.281**	0.334	0.290	0.342	0.286	**0.332**	0.282	0.336	0.296	0.345	0.290	0.338	0.295	0.343	0.296	0.348	0.298	0.344	0.297	0.347	0.297	0.349
	192	0.365	0.392	0.362	0.389	0.362	**0.380**	**0.361**	0.388	0.368	0.389	0.374	0.390	0.367	0.393	0.376	0.396	0.372	0.396	0.373	0.394	0.380	0.400
	336	**0.396**	**0.415**	0.414	0.429	0.409	0.416	0.415	0.428	0.411	0.422	0.406	0.420	0.417	0.427	0.424	0.431	0.431	0.439	0.410	0.426	0.428	0.432
	720	0.408	0.441	0.423	0.442	**0.396**	**0.422**	0.425	0.443	0.412	0.434	0.418	0.437	0.433	0.446	0.426	0.444	0.450	0.458	0.411	0.433	0.427	0.445
	Avg	**0.362**	0.395	0.372	0.400	0.363	**0.388**	0.371	0.399	0.372	0.397	0.372	0.396	0.378	0.402	0.381	0.405	0.388	0.409	0.373	0.400	0.383	0.407
ETTm1	96	**0.303**	0.348	0.305	0.349	0.308	**0.338**	0.310	0.353	0.324	0.354	0.321	0.360	0.311	0.353	0.333	0.368	0.319	0.360	0.325	0.361	0.334	0.368
	192	**0.344**	0.373	0.349	0.375	0.364	**0.368**	0.351	0.374	0.369	0.379	0.363	0.382	0.356	0.378	0.376	0.390	0.360	0.381	0.375	0.389	0.377	0.391
	336	**0.375**	0.397	0.380	0.398	0.394	**0.390**	0.383	0.396	0.404	0.402	0.395	0.403	0.390	0.401	0.408	0.413	0.389	0.403	0.405	0.412	0.426	0.420
	720	**0.427**	0.430	0.435	0.436	0.460	**0.428**	0.446	0.434	0.463	0.437	0.462	0.438	0.452	0.440	0.475	0.448	0.447	0.441	0.466	0.447	0.491	0.459
	Avg	**0.362**	0.387	0.367	0.389	0.381	**0.381**	0.372	0.389	0.390	0.393	0.385	0.396	0.377	0.393	0.398	0.405	0.379	0.396	0.393	0.402	0.407	0.410
ETTm2	96	0.163	0.249	0.168	0.249	0.170	0.248	**0.162**	**0.244**	0.174	0.255	0.172	0.255	0.173	0.256	0.179	0.263	0.163	0.246	0.180	0.261	0.180	0.264
	192	0.227	0.291	0.232	0.292	0.236	0.291	**0.226**	**0.286**	0.243	0.302	0.237	0.299	0.238	0.298	0.250	0.309	0.229	0.290	0.246	0.306	0.250	0.309
	336	0.285	0.329	0.287	0.330	0.295	0.331	0.283	**0.326**	0.304	0.341	0.299	0.398	0.301	0.340	0.312	0.349	0.284	0.327	0.319	0.352	0.311	0.348
	720	**0.381**	0.389	0.386	0.390	0.392	**0.386**	0.383	0.387	0.399	0.397	0.397	0.394	0.397	0.396	0.411	0.406	0.389	0.391	0.405	0.411	0.412	0.407
	Avg	0.264	0.314	0.268	0.315	0.273	0.314	**0.263**	**0.311**	0.280	0.324	0.276	0.321	0.277	0.323	0.288	0.332	0.266	0.314	0.287	0.330	0.288	0.332
Weather	96	**0.153**	**0.198**	0.154	0.200	0.165	**0.198**	0.159	0.203	0.163	0.202	0.162	0.207	0.157	0.200	0.165	0.210	0.158	0.203	0.166	0.208	0.174	0.214
	192	**0.202**	**0.242**	**0.202**	0.244	0.215	0.243	0.208	0.248	0.218	0.252	0.209	0.252	0.206	0.245	0.214	0.252	0.207	0.247	0.217	0.253	0.221	0.254
	336	**0.259**	**0.285**	**0.259**	0.287	0.272	0.286	0.264	0.289	0.274	0.294	0.263	0.292	0.262	0.287	0.274	0.297	0.262	0.289	0.282	0.300	0.278	0.296
	720	**0.340**	**0.340**	0.341	0.342	0.353	**0.340**	0.343	0.342	0.349	0.343	0.344	0.344	0.344	0.342	0.350	0.345	0.344	0.344	0.356	0.351	0.358	0.349
	Avg	**0.238**	**0.266**	0.239	0.268	0.251	0.267	0.243	0.270	0.251	0.273	0.244	0.274	0.242	0.269	0.251	0.276	0.243	0.271	0.255	0.278	0.258	0.278
Electricity	96	**0.129**	**0.225**	0.134	0.229	0.160	0.258	0.140	0.234	0.145	0.233	0.151	0.245	0.134	0.229	0.139	0.235	0.136	0.229	0.143	0.233	0.148	0.240
	192	**0.148**	0.245	0.152	0.245	0.174	0.269	0.154	0.246	0.163	0.248	0.163	0.256	0.154	0.247	0.159	0.255	0.152	**0.244**	0.158	0.248	0.162	0.253
	336	**0.165**	0.265	0.171	0.264	0.189	0.283	0.170	0.264	0.175	0.262	0.180	0.274	0.169	0.264	0.176	0.272	0.170	0.264	0.178	0.269	0.178	0.269
	720	**0.200**	0.298	0.208	0.300	0.225	0.304	0.210	0.297	0.204	**0.291**	0.224	0.311	0.201	0.294	0.204	0.298	0.212	0.299	0.218	0.305	0.225	0.317
	Avg	**0.160**	**0.258**	0.166	0.260	0.187	0.279	0.168	0.260	0.172	**0.258**	0.180	0.271	0.164	0.259	0.170	0.265	0.168	0.259	0.174	0.264	0.178	0.270
Traffic	96	0.429	0.259	0.440	0.263	0.424	0.269	0.438	0.262	0.407	0.252	0.448	0.309	0.413	0.261	0.382	0.261	0.458	0.296	**0.376**	**0.251**	0.395	0.268
	192	0.446	0.269	0.454	0.274	0.448	0.279	0.434	**0.258**	0.431	0.262	0.455	0.307	0.432	0.271	**0.396**	0.267	0.457	0.294	0.398	0.261	0.417	0.276
	336	0.459	0.276	0.472	0.282	0.464	0.284	0.446	**0.258**	0.456	0.269	0.472	0.313	0.450	0.277	0.417	0.276	0.470	0.299	**0.415**	0.269	0.433	0.283
	720	0.486	0.295	0.514	0.301	0.495	0.301	0.476	**0.265**	0.509	0.292	0.508	0.332	0.486	0.295	0.460	0.300	0.502	0.314	**0.447**	0.287	0.467	0.302
	Avg	0.455	0.275	0.470	0.280	0.458	0.283	0.448	**0.261**	0.451	0.269	0.471	0.315	0.445	0.276	0.414	0.276	0.472	0.301	**0.409**	0.267	0.428	0.282
Solar	96	0.174	0.232	0.176	0.235	0.209	0.248	-	-	0.200	**0.207**	-	-	**0.173**	0.233	0.205	0.244	0.190	0.247	0.200	0.230	0.203	0.237
	192	**0.193**	0.242	**0.193**	0.244	0.238	0.272	-	-	0.228	**0.233**	-	-	0.199	0.257	0.237	0.270	0.210	0.266	0.229	0.253	0.233	0.261
	336	0.202	0.249	**0.201**	0.250	0.255	0.277	-	-	0.262	**0.244**	-	-	0.211	0.263	0.258	0.288	0.217	0.266	0.243	0.269	0.248	0.273
	720	0.206	**0.249**	**0.205**	0.251	0.257	0.284	-	-	0.258	**0.249**	-	-	0.209	0.270	0.260	0.288	0.223	0.266	0.245	0.272	0.249	0.275
	Avg	**0.194**	0.243	**0.194**	0.245	0.240	0.270	-	-	0.237	**0.233**	-	-	0.198	0.256	0.240	0.273	0.210	0.261	0.229	0.256	0.233	0.262
1st		**29**		5		12		10		5		0		1		1		1		4		0	

