SNOW: A STABLE SIGNED CHANNEL INTERACTION MODULE FOR HIGH-DIMENSIONAL TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Time series forecasting is crucial across fields such as economics, energy, and traffic. Modeling cross-channel dependencies in high-dimensional time series remains challenging. Current methods are commonly divided into channel-independent (CI) and channel-dependent (CD) models. CI models process each channel independently, which contributes to their robustness. However, they ignore cross-channel dependencies and become less effective as the number of channels increases. CD models explicitly model interactions among channels, yet they may struggle to represent negative cross-channel relationships, produce inconsistent predictions when the channel order is permuted, and amplify perturbations during cross-channel propagation. These limitations motivate us to explore an alternative path: instead of designing complex CD models, we equip CI models with cross-channel capabilities. Based on this motivation, we propose Snow, a plug-and-play channel interaction module. Snow adds cross-channel communication through Sample-adaptive Cluster Representations. It enables signed and low-rank channel mixing to stabilize cross-channel propagation. Snow satisfies three properties: signed-interaction compatibility, channel-permutation equivariance, and interference stability. We provide theoretical analysis of these properties and validate them empirically. Extensive experiments show that inserting Snow into CI models consistently improves forecasting performance on high-dimensional datasets, while replacing interaction modules in CD models also yields consistent gains. **Code is available at:** `https://anonymous.4open.science/r/snow-iclr/README.md`

1 INTRODUCTION

Time series forecasting supports a wide range of applications, including energy management, traffic monitoring, and financial analysis (Lim & Zohren, 2021). As data-collection systems continue to expand, many real-world time series now contain hundreds or even thousands of variables. Figure 1(a) illustrates this rapid growth in channel dimension. Forecasting in this setting requires models to capture informative relationships among many channels. At the same time, the models must remain robust to noisy, irrelevant, or corrupted interactions. Effectively controlling cross-channel information exchange has therefore become a central challenge in high-dimensional time series forecasting (HTSF) (Qiu et al., 2026).

Existing forecasting models generally follow either a channel-independent (CI) or channel-dependent (CD) paradigm. CI models process each channel separately. By avoiding explicit cross-channel interactions, they reduce the influence of unreliable channel relationships. They also provide an efficient and robust modeling path (Han et al., 2024b). This design has proven effective in models such as DLinear (Zeng et al., 2023) and PatchTST (Nie et al., 2023). However, the independence assumption prevents these models from directly exploiting informative relationships among channels. Figure 1(b) shows that the relative advantage of CI models tends to diminish as the number of channels increases. This trend suggests that channel independence becomes increasingly restrictive in high-dimensional settings. CD models address this limitation by explicitly mixing information across channels. They use feature-mixing layers, attention mechanisms, or learned dependency

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

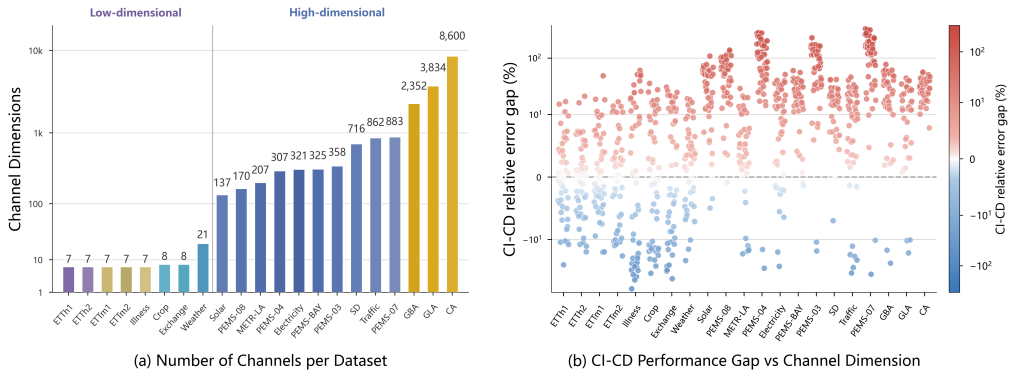


Figure 1: Number of channels across datasets and the CI–CD performance gap. (a) The number of channels ranges from 7 to 8,600 across representative datasets. (b) Pairwise comparison between seven channel-independent (CI) and seven channel-dependent (CD) models. Each point represents the relative error gap of one CI–CD model pair. Red points indicate lower error from the CD model, whereas blue points indicate lower error from the CI model. The relative advantage of CI models gradually diminishes as the number of channels increases.

structures (Chen et al., 2023; Zhang & Yan, 2023; Liu et al., 2023; Qiu et al., 2025; Li et al., 2026b; Liu et al., 2024).

The central question is not simply whether channels should interact. We must also consider how cross-channel information should be propagated. Many existing mechanisms focus on increasing dependency-modeling capacity. Their structural behavior receives less attention. As summarized in Figure 2, three challenges are particularly important in high-dimensional forecasting:

(1) Nonnegative interaction weights cannot explicitly represent negative or inhibitory relationships between channels.

(2) Interaction parameters associated with fixed channel indices may introduce sensitivity to channel ordering. However, channel order should not change the corresponding predictions.

(3) Unconstrained cross-channel propagation may spread local perturbations to other channels. It may also amplify their influence on the learned representations.

These observations motivate a controlled form of channel interaction. A forecasting model need not choose between complete channel independence and unrestricted channel dependence. It should retain the reliable channel-wise temporal modeling path of CI methods. It should also introduce compact and structurally constrained cross-channel communication. Such an interaction mechanism should preserve the sign of channel relationships. It should respond to the current input rather than fixed channel identities and regulate information propagation across channels.

Based on this principle, we propose Snow, a plug-and-play channel interaction module for high-dimensional time series forecasting. Snow routes channel representations through a small set of sample-adaptive Cluster Representations. It aggregates information in this compact space and then redistributes the context to the associated channels. Direction Softmax preserves both the magnitude and sign of the routing coefficients. Cluster-mediated aggregation induces a compact, low-rank channel propagation operator. The routing and output mappings are shared across channels. Snow therefore responds to channel content rather than fixed channel positions. It can be inserted into CI models or used to replace interaction modules in CD models.

Snow satisfies three structural properties. Signed-interaction compatibility allows the propagation operator to realize negative direct cross-channel coefficients. Channel-permutation equivariance ensures that reordering input channels correspondingly reorders the outputs. Interference stability ensures that perturbations are strictly attenuated during cross-channel propagation. Together, these properties characterize Snow as a signed, order-consistent, and stable channel interaction mechanism. Our contributions are summarized as follows:

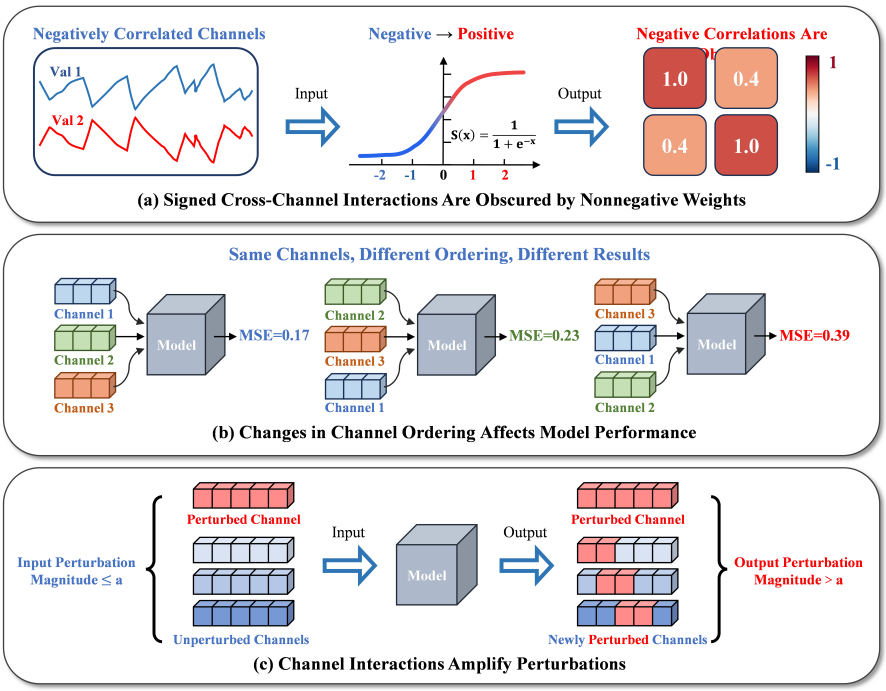


Figure 2: Three structural challenges in cross-channel interaction. (a) Signed cross-channel interactions are obscured by nonnegative weights. (b) Channel-specific designs may introduce dependence on channel ordering. (c) Unconstrained propagation may spread and amplify local perturbations across channels.

- We propose Snow, a plug-and-play module for sample-adaptive, signed, and low-rank channel interaction. It can augment CI models or replace the interaction modules of CD models.
- We establish that Snow satisfies signed-interaction compatibility, channel-permutation equivariance, and interference stability. These properties provide a unified characterization of its cross-channel propagation.
- We evaluate Snow with multiple CI and CD forecasting backbones across high- and low-dimensional datasets. The results show broad forecasting improvements and stable behavior under channel permutation, low-quality data, and different input lengths.

The remainder of this paper is organized as follows. Section 2 presents the problem formulation, Snow workflow, and backbone integration; Section 3 analyzes its three structural properties; Section 4 reports the empirical evaluation; and Section 5 concludes. The appendix provides related work, complete proofs, detailed experimental settings, and extended results.

2 Snow: Stable Cluster-Mediated Channel Interaction

2.1 Problem Formulation

Let $X \in \mathbb{R}^{L\times N}$ denote a historical window of a time series, where L is the input length and N is the number of channels. Let $Y \in \mathbb{R}^{T\times N}$ denote the corresponding future sequence over a prediction horizon of length T. Given X, the forecasting objective is to learn a model f_θ that produces

$$\widehat{Y} = f_\theta(X),$$

where $\widehat{Y}$ approximates the ground-truth target Y. This formulation covers both channel-independent (CI) and channel-dependent (CD) forecasting models. Snow operates on the channel-wise hidden representations produced by the backbone and introduces cross-channel interaction while preserving the correspondence between input and output channels.

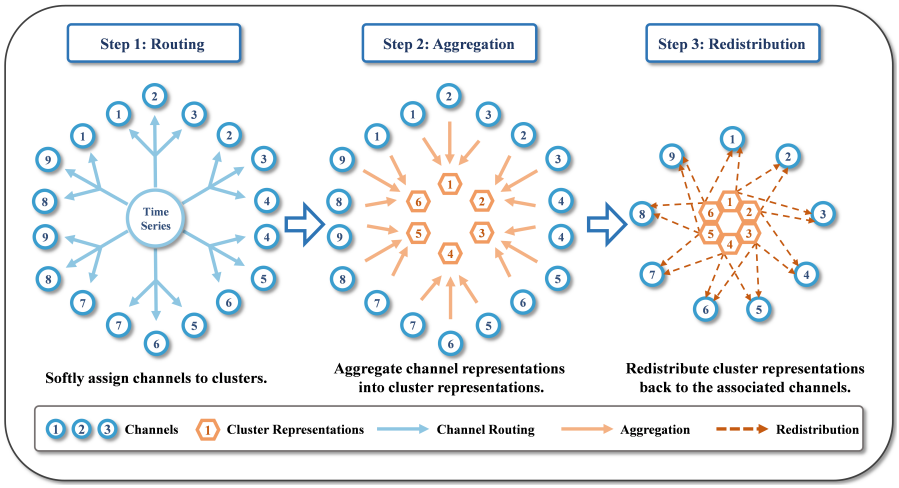


Figure 3: Overview of Snow's cluster-mediated channel interaction. Routing softly assigns channel representations to adaptive clusters, Aggregation forms Cluster Representations, and Redistribution returns the cluster context to the associated channels.

2.2 Overview of Snow

Snow is a plug-and-play module for controlled cross-channel interaction. After channel-wise temporal encoding, it receives the hidden representation $H \in \mathbb{R}^{N \times d_h}$, where each row corresponds to one channel, and routes the channels through a small set of sample-adaptive Cluster Representations. This compact bottleneck avoids unrestricted all-to-all mixing, preserves access to channel-specific information, and induces low-rank cross-channel interaction. Figure 3 summarizes the resulting workflow. The module can be integrated into channel-independent (CI) backbones to introduce cross-channel interaction or replace existing interaction modules in channel-dependent (CD) models. Section 2.3 formalizes the three stages, while Section 2.4 describes its integration with CI and CD forecasting models.

2.3 Workflow of Snow

2.3.1 Step 1: Routing

In the first stage, each channel representation is softly routed to K clusters. Snow first computes a spectral signature for each channel representation:

$$A(H) = |\mathrm{rFFT}(H, \mathrm{dim} = -1)| \in \mathbb{R}^{N \times F}, \qquad F = \left\lfloor \frac{d_h}{2} \right\rfloor + 1. \tag{1}$$

Here, rFFT is applied along the hidden feature dimension d_h to extract the spectral characteristics of each channel representation.

A shared routing network ϕ_r maps the spectral signature to routing logits:

$$Z(H) = \frac{\phi_r(A(H))}{\tau} \in \mathbb{R}^{N \times K}, \tag{2}$$

where $\tau > 0$ is a temperature parameter. Snow then applies **Direction Softmax** along the cluster dimension:

$$S(H) = \mathrm{DSoftmax}_K(Z(H)) := \mathrm{sign}(Z(H)) \odot \mathrm{softmax}(|Z(H)|, \mathrm{dim} = K) \in \mathbb{R}^{N \times K}. \tag{3}$$

Here, $|S_{ik}(H)|$ determines the routing strength from channel i to cluster k, while its sign determines the polarity of the contribution. A channel may participate in multiple clusters, and channels routed to the same cluster may contribute with opposite signs. Direction Softmax gives each row of $S(H)$ an absolute routing mass no larger than one; consequently, the total mass $m_k(H)$ received by any cluster is at most N.

2.3.2 STEP 2: AGGREGATION

In the second stage, Snow aggregates channel information into Cluster Representations. A shared value network first maps each channel representation into the Cluster Representation space:

$$V(H) = \phi_v(H) \in \mathbb{R}^{N \times d_v}, \tag{4}$$

where d_v denotes the value/cluster feature dimension. For each cluster k, we first define its total routing strength as

$$m_k(H) = \sum_{i=1}^{N} |S_{ik}(H)|, \tag{5}$$

and collect the normalization terms into

$$D(H) = \mathrm{diag}\big(m_1(H) + \varepsilon, \ldots, m_K(H) + \varepsilon\big) \in \mathbb{R}^{K \times K}, \tag{6}$$

where $\varepsilon > 0$ stabilizes the normalization when a cluster receives little routing mass. Snow then performs signed aggregation:

$$C(H) = D(H)^{-1} S(H)^\top V(H) \in \mathbb{R}^{K \times d_v}. \tag{7}$$

Its k-th row is

$$C_k(H) = \frac{\sum_{i=1}^{N} S_{ik}(H) V_i(H)}{m_k(H) + \varepsilon}. \tag{8}$$

Normalizing by the sum of routing magnitudes preserves signed contributions while controlling the scale of each Cluster Representation.

2.3.3 STEP 3: REDISTRIBUTION

In the third stage, Snow redistributes the Cluster Representation context back to individual channels using the same signed routing matrix:

$$G(H) = S(H) C(H) \in \mathbb{R}^{N \times d_v}. \tag{9}$$

Each channel thereby receives context aggregated from the clusters selected by its signed routing weights.

Combining aggregation and redistribution yields

$$G(H) = S(H) D(H)^{-1} S(H)^\top V(H) = P(H) V(H), \tag{10}$$

where

$$P(H) := S(H) D(H)^{-1} S(H)^\top \in \mathbb{R}^{N \times N} \tag{11}$$

is the sample-adaptive channel propagation matrix, whose entry $P_{ij}(H)$ is the direct propagation coefficient from source channel j to target channel i. Because this propagation factors through K Cluster Representations, its rank satisfies

$$\mathrm{rank}(P(H)) \le K. \tag{12}$$

Thus, for fixed routing, the aggregation and redistribution operations require $O(NKd_v)$ computation, instead of the $O(N^2 d_v)$ cost of direct dense channel mixing.

Finally, Snow fuses each channel representation with its redistributed context through a shared output network:

$$\mathrm{Snow}(H) = \phi_o([H \| G(H)]) \in \mathbb{R}^{N \times d_h}, \tag{13}$$

where $[\cdot\|\cdot]$ denotes feature concatenation. The shared output mapping preserves the channel-wise organization and restores the hidden dimension d_h.

2.4 PLUG-AND-PLAY INTEGRATION

Snow supports two integration modes: augmenting channel-independent backbones with cross-channel interaction or replacing the interaction module in channel-dependent backbones. We use PatchTST and iTransformer as representative examples (Nie et al., 2023; Liu et al., 2024), with implementation details provided in Appendix C.1 and Appendix C.1.2.

3 THEORETICAL ANALYSIS

Snow module satisfies three structural properties: signed-interaction compatibility, channel-permutation equivariance, and interference stability. We introduce these properties below; complete derivations are provided in Appendices B.1–B.4.

3.1 THEORETICAL PRELIMINARIES

The complete forward process is defined in Section 2. The three properties are analyzed through the cross-channel propagation induced by aggregation and redistribution:

$$G(H) = P(H)V(H), \tag{14}$$

where the core propagation operator is

$$P(H) = S(H)D(H)^{-1}S(H)^{\top}. \tag{15}$$

Here, $S(H)$, $D(H)$, and $V(H)$ denote the signed routing matrix, the normalization matrix formed from the sums of absolute routing weights, and the value representations, respectively. For the strict contraction bound, define

$$\rho(H) = \max_{1\le k\le K} \frac{m_k(H)}{m_k(H)+\varepsilon} < 1, \tag{16}$$

where $m_k(H)$ and ε follow the definitions in Section 2.3.2. The permutation transport law, weighted Gram structure, and spectral bound of $P(H)$ are established in Appendix B.1. The row-wise absolute routing mass is at most one, so $m_k(H) \le N$ for every cluster and $\rho(H) \le N/(N+\varepsilon) < 1$. Thus $\rho(H)$ is the largest cluster-wise propagation ratio that controls the attenuation bound used below.

3.2 SIGNED-INTERACTION COMPATIBILITY

Time series can contain opposite-sign or suppressive interactions between channels. Snow therefore allows a channel representation to contribute to another channel with either sign.

Definition 1 (Signed-Interaction Compatibility). Consider a family of channel propagation matrices $P^{(\theta)}(H)$. It is signed-interaction compatible if, for every pair of distinct channels $i \neq j$, there exist valid parameters θ and an input H such that

$$P_{ij}^{(\theta)}(H) < 0, \tag{17}$$

where $P_{ij}^{(\theta)}(H)$ is the direct propagation coefficient from source channel j to the update of target channel i.

Theorem 1 (Snow Satisfies Signed-Interaction Compatibility). For a fixed valid parameter setting, we suppress the superscript θ in the following expression:

$$P_{ij}(H) = \sum_{k=1}^{K} \frac{S_{ik}(H)S_{jk}(H)}{m_k(H)+\varepsilon}. \tag{18}$$

Thus, for every $i \neq j$,

$$P_{ij}(H) < 0 \quad \Longleftrightarrow \quad \sum_{k:S_{ik}S_{jk}<0} \frac{|S_{ik}S_{jk}|}{m_k+\varepsilon} > \sum_{k:S_{ik}S_{jk}>0} \frac{|S_{ik}S_{jk}|}{m_k+\varepsilon}. \tag{19}$$

Moreover,

$$\boxed{\forall i \neq j, \quad \exists(\theta, H) : P_{ij}^{(\theta)}(H) < 0}. \tag{20}$$

Proof sketch. The contribution of cluster k to P_{ij} is positive when the two routing coefficients have the same sign and negative when their signs differ. The total coefficient is negative exactly when the weighted magnitude of the opposite-sign contributions exceeds that of the same-sign contributions. Direction Softmax preserves routing signs, and its routing network can construct opposite-signed logits for any specified channel pair. Appendix B.2 gives the complete construction and verifies its compatibility with the weighted Gram structure.

3.3 Channel-Permutation Equivariance

The order in which channels are presented is only an organization of the input. Reordering the channels should therefore reorder the corresponding outputs without changing the prediction associated with any channel.