B.2 SHORT-TERM FORECASTING RESULTS

As shown in Table 6, COMPASS (denoted as FCG in the table) demonstrates an overwhelming performance advantage in short-term forecasting tasks across the four PEMS traffic datasets. A detailed summary is as follows:

- **Absolute Leading Win Rate:** Out of a total of 32 independent benchmark evaluations (4 datasets $\times$ 4 forecasting horizons $\times$ 2 metrics), COMPASS secures 31 first-place results, far exceeding the runner-up model GTR (only 4 times).
- **Comprehensive Dominance Across Datasets:** On the PEMS03, PEMS04, and PEMS08 datasets, COMPASS sweeps all best values for both MSE and MAE across all forecasting horizons (12, 24, 48, 96). On the PEMS07 dataset, it also ranks first in all metrics, except for the MSE at $T = 96$, where it takes a close second with a value of 0.104.
- **Strong Robustness at Longer Horizons:** As the forecasting horizon increases (especially when extending to 48 and 96), the errors of other baseline models grow drastically, whereas COMPASS effectively suppresses error accumulation. For instance, in the PEMS08 ($T = 96$) task, COMPASS controls the MSE at 0.151, while the second-place SOFTS model degrades to 0.211, establishing a significant margin of superiority.
- **Complete Sweep of Average Metrics:** In the comprehensive average (Avg) statistics across different horizons for all four datasets, COMPASS's average MSE and average MAE consistently secure the top position across the board, firmly establishing its state-of-the-art (SOTA) status in short-term traffic flow forecasting tasks.

Table 6: Full results on 4 short-term forecasting datasets. Input length is fixed to 96. Avg denotes the average over four forecasting horizons: $\{12, 24, 48, 96\}$. **Bold** and underlined indicate the best and second-best results, respectively. Lower is better.

Dataset	T	FCG (Ours)		GTR (2026)		TQNet (2025)		S-Mamba (2025)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)	
		MSE↓	MAE↓	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS03	12	**0.057**	**0.155**	**0.057**	0.156	0.060	0.161	0.065	0.169	0.066	0.172	0.064	0.165	0.071	0.174
	24	**0.070**	**0.171**	**0.070**	0.172	0.077	0.182	0.087	0.196	0.089	0.201	0.083	0.188	0.093	0.201
	48	**0.093**	**0.194**	0.094	0.198	0.104	0.215	0.133	0.243	0.136	0.247	0.114	0.223	0.125	0.236
	96	**0.125**	**0.222**	0.127	0.228	0.148	0.253	0.201	0.305	0.182	0.282	0.156	0.264	0.164	0.275
	Avg	**0.086**	**0.185**	0.087	0.189	0.097	0.203	0.122	0.228	0.118	0.226	0.104	0.210	0.113	0.222
PEMS04	12	**0.065**	**0.163**	**0.065**	0.164	0.067	0.166	0.076	0.180	0.078	0.186	0.074	0.176	0.078	0.183
	24	**0.074**	**0.175**	0.075	0.177	0.077	0.181	0.084	0.193	0.099	0.212	0.088	0.194	0.095	0.205
	48	**0.090**	**0.196**	0.093	0.197	0.097	0.206	0.115	0.224	0.133	0.248	0.110	0.219	0.120	0.233
	96	**0.111**	**0.217**	0.114	0.219	0.123	0.233	0.137	0.248	0.167	0.281	0.135	0.244	0.150	0.262
	Avg	**0.085**	**0.188**	0.087	0.189	0.091	0.197	0.103	0.211	0.119	0.232	0.102	0.208	0.111	0.221
PEMS07	12	**0.051**	**0.141**	**0.051**	0.142	**0.051**	0.143	0.063	0.159	0.062	0.162	0.057	0.152	0.067	0.165
	24	**0.061**	**0.154**	0.063	0.156	0.063	0.159	0.081	0.183	0.086	0.192	0.073	0.173	0.088	0.190
	48	**0.079**	**0.173**	0.082	0.178	0.081	0.179	0.093	0.192	0.128	0.234	0.096	0.195	0.110	0.215
	96	0.104	**0.194**	0.110	0.202	**0.103**	0.203	0.117	0.217	0.176	0.268	0.120	0.218	0.139	0.245
	Avg	**0.074**	**0.165**	0.076	0.169	0.075	0.171	0.089	0.188	0.113	0.214	0.086	0.184	0.101	0.204
PEMS08	12	**0.069**	**0.162**	0.071	0.168	0.071	0.170	0.076	0.178	0.082	0.185	0.074	0.171	0.079	0.182
	24	**0.086**	**0.177**	0.095	0.192	0.096	0.196	0.104	0.209	0.117	0.226	0.104	0.201	0.115	0.219
	48	**0.110**	**0.196**	0.148	0.234	0.149	0.244	0.167	0.228	0.169	0.268	0.164	0.253	0.186	0.235
	96	**0.151**	**0.223**	0.256	0.293	0.253	0.309	0.245	0.280	0.233	0.306	0.211	0.253	0.221	0.267
	Avg	**0.104**	**0.189**	0.142	0.222	0.142	0.230	0.148	0.224	0.150	0.246	0.138	0.220	0.150	0.226
1st		**31**		4		2		0		0		0		0	

C MORE EXPERIMENTAL DETAILS

C.1 COMPARISON OF ABLATION RESULTS

To systematically evaluate the contribution of each key component in our proposed framework, we design and compare several ablation variants.

COMPASS Ablation Variants.

- **w/o CRC** removes the Channel-wise Relation Calibration (CRC) module, directly feeding the temporally aligned representations into the intermediate refinement layer, i.e., $\pi^c = \pi^{align}, x^c = x^{align}$.
- **w/o DPE** eliminates the Dynamic-guided Prior Enhancement (DPE) module, bypassing the cross-channel enhancement step entirely, i.e., $\pi^e = \pi^r$.
- **w/o AF** removes the Adaptive Fusion (AF) module and instead merges the enhanced periodic prior and refined dynamic representation via direct addition, i.e., $h = \pi^e + x^r$.
- **R-C1** modifies the CRC calibration step by using direct subtraction rather than complementary weighting, i.e., $x_i^c = x_i - \pi_i$.
- **R-C2** alters the CRC consistency score by averaging it across the channel dimension, applying a uniform scalar weight to all channels instead of channel-wise calibration, i.e., $s_i = \text{mean}(s_i, \text{dim} = 1)$.
- **R-C3** extracts the channel-wise inconsistency in CRC by taking the maximum relation change rather than the average, i.e., $d_i = \max_{j \neq i}(D_{ij})$.
- **R-C4** removes the learnable scaling and bias hyperparameters (α, β) in the CRC module, directly setting the consistency score as $s_i = \sigma(-d_i)$.
- **R-C5** bypasses the Pearson correlation matrix calculation in CRC, obtaining the consistency score via direct neural mapping of the aligned features, i.e., $s_i = \sigma(\text{MLP}([\pi^{align}; x^{align}]))$.
- **R-C6** derives the consistency score using cosine similarity between the aligned temporal features instead of structural correlation differences, i.e., $s_i = \text{CosineSimilarity}(\pi^{align}, x^{align})$.
- **R-D1** modifies the attention mechanism in the DPE module by reversing the Query and Key roles, utilizing the refined periodic prior π^r as the Query and the dynamic representation x^r as the Key to compute the attention score, i.e., $A = \text{softmax}(Q_{\pi^r}(K_{x^r})^T/\sqrt{d})$.

Table 7: Comparison of Ablation Results. The input length is fixed to 96. The best results are in **bold**.

Model		w/o CRC		w/o DPE		w/o AF		R-C1		R-C2		R-C3		R-C4		R-C5		R-C6		R-D1		COMPASS	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	**0.362**	0.395	0.364	0.394	0.366	**0.393**	0.364	0.395	0.363	0.394	0.364	0.394	0.365	0.396	0.368	0.396	0.363	0.394	0.364	0.394	0.364	0.394
	192	**0.407**	**0.420**	0.411	0.421	0.410	**0.420**	0.415	0.424	0.409	0.421	0.408	0.421	0.410	0.424	0.410	0.424	0.408	0.421	0.410	**0.420**	0.408	**0.420**
	336	0.444	0.440	0.431	**0.435**	0.448	0.436	0.446	0.442	0.435	**0.435**	0.434	**0.435**	0.436	0.439	0.435	0.437	0.434	**0.435**	**0.430**	**0.435**	0.434	**0.435**
	720	0.455	0.463	0.426	0.449	0.420	0.467	0.465	0.466	0.423	0.449	0.422	0.449	0.450	0.460	0.437	0.456	0.419	0.447	0.459	0.466	0.420	0.448
	720	0.455	0.463	0.426	0.449	0.420	0.467	0.465	0.466	0.423	0.449	0.422	0.449	0.450	0.460	0.437	0.456	**0.419**	**0.447**	0.459	0.466	0.420	0.448
	avg	0.417	0.429	0.408	0.425	0.411	0.429	0.422	0.432	0.407	0.425	0.407	0.425	0.415	0.430	0.412	0.428	**0.406**	**0.424**	0.416	0.429	**0.406**	**0.424**
Weather	96	0.158	0.202	**0.153**	**0.198**	0.155	0.200	0.157	0.200	0.157	0.202	0.156	0.199	0.156	0.201	0.157	0.200	0.155	0.199	0.154	0.199	**0.153**	**0.198**
	192	0.204	0.244	**0.202**	0.243	0.205	0.245	0.207	0.245	0.208	0.246	0.204	0.243	0.207	0.245	0.205	0.244	0.203	0.243	**0.202**	**0.242**	**0.202**	**0.242**
	336	0.262	0.287	**0.258**	**0.285**	0.261	0.286	0.261	0.286	0.262	0.286	0.262	0.288	0.262	0.287	0.264	0.286	0.260	**0.285**	0.259	0.286	0.259	**0.285**
	720	0.342	0.341	**0.340**	0.340	0.341	0.340	0.341	0.341	0.342	0.340	0.344	0.342	0.343	0.341	0.343	0.341	0.341	**0.339**	**0.340**	**0.339**	**0.340**	0.340
	avg	0.241	0.268	**0.238**	**0.266**	0.240	0.268	0.241	0.268	0.242	0.268	0.241	0.268	0.242	0.268	0.242	0.268	0.240	**0.266**	**0.239**	0.266	**0.238**	**0.266**
Solar	96	**0.174**	**0.230**	**0.174**	0.234	0.176	0.248	0.183	0.237	**0.174**	0.232	**0.174**	0.231	0.181	0.236	0.180	0.232	**0.174**	0.232	**0.174**	0.232	**0.174**	0.232
	192	0.193	0.243	0.196	0.248	0.198	0.258	**0.192**	**0.242**	0.193	**0.242**	0.193	**0.242**	0.193	0.243	0.199	0.263	0.193	**0.242**	0.193	**0.242**	0.193	**0.242**
	336	0.204	0.253	0.210	0.257	0.204	0.253	0.203	**0.246**	**0.202**	0.249	**0.202**	0.249	0.205	0.246	0.203	0.253	**0.202**	0.249	**0.202**	0.250	**0.202**	0.249
	720	0.206	0.250	0.206	0.250	0.215	0.262	**0.205**	**0.248**	0.206	0.249	0.207	0.249	0.208	0.252	0.206	0.256	0.206	0.249	0.206	0.249	0.206	0.249
	avg	**0.194**	0.244	0.196	0.247	0.198	0.255	0.196	**0.243**	**0.194**	**0.243**	**0.194**	**0.243**	0.197	0.244	0.197	0.251	**0.194**	**0.243**	**0.194**	**0.243**	**0.194**	**0.243**
PEMS08	12	0.058	0.156	0.058	**0.155**	0.058	0.159	**0.057**	**0.155**	**0.057**	0.156	**0.057**	**0.155**	**0.057**	**0.155**	0.073	0.183	**0.057**	**0.155**	**0.057**	**0.155**	**0.057**	**0.155**
	24	0.070	0.171	0.070	0.170	0.073	0.179	0.070	0.171	0.070	0.171	0.070	0.172	0.069	0.170	**0.068**	0.169	**0.068**	**0.168**	0.070	0.171	0.070	0.171
	48	0.094	0.195	0.093	0.195	0.105	0.216	0.094	0.197	0.093	0.195	**0.090**	**0.194**	0.096	0.196	0.093	0.196	0.103	0.204	0.094	0.196	0.093	**0.194**
	96	0.131	0.225	0.124	0.221	0.140	0.240	0.138	0.227	**0.123**	**0.221**	0.153	0.239	0.139	0.231	0.224	0.331	0.142	0.231	0.132	0.226	0.125	0.222
	avg	0.088	0.187	**0.086**	**0.185**	0.094	0.198	0.090	0.187	**0.086**	0.186	0.092	0.190	0.090	0.188	0.114	0.220	0.092	0.189	0.088	0.187	**0.086**	**0.185**
1st Count		5		9		2		**7**		**7**		**8**		**2**		**1**		**12**		**12**		**14**	