Definition 2 (Channel-Permutation Equivariance). For an arbitrary $N \times N$ permutation matrix Q, a mapping F is channel-permutation equivariant if

$$F(QH) = QF(H). \tag{21}$$

Theorem 2 (Snow Satisfies Channel-Permutation Equivariance). If the routing, value, and output mappings use shared channel-wise parameters and rFFT acts only along the within-channel feature axis, then

$$\begin{aligned} S(QH) &= QS(H), \qquad P(QH) = QP(H)Q^\top, \\ G(QH) &= QG(H), \qquad \mathrm{Snow}(QH) = Q\,\mathrm{Snow}(H). \end{aligned} \tag{22}$$

Proof sketch. Shared channel-wise mappings and the within-channel rFFT commute with row permutations, yielding $S(QH) = QS(H)$ and $V(QH) = QV(H)$. The absolute routing masses are unchanged, so the normalization matrix is unchanged. Substituting these relations into the aggregation and redistribution equations gives the transformation law for P and G. Feature concatenation and the shared output mapping then preserve the same row permutation. The complete algebraic chain and the invariance of the propagation spectrum are given in Appendix B.3.

3.4 Interference Stability

Suppose that the value representations are perturbed by $E \in \mathbb{R}^{N \times d_v}$. For the propagation operator induced by input H, the difference between the perturbed and unperturbed propagated contexts is

$$P(H)\big(V(H) + E\big) - P(H)V(H) = P(H)E. \tag{23}$$

Definition 3 (Interference Stability). Snow satisfies interference stability if every nonzero value perturbation satisfies

$$\|P(H)E\|_F < \|E\|_F. \tag{24}$$

Theorem 3 (Snow Satisfies Interference Stability). For every nonzero value perturbation E,

$$\boxed{\|P(H)E\|_F \le \rho(H)\|E\|_F < \|E\|_F}. \tag{25}$$

The corresponding energy decrease satisfies

$$\boxed{\|E\|_F^2 - \|P(H)E\|_F^2 \ge \big(1 - \rho(H)^2\big)\|E\|_F^2 > 0}. \tag{26}$$

Proof sketch. Cluster k receives an absolute routing mass $m_k(H)$ and normalizes it by $m_k(H)+\varepsilon$. Its propagation ratio is therefore at most $m_k(H)/(m_k(H)+\varepsilon)$, and the largest such ratio is $\rho(H) < 1$. The spectral bound in Appendix B.1 then gives strict Frobenius-norm contraction for every nonzero perturbation. Appendix B.4 further derives the modal and repeated-propagation behavior and provides a finite counterexample for standard row-normalized attention.

4 EXPERIMENTS

The goal of this section is to evaluate Snow across multiple backbone families and dataset regimes through the following research questions (RQs). The main text presents the primary comparisons, while detailed configurations and complementary evaluations are provided in Appendix C.1. We organize the evaluation around the following RQs:

- **RQ1:** Can Snow improve channel-independent (CI) models on high-dimensional time series?
- **RQ2:** Can Snow replace cross-channel interaction modules in channel-dependent (CD) models?
- **RQ3:** Which components contribute to the effectiveness of Snow?
- **RQ4:** Does empirical evidence show that Snow satisfies signed-interaction compatibility, channel-permutation equivariance, and interference stability?
- **RQ5:** Can Snow adapt to low-dimensional time series?
- **RQ6:** Is the improvement of Snow stable across different input lengths?
- **RQ7:** How does Snow compare with closely related enhancement strategies?

4.1 EXPERIMENTAL SETUP

We evaluate Snow on 8 CI and 4 CD backbones across 12 high-dimensional and 8 low-dimensional datasets using matched Origin/+Snow pairs under standard benchmark and large-scale traffic protocols (Qiu et al., 2024; Fang et al., 2025). Standard runs use 96-step inputs; dataset splits, traffic horizons, training settings, and random seeds are detailed in Appendix C.1. We report MSE/MAE, and RMSE/MAPE for the four high-dimensional traffic datasets.

4.2 MAIN RESULTS

RQ1 evaluates whether Snow improves CI models while retaining their channel-wise modeling pathway. Tables 1 and 2 show that Snow improves 62 of 64 paired metric comparisons in the 6-to-6 high-dimensional traffic setting and 126 of 128 comparisons under the standard four-horizon protocol; the exceptions are two traffic MAPE comparisons and both metrics for SegRNN on Electricity. The gains are especially pronounced on the PEMS traffic datasets and remain positive for nearly all backbones on Solar, METR-LA, and Electricity, demonstrating that Snow complements diverse CI architectures across high-dimensional temporal patterns.

Appendix C.3–C.8 provides the remaining answers: Snow improves most matched CD comparisons (RQ2), both Direction Softmax and Sample-adaptive Cluster Representations contribute to the gains (RQ3), and the three structural properties are tested through interaction, permutation, corruption, and low-quality evaluations (RQ4). Snow also improves most low-dimensional comparisons (RQ5) and remains effective with longer inputs (RQ6). Against related enhancement strategies, Snow is best in 20 of 24 metric entries (RQ7); hyperparameter sensitivity and efficiency are reported in Appendix C.1.

5 CONCLUSION

In high-dimensional time-series forecasting, channel-independent (CI) models struggle to capture channel dependencies, while channel-dependent (CD) models remain limited in expressing negative correlations, maintaining stability under channel permutations, and resisting interference. To address these issues, we propose Snow, a plug-and-play module that realizes signed channel interaction in a low-rank space through sample-adaptive Cluster Representations. Snow preserves the channel-wise temporal modeling capability of CI models while introducing the cross-channel expressiveness

Table 1: CI traffic results under the 6-to-6 setting (RMSE / MAPE (%)); lower is better.

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)	
CA	52.18 / 21.73	**46.11 / 18.25**	54.05 / 21.06	**43.15 / 17.60**	51.12 / **20.04**	**43.16** / 20.22	52.37 / 20.66	**44.82 / 17.04**
GBA	49.74 / 23.24	**47.62 / 21.79**	52.72 / 23.65	**46.69 / 21.67**	52.74 / 23.58	**46.81 / 20.17**	50.92 / 22.99	**47.39 / 21.02**
GLA	51.97 / 17.16	**45.22 / 14.92**	55.86 / 21.31	**45.84 / 15.94**	52.14 / 22.98	**46.15 / 18.98**	54.13 / 17.34	**47.11 / 15.61**
SD	56.87 / 23.13	**46.91 / 17.65**	59.52 / 21.71	**49.50 / 17.77**	55.59 / 19.68	**48.34 / 19.44**	55.20 / 20.06	**46.90 / 16.68**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)	
CA	54.07 / 21.09	**45.44 / 17.86**	50.08 / 21.74	**38.64 / 18.58**	72.78 / 53.02	**62.60 / 44.16**	50.42 / 20.64	**38.16 / 15.96**
GBA	52.72 / 23.73	**48.06 / 21.79**	49.96 / **22.96**	**42.16** / 23.32	46.10 / 22.93	**40.29 / 19.50**	54.38 / 25.20	**43.80 / 21.48**
GLA	54.74 / 27.77	**49.05 / 16.22**	49.67 / 23.34	**41.75 / 16.94**	81.35 / 41.78	**69.06 / 34.39**	58.21 / 20.76	**46.92 / 16.44**
SD	59.60 / 23.38	**50.49 / 17.76**	55.33 / 21.34	**46.29 / 19.46**	51.30 / 15.12	**42.17 / 12.60**	49.50 / 19.56	**40.56 / 15.60**

Table 2: Four-horizon average CI results on high-dimensional datasets (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
Solar	0.262 / 0.287	**0.227 / 0.257**	0.356 / 0.353	**0.242 / 0.274**	0.369 / 0.368	**0.215 / 0.288**	0.253 / 0.294	**0.252 / 0.289**
PEMS08	0.838 / 0.568	**0.607 / 0.468**	1.061 / 0.688	**0.616 / 0.478**	1.069 / 0.699	**0.542 / 0.441**	0.658 / 0.497	**0.529 / 0.430**
METR-LA	1.444 / 0.740	**1.373 / 0.717**	1.503 / 0.768	**1.401 / 0.734**	1.521 / 0.804	**1.277 / 0.743**	1.401 / 0.758	**1.344 / 0.728**
PEMS04	0.685 / 0.607	**0.466 / 0.500**	1.113 / 0.797	**0.515 / 0.523**	1.102 / 0.794	**0.395 / 0.458**	0.605 / 0.560	**0.406 / 0.455**
Electricity	0.205 / 0.276	**0.197 / 0.267**	0.231 / 0.306	**0.210 / 0.288**	0.240 / 0.316	**0.219 / 0.299**	0.203 / 0.277	**0.190 / 0.266**
PEMS-BAY	1.200 / 0.561	**0.992 / 0.504**	1.386 / 0.624	**1.122 / 0.530**	1.541 / 0.726	**0.909 / 0.504**	1.165 / 0.565	**1.034 / 0.520**
PEMS03	0.470 / 0.489	**0.337 / 0.419**	1.014 / 0.742	**0.413 / 0.455**	1.010 / 0.746	**0.323 / 0.405**	0.461 / 0.479	**0.337 / 0.404**
PEMS07	0.552 / 0.522	**0.309 / 0.382**	1.042 / 0.751	**0.383 / 0.444**	1.085 / 0.783	**0.328 / 0.403**	0.486 / 0.490	**0.309 / 0.389**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
Solar	0.362 / 0.361	**0.241 / 0.276**	0.266 / 0.327	**0.244 / 0.323**	0.361 / 0.427	**0.235 / 0.300**	0.331 / 0.441	**0.328 / 0.420**
PEMS08	1.074 / 0.702	**0.619 / 0.480**	0.699 / 0.545	**0.504 / 0.446**	0.971 / 0.733	**0.673 / 0.568**	0.836 / 0.736	**0.598 / 0.566**
METR-LA	1.504 / 0.776	**1.424 / 0.733**	1.265 / 0.755	**1.247 / 0.722**	1.751 / 0.924	**1.451 / 0.838**	1.957 / 1.012	**1.776 / 0.942**
PEMS04	1.127 / 0.807	**0.532 / 0.532**	0.710 / 0.681	**0.455 / 0.534**	0.652 / 0.664	**0.418 / 0.504**	0.595 / 0.735	**0.409 / 0.540**
Electricity	0.231 / 0.306	**0.212 / 0.289**	**0.269 / 0.345**	0.279 / 0.352	0.313 / 0.415	**0.264 / 0.358**	0.537 / 0.561	**0.492 / 0.524**
PEMS-BAY	1.392 / 0.633	**1.125 / 0.526**	1.161 / 0.596	**0.941 / 0.515**	1.530 / 0.787	**1.122 / 0.628**	1.888 / 0.851	**1.381 / 0.706**
PEMS03	1.025 / 0.753	**0.428 / 0.464**	0.566 / 0.582	**0.373 / 0.475**	0.566 / 0.612	**0.377 / 0.469**	0.517 / 0.605	**0.358 / 0.476**
PEMS07	1.061 / 0.766	**0.429 / 0.468**	0.449 / 0.477	**0.298 / 0.383**	0.404 / 0.517	**0.276 / 0.405**	0.612 / 0.661	**0.276 / 0.407**

of CD models. It also satisfies signed-interaction compatibility, channel-permutation equivariance, and interference stability. Experiments on a range of CI and CD models show that Snow improves forecasting performance in nearly all matched comparisons while maintaining robust behavior under channel permutations and low-quality data.

Snow incurs additional routing and projection costs despite its $O(NKd_v)$ interaction scaling; the practical overhead depends on the channel count, K, d_v, and the backbone (Table 6). Its routing scores and signed coefficients are inspectable but sample-adaptive, so deeper domain-grounded interpretability remains open.

AI USE STATEMENT

We used GPT-5.6 solely for manuscript formatting assistance and language checking. No generative AI tool was used to develop the research problem, methodology, theoretical analysis, experiments,

data, results, or scientific claims. The authors reviewed all AI-assisted edits and take responsibility for the final content of this paper.

Ethics Statement

This work uses publicly available benchmark datasets and does not involve human participants, personal data, or sensitive information. We are not aware of any ethical concerns requiring additional safeguards for this study.

Reproducibility Statement

The paper provides the model formulation, theoretical derivations, experimental settings, dataset descriptions, and detailed results in the main text and appendices. The organized implementation, configuration files, and experiment scripts are prepared for release with the submission to support independent reproduction of the reported results. Code and data availability information is included with the supplementary materials.

References

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018. doi: 10.48550/arXiv.1803.01271. URL https://arxiv.org/abs/1803.01271.

Valérie Castin, Pierre Ablin, and Gabriel Peyré. How smooth is attention? In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 5817–5840. PMLR, 2024. URL https://proceedings.mlr.press/v235/castin24a.html.

Jialin Chen, Jan Eric Lenssen, Aosong Feng, Weihua Hu, Matthias Fey, Leandros Tassiulas, Jure Leskovec, and Rex Ying. From similarity to superiority: Channel clustering for time series forecasting. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024a. doi: 10.52202/079017-4152. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/eb9b18ccb76a1156af5779ffdca1d91f-Abstract-Conference.html.

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

Si-An Chen, Chun-Liang Li, Nathanael C. Yoder, Sercan O. Arik, and Tomas Pfister. TSMixer: An all-MLP architecture for time series forecasting. *Transactions on Machine Learning Research*, 2023. URL https://openreview.net/forum?id=wbpxTuXgm0.

Xinyang Chen, Huidong Jin, Yu Huang, and Zaiwen Feng. XLinear: A lightweight and accurate mlp-based model for long-term time series forecasting with exogenous inputs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20325–20335. AAAI Press, 2026. doi: 10.1609/aaai.v40i24.39121. URL https://ojs.aaai.org/index.php/AAAI/article/view/39121.

Hanyin Cheng, Xingjian Wu, Yang Shu, Zhongwen Rao, Lujia Pan, Bin Yang, and Chenjuan Guo. CoRA: Boosting time series foundation models for multivariate forecasting through correlation-aware adapter. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ae3e173398d6e43fea63cbcc16fbaa98-Abstract-Conference.html.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734. Association for Computational Linguistics, 2014. doi: 10.3115/v1/D14-1179. URL https://aclanthology.org/D14-1179/.

Yuchen Fang, Yuxuan Liang, Bo Hui, Zezhi Shao, Liwei Deng, Xu Liu, Xinke Jiang, and Kai Zheng. Efficient large-scale traffic forecasting with transformers: A spatial data management perspective. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 307–317. Association for Computing Machinery, 2025. doi: 10.1145/3690624.3709177.

Lu Han, Xu-Yang Chen, Han-Jia Ye, and De-Chuan Zhan. SOFTS: Efficient multivariate time series forecasting with series-core fusion. In *Advances in Neural Information Processing Systems*, volume 37. Curran Associates, Inc., 2024a. doi: 10.52202/079017-2046. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/754612bde73a8b65ad8743f1f6d8ddf6-Abstract-Conference.html.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7129–7142, 2024b. doi: 10.1109/TKDE.2024.3400008.

Qihe Huang, Lei Shen, Ruixin Zhang, Jiahuan Cheng, Shouhong Ding, Zhengyang Zhou, and Yang Wang. HDMixer: Hierarchical dependency with extendable patch for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 12608–12616. AAAI Press, 2024. doi: 10.1609/aaai.v38i11.29155. URL https://ojs.aaai.org/index.php/AAAI/article/view/29155.

Dongyuan Li, Shun Zheng, Chang Xu, Jiang Bian, and Renhe Jiang. Routing channel-patch dependencies in time series forecasting with graph spectral decomposition. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/494ad6d9b36d9a1528dbf3baeb4e8a72-Abstract-Conference.html.

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

Zhe Li, Shiyi Qi, Yiduo Li, and Zenglin Xu. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv preprint arXiv:2305.10721*, 2023. doi: 10.48550/arXiv.2305.10721. URL https://arxiv.org/abs/2305.10721.

Zhengyu Li, Xiangfei Qiu, Yuhan Zhu, Xingjian Wu, Jilin Hu, Chenjuan Guo, and Bin Yang. GCGNet: Graph-consistent generative network for time series forecasting with exogenous variables. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/2f27964513a28d034530bfdd117ea31d-Abstract-Conference.html.

Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: A survey. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379 (2194):20200209, 2021. doi: 10.1098/rsta.2020.0209.

Shengsheng Lin, Weiwei Lin, Xinyi Hu, Wentai Wu, Ruichao Mo, and Haocheng Zhong. CycleNet: Enhancing time series forecasting through modeling periodic patterns. In *Advances in Neural Information Processing Systems*, volume 37, pp. 106315–106345, 2024. doi: 10.52202/079017-3373.

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*, 13(5):9861–9871, 2026. doi: 10.1109/JIOT.2025.3647705.

Linbo Liu, Youngsuk Park, Trong Nghia Hoang, Hilaf Hasson, and Jun Huan. Robust multivariate time-series forecasting: Adversarial attacks and defense mechanisms. In *The Eleventh International Conference on Learning Representations*. OpenReview.net, 2023. URL https://openreview.net/forum?id=ctmLBs8lITa.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*,

2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/2ea18fdc667e0ef2ad82b2b4d65147ad-Abstract-Conference.html.

Donghao Luo and Xue Wang. ModernTCN: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*. OpenReview.net, 2024. URL https://openreview.net/forum?id=vpJMJerXHU.

Shusen Ma, Yunbo Zhao, and Yu Kang. C3RL: Rethinking the combination of channel-independence and channel-mixing from representation learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(29):24281–24289, 2026. doi: 10.1609/aaai.v40i29.39608.

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

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S. Jensen, Zhenli Sheng, and Bin Yang. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17 (9):2363–2377, 2024. doi: 10.14778/3665844.3665863.

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*, pp. 1185–1196. Association for Computing Machinery, 2025. doi: 10.1145/3690624.3709325.

Xiangfei Qiu, Hanyin Cheng, Xingjian Wu, Junkai Lu, Jilin Hu, Chenjuan Guo, Christian S. Jensen, and Bin Yang. A comprehensive survey of deep learning for multivariate time series forecasting: A channel strategy perspective. In *Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 2026.

Chao Song, Youfang Lin, Shengnan Guo, and Huaiyu Wan. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 914–921. AAAI Press, 2020. doi: 10.1609/aaai.v34i01.5438. URL https://ojs.aaai.org/index.php/AAAI/article/view/5438.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Huiqiang Wang, Jieming Shi, and Qing Li. FACT: Fine-grained across-variable convolution for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/090d636f951ac0bb18fd5610c95c6a76-Abstract-Conference.html.

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

Xue Wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. CARD: Channel aligned robust blend transformer for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/2f4d6f8e0f4f543db12260696b2a3551-Abstract-Conference.html.

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

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

Jiyuan Xu, Wenyu Zhang, Xin Jing, Jiahao Nie, Shuai Chen, and Shuai Zhang. CPiRi: Channel permutation-invariant relational interaction for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ebba182cb97864368fdb6ae00773a5e4-Abstract-Conference.html.

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

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(9):11121–11128, 2023. doi: 10.1609/aaai.v37i9.26317.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*. OpenReview.net, 2023. URL https://openreview.net/forum?id=vSVLM2j9eie.

Lifan Zhao and Yanyan Shen. Rethinking channel dependence for multivariate time series forecasting: Learning from leading indicators. In *The Twelfth International Conference on Learning Representations*. OpenReview.net, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/b52b07a239a7afa155ca25cf17a55074-Abstract-Conference.html.

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. AAAI Press, 2021. doi: 10.1609/aaai.v35i12.17325. URL https://ojs.aaai.org/index.php/AAAI/article/view/17325.

Appendix

Appendix Roadmap

A. Related Work

- Related Work

B. Theory Analysis

- B.1 Propagation Operator Lemmas
- B.2 Signed-Interaction Compatibility
- B.3 Channel-Permutation Equivariance
- B.4 Interference Stability

C. Experiments

- C.1 Experimental Setup
- C.2 RQ1: CI Backbone Gains
- C.3 RQ2: CD Module Replacement
- C.4 RQ3: Component Contributions
- C.5 RQ4: Structural Properties
- C.6 RQ5: Low-Dimensional Adaptation
- C.7 RQ6: Input-Length Sensitivity
- C.8 RQ7: Related Enhancement Strategies
- C.9 Hyperparameter Sensitivity
- C.10 Extended Results

Complete tables, ablations, robustness tests, and extended evaluation settings.

D. Limitations

Limitations are summarized in the Conclusion.