The complete COMPASS model achieves 14 best evaluations across the forecasting tasks on the four datasets, consistently maintaining the lowest overall forecasting errors, which verifies the indispensability of each architectural component design.

- **Effectiveness of Core Modules:** Removing Adaptive Fusion (w/o AF) leads to the most significant overall performance degradation, especially in the PEMS08 and Solar datasets where the average MSE increases substantially. This indicates that simple feature addition cannot effectively handle the distribution discrepancies between prior and dynamic features. The absence of Channel-wise Relationship Calibration (w/o CRC) or Dynamic-guided Prior Enhancement (w/o DPE) similarly causes consistent error increases across datasets, proving that feature decoupling and cross-channel enhancement mechanisms are cornerstones for ensuring forecasting accuracy.

- **Precision of the CRC Mechanism:** Degrading channel-wise independent calibration to a globally uniform weight (R-C2) or removing the learnable scaling parameters (R-C4) both result in decreased forecasting performance, confirming the necessity of channel-wise adaptive regulation. Regarding the choice of the consistency scoring function, MLP-based black-box mapping (R-C5) causes a severe performance collapse on the PEMS08 dataset (with the average MSE surging to 0.114), and cosine similarity (R-C6) also underperforms slightly. This demonstrates that explicitly measuring cross-channel structural correlations using the Pearson correlation coefficient can more robustly and accurately evaluate the matching degree between the prior and current input.
- **Rationality of the DPE Attention Direction:** Reversing the Query and Key roles in the attention mechanism (R-D1, i.e., using the historical prior to query current dynamics) brings noticeable performance degradation. This comparison confirms that using the current dynamic features as queries to actively retrieve and aggregate relevant historical periodic prior information is the optimal information flow for achieving efficient feature alignment and enhancement.

C.2 Model Complexity.

The COMPASS Model. The COMPASS model achieves a strong performance-efficiency trade-off on the Electricity (Elec) dataset. It attains the best forecasting accuracy (MSE = 0.200), outperforming TQNet (0.201), S-Mamba, and DUET (both 0.204). With only 1.52M parameters, COMPASS is substantially lighter than iTransformer (5.15M) and S-Mamba (2.0M), while requiring only 17.3s per epoch. Although slightly slower than lightweight models such as CycleNet (11.7s) and GTR (14.3s), it is considerably faster than S-Mamba (39.6s) and FilterTS (32.3s).

Overall, COMPASS achieves competitive computational efficiency while delivering the lowest forecasting error, demonstrating an effective balance between accuracy and resource consumption.

Table 8: Performance Comparison on Elec Dataset

Metric	PatchMLP	iTransformer	S-Mamba	CycleNet	TQNet	FilterTS	DUET	TImeMosaic	FreqCycle	GTR	COMPASS
Params	1.9M	5.15M	2M	419k	944k	1.46M	692.69k	OOM	105k	1M	1.52M
Time(s)	20.2	26.8	39.6	11.7	13.9	32.3	17.9	OOM	12.5	14.3	17.3
MSE	0.242	0.225	0.204	0.212	0.201	0.224	0.204	0.225	0.21	0.208	0.2

Table 9: Robustness across random seeds. For each method, results under each seed are first averaged over four forecasting horizons, and we then report the mean ± standard deviation across four seeds (2024–2027). Error bars denote standard deviation. FSR achieves lower errors with reduced variability on most datasets, indicating strong stability.

Model		CycleNet(2024)		TQNet(2025)		GTR(2026)		COMPASS(ours)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.3756±0.0018	0.3954±0.0013	0.3713±0.0006	0.3933±0.0004	0.3695±0.0010	**0.3920**±0.0005	**0.3640**±0.0019	0.3943±0.0011
	192	0.4369±0.0032	0.4284±0.0024	0.4295±0.0011	0.4241±0.0014	0.4253±0.0016	**0.4231**±0.0011	**0.4098**±0.0027	0.4215±0.0010
	336	0.4965±0.0017	0.4559±0.0036	0.4782±0.0027	0.4489±0.0034	0.4753±0.0076	0.4467±0.0030	**0.4354**±0.0022	**0.4367**±0.0011
	720	0.5181±0.0266	0.4836±0.0157	0.5005±0.0125	0.4799±0.0099	0.5079±0.0173	0.4829±0.0092	**0.4343**±0.0099	**0.4551**±0.0052
	avg	0.4568±0.0062	0.4408±0.0035	0.4449±0.0038	0.4366±0.0032	0.4445±0.0063	0.4362±0.0033	**0.4109**±0.0031	**0.4269**±0.0018
Solar	96	0.1912±0.0083	0.2474±0.0033	**0.1781**±0.0049	0.2399±0.0066	0.1797±0.0036	0.2412±0.0062	0.1800±0.0053	**0.2321**±0.0023
	192	0.2025±0.0021	0.2567±0.0025	0.1957±0.0067	0.2546±0.0086	0.1965±0.0033	0.2467±0.0043	**0.1924**±0.0013	**0.2438**±0.0025
	336	0.2134±0.0017	0.2608±0.0038	0.2049±0.0017	0.2643±0.0028	0.2046±0.0040	0.2533±0.0078	**0.2031**±0.0038	**0.2484**±0.0029
	720	0.2232±0.0026	0.2687±0.0021	0.2119±0.0024	0.2682±0.0019	0.2171±0.0132	0.2621±0.0147	**0.2117**±0.0067	**0.2548**±0.0058
	avg	0.2076±0.0016	0.2584±0.0007	0.1977±0.0027	0.2568±0.0036	0.1995±0.0044	0.2508±0.0040	**0.1968**±0.0024	**0.2448**±0.0018
PEMS04	12	0.0782±0.0002	0.1861±0.0003	0.0665±0.0001	0.1665±0.0004	0.0657±0.0001	0.1645±0.0003	**0.0648**±0.0001	**0.1635**±0.0003
	24	0.0988±0.0001	0.2124±0.0003	0.0772±0.0005	0.1815±0.0005	0.0758±0.0005	0.1775±0.0014	**0.0737**±0.0004	**0.1756**±0.0003
	48	0.1321±0.0003	0.2480±0.0003	0.0964±0.0011	0.2057±0.0009	0.0945±0.0011	0.2006±0.0019	**0.0900**±0.0010	**0.1968**±0.0013
	96	0.1679±0.0022	0.2814±0.0025	0.1234±0.0012	0.2333±0.0012	0.1167±0.0009	0.2235±0.0018	**0.1112**±0.0008	**0.2174**±0.0010
	avg	0.1192±0.0005	0.2320±0.0006	0.0909±0.0005	0.1968±0.0003	0.0882±0.0005	0.1917±0.0010	**0.0852**±0.0002	**0.1883**±0.0001

C.3 Robustness Analysis.

Based on the experimental results across multiple random seeds (2024–2027), the **COMPASS** model demonstrates superior forecasting accuracy and strong resilience to random initialization across all evaluation benchmarks. The detailed robustness analysis is as follows:

- **Lowest Overall Forecasting Error:** Across the three datasets (ETTh1, Solar, and PEMS04), COMPASS achieves the lowest average forecasting errors (MSE and MAE), comprehensively outperforming recent baseline models (CycleNet, TQNet, GTR).

- **Significantly Reduced Variability (Low Variance):** The standard deviations of the predictions produced by COMPASS (the values following $\pm$) are generally smaller than those of the baseline models, indicating extremely low sensitivity to random seeds. Particularly on the PEMS04 dataset, COMPASS exhibits highly stable performance, with the standard deviations of its average metrics compressed to an exceptionally low 0.0002 (MSE) and 0.0001 (MAE).
- **Strong Stability in Long-Term Forecasting:** Uncertainty typically amplifies as the forecasting horizon increases. In the more challenging long-term forecasting scenarios (e.g., the 720-step horizon in ETTh1), baseline models show drastic error fluctuations (e.g., the standard deviation of CycleNet's MSE reaches up to ± 0.0266). In contrast, COMPASS successfully restricts this fluctuation to a much narrower range of ± 0.0099 while maintaining the optimal absolute accuracy.

Overall, the extremely low standard deviations combined with consistently leading mean errors demonstrate that COMPASS not only achieves high forecasting accuracy but also exhibits exceptional training consistency. This enables the model to provide highly stable and reliable forecasting outputs in complex real-world scenarios.

C.4 Detailed Analysis on the Impact of d_{model}

As shown in Table 10, COMPASS consistently achieves lower MSE and MAE than other methods across the three datasets. Its performance also remains more stable across different d_{model} sizes, demonstrating robustness to the choice of hidden dimension.

Table 10: Performance comparison of different models across three datasets with varying d_{model} sizes. Best results are highlighted in **bold**. Results on PeMS03 are averaged over prediction lengths $12, 24, 48, 96$, while those on the other datasets are averaged over $96, 192, 336, 720$.

Dataset	Model	MSE (d_{model})					MAE (d_{model})				
		64	128	256	512	1024	64	128	256	512	1024
ETTh1	COMPASS	**0.413**	**0.414**	**0.406**	**0.414**	**0.423**	**0.428**	**0.428**	**0.424**	**0.428**	**0.436**
	GTR	0.445	0.437	0.444	0.449	0.453	0.442	0.434	0.437	0.439	0.442
	TQNet	0.450	0.445	0.447	0.440	0.455	0.442	0.436	0.436	0.434	0.443
	CycleNet	0.452	0.456	0.445	0.454	0.451	0.438	0.439	0.437	0.438	0.439
Weather	COMPASS	**0.240**	**0.239**	**0.238**	**0.239**	**0.240**	**0.267**	**0.266**	**0.266**	**0.267**	**0.268**
	GTR	0.245	0.243	0.241	0.241	**0.240**	0.272	0.270	0.269	0.269	**0.268**
	TQNet	0.247	0.245	0.243	0.242	0.243	0.273	0.271	0.269	0.268	0.269
	CycleNet	0.247	0.245	0.245	0.243	0.242	0.273	0.272	0.273	0.271	0.270
PeMS03	COMPASS	**0.090**	**0.087**	**0.086**	**0.086**	**0.086**	**0.193**	**0.188**	**0.187**	**0.185**	**0.185**
	GTR	0.097	0.094	0.090	0.088	0.087	0.202	0.198	0.194	0.191	0.189
	TQNet	0.106	0.101	0.098	0.097	0.098	0.212	0.206	0.203	0.203	0.202
	CycleNet	0.126	0.122	0.120	0.118	0.117	0.233	0.229	0.227	0.225	0.225

D Theoretical Analysis of Channel-wise Calibration

We provide a theoretical analysis to motivate channel-wise prior adaptation in COMPASS. The analysis focuses on two questions: why different channels require different contributions from the periodic prior, and why the channel-wise relationship mismatch can serve as a signal for adjusting such contributions.