A Related Work

Channel-independent models. Channel-independent (CI) methods model each channel independently. DLinear and PatchTST are representative examples (Zeng et al., 2023; Nie et al., 2023). By avoiding explicit cross-channel interactions, CI models are computationally efficient and less affected by weakly related channels or channel-specific noise (Wang et al., 2024b; 2026). Recent studies have extended the CI paradigm with additional temporal inductive biases: CycleNet models periodic patterns through learnable recurrent cycles, SegRNN improves recurrent forecasting through segment-wise iteration and parallel multi-step prediction, and FITS introduces lightweight frequency-domain interpolation for compact temporal modeling (Lin et al., 2024; 2026; Xu et al., 2024). Despite their efficiency and strong performance on low-dimensional datasets, CI models may become restrictive in high-dimensional time-series forecasting, where cross-channel dependencies can contain important predictive information (Wang et al., 2026). Our Snow module is designed to address this limitation by injecting cross-channel information into CI backbones and unlocking their potential on high-dimensional time series.

Recent general-purpose forecasting architectures further explore temporal variation modeling, multiscale mixing, convolutional structure, and adaptive pathways (Wu et al., 2023; Wang et al., 2024a; Luo & Wang, 2024; Chen et al., 2024b). These models improve temporal representation, but do not directly resolve the robustness and structural issues of cross-channel propagation considered in Snow.

Channel-dependent models. Channel-dependent (CD) methods explicitly model dependencies among channels. Early graph-based methods such as MTGNN learn directed relations for cross-channel propagation, while Transformer-based methods such as Crossformer use dimension-aware embeddings and cross-dimension attention to capture inter-channel dependencies (Wu et al., 2020; Zhang & Yan, 2023). More recent architectures seek to combine cross-channel modeling with improved robustness or adaptive interaction: CARD introduces channel-aligned attention and a robust forecasting loss, TSMixer performs efficient feature mixing with MLPs, and C3RL and xCPD explore the complementary strengths of channel independence and channel mixing (Wang et al., 2024b; Chen et al., 2023; Ma et al., 2026; Li et al., 2026a). Other methods leverage leading indicators, series-core fusion, channel clustering, or hierarchical dependency modeling; in particular, CCM organizes cross-variable information through channel clustering (Zhao & Shen, 2024; Han et al., 2024a; Chen et al., 2024a; Huang et al., 2024).

These advances strengthen cross-channel modeling, but clustering, routing, and feature mixing alone do not specify the sign of an effective pairwise propagation coefficient, the response of the interaction to a channel permutation, or its propagation gain. Interaction mechanisms based on nonnegative normalizations, such as Softmax attention, cannot directly assign negative coefficients to cross-channel interactions (Vaswani et al., 2017; Cheng et al., 2026); models that associate interaction parameters with fixed channel positions can yield order-sensitive interaction patterns (Zhang & Yan, 2023; Xu et al., 2026); and unconstrained propagation can spread local perturbations across output channels (Castin et al., 2024). Snow instead formulates cross-channel exchange as a sample-adaptive propagation operator $P(H) = S(H)D(H)^{-1}S(H)^{\top}$. Its signed routing permits positive and negative contributions, the shared routing used for aggregation and redistribution preserves channel-order consistency, and the normalization controls the propagation gain. This construction yields signed-interaction compatibility, channel-permutation equivariance, and interference stability within a unified interaction mechanism.

B Theory Analysis

B.1 Common Lemmas for the Snow Propagation Operator

Snow aggregates channel representations into Cluster Representations and writes the aggregated context back to the channels using the same signed routing matrix. This aggregation–redistribution operation induces the sample-adaptive propagation operator

$$P = SD^{-1}S^{\top}. \tag{27}$$

This appendix establishes the three facts used by the proofs of the structural properties: how P transforms under channel permutations, the weighted Gram structure of P, and its strict spectral bound.

B.1.1 Notation and propagation operator

For an input with N channels, let

$$H \in \mathbb{R}^{N \times d_h}. \tag{28}$$

The shared routing network produces

$$\begin{aligned} Z(H) &= \frac{\phi_r(|\mathrm{rFFT}(H)|)}{\tau}, \\ S(H) &= \mathrm{DSoftmax}_K(Z(H)) \\ &:= \mathrm{sign}(Z(H)) \odot \mathrm{softmax}_K(|Z(H)|) \in \mathbb{R}^{N \times K}, \end{aligned} \tag{29}$$

where $\tau > 0$. The absolute routing mass of channel i and the absolute mass received by Cluster Representation k are

$$q_i(H) = \sum_{k=1}^{K} |S_{ik}(H)|, \qquad m_k(H) = \sum_{i=1}^{N} |S_{ik}(H)|. \tag{30}$$

Direction Softmax gives $0 \le q_i(H) \le 1$. With $\varepsilon > 0$, define

$$d_k(H) = m_k(H) + \varepsilon, \qquad D(H) = \mathrm{diag}\big(d_1(H), \dots, d_K(H)\big). \tag{31}$$

Then $D(H)^{-1}$ and $D(H)^{-1/2}$ are well-defined. Let the shared value mapping be

$$V(H) = \phi_v(H) \in \mathbb{R}^{N \times d_v}. \tag{32}$$

The Cluster Representations, propagation operator, and redistributed context are

$$\begin{aligned} C(H) &= D(H)^{-1} S(H)^\top V(H), \\ P(H) &= S(H) D(H)^{-1} S(H)^\top, \\ G(H) &= S(H) C(H) = P(H) V(H). \end{aligned} \tag{33}$$

For a fixed H, we suppress the argument and write S, V, D, C, P, G. Define

$$\rho(H) = \max_{1 \le k \le K} \frac{m_k(H)}{m_k(H) + \varepsilon}. \tag{34}$$

Since $m_k(H) \ge 0$ and $\varepsilon > 0$, $0 \le \rho(H) < 1$. Also $m_k(H) \le N$, so

$$\rho(H) \le \frac{N}{N + \varepsilon} < 1. \tag{35}$$

B.1.2 Permutation transport lemma

Lemma A.1 (Propagation algebra under channel permutation). Let $Q \in \{0,1\}^{N \times N}$ be a channel-permutation matrix. Suppose the routing matrix and value representations transform as

$$S' = QS, \qquad V' = QV. \tag{36}$$

Recompute D', C', P', G' from S' and V' using the definitions above. Then

$$\boxed{D' = D, \qquad C' = C, \qquad P' = QPQ^\top, \qquad G' = QG.} \tag{37}$$

Proof. Because Q only permutes rows,

$$|S'| = |QS| = Q|S|. \tag{38}$$

For the all-ones vector $\mathbf{1}_N$, $Q^\top \mathbf{1}_N = \mathbf{1}_N$. Hence

$$|S'|^\top \mathbf{1}_N = |S|^\top Q^\top \mathbf{1}_N = |S|^\top \mathbf{1}_N, \tag{39}$$

so $m'_k = m_k$, $d'_k = d_k$, and $D' = D$. Since $Q^\top Q = I$,

$$\begin{aligned} C' &= D'^{-1} S'^\top V' = D^{-1} (QS)^\top (QV) = D^{-1} S^\top V = C, \\ P' &= S' D'^{-1} S'^\top = (QS) D^{-1} (QS)^\top = QPQ^\top, \\ G' &= S' C' = (QS) C = QG. \end{aligned} \tag{40}$$

This proves the claim. □

B.1.3 WEIGHTED GRAM STRUCTURE LEMMA

Lemma A.2 (Weighted Gram structure). Define

$$U = SD^{-1/2} \in \mathbb{R}^{N\times K}. \tag{41}$$

Then

$$\boxed{P = UU^\top}, \qquad \boxed{P_{ij} = \langle u_i, u_j\rangle}, \tag{42}$$

where $u_i^\top$ is the i-th row of U. Moreover,

$$\boxed{P = P^\top, \qquad P \succeq 0, \qquad \operatorname{rank}(P) \le \operatorname{rank}(S) \le K.} \tag{43}$$

Proof. Since $D^{-1/2}$ is a real diagonal matrix,

$$UU^\top = SD^{-1/2}(SD^{-1/2})^\top = SD^{-1}S^\top = P. \tag{44}$$

Thus P is a Gram matrix, and

$$P_{ij} = e_i^\top UU^\top e_j = u_i^\top u_j = \langle u_i, u_j\rangle = \sum_{k=1}^{K} \frac{S_{ik}S_{jk}}{d_k}. \tag{45}$$

The last expression shows that each Cluster Representation contributes positively when the two routing coefficients have the same sign and negatively when they have opposite signs. Symmetry follows from $P = UU^\top$. For any $x \in \mathbb{R}^N$,

$$x^\top Px = x^\top UU^\top x = \|U^\top x\|_2^2 \ge 0, \tag{46}$$

so $P \succeq 0$. Finally, multiplication by the invertible matrix $D^{-1/2}$ preserves rank, and therefore

$$\operatorname{rank}(P) = \operatorname{rank}(UU^\top) = \operatorname{rank}(U) = \operatorname{rank}(S) \le K. \tag{47}$$

□

B.1.4 PROPAGATION SPECTRAL BOUND LEMMA

Lemma A.3 (Spectral bound). Assume $\sum_{k=1}^{K} |S_{ik}| \le 1$ for every channel i, and let $d_k = m_k + \varepsilon$ with $\varepsilon > 0$. For

$$\rho = \max_{1\le k\le K} \frac{m_k}{m_k + \varepsilon}, \tag{48}$$

the propagation operator satisfies

$$\boxed{\|P\|_\infty \le \rho, \qquad \|P\|_1 \le \rho, \qquad \|P\|_2 \le \rho < 1.} \tag{49}$$

Consequently,

$$\boxed{0 \le \lambda_\ell(P) \le \rho < 1 \quad (\ell = 1, \dots, N),} \qquad \boxed{0 \preceq P \preceq \rho I.} \tag{50}$$

Proof. For row i, the triangle inequality gives

$$\begin{aligned}
\sum_{j=1}^{N} |P_{ij}| &= \sum_{j=1}^{N} \left| \sum_{k=1}^{K} \frac{S_{ik}S_{jk}}{d_k} \right| \\
&\le \sum_{k=1}^{K} |S_{ik}| \frac{\sum_{j=1}^{N} |S_{jk}|}{d_k} \\
&= \sum_{k=1}^{K} |S_{ik}| \frac{m_k}{m_k + \varepsilon} \\
&\le \rho \sum_{k=1}^{K} |S_{ik}| \le \rho.
\end{aligned} \tag{51}$$

Thus $\|P\|_\infty \le \rho$. By Lemma A.2, P is symmetric, and hence

$$\|P\|_1 = \|P^\top\|_\infty = \|P\|_\infty \le \rho. \tag{52}$$

The induced-norm inequality then yields

$$\|P\|_2 \le \sqrt{\|P\|_1 \|P\|_\infty} \le \rho < 1. \tag{53}$$

Lemma A.2 also gives that P is symmetric positive semidefinite. Therefore all eigenvalues are nonnegative, while for a symmetric matrix $\|P\|_2 = \max_\ell |\lambda_\ell(P)|$, which proves $0 \le \lambda_\ell(P) \le \rho$. Since the eigenvalues of $\rho I - P$ are $\rho - \lambda_\ell(P) \ge 0$, $\rho I - P \succeq 0$, or equivalently $0 \preceq P \preceq \rho I$. Finally, $m_k \le N$ and the function $x/(x+\varepsilon)$ is increasing on $x \ge 0$, so $\rho \le N/(N+\varepsilon) < 1$. □

B.1.5 How the lemmas support the three properties

Lemma A.1 provides the transport rule needed to extend channel-wise equivariance to the propagation result. Lemma A.2 identifies P as a low-rank weighted Gram matrix, which explains how negative off-diagonal coefficients can coexist with positive semidefiniteness. Lemma A.3 bounds the propagation gain and gives strict attenuation of every nonzero value perturbation. The three property appendices combine these common facts with their respective structural arguments.

B.2 Signed-Interaction Compatibility

High-dimensional time series can contain opposite-sign, suppressive, or cancelling interactions. Snow represents such interactions through signed routing: the value representation of a source channel may enter the update of a target channel with a negative direct propagation coefficient. This appendix establishes the result in four steps: (1) characterize when a coefficient is negative; (2) identify the limitation of nonnegative routing; (3) construct a strictly negative coefficient under Direction Softmax; and (4) verify compatibility with the global propagation structure.

B.2.1 Definition and proof target

We use the notation introduced in Appendix B.1. For a family of channel propagation matrices $P^{(\theta)}(H)$, define signed-interaction compatibility by

$$\forall i \ne j, \qquad \exists(\theta, H): \quad P_{ij}^{(\theta)}(H) < 0, \tag{54}$$

where $P_{ij}^{(\theta)}(H)$ is the direct coefficient through which the representation of source channel j enters the update of target channel i. Snow realizes this family with

$$P^{(\theta)}(H) = S^{(\theta)}(H) D^{(\theta)}(H)^{-1} S^{(\theta)}(H)^\top. \tag{55}$$

Suppressing parameters and the input argument, $G = PV$ gives

$$G_i = \sum_{j=1}^{N} P_{ij} V_j. \tag{56}$$

Thus $P_{ij} < 0$ means that the source value V_j contributes with the opposite sign to the direct interaction update of channel i.

The goal is to prove

$$\boxed{\forall i \ne j, \qquad \exists(\theta, H): \quad P_{ij}^{(\theta)}(H) < 0.} \tag{57}$$

B.2.2 A necessary and sufficient sign criterion

For $i \ne j$, write

$$P_{ij} = \sum_{k=1}^{K} \frac{S_{ik} S_{jk}}{d_k}. \tag{58}$$

Partition the Cluster Representations according to whether the two routing coefficients have the same or opposite signs:

$$\mathcal{K}_{ij}^{+} = \{k : S_{ik} S_{jk} > 0\}, \qquad \mathcal{K}_{ij}^{-} = \{k : S_{ik} S_{jk} < 0\}. \tag{59}$$

Define the corresponding weighted magnitudes

$$A^+_{ij} = \sum_{k \in \mathcal{K}^+_{ij}} \frac{|S_{ik} S_{jk}|}{d_k}, \qquad A^-_{ij} = \sum_{k \in \mathcal{K}^-_{ij}} \frac{|S_{ik} S_{jk}|}{d_k}. \tag{60}$$

Theorem B.1 (Sign criterion for negative propagation). For every $i \neq j$,

$$\boxed{P_{ij} = A^+_{ij} - A^-_{ij}}, \qquad \boxed{P_{ij} < 0 \iff A^-_{ij} > A^+_{ij}.} \tag{61}$$

Proof. Since $d_k > 0$, division by d_k preserves the sign of $S_{ik} S_{jk}$. Separating the positive, negative, and zero products gives

$$\begin{aligned} P_{ij} &= \sum_{k \in \mathcal{K}^+_{ij}} \frac{S_{ik} S_{jk}}{d_k} + \sum_{k \in \mathcal{K}^-_{ij}} \frac{S_{ik} S_{jk}}{d_k} \\ &= \sum_{k \in \mathcal{K}^+_{ij}} \frac{|S_{ik} S_{jk}|}{d_k} - \sum_{k \in \mathcal{K}^-_{ij}} \frac{|S_{ik} S_{jk}|}{d_k} \\ &= A^+_{ij} - A^-_{ij}. \end{aligned} \tag{62}$$

Because both weighted magnitudes are nonnegative, $P_{ij} < 0$ holds exactly when $A^-_{ij} > A^+_{ij}$. In other words, the total weighted magnitude of contributions generated by opposite routing signs must exceed that of contributions generated by equal routing signs. □

B.2.3 Limitation of nonnegative routing

Suppose routing is constrained to be nonnegative:

$$S_{ik} \geq 0, \qquad \forall i, k. \tag{63}$$

Then $S_{ik} S_{jk} \geq 0$ for every i, j, k. Since $d_k > 0$,

$$P_{ij} = \sum_{k=1}^{K} \frac{S_{ik} S_{jk}}{d_k} \geq 0. \tag{64}$$

Therefore, under the shared aggregation–redistribution form $P = S D^{-1} S^\top$, nonnegative routing cannot express a negative direct cross-channel coefficient. Signed routing supplies the sign degrees of freedom required by the criterion in Theorem B.1.

B.2.4 Strict realizability under Direction Softmax

Snow generates routing with Direction Softmax,

$$S = \mathrm{DSoftmax}_K(Z) := \mathrm{sign}(Z) \odot \mathrm{softmax}_K(|Z|). \tag{65}$$

The absolute logits determine the routing magnitudes, while their signs determine routing polarity.

Lemma B.2 (Realizability of opposite-signed logits). Let

$$Z(H) = \frac{\phi_r(A(H))}{\tau}, \qquad A(H) = |\mathrm{rFFT}(H)|, \tag{66}$$

where ϕ_r is a two-layer MLP with at least one hidden unit and $\tau > 0$. For any $i \neq j$ and any $a \in (0, \infty)^K$, there exist an input H and finite routing-network parameters such that

$$Z_{i,:} = a^\top, \qquad Z_{j,:} = -a^\top. \tag{67}$$

Proof. Choose two channel representations with distinct spectral features, so $A_i(H) \neq A_j(H)$. The first affine layer and its activation can be chosen so that one hidden unit takes two distinct values $g_i \neq g_j$ on these representations. Use only this hidden unit in the output layer. For output coordinate k, set

$$u_k = \frac{2\tau a_k}{g_i - g_j}, \qquad b_k = \tau a_k - u_k g_i. \tag{68}$$

Then

$$\frac{u_k g_i + b_k}{\tau} = a_k, \qquad \frac{u_k g_j + b_k}{\tau} = -a_k. \tag{69}$$

Setting the weights from all other hidden units to zero gives the required two logit rows. □

Theorem B.3 (Strict negative propagation for any specified channel pair). Let $N \geq 2$, $K \geq 1$, and fix $i \neq j$. For any $a = (a_1, \ldots, a_K)^\top \in (0, \infty)^K$, set

$$Z_{i,:} = a^\top, \qquad Z_{j,:} = -a^\top. \tag{70}$$

Then Direction Softmax gives

$$S_{i,:} = p^\top, \qquad S_{j,:} = -p^\top, \qquad p = \mathrm{softmax}_K(a), \tag{71}$$

and, regardless of the routing of the remaining channels,

$$\boxed{P_{ij} < 0.} \tag{72}$$

Proof. Because $a_k > 0$,

$$S_{ik} = p_k, \qquad S_{jk} = -p_k, \qquad p_k = \frac{\exp(a_k)}{\sum_{\ell=1}^K \exp(a_\ell)} > 0. \tag{73}$$

Hence $S_{ik}S_{jk} = -p_k^2 < 0$ for every k. Other channels can only increase the positive denominator:

$$d_k = \sum_{n=1}^N |S_{nk}| + \varepsilon \geq 2p_k + \varepsilon > 0. \tag{74}$$

Consequently,

$$P_{ij} = \sum_{k=1}^K \frac{S_{ik}S_{jk}}{d_k} = -\sum_{k=1}^K \frac{p_k^2}{d_k} < 0, \tag{75}$$

because every term is strictly negative and $K \geq 1$. □

Lemma B.2 shows that the required opposite-signed logits are realizable by the shared routing MLP. Applying Theorem B.3 to any specified pair therefore proves

$$\boxed{\forall i \neq j, \qquad \exists(\theta, H): \quad P_{ij}^{(\theta)}(H) < 0.} \tag{76}$$

B.2.5 Compatibility with the weighted Gram structure

By Lemma A.2, $U = SD^{-1/2}$ and $P = UU^\top$. If $u_i^\top$ is row i of U, then

$$P_{ij} = \langle u_i, u_j \rangle. \tag{77}$$

For nonzero u_i and u_j,

$$P_{ij} = \|u_i\|_2 \|u_j\|_2 \cos\theta_{ij}, \qquad P_{ij} < 0 \iff \theta_{ij} \in \left(\frac{\pi}{2}, \pi\right]. \tag{78}$$

Thus a negative off-diagonal coefficient describes an obtuse angle between two channel coordinates in the weighted signed-routing space. At the same time, for every $x \in \mathbb{R}^N$,

$$x^\top P x = \|U^\top x\|_2^2 \geq 0, \tag{79}$$

so $P_{ij} < 0$ and $P \succeq 0$ are compatible: the former concerns one off-diagonal interaction, while the latter concerns the quadratic form of the full matrix.

B.2.6 Routing-sign flips and the spectral bound

Let

$$J = \mathrm{diag}(\eta_1, \ldots, \eta_N), \qquad \eta_i \in \{-1, +1\}, \qquad S' = JS. \tag{80}$$

Since $|S'| = |S|$, we have $D' = D$, $m'_k = m_k$, and $\rho' = \rho$. The transformed propagation operator is

$$\begin{aligned} P' &= S' D'^{-1} S'^\top \\ &= (JS) D^{-1} (JS)^\top = JPJ. \end{aligned} \tag{81}$$

Because $J^{-1} = J$, P' is similar to P and has the same eigenvalues, rank, and spectral norm. Entrywise,

$$P'_{ij} = \eta_i \eta_j P_{ij}. \tag{82}$$

Flipping the routing sign of one channel changes the signs of its pairwise coefficients while preserving the global spectrum. In conjunction with Lemma A.3,

$$0 \preceq P \preceq \rho I, \qquad \|P\|_2 \leq \rho < 1, \tag{83}$$

negative off-diagonal coefficients, positive semidefiniteness, and strict spectral contraction can hold simultaneously.