Channel-specific optimal fusion. Consider channel i and let z_i denote its underlying current state. We model the periodic prior and current observation as two noisy estimates of z_i:

$$\pi_i = z_i + \epsilon_i^{\pi}, \qquad x_i = z_i + \epsilon_i^{x}, \tag{16}$$

where the estimation errors are assumed to be zero-mean and uncorrelated, with variances

$$\mathbb{E}\left[(\epsilon_i^{\pi})^2\right] = \sigma_{\pi,i}^2, \qquad \mathbb{E}\left[(\epsilon_i^{x})^2\right] = \sigma_{x,i}^2. \tag{17}$$

Consider a weighted fusion of the two representations,

$$h_i = s_i \pi_i + (1 - s_i) x_i. \tag{18}$$

Its expected squared error with respect to z_i is

$$\begin{aligned} \mathcal{L}_i(s_i) &= \mathbb{E}\left[(h_i - z_i)^2\right] \\ &= s_i^2 \sigma_{\pi,i}^2 + (1 - s_i)^2 \sigma_{x,i}^2. \end{aligned} \tag{19}$$

Minimizing $\mathcal{L}_i(s_i)$ with respect to s_i yields the channel-specific optimal weight

$$s_i^* = \frac{\sigma_{x,i}^2}{\sigma_{\pi,i}^2 + \sigma_{x,i}^2}. \tag{20}$$

Therefore, the optimal contribution of the periodic prior depends on the relative uncertainty of the prior and current dynamics for each channel. Since these uncertainties can vary across channels, generally

$$s_i^* \neq s_j^*, \qquad i \neq j. \tag{21}$$

In contrast, a uniform fusion strategy uses a shared weight λ for all channels,

$$h_i^{\text{uni}} = \lambda \pi_i + (1-\lambda) x_i, \qquad \forall i. \tag{22}$$

Its minimum average error therefore satisfies

$$\min_{\lambda} \frac{1}{C} \sum_{i=1}^{C} \mathcal{L}_i(\lambda) \geq \frac{1}{C} \sum_{i=1}^{C} \min_{s_i} \mathcal{L}_i(s_i), \tag{23}$$

with strict inequality when the channel-specific optimal weights are not identical. This formally shows that channel heterogeneity can make uniform prior utilization suboptimal.

Relationship mismatch as a calibration signal. CRC measures the channel-wise mismatch between the periodic prior and current dynamics as

$$d_i = \frac{1}{C-1} \sum_{j \neq i} \left| R_{\pi,ij} - R_{x,ij} \right|, \tag{24}$$

where R_π and R_x denote the channel-wise Pearson relationship matrices of the periodic prior and current input, respectively. We assume that a larger structural mismatch between the historical periodic prior and current dynamics indicates greater uncertainty of the periodic prior for the current window. Under this assumption,

$$d_i \uparrow \quad \Rightarrow \quad \sigma_{\pi,i}^2 \uparrow. \tag{25}$$

From the optimal weight in Eq. (20), it follows that

$$d_i \uparrow \quad \Rightarrow \quad s_i^* \downarrow. \tag{26}$$

This agrees with the CRC design:

$$s_i = \sigma(\beta - \alpha d_i), \qquad \alpha > 0, \tag{27}$$

whose sensitivity to the mismatch is

$$\frac{\partial s_i}{\partial d_i} = -\alpha s_i (1 - s_i) < 0. \tag{28}$$

Hence, CRC implements the desired monotonic behavior: channels with larger prior–dynamics mismatch receive less contribution from the periodic prior, whereas more consistent channels retain more historical periodic information.

Overall, the analysis establishes two properties of channel-wise prior adaptation: (i) heterogeneous channels generally require different prior weights, making uniform fusion suboptimal; and (ii) the channel-wise relationship mismatch provides a principled signal for adjusting these weights, with larger mismatch corresponding to a smaller contribution from the periodic prior.

E More Visualization

E.1 CRC Analysis.

To further examine the effect of Channel-wise Relation Calibration (CRC), we visualize the Pearson correlation between the original periodic prior π and the input x before and after calibration on representative datasets, including ETTm1, PEMS04, PEMS08, and Electricity. As shown in Fig. 8, CRC consistently reduces the correlation between π and x, indicating that the calibrated prior becomes more decoupled from the input representation. This observation provides further evidence that CRC effectively reduces the redundancy between the periodic prior and the current input dynamics.

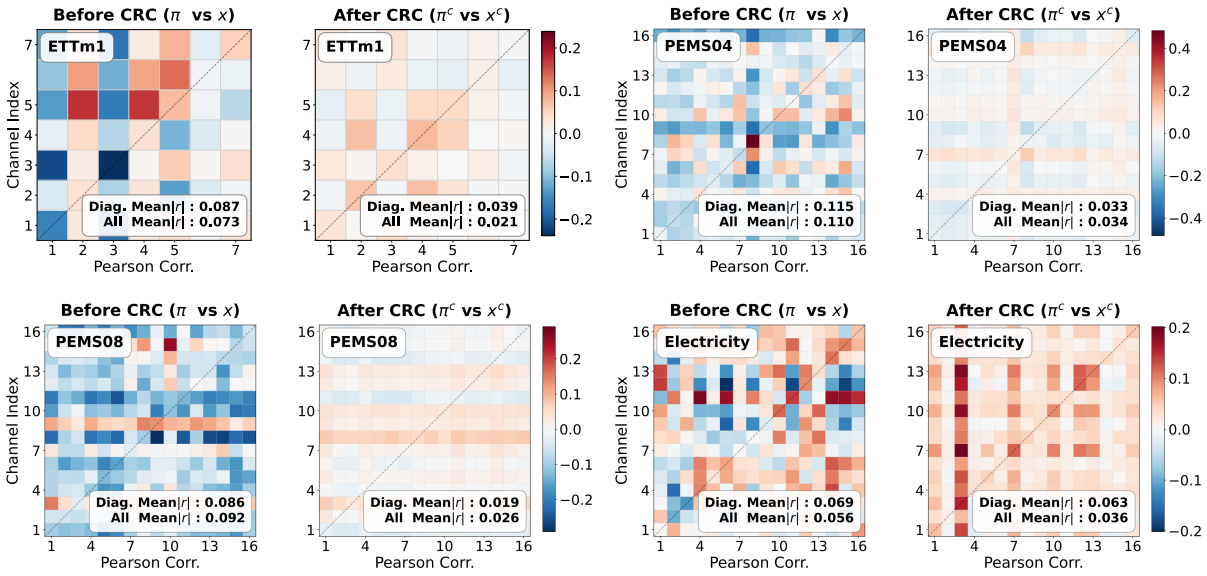

Figure 8: **Pearson Correlation Before and After CRC Calibration.** Comparison of the correlation between the original periodic prior π and input x before and after CRC calibration on ETTm1, PEMS04, PEMS08, and Electricity. CRC reduces the mean absolute Pearson correlation (Mean $|r|$), indicating improved decoupling between the calibrated prior and input.

E.2 DPE Analysis.

We further visualize the effect of Dynamic-guided Prior Enhancement (DPE) on the calibrated prior. As shown in Fig. 9, we compare the Pearson correlation matrices of the calibrated prior π^r, enhanced prior π^e, and ground truth Y on ETTm1, PEMS04, PEMS08, and Electricity. Across these datasets, the correlation contours of π^e become more consistent with those of Y after DPE, providing further evidence that DPE incorporates dynamic information into the calibrated prior and improves its alignment with the underlying target.