B.2.7 CONCLUSION

The result has three components. First, $P_{ij} < 0$ exactly when the weighted magnitude of opposite-sign routing contributions exceeds that of same-sign contributions. Second, Direction Softmax and the shared routing MLP can realize such a configuration for every specified channel pair. Third, the resulting negative off-diagonal coefficients remain compatible with

$$P = UU^\top, \qquad 0 \preceq P \preceq \rho I, \qquad \|P\|_2 \le \rho < 1. \tag{84}$$

Snow therefore supports direct opposite-sign or suppressive cross-channel propagation while retaining the global structural bounds of its propagation operator.

B.3 CHANNEL-PERMUTATION EQUIVARIANCE

The order of input channels is an organization of the data. Channel-permutation equivariance means that reordering the input channels produces the same reordering of the corresponding outputs. This appendix proves the property by tracing the permutation through shared mappings, routing, aggregation, redistribution, and output projection. We then show that the propagation spectrum and contraction factor are unchanged, and use Crossformer to illustrate how channel-index parameters can break the structural guarantee.

B.3.1 PERMUTATION AND EQUIVARIANCE TARGET

Let $Q \in \{0,1\}^{N\times N}$ be an arbitrary channel-permutation matrix. There is a permutation π such that

$$[QH]_{i,:} = H_{\pi(i),:}, \tag{85}$$

and

$$Q^\top Q = QQ^\top = I, \qquad Q^\top \mathbf{1}_N = \mathbf{1}_N. \tag{86}$$

We prove

$$\boxed{\mathrm{Snow}(QH) = Q\,\mathrm{Snow}(H).} \tag{87}$$

B.3.2 SHARED ROW-WISE MAPPINGS

Lemma C.1 (Equivariance of a shared row-wise mapping). Suppose Φ acts row-wise as

$$[\Phi(H)]_{i,:} = f(H_{i,:}), \tag{88}$$

where f is independent of the channel index i. Then, for every channel-permutation matrix Q,

$$\boxed{\Phi(QH) = Q\Phi(H).} \tag{89}$$

Proof. For any channel i,

$$\begin{aligned}[\Phi(QH)]_{i,:} &= f([QH]_{i,:}) = f(H_{\pi(i),:}) \\ &= [\Phi(H)]_{\pi(i),:} = [Q\Phi(H)]_{i,:}.\end{aligned} \tag{90}$$

Thus $\Phi(QH) = Q\Phi(H)$. □

Applying Lemma C.1 gives

$$\phi_v(QH) = Q\phi_v(H). \tag{91}$$

For any row-feature matrix Y, the shared output mapping similarly satisfies

$$\phi_o(QY) = Q\phi_o(Y). \tag{92}$$

B.3.3 EQUIVARIANCE OF THE ROUTING MAPPING

The rFFT acts along the within-channel feature axis. Therefore,

$$\begin{aligned}[\mathrm{rFFT}(QH)]_{i,:} &= \mathrm{rFFT}(H_{\pi(i),:}) \\ &= [Q\,\mathrm{rFFT}(H)]_{i,:},\end{aligned} \tag{93}$$

which gives

$$\mathrm{rFFT}(QH) = Q\,\mathrm{rFFT}(H). \tag{94}$$

Taking magnitudes, applying the shared ϕ_r, and scaling by τ do not mix channel rows. Hence

$$Z(QH) = QZ(H). \tag{95}$$

The sign and absolute-value operations in Direction Softmax are elementwise, while Softmax is applied along the Cluster Representation dimension within each row. Thus

$$\boxed{S(QH) = QS(H).} \tag{96}$$

Together with Lemma C.1,

$$\boxed{V(QH) = QV(H).} \tag{97}$$

B.3.4 Complete equivariance chain

Set $S' = S(QH) = QS(H)$ and $V' = V(QH) = QV(H)$. Lemma A.1 gives

$$D' = D, \qquad C' = C, \qquad P' = QPQ^\top, \qquad G' = QG. \tag{98}$$

Equivalently,

$$\boxed{D(QH) = D(H), \qquad C(QH) = C(H),} \tag{99}$$

$$\boxed{P(QH) = QP(H)Q^\top, \qquad G(QH) = QG(H).} \tag{100}$$

Feature concatenation preserves the row permutation:

$$[QH \| QG(H)] = Q[H \| G(H)]. \tag{101}$$

Using the shared output mapping,

$$\begin{aligned} \mathrm{Snow}(QH) &= \phi_o([QH \| G(QH)]) \\ &= \phi_o(Q[H \| G(H)]) \\ &= Q\phi_o([H \| G(H)]) \\ &= Q\,\mathrm{Snow}(H). \end{aligned} \tag{102}$$

Therefore,

$$\boxed{\mathrm{Snow}(QH) = Q\,\mathrm{Snow}(H).} \tag{103}$$

B.3.5 Invariance of the propagation spectrum

Since $S(QH) = QS(H)$ only reorders the terms in each Cluster Representation mass,

$$m_k(QH) = m_k(H), \qquad \boxed{\rho(QH) = \rho(H).} \tag{104}$$

Moreover,

$$P(QH) = QP(H)Q^\top \tag{105}$$

is an orthogonal similarity transformation. Hence the two propagation operators have the same characteristic polynomial, eigenvalues, rank, trace, singular values, and spectral norm. In particular,

$$\|P(QH)\|_2 = \|P(H)\|_2 \le \rho(H) = \rho(QH) < 1. \tag{106}$$

Thus channel permutation changes the indexing of propagation modes but not their algebraic or spectral structure.

Table 3: Numerical channel-permutation equivariance check under a three-channel synthetic setting. Lower values indicate closer agreement with the equivariance relation.

Model or setting	Maximum equivariance error
Original Crossformer	2.2182
Crossformer with shared encoder and decoder channel embeddings	1.19×10^{-7}
iTransformer forecasting model	1.34×10^{-7}

B.3.6 WHY CHANNEL-INDEX EMBEDDINGS CAN BREAK EQUIVARIANCE

For per-channel prediction, imputation, and anomaly-detection outputs, consider a representation $E(X) \in \mathbb{R}^{N\times M\times d}$ obtained from channel-shared operations. These operations satisfy

$$E(QX) = QE(X). \tag{107}$$

Suppose the model then adds a channel-indexed embedding $A \in \mathbb{R}^{N\times M\times d}$:

$$\widetilde{E}(X) = E(X) + A. \tag{108}$$

After a channel permutation,

$$\widetilde{E}(QX) = QE(X) + A, \qquad Q\widetilde{E}(X) = QE(X) + QA. \tag{109}$$

Therefore,

$$\widetilde{E}(QX) = Q\widetilde{E}(X) \quad \Longleftrightarrow \quad A = QA. \tag{110}$$

If this condition is required for every permutation, all channel slices of A must be shared. A channel-indexed encoder embedding such as `enc_pos_embedding` does not impose this sharing, so the encoder representation is not structurally equivariant in general.

The decoder has the same issue. For a channel-indexed decoder embedding $B \in \mathbb{R}^{N\times M_{\text{out}}\times d}$, equivariance requires

$$B = QB \tag{111}$$

for every Q. Thus an embedding such as `dec_pos_embedding` provides a second channel-indexed source of order dependence. Even if the attention, feed-forward, and shared prediction mappings preserve row permutations, these unconstrained embeddings remove the structural guarantee $F(QX) = QF(X)$.

B.3.7 NUMERICAL CHECK

For a three-channel random input with fixed random seed, use seq_len $= 8$, pred_len $= 4$, seg_len $= 2$, $d_{\text{model}} = 8$, one encoder layer, two attention heads, and zero dropout. For a channel permutation Q, measure

$$\delta_Q(F, X) = \|F(QX) - QF(X)\|_{\max}. \tag{112}$$

The resulting structural check is summarized in Table 3.

The check illustrates that channel-indexed embeddings can produce a finite equivariance error, whereas tying those embeddings restores the row-permutation relation up to numerical precision.

B.3.8 CONCLUSION

The proof forms the closed chain

$$\begin{aligned} &\text{shared row-wise mappings and within-channel rFFT} \\ &\Longrightarrow S(QH) = QS(H), \quad V(QH) = QV(H) \\ &\Longrightarrow D(QH) = D(H), \quad C(QH) = C(H) \\ &\Longrightarrow P(QH) = QP(H)Q^{\top}, \quad G(QH) = QG(H) \\ &\Longrightarrow \mathrm{Snow}(QH) = Q\,\mathrm{Snow}(H), \quad \rho(QH) = \rho(H). \end{aligned} \tag{113}$$

Snow therefore preserves the correspondence between channel order and channel-specific outputs while keeping the propagation spectrum unchanged under permutation.

B.4 Interference Stability

Input channel representations may contain noise or anomalous perturbations. Snow's interference stability follows from the strict spectral contraction of its core channel propagation: a value perturbation is attenuated after propagation. This appendix defines the perturbation map, proves the one-step energy bound, derives modal and repeated-propagation behavior, and gives a finite attention counterexample for comparison.

B.4.1 Perturbation model

Fix an input H and its induced propagation operator $P(H)$. If the value representation is perturbed as

$$V \longmapsto V + E, \qquad E \in \mathbb{R}^{N\times d_v}, \tag{114}$$

then the change in the propagated context is

$$\Delta G = P(H)E. \tag{115}$$

Define the linear perturbation map

$$\mathcal{T}_H(E) = P(H)E. \tag{116}$$

By Lemmas A.2 and A.3, after suppressing the fixed input argument,

$$P = SD^{-1}S^\top = UU^\top, \qquad U = SD^{-1/2}, \tag{117}$$

and

$$0 \preceq P \preceq \rho I, \qquad \|P\|_2 \le \rho < 1. \tag{118}$$

B.4.2 Strict contraction of propagated perturbations

Theorem D.1 (Strict contraction bound). For every input H and every nonzero perturbation $E \in \mathbb{R}^{N\times d_v}$,

$$\boxed{\|P(H)E\|_F \le \rho(H)\|E\|_F < \|E\|_F.} \tag{119}$$

Moreover,

$$\boxed{\|E\|_F^2 - \|P(H)E\|_F^2 \ge \big(1-\rho(H)^2\big)\|E\|_F^2 > 0.} \tag{120}$$

Proof. For fixed H, take the orthogonal eigendecomposition

$$P = R\Lambda R^\top, \qquad 0 \le \lambda_\ell \le \rho < 1. \tag{121}$$

Let $\widehat{E} = R^\top E$. Orthogonality gives $\|\widehat{E}\|_F = \|E\|_F$, and

$$PE = R\Lambda\widehat{E}, \qquad \|PE\|_F^2 = \|\Lambda\widehat{E}\|_F^2 = \sum_{\ell=1}^{N} \lambda_\ell^2 \|\widehat{E}_{\ell,:}\|_2^2. \tag{122}$$

Since

$$\|E\|_F^2 = \sum_{\ell=1}^{N} \|\widehat{E}_{\ell,:}\|_2^2, \tag{123}$$

subtracting the two expressions yields

$$\begin{aligned} \|E\|_F^2 - \|PE\|_F^2 &= \sum_{\ell=1}^{N} (1-\lambda_\ell^2)\|\widehat{E}_{\ell,:}\|_2^2 \\ &\ge (1-\rho^2)\|E\|_F^2 > 0. \end{aligned} \tag{124}$$

Equivalently, $\|PE\|_F^2 \le \rho^2\|E\|_F^2$. Taking square roots proves the strict contraction bound. □

Corollary D.2 (Contraction between perturbation states). For any $E_1, E_2 \in \mathbb{R}^{N\times d_v}$,

$$\begin{aligned} \|\mathcal{T}_H(E_1) - \mathcal{T}_H(E_2)\|_F &= \|P(H)(E_1 - E_2)\|_F \\ &\le \rho(H)\|E_1 - E_2\|_F. \end{aligned} \tag{125}$$

When $E_1 \neq E_2$, the factor $\rho(H) < 1$ makes the distance strictly smaller.

B.4.3 MODAL ATTENUATION

Theorem D.3 (Energy behavior of propagation modes). Let r_ℓ be a unit eigenvector of P with $Pr_\ell = \lambda_\ell r_\ell$. If the perturbation lies in this mode,

$$E = r_\ell z^\top, \qquad z \in \mathbb{R}^{d_v}, \tag{126}$$

then

$$PE = \lambda_\ell r_\ell z^\top, \qquad \|PE\|_F = \lambda_\ell \|E\|_F. \tag{127}$$

Consequently, $\lambda_\ell = 0$ gives complete filtering, while $0 < \lambda_\ell \le \rho < 1$ gives strict attenuation.

Proof. The eigenvector relation directly gives

$$P(r_\ell z^\top) = (Pr_\ell)z^\top = \lambda_\ell r_\ell z^\top. \tag{128}$$

Because r_ℓ is a unit vector, the Frobenius norm is multiplied by λ_ℓ. □

The modes therefore have different propagation gains, but all gains lie in $[0, \rho]$. In particular, the zero-eigenvalue modes disappear after one propagation, and every nonzero mode decays.

B.4.4 REPEATED PROPAGATION

For a fixed propagation matrix, m successive propagation steps give

$$P^m E = R\Lambda^m R^\top E, \qquad \|P^m E\|_F \le \rho^m \|E\|_F. \tag{129}$$

For $0 < \lambda_\ell \le \rho < 1$, the modal amplitude decays as λ_ℓ^m and its squared energy as λ_ℓ^{2m}. A $\lambda_\ell = 0$ mode is filtered at the first step.

B.4.5 COMPARISON WITH STANDARD ATTENTION PROPAGATION

iTransformer represents variables as tokens and applies standard multi-head self-attention across variable tokens (Liu et al., 2024; Vaswani et al., 2017). The value propagation of one attention head can be written as

$$Y = AV, \qquad A = \mathrm{softmax}\left(\frac{QK^\top}{\sqrt{d}}\right). \tag{130}$$

Softmax guarantees

$$A_{ij} \ge 0, \qquad \sum_{j=1}^{N} A_{ij} = 1, \tag{131}$$

so A is row-stochastic. Row stochasticity alone does not imply $\|A\|_2 \le 1$, whereas Snow's bound follows from the symmetric weighted Gram structure, absolute-mass normalization, and positive stabilizing term.

A finite softmax counterexample. For $0 < \eta < 1$, consider

$$A_\eta = \begin{bmatrix} 1-\eta & \eta \\ 1-\eta & \eta \end{bmatrix}. \tag{132}$$

This is a valid row-stochastic attention matrix. It is produced by finite logits $[\log((1-\eta)/\eta), 0]$ in each row, since

$$\mathrm{softmax}\left(\left[\log\frac{1-\eta}{\eta}, 0\right]\right) = [1-\eta, \eta]. \tag{133}$$

Take a perturbation in the first channel,

$$E = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{134}$$

Then

$$A_\eta E = \begin{bmatrix} 1-\eta \\ 1-\eta \end{bmatrix}, \qquad \|E\|_2^2 = 1, \qquad \|A_\eta E\|_2^2 = 2(1-\eta)^2. \tag{135}$$

When $0 < \eta < 1 - 1/\sqrt{2}$, the propagated energy is larger than the input energy. For example, $\eta = 0.01$ gives

$$\|A_\eta E\|_2^2 = 2(0.99)^2 = 1.9602 > 1. \tag{136}$$

Thus a row-normalized attention operator can copy a perturbation to multiple output channels and amplify its total energy.

B.4.6 STRUCTURAL COMPARISON

The example satisfies $A_\eta \mathbf{1} = \mathbf{1}$, but A_η need not be symmetric or positive semidefinite and can have spectral norm greater than one. Snow instead satisfies

$$P = UU^\top, \qquad P = P^\top, \qquad P \succeq 0, \qquad \|P\|_2 \le \rho < 1. \tag{137}$$

These structural properties limit every propagation mode to the interval $[0, \rho]$, yielding strict attenuation for every nonzero value-path perturbation.

B.4.7 CONCLUSION

Snow's interference stability is a direct consequence of the strict spectral bound of its core propagation operator. Every nonzero value perturbation satisfies

$$\|P(H)E\|_F \le \rho(H)\|E\|_F < \|E\|_F, \tag{138}$$

with complete filtering or strict attenuation at the modal level and at most $\rho(H)^m$ attenuation after m propagation steps. The finite attention example shows why row normalization alone does not provide the same structural guarantee.

C EXPERIMENTS

This appendix documents the experimental protocol and supplies extended evidence for RQ1–RQ7. It first specifies the datasets, backbones, metrics, forecasting settings, and integration choices. It then reports detailed comparisons, repeated-seed statistics, ablations, robustness evaluations, the comparison with related enhancement strategies, and additional settings that support and extend the compact results presented in the main text.

C.1 EXPERIMENTAL SETUP

We report the dataset, model, metric, forecasting-horizon, and input-length configurations used in the experiments. For RQ1 and RQ2, each original backbone is compared with its corresponding "+Snow" variant under the same protocol. Ordinary forecasting results use MSE and MAE, whereas traffic-setting results use RMSE and MAPE. Unless otherwise noted, each dataset follows its standard prediction-horizon protocol and each Origin/+Snow pair uses the same backbone-specific settings.

Forecasting settings. The standard long-horizon experiments use an input length of 96 and prediction horizons of 96, 192, 336, and 720. Datasets with short-horizon TFB protocols retain their dataset-specific horizons, including the additional horizons reported for Crop, Illness, NASDAQ, FRED-MD, Wike2000, USHCN, and Activity in Table 5. For the four high-dimensional traffic datasets (CA, GBA, GLA, and SD), we evaluate both 6-to-6 and 12-to-12 input–output settings; the latter is reported in Table 28. RQ6 additionally evaluates input lengths of 336 and 512 while averaging over prediction lengths 336 and 720.

Training protocol and implementation. Batch size, learning rate, training epochs, early-stopping patience, and related training choices follow the corresponding dataset- and backbone-specific TFB configurations. Standard datasets use the chronological train/validation/test protocols defined by TFB. The four high-dimensional traffic datasets use a chronological 60%/20%/20% train/validation/test split, and normalization statistics are computed from the training portion only. All experiments reported in this paper use the Adam optimizer. Snow uses $K = 8$, $\tau = 1.0$, $d_v = 32$, and $\varepsilon = 10^{-6}$ unless stated otherwise. The default single-run experiments and RQ6 use random seed 2021; the repeated-seed traffic analysis uses the ten seeds 2021, 2022, 2023, 2024, 2025, 3407, 3408, 3409, 3410, and 6666. Experiments are implemented in PyTorch 2.12.0 and run on a single NVIDIA RTX 5880 Ada GPU.

The CI backbones are FITS (Xu et al., 2024), CycleNet (Lin et al., 2024), NLinear (Zeng et al., 2023), PatchTST (Nie et al., 2023), RLinear (Li et al., 2023), SegRNN (Lin et al., 2026), TCN (Bai et al., 2018), and GRU-CI (Cho et al., 2014); the CD backbones are TSMixer (Chen et al., 2023),

Table 4: Statistics of the standard high-dimensional, standard low-dimensional, high-dimensional traffic, and low-quality datasets used in the study.

Dataset	Channels	Application domain	Timesteps	Sampling interval
ETTh1	7	Electricity	17,420	1 hour
ETTh2	7	Electricity	17,420	1 hour
ETTm1	7	Electricity	69,680	15 minutes
ETTm2	7	Electricity	69,680	15 minutes
Illness	7	Epidemiology	966	1 week
Crop	8	Agriculture	8,913	5 minutes
Exchange	8	Economics	7,588	1 day
Weather	21	Environment	52,696	10 minutes
Solar	137	Energy	52,560	10 minutes
PEMS08	170	Traffic	17,856	5 minutes
METR-LA	207	Traffic	34,272	5 minutes
PEMS04	307	Traffic	16,992	5 minutes
Electricity (ECL)	321	Electricity	26,304	1 hour
PEMS-BAY	325	Traffic	52,116	5 minutes
PEMS03	358	Traffic	26,208	5 minutes
Traffic	862	Traffic	17,544	1 hour
PEMS07	883	Traffic	28,224	5 minutes
SD	716	Traffic	525,888	5 minutes
GBA	2,352	Traffic	525,888	5 minutes
GLA	3,834	Traffic	525,888	5 minutes
CA	8,600	Traffic	525,888	5 minutes

XLinear (Chen et al., 2026), GRU-CD (Cho et al., 2014), and iTransformer (Liu et al., 2024). The main text emphasizes representative high-dimensional comparisons, while this appendix provides the full low-dimensional, horizon-specific, traffic, robustness, and extended-setting results.

For RQ4, we examine learned interaction matrices, channel permutations, partial-channel Gaussian corruption, and four real-world low-quality conditions. These robustness evaluations use the settings and datasets described in the corresponding subsections below.

Datasets. The standard benchmark suite follows commonly used forecasting and fair-benchmark protocols (Zhou et al., 2021; Zeng et al., 2023; Qiu et al., 2024). It contains eight low-dimensional datasets (ETTh1, ETTh2, ETTm1, ETTm2, Illness, Crop, Exchange, and Weather) and eight high-dimensional datasets (Solar, PEMS08, METR-LA, PEMS04, Electricity/ECL, PEMS-BAY, PEMS03, and PEMS07). METR-LA and PEMS-BAY are standard traffic benchmarks introduced with DCRNN, while PEMS03, PEMS04, PEMS07, and PEMS08 are associated with STSGCN (Li et al., 2018; Song et al., 2020). The large-scale traffic evaluation uses CA, GBA, GLA, and SD, with channel counts ranging from 716 to 8,600 (Fang et al., 2025). The controlled RQ4 perturbation experiment additionally uses Traffic, while the real-world low-quality evaluation covers NASDAQ, Illness, FRED-MD, Wike2000, SMD, PSM, USHCN, and Activity (Qiu et al., 2024). Dataset-specific channel counts and temporal metadata follow the matched entries in the accompanying dataset notes.

C.1.1 Dataset Descriptions

Table 4 summarizes the standard high-dimensional, standard low-dimensional, high-dimensional traffic, and low-quality datasets used in the study. The metadata and standard benchmark protocol follow TFB (Qiu et al., 2024), while the four large-scale traffic datasets follow the corresponding large-scale traffic protocol (Fang et al., 2025). "Timesteps" denotes the number of observations in the original series, and the sampling interval specifies the temporal granularity.

The eight low-quality datasets in RQ4 are summarized in Table 5. The distribution-shift and white-noise settings follow the low-quality benchmark characterization of TFB (Qiu et al., 2024); SMD, PSM, USHCN, and Activity provide additional real-world stress-test cases used in our experiments.

C.1.2 Model Descriptions

To contextualize the CI–CD comparison used throughout the paper, Figure 1(b) compares seven CI and seven CD models using their corresponding server-side results. The CI group comprises NLinear, RLinear, TCN, CycleNet, PatchTST, FITS, and SegRNN, while the CD group comprises

Table 5: Real-world low-quality datasets used in RQ4.

Dataset	Application domain	Low-quality condition	Forecast horizons (steps)
NASDAQ	Stock market	Distribution shift	$\{24, 36, 48, 60\}$
Illness (ILI)	Epidemiology	Distribution shift	$\{24, 36, 48, 60\}$
FRED-MD	Macroeconomics	White noise	$\{24, 36, 48, 60\}$
Wike2000	Web traffic	White noise	$\{24, 36, 48, 60\}$
SMD	Server monitoring	Anomalies	$\{96, 192, 336, 720\}$
PSM	Server monitoring	Anomalies	$\{96, 192, 336, 720\}$
USHCN	Climate	Missing values	$\{1, 6, 12, 24\}$
Activity	Human activity	Missing values	$\{200, 1200, 3000, 6000\}$

GRU-CD, TSMixer, XLinear, iTransformer, DLinear, Autoformer, and Informer. The grouping follows the model configurations used in this analysis.

The following list briefly describes the backbones used in our CI and CD experiments. For GRU, the "CI" and "CD" labels denote two different channel-processing configurations of the same recurrent architecture.

- **FITS (Xu et al., 2024).** FITS performs lightweight frequency-domain interpolation to reconstruct future signals. We use its channel-wise configuration as a CI backbone.
- **CycleNet (Lin et al., 2024).** CycleNet explicitly learns periodic patterns and models the remaining temporal variation. We use its channel-wise configuration to evaluate Snow on periodic series.
- **NLinear (Zeng et al., 2023).** NLinear subtracts the last observation before applying a linear history-to-future projection. Each channel is forecast independently in our CI setting.
- **PatchTST (Nie et al., 2023).** PatchTST divides a time series into patches and applies Transformer layers to patch tokens. We use its channel-independent configuration as a CI backbone.
- **RLinear (Li et al., 2023).** RLinear combines reversible normalization with a simple linear forecasting map. Its per-channel implementation provides a lightweight CI baseline.
- **SegRNN (Lin et al., 2026).** SegRNN partitions the input into temporal segments and uses recurrent updates for multi-step prediction. We adopt its channel-wise configuration for CI forecasting.
- **TCN (Bai et al., 2018).** TCN uses causal dilated convolutions and residual connections to capture multi-scale temporal dependencies. We apply the temporal convolution independently to each channel in the CI setting.
- **GRU-CI (Cho et al., 2014).** GRU-CI applies a gated recurrent unit separately to each channel sequence. This configuration preserves the independent channel-processing path while allowing Snow to provide cross-channel context.
- **TSMixer (Chen et al., 2023).** TSMixer alternates temporal mixing and feature mixing through MLP layers. Its feature-mixing path serves as a CD backbone in our experiments.
- **XLinear (Chen et al., 2026).** XLinear uses lightweight temporal and variable-wise transformations to combine information across dimensions. We evaluate it as a CD backbone with its original cross-channel path.
- **GRU-CD (Cho et al., 2014).** GRU-CD jointly processes all channels with a shared recurrent hidden state. It therefore provides a recurrent CD counterpart to GRU-CI.
- **iTransformer (Liu et al., 2024).** iTransformer treats variates as tokens and applies attention across variables. We use it as a representative CD model whose original interaction module can be replaced by Snow.

Plug-and-Play Integration. Figure 4 makes the two implementation modes concrete. In PatchTST, Snow is inserted between temporal encoder blocks and augments the channel-independent backbone with cross-channel context. In iTransformer, Snow replaces the original multivariate interaction module and serves as the channel-dependent interaction mechanism. These examples define the integration settings used in RQ1 and RQ2.

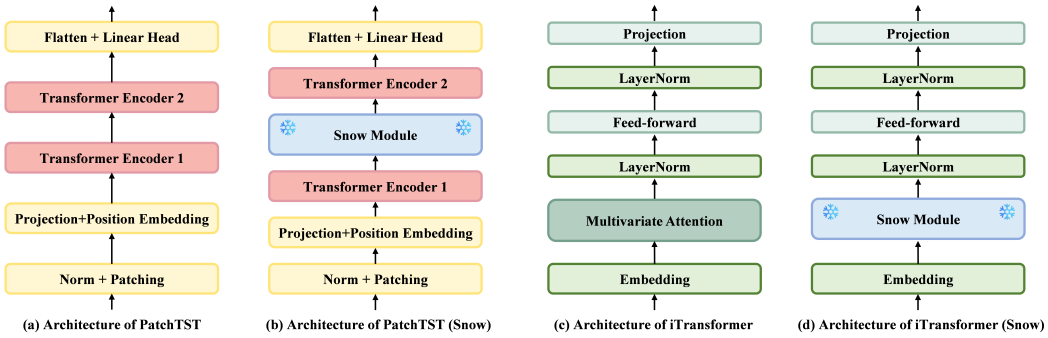


Figure 4: Examples of Snow integration into representative forecasting backbones. Snow is inserted into PatchTST to augment a channel-independent model and replaces the multivariate interaction module in iTransformer to provide a cluster-mediated channel-dependent interaction.

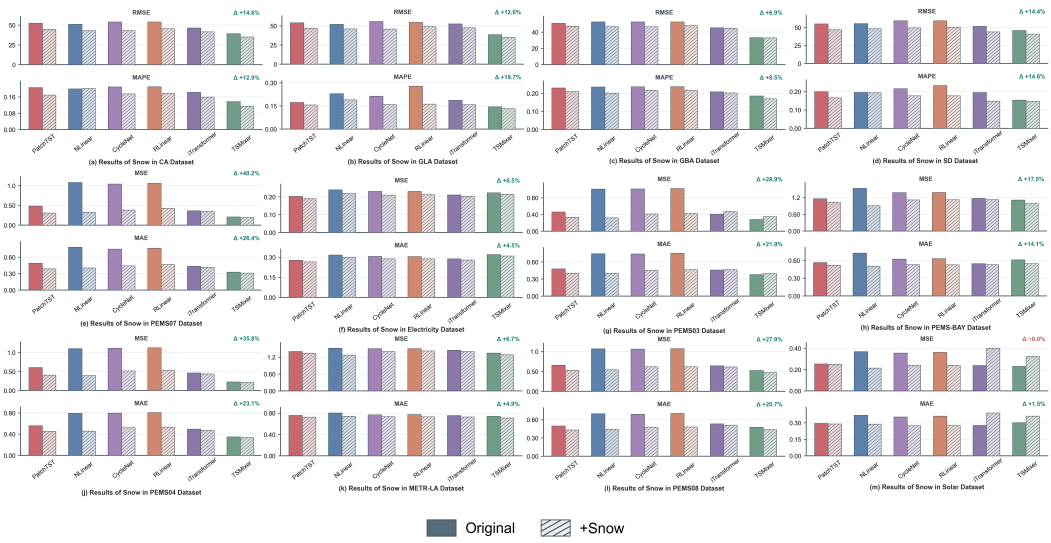


Figure 5: Performance overview of Snow across high-dimensional datasets. Each pair compares an original backbone with its Snow-enhanced counterpart under the same evaluation protocol; lower bars indicate lower forecasting error.

C.1.3 Performance Overview

Before the question-specific tables, Figure 5 provides a visual summary of Snow's effect across the evaluated high-dimensional datasets. The paired results compare each backbone with and without Snow; traffic panels report RMSE and MAPE, while the remaining panels report MSE and MAE. The figure is intended as a navigation view, whereas the detailed numerical conclusions are drawn from the RQ1 and RQ2 tables.

Efficiency Diagnostic. The following comparison reports trainable parameter counts and per-epoch training time for the SD and CA traffic datasets under the reported hardware and training protocol. Parameter counts are dataset-independent for channel-shared backbones such as PatchTST and iTransformer, whereas channel-dependent backbones have dataset-specific counts.

C.2 RQ1: Snow Improves CI Backbones

The main text summarizes the high-dimensional CI comparisons. This subsection provides the corresponding four-horizon averages on all eight low-dimensional datasets, including Crop and Illness, together with repeated-seed variability for the traffic setting. Each dataset follows its standard forecasting horizons, and each Origin/+Snow pair uses the same backbone and dataset; the lower value

Table 6: Training-efficiency overhead of Snow on the SD and CA traffic datasets. "Params" denotes the number of trainable parameters, and "Time/Epoch" denotes the training time per epoch in seconds. Bold variant labels identify the Snow-enhanced runs.

		SD		CA	
Model	Variant	Params	Time/Epoch (s)	Params	Time/Epoch (s)
PatchTST	Origin	9,126	21.296	9,126	227.272
	+Snow	13,388	26.249	13,388	263.736
SegRNN	Origin	18,051	21.401	144,195	187.371
	+Snow	22,313	26.224	148,457	240.376
TCN	Origin	517,710	12.977	74,020,242	232.552
	+Snow	518,387	18.118	74,020,919	293.867
TSMixer	Origin	2,056,100	14.331	295,900,664	930.238
	+Snow	2,056,777	19.377	295,901,341	1,004.903
XLinear	Origin	216,638	16.931	2,502,998	149.012
	+Snow	229,211	21.757	2,515,571	205.887
iTransformer	Origin	9,030	20.136	9,030	832.491
	+Snow	13,292	25.404	13,292	1,050.298

Table 7: Four-horizon average results on low-dimensional datasets for CI models (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.452 / 0.436	**0.435 / 0.420**	**0.459** / 0.446	0.463 / **0.445**	**0.464 / 0.448**	0.469 / 0.462	0.493 / 0.474	**0.486 / 0.462**
ETTh2	0.380 / 0.403	**0.360 / 0.383**	0.398 / 0.420	**0.382 / 0.403**	0.395 / **0.417**	**0.394** / 0.421	0.419 / 0.429	**0.389 / 0.406**
ETTm1	0.382 / 0.393	**0.380 / 0.385**	0.421 / 0.412	**0.405 / 0.406**	0.413 / 0.411	**0.386 / 0.404**	0.408 / 0.417	**0.386 / 0.399**
ETTm2	0.277 / 0.322	**0.271 / 0.311**	0.295 / 0.339	**0.285 / 0.325**	**0.288 / 0.330**	0.302 / 0.342	0.297 / 0.341	**0.281 / 0.325**
Illness	2.505 / 1.084	**2.427 / 1.059**	2.073 / 0.984	**2.022 / 0.967**	1.941 / 0.922	**1.896 / 0.905**	1.896 / 0.881	**1.854 / 0.872**
Crop	0.253 / **0.302**	**0.250** / 0.305	**0.271** / 0.342	0.278 / **0.336**	**0.237 / 0.300**	0.261 / 0.311	0.243 / 0.312	**0.241 / 0.302**
Exchange	0.365 / 0.402	**0.331 / 0.377**	0.385 / 0.421	**0.351 / 0.400**	0.303 / 0.362	**0.293 / 0.278**	0.458 / 0.451	**0.423 / 0.423**
Weather	0.247 / 0.273	**0.236 / 0.262**	0.277 / 0.297	**0.249 / 0.273**	0.254 / **0.281**	**0.247** / 0.295	0.252 / 0.283	**0.243 / 0.272**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.462 / 0.445	**0.429 / 0.411**	**0.429 / 0.447**	0.438 / 0.452	0.546 / 0.533	**0.508 / 0.471**	**0.643** / 0.664	0.660 / **0.655**
ETTh2	**0.393 / 0.414**	0.403 / 0.417	0.396 / 0.426	**0.385 / 0.418**	0.466 / 0.497	**0.429 / 0.457**	0.565 / 0.592	**0.545 / 0.571**
ETTm1	0.409 / **0.407**	**0.402 / 0.407**	0.411 / 0.424	**0.406 / 0.416**	0.506 / 0.497	**0.442 / 0.447**	0.575 / 0.604	**0.563 / 0.580**
ETTm2	0.289 / 0.331	**0.286 / 0.324**	0.292 / 0.343	**0.289 / 0.333**	0.349 / 0.402	**0.336 / 0.386**	0.431 / 0.481	**0.390 / 0.460**
Illness	2.008 / 0.956	**1.954 / 0.940**	1.882 / 0.918	**1.821 / 0.892**	5.030 / 1.571	**4.580 / 1.433**	5.900 / 1.770	**5.826 / 1.752**
Crop	**0.246 / 0.312**	0.278 / 0.315	0.278 / 0.329	**0.239 / 0.306**	0.322 / 0.403	**0.315 / 0.367**	0.418 / 0.464	**0.314 / 0.404**
Exchange	**0.368 / 0.411**	0.420 / 0.423	**0.442** / 0.445	0.444 / **0.434**	**0.462** / 0.504	0.474 / **0.475**	**0.584** / 0.610	0.610 / **0.523**
Weather	0.250 / 0.276	**0.243 / 0.270**	0.254 / 0.300	**0.242 / 0.298**	0.341 / 0.364	**0.279 / 0.315**	**0.340** / 0.420	0.341 / **0.409**

is preferred for both metrics. For the 6-to-6 traffic protocol, the exact CI results are reported in Table 1 in the main text. The relative improvement is computed as $100(E_{\mathrm{Origin}} - E_{+\mathrm{Snow}})/E_{\mathrm{Origin}}$, so positive values indicate lower forecasting error after adding Snow.

Repeated-Seed Analysis. To quantify run-to-run variability, we repeat the reported 6-to-6 traffic evaluation with the ten fixed random seeds 2021, 2022, 2023, 2024, 2025, 3407, 3408, 3409, 3410, and 6666. Table 8 reports the mean and sample standard deviation over the same ten paired runs for each Origin/+Snow comparison. The results include two representative CI backbones (PatchTST and FITS) and two CD backbones (iTransformer and XLinear), covering CA, GBA, GLA, and SD. Table 9 further reports paired statistical tests based on these same runs.

C.2.1 Paired Statistical Tests

We assess whether the repeated-seed improvements are statistically reliable using matched Origin/+Snow runs. For an error metric e and seed s, let e_s^{O} and e_s^{S} denote the errors of the original and

Table 8: Ten-seed results on the CA, GBA, GLA, and SD traffic datasets under the 6-to-6 forecasting setting. Each entry reports the mean ± sample standard deviation over the same ten random seeds 2021, 2022, 2023, 2024, 2025, 3407, 3408, 3409, 3410, and 6666. Each Origin/+Snow pair uses identical seeds and experimental configurations. Values are RMSE / MAPE (%), lower is better, and bold indicates the lower mean for the corresponding metric within each pair.

Dataset	Backbone	Origin	+Snow
CA	PatchTST	52.30±0.38 / 20.61±0.23	**44.75±0.37 / 17.04±0.13**
CA	FITS	52.35±0.51 / 21.75±0.22	**46.12±0.47 / 18.27±0.12**
CA	iTransformer	46.36±0.50 / 18.33±0.11	**41.48±0.35 / 16.03±0.15**
CA	XLinear	38.65±0.41 / 18.46±0.17	**34.80±0.34 / 14.22±0.09**
GBA	PatchTST	50.92±0.51 / 23.05±0.21	**47.39±0.43 / 20.96±0.19**
GBA	FITS	49.68±0.28 / 23.21±0.14	**47.53±0.41 / 21.80±0.21**
GBA	iTransformer	45.31±0.47 / 20.88±0.17	**44.67±0.32 / 20.28±0.13**
GBA	XLinear	39.62±0.33 / 18.89±0.11	**38.00±0.35 / 18.36±0.09**
GLA	PatchTST	53.93±0.59 / 17.34±0.14	**47.16±0.44 / 15.61±0.10**
GLA	FITS	52.14±0.54 / 17.13±0.12	**45.18±0.47 / 14.97±0.14**
GLA	iTransformer	52.52±0.36 / 18.72±0.19	**47.56±0.41 / 15.89±0.13**
GLA	XLinear	42.73±0.41 / 15.81±0.18	**37.98±0.39 / 13.46±0.08**
SD	PatchTST	55.31±0.31 / 20.08±0.22	**46.88±0.22 / 16.67±0.08**
SD	FITS	56.86±0.52 / 23.20±0.21	**47.03±0.36 / 17.63±0.13**
SD	iTransformer	51.67±0.59 / 19.50±0.15	**43.93±0.45 / 14.92±0.12**
SD	XLinear	45.92±0.40 / 17.11±0.17	**37.46±0.18 / 15.00±0.11**

Snow-enhanced models, respectively. We report the relative reduction in the mean error,

$$\text{Imp.} = 100 \times \frac{\bar{e}^{\text{O}} - \bar{e}^{\text{S}}}{\bar{e}^{\text{O}}}, \tag{139}$$

where the bar denotes the mean across the ten seeds. For each dataset–backbone–metric comparison, we form paired differences $d_s = e_s^{\text{O}} - e_s^{\text{S}}$ and conduct a two-sided paired t-test of $H_0 : \mathbb{E}[d_s] = 0$ using

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \qquad n = 10, \tag{140}$$

where $\bar{d}$ and s_d are the sample mean and sample standard deviation of the paired differences. We correct the resulting $m = 32$ raw p-values (four datasets, four backbones, and two metrics) using the Holm–Bonferroni procedure. For ordered raw p-values $p_{(1)} \le \cdots \le p_{(m)}$, the adjusted values are

$$\tilde{p}_{(i)} = \max_{j \le i} \left\{ \min \left[1, (m - j + 1) p_{(j)} \right] \right\}. \tag{141}$$

Table 9 gives the paired-test results. All 32 comparisons remain significant after Holm correction ($\tilde{p} \le 0.00212$), supporting consistent error reductions across the repeated runs.