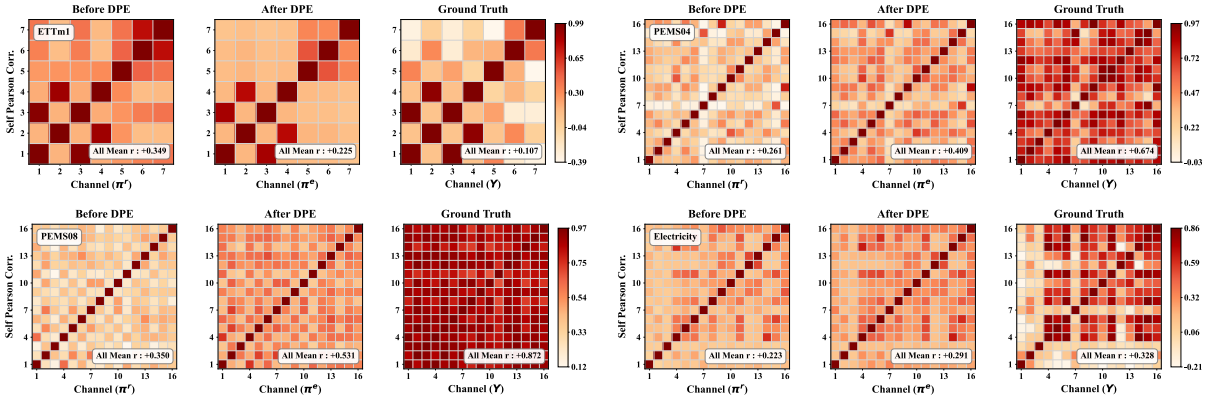

Figure 9: **Visualization of DPE.** Pearson correlation matrices of the calibrated prior π^r, enhanced prior π^e, and ground truth Y on ETTm1, PEMS04, PEMS08, and Electricity. The enhanced prior exhibits correlation contours that are more consistent with those of the ground truth after DPE.

E.3 Forecasting Visualization.

We provide additional visualizations of the forecasting results on ETTh1, Electricity, PEMS03, and PEMS04. As shown in Figs. 10–13, periodic methods such as CycleNet and GTR exhibit noticeable prediction lag, whereas COMPASS more closely follows the ground-truth sequences across all four datasets. This observation is consistent with the design of COMPASS, where CRC reduces redundancy between the periodic prior and current input, while DPE incorporates dynamic information to refine the calibrated prior.

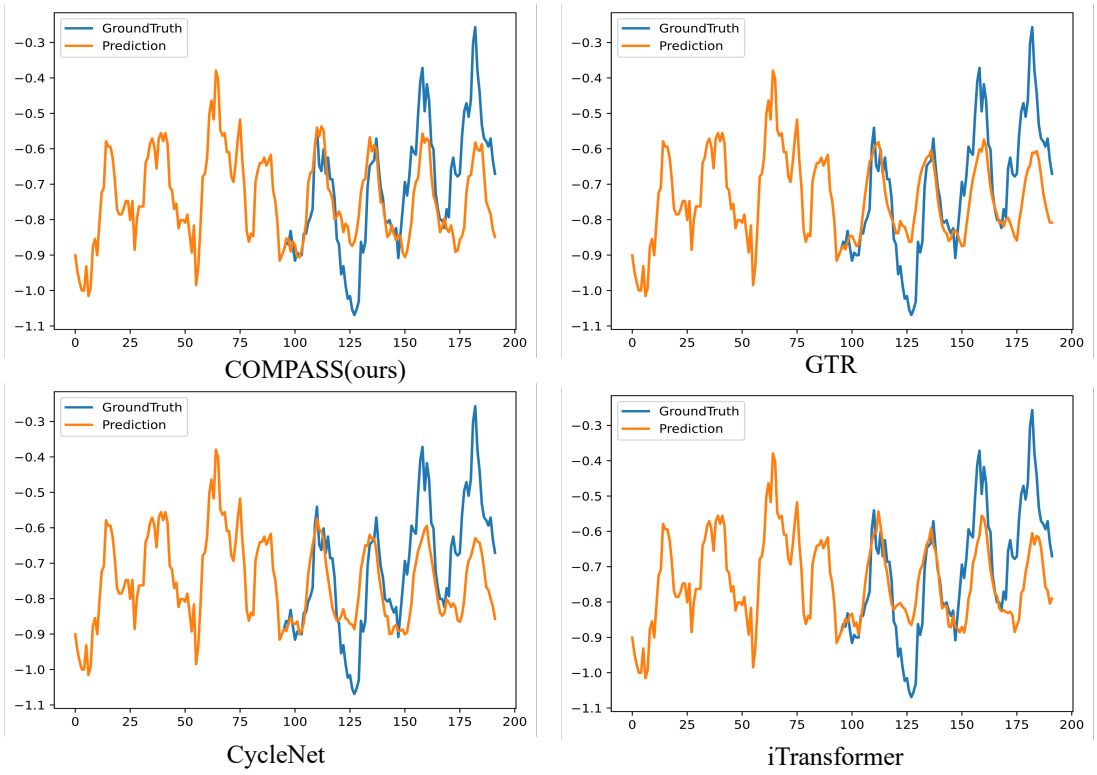

Figure 10: Visualization with input length 96 and forecasting length 96 on the ETTh1 dataset.