C.3 RQ2: Snow Replaces Cross-Channel Interaction Modules

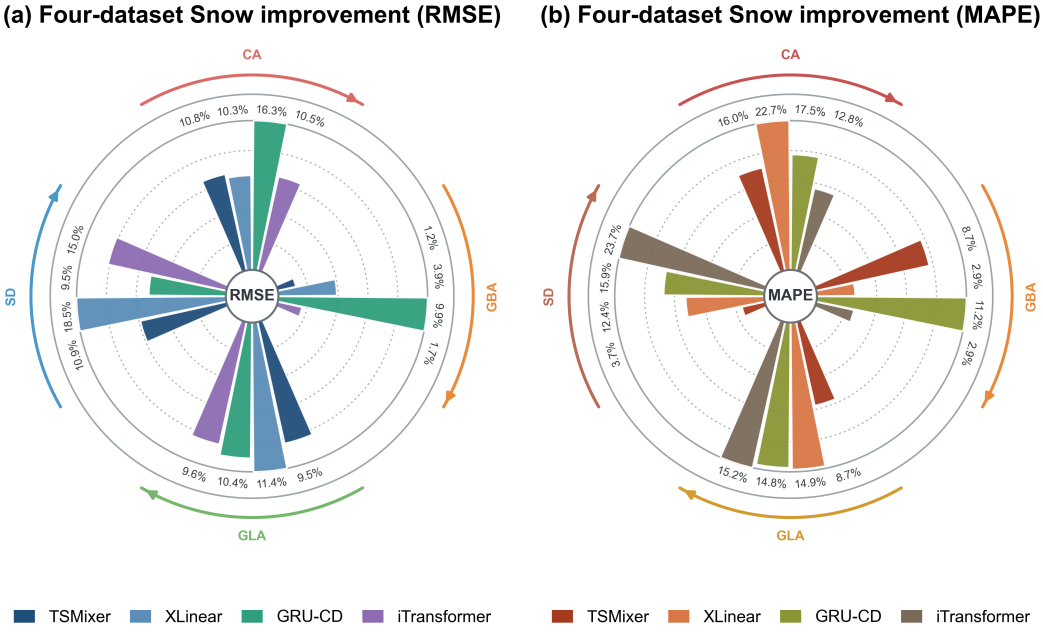
The following tables provide detailed CD comparisons that expand the compact results reported in the main text. Table 10 gives the four-horizon averages on the high-dimensional datasets, while Table 11 reports the dedicated 6-to-6 traffic evaluation. Figure 6 visualizes the corresponding traffic improvements for the four CD backbones. All tables use the same Origin/+Snow pairing, so the effect of replacing the original cross-channel interaction module can be read under a matched protocol.

Traffic Results. Table 11 isolates the CD portion of the 6-to-6 traffic evaluation; the corresponding CI comparison is retained in the main text.

Representation Analysis. To provide an intuitive view of how Snow affects the representations learned by iTransformer, we use UMAP to project the high-dimensional variable representations into a two-dimensional space. As shown in Figure 7, the representations produced by the original iTransformer are concentrated mainly around the center, whereas adding Snow preserves the main

Table 9: Paired statistical tests for the ten-seed 6-to-6 traffic evaluation. Improvement is the relative reduction in the mean error from Origin to +Snow. Each $p/\tilde{p}$ cell reports the raw two-sided paired-t-test p-value followed by its Holm–Bonferroni-adjusted value. RMSE and MAPE (%) are lower is better.

		RMSE			MAPE (%)		
Dataset	Backbone	Imp.	t	$p/\tilde{p}$	Imp.	t	$p/\tilde{p}$
CA	PatchTST	14.45%	40.06	$< 10^{-8} / < 10^{-7}$	17.49%	48.20	$< 10^{-9} / < 10^{-8}$
CA	FITS	11.59%	29.39	$< 10^{-8} / < 10^{-7}$	16.02%	62.70	$< 10^{-10} / < 10^{-9}$
CA	iTransformer	10.38%	38.22	$< 10^{-8} / < 10^{-7}$	12.74%	37.44	$< 10^{-8} / < 10^{-7}$
CA	XLinear	10.20%	31.47	$< 10^{-8} / < 10^{-7}$	22.58%	114.98	$< 10^{-12} / < 10^{-11}$
GBA	PatchTST	6.81%	13.38	$< 10^{-6} / 2 \times 10^{-6}$	8.61%	26.63	$< 10^{-8} / < 10^{-7}$
GBA	FITS	4.32%	12.52	$10^{-6} / 3 \times 10^{-6}$	6.28%	14.46	$< 10^{-6} / < 10^{-6}$
GBA	iTransformer	1.64%	4.26	$0.00212 / 0.00212$	3.02%	11.68	$10^{-6} / 3 \times 10^{-6}$
GBA	XLinear	3.98%	9.52	$5 \times 10^{-6} / 1.1 \times 10^{-5}$	2.87%	12.49	$10^{-6} / 3 \times 10^{-6}$
GLA	PatchTST	12.99%	39.14	$< 10^{-8} / < 10^{-7}$	10.04%	29.93	$< 10^{-8} / < 10^{-7}$
GLA	FITS	13.00%	33.27	$< 10^{-8} / < 10^{-7}$	12.95%	37.22	$< 10^{-8} / < 10^{-7}$
GLA	iTransformer	9.61%	24.06	$< 10^{-8} / < 10^{-7}$	15.31%	42.84	$< 10^{-9} / < 10^{-8}$
GLA	XLinear	11.42%	36.98	$< 10^{-8} / < 10^{-7}$	14.88%	38.98	$< 10^{-8} / < 10^{-7}$
SD	PatchTST	15.10%	37.85	$< 10^{-8} / < 10^{-7}$	16.82%	43.81	$< 10^{-9} / < 10^{-8}$
SD	FITS	17.43%	57.10	$< 10^{-10} / < 10^{-9}$	23.66%	84.26	$< 10^{-11} / < 10^{-10}$
SD	iTransformer	14.97%	30.35	$< 10^{-8} / < 10^{-7}$	23.69%	121.98	$< 10^{-12} / < 10^{-11}$
SD	XLinear	18.60%	41.60	$< 10^{-9} / < 10^{-8}$	12.58%	36.75	$< 10^{-8} / < 10^{-7}$



(a) RMSE improvement.

(b) MAPE improvement.

Figure 6: Relative improvements of CD backbones on the four high-dimensional traffic datasets. Each sector represents one backbone, and the four bars correspond to CA, GBA, GLA, and SD. Positive values denote percentage reductions relative to the original model.

structure while moderately expanding the representation toward the surrounding space. This pattern is consistent across PEMS04, PEMS-BAY, PEMS08, and SD, suggesting that Snow alleviates excessive clustering of variable representations and improves their diversity and separability. The effect is especially pronounced on SD, where the Snow-enhanced representations cover a substantially broader region, indicating richer modeling of latent patterns.

Table 10: Four-horizon average results on high-dimensional datasets for CD models (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
Solar	**0.231 / 0.301**	0.324 / 0.361	**0.240 / 0.277**	0.376 / 0.375	0.497 / 0.563	**0.378 / 0.432**	**0.239 / 0.276**	0.402 / 0.389
PEMS08	0.523 / 0.474	**0.471 / 0.436**	0.659 / 0.541	**0.561 / 0.479**	0.676 / 0.630	**0.560 / 0.534**	0.641 / 0.531	**0.613 / 0.509**
METR-LA	1.346 / 0.741	**1.285 / 0.711**	1.456 / 0.728	**1.377 / 0.718**	1.738 / 0.928	**1.650 / 0.883**	1.449 / 0.755	**1.401 / 0.728**
PEMS04	0.228 / 0.350	**0.216 / 0.335**	0.418 / 0.475	**0.387 / 0.448**	0.501 / 0.613	**0.383 / 0.519**	0.465 / 0.496	**0.437 / 0.473**
Electricity	0.223 / 0.320	**0.215 / 0.309**	0.203 / 0.285	**0.197 / 0.278**	0.489 / 0.529	**0.461 / 0.514**	0.211 / 0.288	**0.204 / 0.280**
PEMS-BAY	1.116 / 0.613	**0.996 / 0.546**	1.091 / 0.536	**0.942 / 0.486**	1.511 / 0.719	**1.314 / 0.675**	1.179 / 0.548	**1.133 / 0.529**
PEMS03	**0.281 / 0.378**	0.348 / 0.397	**0.411** / 0.454	0.435 / **0.446**	0.567 / 0.634	**0.448 / 0.524**	**0.410 / 0.458**	0.473 / 0.467
PEMS07	0.212 / 0.328	**0.197 / 0.311**	0.332 / 0.414	**0.285 / 0.377**	0.428 / 0.490	**0.279 / 0.401**	0.366 / 0.437	**0.347 / 0.418**

Table 11: Results on the CA, GBA, GLA, and SD traffic datasets with the 6-to-6 setting for CD models (RMSE / MAPE (%); lower is better). Each Origin/+Snow pair uses the same backbone, and bold indicates the lower value.

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)		RMSE / MAPE (%)	
CA	39.21 / 13.76	**34.99 / 11.56**	38.70 / 18.41	**34.71 / 14.23**	44.11 / 18.54	**36.92 / 15.29**	46.43 / 18.34	**41.56 / 16.00**
GBA	33.30 / 18.60	**32.89 / 16.99**	39.61 / 18.89	**38.07 / 18.35**	47.65 / 23.76	**42.91 / 21.10**	45.44 / 20.86	**44.69 / 20.26**
GLA	38.71 / 14.46	**35.03 / 13.20**	42.79 / 15.86	**37.91 / 13.49**	51.34 / 18.68	**46.02 / 15.92**	52.54 / 18.69	**47.51 / 15.85**
SD	46.07 / 15.41	**41.04 / 14.84**	45.96 / 17.15	**37.45 / 15.02**	45.79 / 18.84	**41.42 / 15.84**	51.86 / 19.54	**44.07 / 14.91**

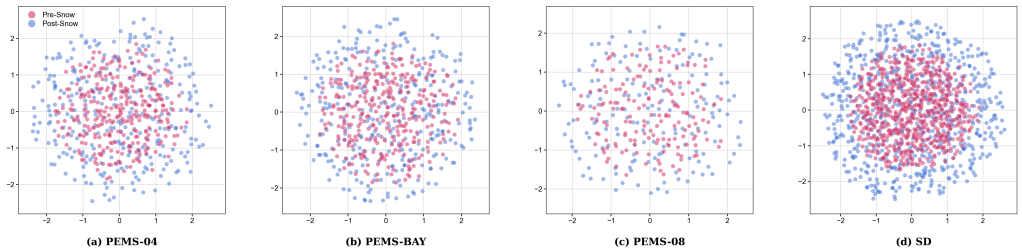


Figure 7: UMAP visualization of the variable representations learned by iTransformer with and without Snow on PEMS04, PEMS-BAY, PEMS08, and SD. Snow preserves the main representation structure while expanding the occupied region and improving the separation of variable representations.

C.4 RQ3: Contributions of Snow Components

RQ3 isolates the two design components that determine Snow's interaction behavior: Direction Softmax (DS), which preserves the sign of routed contributions, and Sample-adaptive Cluster Representations (SACR), which allow the routing pattern to depend on the input instance. "Full Snow" retains both components, "w/o DS" removes signed routing, "w/o SACR" removes sample adaptation, and "w/o both" removes both. Table 12 compares RMSE and MAPE for each variant under the same backbone and dataset settings.

C.5 RQ4: Empirical Evaluation of Three Structural Properties

This subsection evaluates the three structural properties analyzed theoretically through signed channel interactions, channel-order perturbations, and input corruption.

C.5.1 Signed-Interaction Compatibility

To examine how Snow models variable dependencies, we replace the channel-attention module in iTransformer with Snow and visualize the resulting channel interaction matrix. For Snow, the dis-

Table 12: Ablation results on the CA, GBA, GLA, and SD traffic datasets with the 6-to-6 setting (RMSE / MAPE (%); lower is better). DS and SACR denote Direction Softmax and Sample-adaptive Cluster Representations, respectively. Each variant uses the same backbone and dataset, and bold indicates the best variant for each backbone–dataset–metric combination.

Backbone	Dataset	Full Snow		w/o DS		w/o SACR		w/o both	
		RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)	RMSE	MAPE (%)
PatchTST	CA	**44.82**	**17.00**	51.31	19.00	50.10	19.00	58.93	22.00
	GBA	**47.39**	**21.00**	52.32	23.00	53.79	24.00	60.90	27.00
	GLA	**47.11**	**16.00**	52.85	18.00	51.63	17.00	60.11	20.00
	SD	**46.90**	**17.00**	52.29	19.00	53.32	19.00	61.06	22.00
RLinear	CA	**45.44**	**18.00**	51.40	20.00	49.76	20.00	58.53	23.00
	GBA	**48.06**	**22.00**	53.30	24.00	54.16	25.00	60.79	28.00
	GLA	**49.05**	**16.00**	56.01	19.00	55.08	18.00	64.79	21.00
	SD	**50.49**	**18.00**	56.45	20.00	57.77	20.00	65.54	23.00
iTransformer	CA	**41.56**	**16.00**	45.38	17.00	46.33	18.00	52.11	20.00
	GBA	**44.69**	**20.00**	50.10	23.00	48.89	22.00	56.80	26.00
	GLA	**47.51**	**16.00**	52.64	18.00	53.78	18.00	61.52	20.00
	SD	**44.07**	**15.00**	49.09	17.00	48.39	16.00	54.91	19.00
XLinear	CA	**34.71**	**14.00**	39.40	16.00	38.25	16.00	45.30	18.00
	GBA	**38.07**	**18.00**	41.46	20.00	42.34	20.00	48.01	23.00
	GLA	**37.91**	**13.00**	42.68	15.00	43.25	15.00	49.73	18.00
	SD	**37.45**	**15.00**	41.12	16.00	40.63	16.00	47.11	19.00

played quantity is the normalized propagation coefficient $\widetilde{P}_{ij} = P_{ij}/\sqrt{P_{ii}P_{jj}}$, whose sign indicates the polarity of the interaction. Figure 8 compares this signed matrix with the nonnegative channel-attention matrix of the original iTransformer on ETTh2, Weather, PEMS-BAY, and ECL.

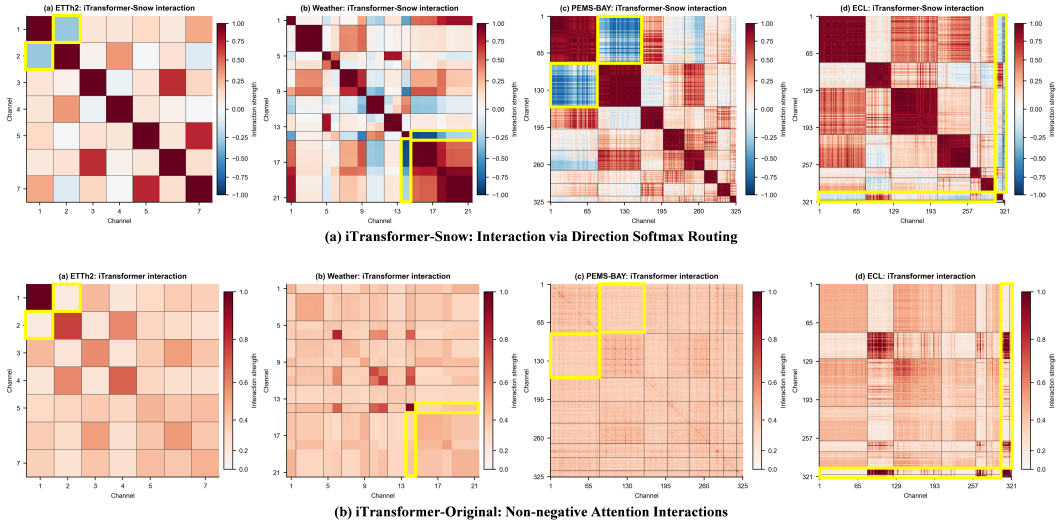
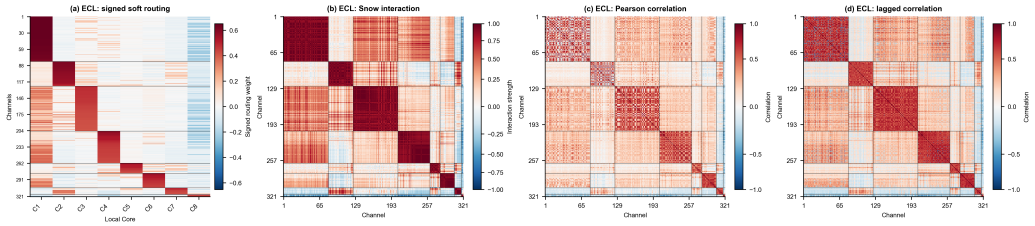
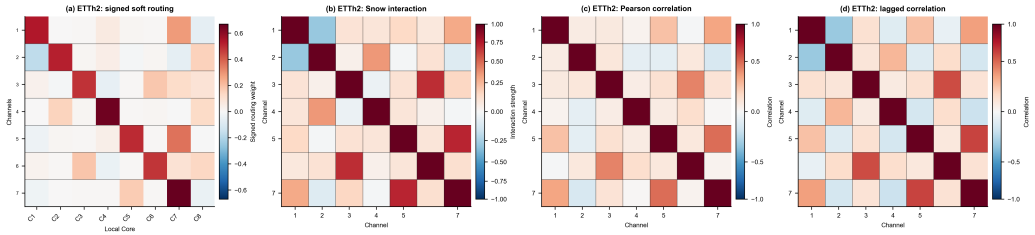


Figure 8: Visualization of variable interaction mechanisms on ETTh2, Weather, PEMS-BAY, and ECL. The upper row shows the signed interaction matrices induced by Snow, where red and blue denote positive and negative interactions, respectively. The lower row shows the nonnegative channel-attention matrices of the original iTransformer. Yellow boxes highlight regions in which Snow identifies clear negative interactions while Softmax attention assigns relatively small nonnegative weights.

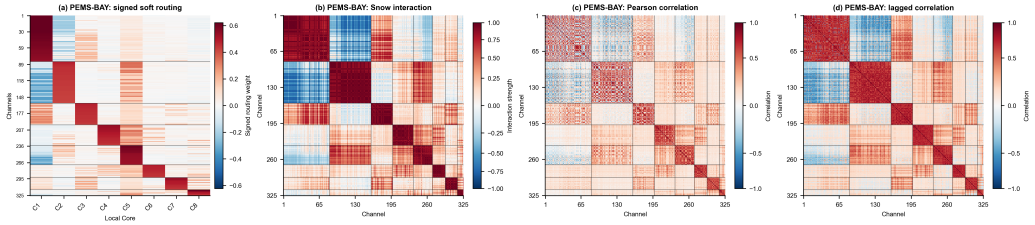
The yellow regions show that Snow assigns negative propagation coefficients to several variable pairs, whereas the corresponding iTransformer attention weights remain small and nonnegative. Softmax attention measures relative attention intensity through nonnegative weights. Direction Softmax additionally retains the sign of the routing coefficient, enabling Snow to distinguish positive and negative cross-variable interactions. These observations are consistent with Snow's signed-interaction compatibility property; the formal guarantee is established by the theorem and proof in Appendix B.2.



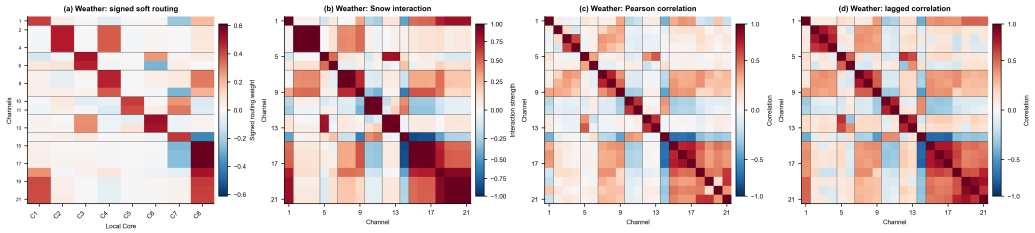
(a) ECL channel-interaction visualization.



(b) ETTh2 channel-interaction visualization.



(c) PEMS-BAY channel-interaction visualization.



(d) Weather channel-interaction visualization.

Figure 9: Additional visualizations of Snow's learned channel-interaction patterns. The four subfigures correspond to ECL, ETTh2, PEMS-BAY, and Weather, respectively. Within each image, the panels show signed routing weights, the learned channel-interaction matrix, the Pearson correlation matrix, and the lagged correlation matrix. The learned interaction patterns exhibit structured groups and signed effects, while the correlation matrices provide data-level reference structures.

These visualizations provide additional qualitative context for the learned interaction patterns (Figure 9). In each dataset, the routing weights show how channels are assigned to local cores, and the resulting interaction matrix displays structured positive and negative effects. The Pearson and lagged correlation matrices serve as data-level reference views, allowing the learned, prediction-oriented interaction pattern to be compared with observed correlations.

C.5.2 Channel-Permutation Equivariance

We train all models using the original channel order and evaluate them after permuting 0%, 50%, or 100% of the input channels. Figure 10 reports the resulting MSE on Solar and ECL.

The original TSMixer and XLinear models become less accurate as the permutation intensity increases. For TSMixer, the MSE rises from 0.261 to 0.351 on Solar and from 0.2288 to 0.457 on ECL. XLinear exhibits the same upward trend with a smaller magnitude. By comparison, TSMixer+Snow

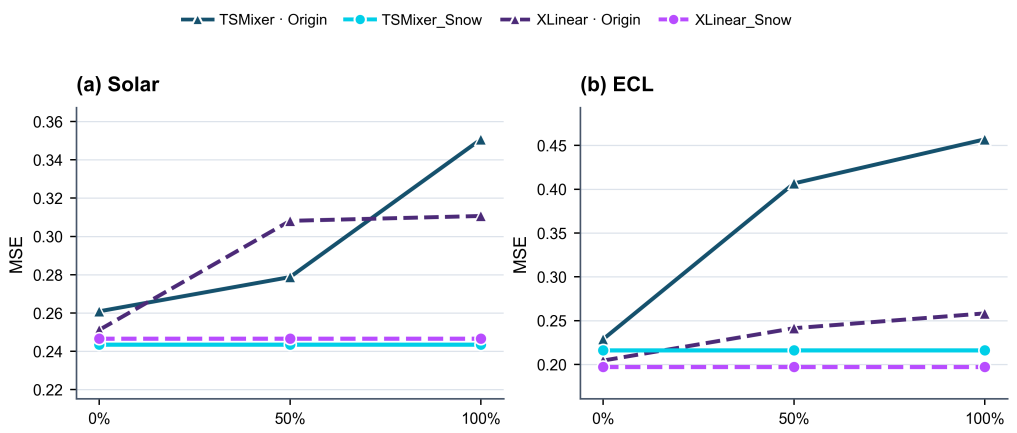


Figure 10: Effect of channel-permutation intensity on forecasting error. Models are trained using the original channel order and tested after permuting 0%, 50%, or 100% of the channels. Triangles denote the original models, and circles denote their Snow-enhanced variants. The Solar and ECL panels use independent vertical-axis ranges.

Table 13: Traffic forecasting under partial-channel Gaussian corruption. Results are averaged over prediction lengths 192 and 720. The noise standard deviation is 0.3 times each channel's training standard deviation. Lower values are better.

Backbone	Noisy channels	MSE		MAE	
		Origin	+Snow	Origin	+Snow
iTransformer	0%	0.556	**0.555**	0.365	**0.364**
	30%	0.574	**0.566**	0.376	**0.370**
	50%	0.591	**0.576**	0.386	**0.376**
	70%	0.610	**0.587**	0.397	**0.383**
	100%	0.632	**0.597**	0.410	**0.389**
RLinear	0%	0.681	**0.660**	0.411	**0.401**
	30%	0.686	**0.664**	0.414	**0.403**
	50%	0.691	**0.667**	0.416	**0.404**
	70%	0.697	**0.670**	0.419	**0.406**
	100%	0.707	**0.676**	0.424	**0.409**
TSMixer	0%	0.870	**0.851**	0.544	**0.534**
	30%	0.895	**0.868**	0.557	**0.544**
	50%	0.918	**0.882**	0.571	**0.551**
	70%	0.947	**0.897**	0.585	**0.560**
	100%	0.974	**0.912**	0.602	**0.570**
XLinear	0%	0.579	**0.571**	0.377	**0.373**
	30%	0.593	**0.581**	0.385	**0.379**
	50%	0.608	**0.589**	0.393	**0.383**
	70%	0.624	**0.599**	0.403	**0.389**
	100%	0.640	**0.608**	0.413	**0.395**

and XLinear+Snow remain stable across the three permutation levels. This pattern is empirically consistent with the channel-permutation equivariance established theoretically in Appendix B.3.

C.5.3 Interference Stability

We evaluate the empirical behavior associated with interference stability from two complementary perspectives. First, Gaussian noise with a standard deviation of 0.3 times each channel's training standard deviation is injected into a controlled fraction of the Traffic training-input channels. The affected channel ratio ranges from 0% to 100%, while evaluation is performed on clean test data. Table 13 reports the exact results averaged over prediction lengths 192 and 720, while Figure 11 visualizes the corresponding MSE comparisons. Snow obtains lower MSE and MAE for every backbone and noise ratio, and its advantage generally increases as more channels are corrupted, consistent with the non-expansive propagation behavior established theoretically.

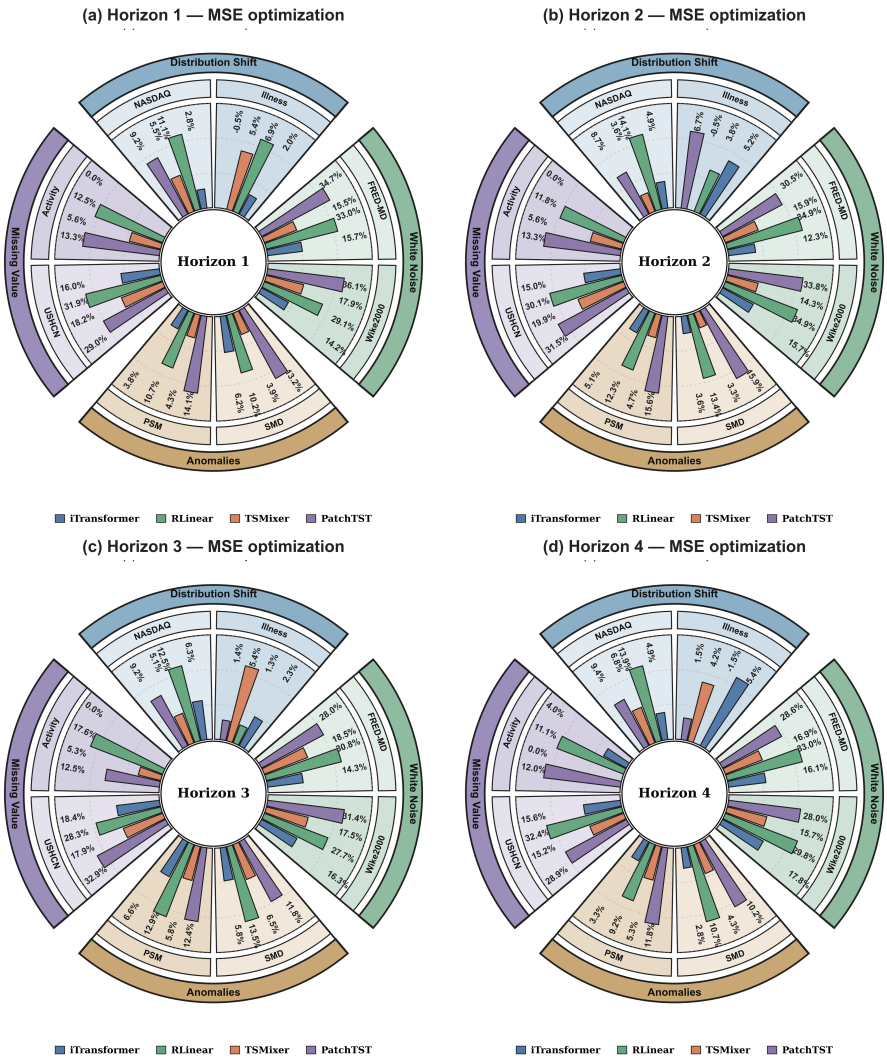


Figure 11: MSE under partial-channel Gaussian corruption for the four evaluated backbones. Each panel compares the original model with its Snow-enhanced variant as the corrupted-channel ratio increases from 0% to 100%. Results are averaged over prediction lengths 192 and 720.

To provide a scale-normalized view of the same experiment, Figure 12 summarizes the relative error reductions for MSE and MAE at the four non-zero corruption ratios. The 0% condition is omitted because it does not represent a perturbed input.

Second, we evaluate four real-world low-quality conditions: distribution shift, white noise, anomalies, and missing values. Figure 13 illustrates the four conditions in the MNAD order (missing values, white noise, anomalies, and distribution shift). Table 14 reports the average MSE and MAE in separate columns; within every Origin/+Snow pair, boldface marks the lower value for that metric. The corresponding horizon-specific results are reported in Table 15.

C.5.4 RQ4 Detailed Low-Quality Results

Table 15 expands the real-world low-quality evaluation across all reported horizons. MSE and MAE are compared independently within each Origin/+Snow pair.

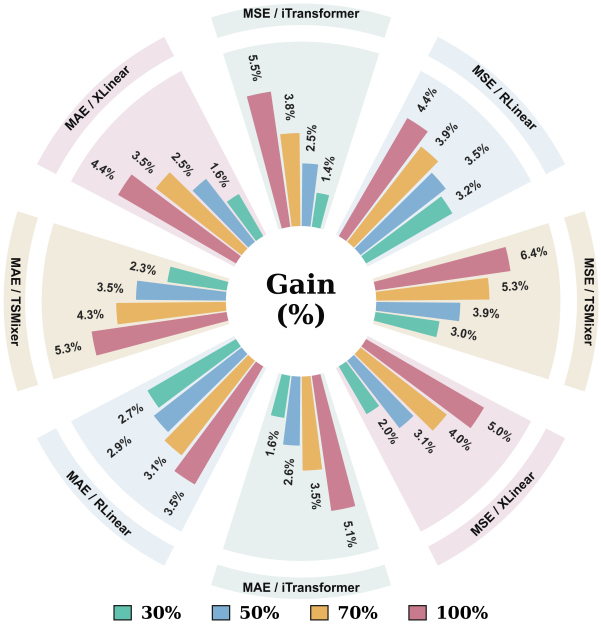


Figure 12: Relative MSE and MAE improvements under partial-channel Gaussian corruption. Each sector represents one backbone–metric pair, and the four bars correspond to 30%, 50%, 70%, and 100% corrupted channels. The 0% condition is omitted.

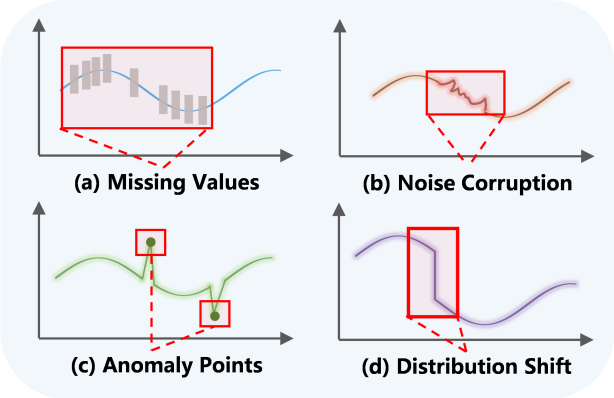


Figure 13: Illustration of the four real-world low-quality conditions evaluated in RQ4, shown in the MNAD order: missing values, white noise, anomalies, and distribution shift.

C.6 RQ5: Adaptation to Low-Dimensional Time Series

RQ5 examines whether Snow remains effective when applied to low-dimensional time series. The low-dimensional CI and CD results are reported in Tables 7 and 16, respectively. Across both model families, Snow improves most paired comparisons, indicating that the proposed interaction mechanism remains applicable beyond the high-dimensional settings emphasized in the main text.

The low-dimensional CD results are reported below using the same Origin/+Snow comparison.

C.7 RQ6: Sensitivity to Input Length

RQ6 tests whether the observed gains persist when the historical input window changes. We compare input lengths 336 and 512 under matched model, dataset, and prediction-length settings, using

Table 14: Average forecasting results under real-world low-quality scenarios. Each backbone is evaluated with and without Snow; MSE and MAE are reported in separate columns, with the lower value in each Origin/+Snow pair shown in bold.

Model	Version	Distribution Shift				White Noise				Anomalies				Missing Value			
		NASDAQ		Illness		FRED-MD		Wike2000		SMD		PSM		USHCN		Activity	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
iTransformer	Original	1.063	0.729	1.849	0.892	69.704	1.392	540.409	1.166	0.655	0.380	1.284	0.739	0.776	0.506	0.018	0.107
	+Snow	**1.009**	**0.712**	**1.779**	**0.871**	**59.315**	**1.229**	**453.270**	**1.052**	**0.627**	**0.373**	**1.226**	**0.725**	**0.650**	**0.445**	**0.017**	**0.104**
RLinear	Original	1.388	0.845	2.008	0.956	77.480	1.525	673.946	1.409	0.565	0.394	1.272	0.753	0.767	0.502	0.017	0.117
	+Snow	**1.207**	**0.783**	**1.954**	**0.940**	**52.154**	**1.107**	**469.295**	**1.016**	**0.497**	**0.359**	**1.135**	**0.680**	**0.532**	**0.368**	**0.015**	**0.107**
TSMixer	Original	1.097	0.739	1.833	0.900	74.232	1.482	560.544	1.171	0.575	0.347	1.294	0.782	0.763	0.494	0.019	0.106
	+Snow	**1.038**	**0.716**	**1.768**	**0.882**	**61.640**	**1.303**	**468.842**	**1.046**	**0.548**	**0.335**	**1.226**	**0.757**	**0.627**	**0.432**	**0.018**	**0.104**
PatchTST	Original	1.051	0.737	1.896	0.881	77.654	1.492	667.302	1.410	0.666	0.378	1.279	0.746	0.782	0.506	0.018	0.107
	+Snow	**0.955**	**0.692**	**1.854**	**0.872**	**54.722**	**1.068**	**453.254**	**1.027**	**0.584**	**0.340**	**1.115**	**0.666**	**0.543**	**0.376**	**0.015**	**0.097**

Table 15: Four-horizon forecasting results under real-world low-quality scenarios. Each backbone is evaluated with and without Snow; MSE and MAE are reported in separate columns, with the lower value in each Origin/+Snow pair shown in bold.

Dataset	Horizon	iTransformer				RLinear				TSMixer				PatchTST			
		Original		+Snow		Original		+Snow		Original		+Snow		Original		+Snow	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NASDAQ	24	0.615	0.550	**0.598**	**0.527**	0.805	0.647	**0.716**	**0.598**	0.650	0.565	**0.614**	**0.544**	0.601	0.566	**0.546**	**0.531**
	36	0.984	0.740	**0.936**	**0.725**	1.231	0.804	**1.058**	**0.750**	0.965	0.715	**0.930**	**0.702**	0.981	0.754	**0.896**	**0.712**
	48	1.327	0.808	**1.243**	**0.792**	1.823	0.998	**1.596**	**0.921**	1.371	0.828	**1.301**	**0.808**	1.275	0.827	**1.158**	**0.773**
	60	1.324	0.819	**1.259**	**0.802**	1.692	0.933	**1.457**	**0.862**	1.404	0.850	**1.308**	**0.811**	1.347	0.803	**1.220**	**0.753**
Illness	24	1.781	0.844	**1.746**	**0.833**	2.078	0.941	**1.935**	**0.896**	1.790	0.824	**1.694**	**0.797**	**1.944**	**0.873**	1.954	0.893
	36	1.728	0.858	**1.638**	**0.823**	1.917	0.927	**1.844**	**0.903**	**1.905**	**0.942**	1.914	0.958	1.864	0.879	**1.740**	**0.843**
	48	1.738	0.888	**1.698**	**0.878**	2.001	0.976	**1.974**	**0.962**	1.768	0.885	**1.672**	**0.856**	1.876	0.886	**1.850**	**0.878**
	60	2.149	0.977	**2.033**	**0.948**	**2.035**	**0.981**	2.065	1.000	1.871	0.950	**1.792**	**0.917**	1.901	0.888	**1.872**	**0.875**
FRED-MD	24	27.069	0.889	**22.807**	**0.768**	37.708	1.047	**25.266**	**0.764**	35.833	1.067	**30.264**	**0.927**	37.438	1.060	**24.445**	**0.778**
	36	46.841	1.220	**41.076**	**1.077**	59.473	1.388	**38.722**	**0.991**	59.962	1.362	**50.445**	**1.202**	59.260	1.344	**41.173**	**0.937**
	48	79.253	1.546	**67.905**	**1.358**	96.908	1.703	**67.052**	**1.214**	81.441	1.608	**66.403**	**1.376**	94.763	1.712	**68.220**	**1.237**
	60	125.653	1.914	**105.470**	**1.711**	115.830	1.961	**77.575**	**1.459**	119.691	1.891	**99.448**	**1.707**	119.155	1.854	**85.051**	**1.320**
Wike2000	24	470.615	1.018	**403.678**	**0.884**	600.478	1.345	**425.729**	**0.981**	519.361	1.080	**426.260**	**0.982**	591.236	1.290	**377.767**	**0.909**
	36	524.282	1.146	**442.128**	**1.049**	664.196	1.385	**432.701**	**0.965**	524.308	1.113	**449.195**	**0.982**	641.435	1.382	**424.554**	**1.007**
	48	574.354	1.220	**480.464**	**1.107**	686.262	1.421	**496.228**	**1.049**	573.847	1.211	**473.210**	**1.100**	678.973	1.462	**465.453**	**1.101**
	60	592.384	1.281	**486.810**	**1.167**	744.847	1.487	**522.522**	**1.067**	624.660	1.279	**526.704**	**1.120**	757.565	1.506	**545.241**	**1.093**
SMD	96	0.449	0.282	**0.421**	**0.277**	0.451	0.324	**0.405**	**0.303**	0.431	0.213	**0.414**	**0.208**	0.468	0.286	**0.406**	**0.260**
	192	0.530	0.328	**0.511**	**0.324**	0.553	0.399	**0.479**	**0.369**	0.461	0.302	**0.446**	**0.291**	0.529	0.319	**0.445**	**0.292**
	336	0.692	0.399	**0.652**	**0.389**	0.617	0.415	**0.534**	**0.371**	0.614	0.345	**0.574**	**0.331**	0.704	0.402	**0.621**	**0.356**
	720	0.949	0.510	**0.922**	**0.503**	0.637	0.437	**0.569**	**0.393**	0.793	0.526	**0.759**	**0.510**	0.963	0.504	**0.865**	**0.452**
PSM	96	0.417	0.425	**0.401**	**0.418**	0.430	0.479	**0.384**	**0.430**	0.423	0.424	**0.405**	**0.411**	0.427	0.429	**0.367**	**0.394**
	192	0.819	0.597	**0.777**	**0.585**	0.774	0.606	**0.679**	**0.561**	0.793	0.620	**0.756**	**0.598**	0.823	0.606	**0.695**	**0.532**
	336	1.450	0.815	**1.354**	**0.804**	1.360	0.787	**1.184**	**0.715**	1.318	0.848	**1.241**	**0.822**	1.498	0.860	**1.312**	**0.780**
	720	2.451	1.120	**2.370**	**1.092**	2.525	1.140	**2.293**	**1.016**	2.642	1.236	**2.503**	**1.197**	**2.367**	1.089	**2.088**	**0.957**
USHCN	1	0.825	0.518	**0.693**	**0.452**	0.787	0.488	**0.536**	**0.341**	0.763	0.490	**0.624**	**0.428**	0.817	0.506	**0.580**	**0.385**
	6	0.745	0.499	**0.633**	**0.439**	0.760	0.482	**0.531**	**0.375**	0.759	0.491	**0.608**	**0.439**	0.753	0.490	**0.516**	**0.366**
	12	0.757	0.498	**0.618**	**0.432**	0.770	0.515	**0.552**	**0.386**	0.764	0.490	**0.627**	**0.417**	0.757	0.512	**0.508**	**0.389**
	24	0.777	0.510	**0.656**	**0.455**	0.751	0.523	**0.508**	**0.369**	0.765	0.503	**0.649**	**0.443**	0.802	0.518	**0.570**	**0.364**
Activity	200	**0.016**	0.100	**0.016**	**0.097**	0.016	0.102	**0.014**	**0.092**	0.018	0.103	**0.017**	**0.099**	0.015	0.104	**0.013**	**0.095**
	1200	**0.015**	0.097	**0.015**	**0.095**	0.017	0.102	**0.015**	**0.091**	0.018	0.107	**0.017**	**0.105**	0.015	0.098	**0.013**	**0.088**
	3000	**0.015**	0.097	**0.015**	**0.095**	0.017	0.107	**0.014**	**0.099**	0.019	0.107	**0.018**	**0.105**	0.016	0.096	**0.014**	**0.088**
	6000	0.025	0.134	**0.024**	**0.128**	0.018	0.157	**0.016**	**0.147**	**0.019**	0.109	**0.019**	**0.106**	0.025	0.128	**0.022**	**0.118**

Table 16: Four-horizon average results on low-dimensional datasets for CD models (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.466/0.464	**0.454/0.459**	0.461/0.446	**0.448/0.433**	**0.575**/0.586	0.582/**0.581**	**0.444**/0.456	0.446/**0.441**
ETTh2	0.417/0.443	**0.403/0.430**	0.390/0.412	**0.374/0.396**	0.502/0.526	**0.482/0.512**	0.408/0.423	**0.382/0.401**
ETTm1	0.384/0.413	**0.377/0.403**	0.417/0.417	**0.397/0.398**	0.506/0.530	**0.502/0.514**	0.422/0.423	**0.413/0.411**
ETTm2	**0.297**/0.348	0.299/**0.346**	**0.280**/0.326	0.283/**0.320**	0.382/0.428	**0.348/0.407**	0.297/0.340	**0.284/0.326**
Illness	1.833/0.900	**1.768/0.882**	1.948/0.928	**1.879/0.903**	5.318/1.604	**5.186/1.576**	1.849/0.892	**1.779/0.871**
Crop	0.230/0.310	**0.211/0.295**	0.271/0.326	**0.255/0.308**	0.377/0.409	**0.276/0.364**	0.283/0.343	**0.265/0.332**
Exchange	0.433/0.445	**0.419/0.430**	0.392/0.421	**0.328/0.384**	**0.509**/0.546	0.565/**0.491**	0.422/0.440	**0.396/0.420**
Weather	0.267/0.327	**0.253/0.312**	0.247/0.278	**0.241/0.269**	0.308/0.376	**0.302/0.366**	0.264/0.286	**0.255/0.275**

Table 17: Input length 336: average results over prediction lengths 336 and 720 (MSE / MAE).