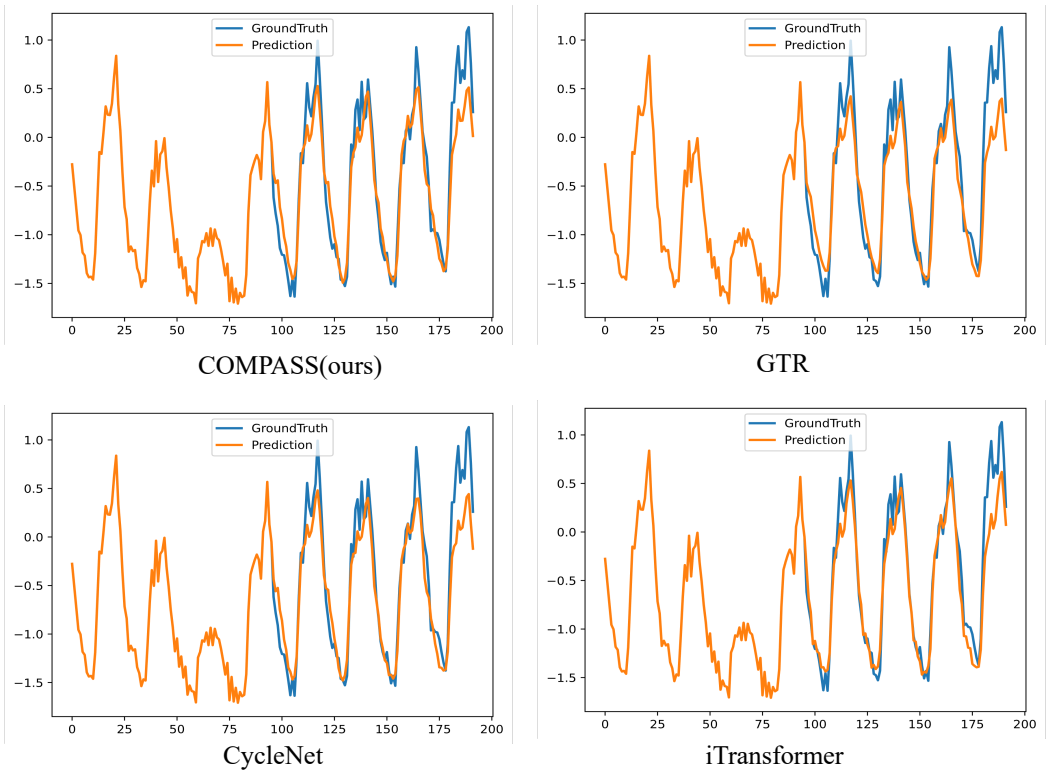

Figure 11: Visualization with input length 96 and forecasting length 96 on the Electricity dataset.

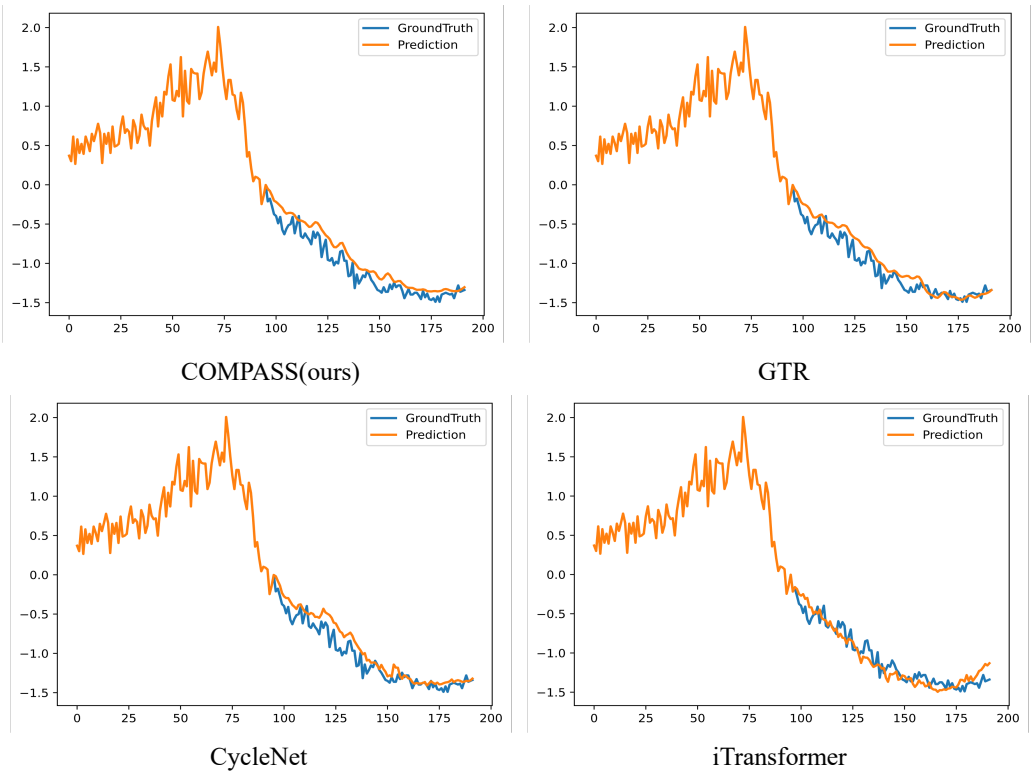

Figure 12: Visualization with input length 96 and forecasting length 96 on the PEMS03 dataset.

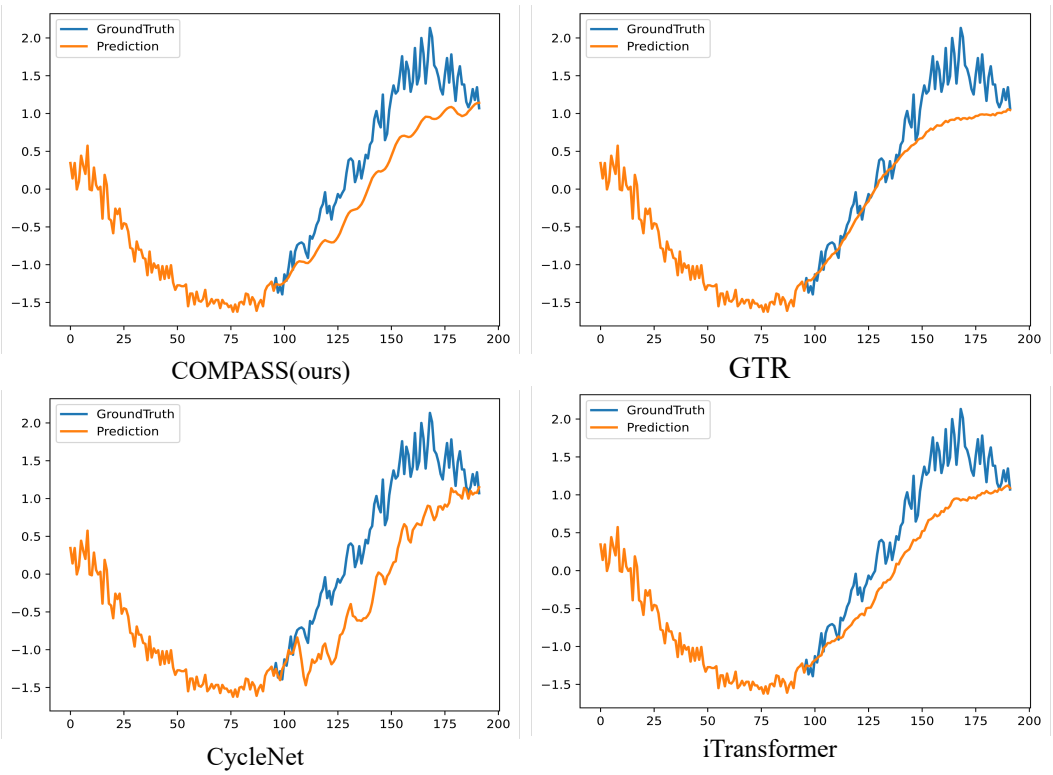

Figure 13: Visualization with input length 96 and forecasting length 96 on the PEMS04 dataset.