Dataset	NLinear		RLinear		CycleNet	
	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	**0.640 / 0.585**	0.641 / 0.586	**0.623 / 0.576**	0.626 / 0.577	**0.624 / 0.576**	0.626 / 0.578
ETTh2	0.374 / 0.419	**0.371 / 0.417**	**0.368 / 0.415**	0.373 / 0.418	**0.369 / 0.415**	0.375 / 0.419
Solar	0.275 / 0.288	**0.195 / 0.258**	0.274 / 0.286	**0.225 / 0.268**	0.274 / 0.286	**0.224 / 0.268**
PEMS08	0.508 / 0.411	**0.352 / 0.334**	0.509 / 0.409	**0.370 / 0.325**	0.508 / 0.409	**0.370 / 0.324**
PEMS04	0.326 / 0.379	**0.212 / 0.316**	0.327 / 0.378	**0.217 / 0.313**	0.327 / 0.378	**0.215 / 0.312**
PEMS-BAY	0.931 / 0.496	**0.751 / 0.423**	0.915 / 0.464	**0.791 / 0.425**	0.915 / 0.464	**0.773 / 0.426**
PEMS03	0.278 / 0.339	**0.226 / 0.317**	0.278 / 0.338	**0.251 / 0.325**	0.278 / 0.338	**0.259 / 0.329**
ECL	0.207 / 0.297	**0.198 / 0.292**	0.205 / 0.293	**0.195 / 0.285**	0.206 / 0.295	**0.198 / 0.289**
Dataset	TSMixer		GRU-CD		iTransformer	
	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	**0.648 / 0.598**	0.653 / 0.602	0.935 / 0.704	**0.931 / 0.703**	0.636 / 0.587	**0.602 / 0.565**
ETTh2	0.383 / 0.437	**0.380 / 0.435**	0.438 / 0.464	**0.433 / 0.461**	0.367 / 0.420	**0.361 / 0.410**
Solar	0.235 / 0.307	**0.225 / 0.304**	0.294 / 0.381	**0.256 / 0.345**	0.223 / 0.284	**0.218 / 0.279**
PEMS08	0.414 / 0.377	**0.394 / 0.357**	0.535 / 0.452	**0.484 / 0.428**	0.396 / 0.348	**0.380 / 0.335**
PEMS04	0.181 / 0.303	**0.165 / 0.283**	0.337 / 0.435	**0.316 / 0.424**	0.230 / 0.327	**0.221 / 0.317**
PEMS-BAY	0.876 / 0.530	**0.843 / 0.501**	**1.193 / 0.569**	1.221 / 0.600	0.789 / 0.427	**0.763 / 0.413**
PEMS03	0.196 / 0.306	**0.192 / 0.297**	0.336 / 0.424	**0.288 / 0.395**	0.225 / 0.315	**0.218 / 0.304**
ECL	0.228 / 0.333	**0.220 / 0.321**	0.424 / 0.461	**0.410 / 0.458**	0.206 / 0.296	**0.201 / 0.289**

Table 18: Input length 512: average results over prediction lengths 336 and 720 (MSE / MAE).

Dataset	NLinear		RLinear		CycleNet	
	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.694 / 0.609	**0.657 / 0.597**	0.721 / 0.626	**0.712 / 0.624**	0.718 / 0.623	**0.696 / 0.617**
ETTh2	0.406 / 0.441	**0.385 / 0.432**	0.424 / 0.445	**0.414 / 0.440**	0.422 / 0.440	**0.411 / 0.438**
Solar	0.199 / 0.257	**0.190 / 0.251**	0.209 / 0.259	**0.203 / 0.255**	0.207 / 0.259	**0.203 / 0.256**
PEMS08	0.339 / 0.325	**0.321 / 0.317**	0.378 / 0.316	**0.372 / 0.313**	0.389 / 0.319	**0.380 / 0.316**
PEMS04	0.203 / 0.307	**0.192 / 0.300**	0.200 / 0.302	**0.196 / 0.298**	0.198 / 0.299	**0.191 / 0.295**
PEMS-BAY	0.770 / 0.421	**0.730 / 0.409**	0.754 / 0.419	**0.746 / 0.416**	0.737 / 0.422	**0.717 / 0.415**
PEMS03	0.237 / 0.319	**0.227 / 0.314**	0.255 / 0.322	**0.249 / 0.319**	0.258 / 0.324	**0.250 / 0.321**
ECL	0.208 / 0.300	**0.191 / 0.284**	0.204 / 0.294	**0.194 / 0.282**	0.204 / 0.296	**0.194 / 0.284**
Dataset	TSMixer		GRU-CD		iTransformer	
	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.682 / 0.611	**0.661 / 0.601**	1.022 / 0.734	**0.959 / 0.726**	0.634 / 0.590	**0.614 / 0.571**
ETTh2	0.403 / 0.459	**0.390 / 0.447**	0.507 / 0.499	**0.475 / 0.492**	0.388 / 0.433	**0.371 / 0.416**
Solar	0.235 / 0.296	**0.228 / 0.289**	0.273 / 0.354	**0.259 / 0.349**	0.211 / 0.279	**0.205 / 0.273**
PEMS08	0.381 / 0.355	**0.369 / 0.346**	0.514 / 0.435	**0.483 / 0.428**	0.362 / 0.335	**0.345 / 0.321**
PEMS04	0.169 / 0.289	**0.164 / 0.281**	0.339 / 0.435	**0.315 / 0.425**	0.218 / 0.322	**0.208 / 0.309**
PEMS-BAY	0.841 / 0.493	**0.817 / 0.480**	1.290 / 0.615	**1.231 / 0.607**	0.755 / 0.418	**0.734 / 0.406**
PEMS03	0.207 / 0.310	**0.201 / 0.302**	0.311 / 0.405	**0.292 / 0.398**	0.227 / 0.315	**0.218 / 0.304**
ECL	0.235 / 0.339	**0.220 / 0.319**	0.447 / 0.476	**0.413 / 0.457**	0.204 / 0.295	**0.198 / 0.288**

the fixed random seed 2021 for all runs. Tables 17 and 18 report the corresponding results. The entries average the results for prediction lengths 336 and 720, allowing the effect of input length to be separated from changes in the evaluation protocol.

C.8 RQ7: Comparison with Related Enhancement Strategies

RQ7 compares Snow with closely related enhancement strategies under a matched experimental protocol. We apply LIFT, CCM, xCPD, DBLoss, and Snow to the same four CI backbones and evaluate all variants with an input length of 96 and a prediction horizon of 720. Dataset splits, training configurations, software environment, and hardware are kept identical across all variants

Table 19: Comparison of representative enhancement strategies at prediction horizon 720 (MSE / MAE; lower is better). All strategies are evaluated with the same backbone and experimental protocol. Bold and underlined values denote the best and second-best results within each row and metric, respectively.

Dataset	Backbone	Metric: MSE / MAE					
		Origin	+LIFT	+CCM	+xCPD	+DBLoss	+Snow
METR-LA	FITS	1.718 / 0.814	1.706 / 0.811	1.672 / 0.798	1.628 / 0.785	1.701 / 0.808	**1.600 / 0.778**
	CycleNet	1.829 / 0.870	1.804 / 0.862	1.754 / 0.842	1.681 / **0.811**	1.811 / 0.866	**1.658** / 0.813
	NLinear	1.864 / 0.920	1.826 / 0.910	1.691 / 0.866	1.522 / 0.823	1.838 / 0.913	**1.447 / 0.804**
	PatchTST	1.646 / 0.846	1.651 / 0.845	1.629 / 0.836	1.608 / 0.825	1.644 / 0.847	**1.598 / 0.818**
PEMS-BAY	FITS	1.164 / 0.554	1.142 / 0.545	1.075 / 0.521	0.891 / 0.493	1.151 / 0.548	**0.854 / 0.475**
	CycleNet	1.411 / 0.625	1.395 / 0.628	1.332 / 0.603	1.261 / 0.579	1.402 / 0.622	**1.233 / 0.564**
	NLinear	1.564 / 0.728	1.516 / 0.714	1.384 / 0.658	1.188 / 0.603	1.548 / 0.721	**1.036 / 0.556**
	PatchTST	1.235 / 0.576	1.219 / 0.573	1.177 / 0.562	**1.087** / 0.544	1.239 / 0.577	1.091 / **0.534**
Electricity	FITS	0.246 / 0.309	0.246 / 0.308	0.242 / 0.305	0.238 / **0.298**	0.245 / 0.309	**0.236** / 0.299
	CycleNet	0.271 / 0.337	0.267 / 0.333	0.260 / 0.326	0.251 / 0.317	0.268 / 0.334	**0.244 / 0.312**
	NLinear	0.280 / 0.347	0.275 / 0.345	0.266 / 0.335	0.254 / 0.327	0.278 / 0.346	**0.248 / 0.322**
	PatchTST	0.245 / 0.311	0.243 / 0.311	0.237 / 0.306	**0.227** / 0.300	0.246 / 0.312	0.229 / **0.297**

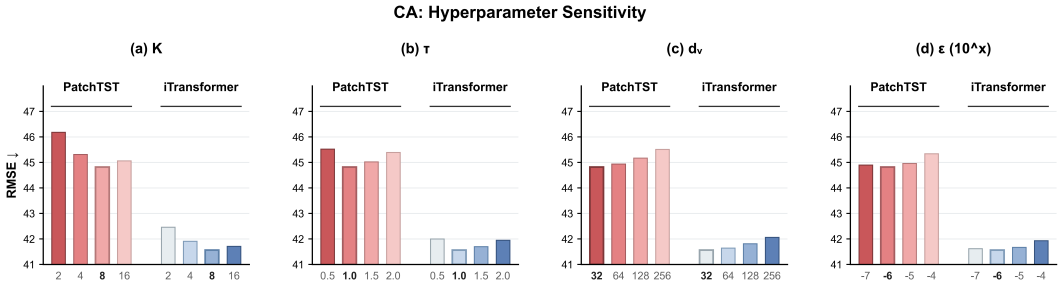


Figure 14: Hyperparameter sensitivity of Snow on CA in the 6-to-6 traffic setting. Each panel reports RMSE (lower is better) over four settings of one hyperparameter. The left and right bar groups correspond to Snow-enhanced PatchTST and iTransformer, respectively; bold tick labels mark the configurations used in the reported experiments.

and follow Appendix C.1. Table 19 reports the MSE and MAE results for all backbone–dataset combinations. Bold and underlined values denote the best and second-best results within each row and metric, respectively.

Snow achieves the lowest MSE in 10 of 12 backbone–dataset comparisons and the lowest MAE in 10 of 12 comparisons. Overall, Snow obtains the best result in 20 of 24 metric entries, while xCPD is the closest competing strategy in the remaining cases. These results show that Snow remains competitive with related enhancement strategies under the fixed 96-to-720 forecasting protocol.

C.9 Hyperparameter Sensitivity

We examine the sensitivity of Snow to the number of clusters K, routing temperature τ, value dimension d_v, and normalization constant ε on the CA and SD traffic datasets. The analysis uses the 6-to-6 setting and evaluates the Snow-enhanced PatchTST and iTransformer backbones using RMSE. Figures 14 and 15 show that performance changes are generally smooth across the investigated ranges. Multiple values near the configurations used in the reported experiments remain competitive, while smaller cluster counts or larger value dimensions lead to some performance degradation in selected settings. Overall, Snow exhibits stable performance trends over reasonable hyperparameter ranges, and its gains are not confined to a single isolated setting.

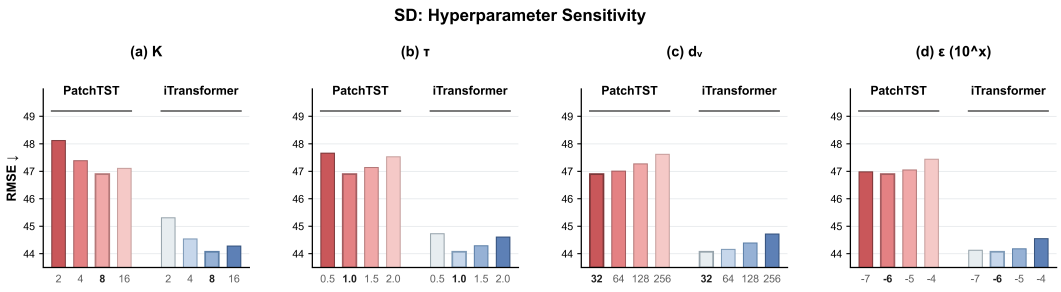


Figure 15: Hyperparameter sensitivity of Snow on SD in the 6-to-6 traffic setting. Each panel reports RMSE (lower is better) over four settings of one hyperparameter. The left and right bar groups correspond to Snow-enhanced PatchTST and iTransformer, respectively; bold tick labels mark the configurations used in the reported experiments.

Table 20: RQ1 CI results at prediction horizon 96 (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.382 / 0.396	**0.380 / 0.387**	**0.399** / 0.407	**0.399 / 0.402**	0.405 / **0.413**	**0.402** / 0.417	0.403 / 0.419	**0.392 / 0.406**
ETTh2	0.295 / 0.344	**0.284 / 0.332**	**0.302** / 0.352	**0.302 / 0.350**	**0.302 / 0.352**	0.320 / 0.369	0.321 / 0.363	**0.300 / 0.348**
ETTm1	0.320 / 0.359	**0.310 / 0.346**	0.360 / 0.379	**0.345 / 0.373**	0.350 / 0.379	**0.327 / 0.368**	0.340 / 0.373	**0.317 / 0.356**
ETTm2	0.178 / 0.259	**0.174 / 0.251**	0.188 / 0.273	**0.182 / 0.264**	0.181 / **0.263**	**0.180** / 0.265	0.180 / 0.265	**0.179 / 0.260**
Exchange	0.085 / 0.202	**0.082 / 0.196**	0.087 / 0.209	**0.084 / 0.205**	0.100 / 0.231	**0.078 / 0.170**	0.102 / 0.225	**0.088 / 0.208**
Weather	0.168 / 0.211	**0.160 / 0.202**	0.198 / 0.239	**0.165 / 0.208**	0.169 / **0.214**	**0.156** / 0.220	0.168 / 0.218	**0.160 / 0.207**
Solar	0.229 / 0.270	**0.195 / 0.233**	0.306 / 0.333	**0.209 / 0.258**	0.317 / 0.346	**0.188 / 0.256**	0.225 / 0.277	**0.213 / 0.269**
PEMS08	0.698 / 0.486	**0.503 / 0.404**	0.924 / 0.606	**0.534 / 0.420**	0.902 / 0.605	**0.505 / 0.393**	0.554 / 0.434	**0.440 / 0.373**
METR-LA	1.283 / 0.712	**1.207 / 0.704**	1.260 / 0.694	**1.181 / 0.647**	1.266 / 0.718	**1.128 / 0.667**	1.190 / 0.691	**1.130 / 0.652**
PEMS04	0.593 / 0.590	**0.378 / 0.472**	1.220 / 0.862	**0.540 / 0.547**	1.170 / 0.833	**0.361 / 0.442**	0.616 / 0.570	**0.385 / 0.446**
Electricity	0.185 / 0.259	**0.179 / 0.252**	0.213 / 0.289	**0.191 / 0.274**	0.221 / 0.300	**0.205 / 0.287**	0.181 / 0.258	**0.166 / 0.245**
PEMS-BAY	1.163 / 0.539	**0.928 / 0.496**	1.390 / 0.628	**1.247 / 0.575**	1.556 / 0.736	**0.959 / 0.541**	1.141 / 0.551	**0.987 / 0.496**
PEMS03	0.445 / 0.470	**0.278 / 0.391**	1.117 / 0.819	**0.443 / 0.483**	1.090 / 0.804	**0.289 / 0.388**	0.453 / 0.482	**0.303 / 0.390**
PEMS07	0.464 / 0.467	**0.223 / 0.318**	1.110 / 0.799	**0.336 / 0.422**	1.144 / 0.825	**0.259 / 0.366**	0.476 / 0.487	**0.281 / 0.374**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.398 / 0.404	**0.365 / 0.369**	**0.379 / 0.408**	0.382 / **0.408**	0.481 / 0.493	**0.391 / 0.405**	**0.581** / 0.622	**0.581 / 0.598**
ETTh2	**0.299 / 0.349**	0.335 / 0.370	0.296 / 0.354	**0.285 / 0.336**	0.361 / 0.423	**0.328 / 0.387**	**0.418 / 0.489**	0.429 / 0.496
ETTm1	0.346 / 0.374	**0.341 / 0.373**	0.359 / 0.390	**0.347 / 0.378**	0.432 / 0.456	**0.369 / 0.403**	0.505 / 0.560	**0.481 / 0.532**
ETTm2	0.181 / 0.263	**0.180 / 0.260**	0.183 / 0.269	**0.181 / 0.262**	**0.225 / 0.323**	0.231 / 0.325	0.270 / 0.388	**0.258 / 0.376**
Exchange	0.090 / 0.212	**0.088 / 0.207**	0.091 / 0.210	**0.080 / 0.196**	0.106 / 0.251	**0.101 / 0.231**	0.127 / 0.301	**0.102 / 0.235**
Weather	0.167 / 0.211	**0.160 / 0.205**	0.165 / **0.224**	**0.164** / 0.231	0.242 / 0.291	**0.188 / 0.244**	0.224 / 0.319	**0.218 / 0.306**
Solar	0.313 / 0.342	**0.206 / 0.243**	0.242 / 0.309	**0.199 / 0.281**	0.314 / 0.403	**0.218 / 0.280**	0.313 / 0.399	**0.284 / 0.365**
PEMS08	0.889 / 0.606	**0.526 / 0.419**	0.503 / 0.421	**0.361 / 0.347**	0.838 / 0.649	**0.570 / 0.501**	0.657 / 0.714	**0.390 / 0.454**
METR-LA	1.261 / 0.702	**1.185 / 0.650**	1.126 / 0.679	**1.037 / 0.640**	1.407 / 0.810	**1.209 / 0.726**	1.798 / 0.853	**1.377 / 0.767**
PEMS04	1.236 / 0.871	**0.561 / 0.559**	0.592 / 0.604	**0.333 / 0.457**	0.484 / 0.566	**0.319 / 0.442**	0.609 / 0.764	**0.314 / 0.441**
Electricity	0.213 / 0.290	**0.195 / 0.275**	**0.240 / 0.322**	0.247 / 0.327	0.283 / 0.393	**0.242 / 0.341**	0.502 / 0.540	**0.457 / 0.507**
PEMS-BAY	1.408 / 0.647	**1.252 / 0.576**	1.168 / 0.599	**0.905 / 0.511**	1.329 / 0.743	**1.033 / 0.602**	2.022 / 0.883	**1.171 / 0.654**
PEMS03	1.120 / 0.821	**0.459 / 0.492**	0.507 / 0.550	**0.311 / 0.434**	0.502 / 0.574	**0.307 / 0.422**	0.416 / 0.539	**0.304 / 0.420**
PEMS07	1.131 / 0.818	**0.409 / 0.468**	0.438 / 0.470	**0.251 / 0.365**	0.303 / 0.436	**0.220 / 0.353**	0.562 / 0.651	**0.201 / 0.333**

C.10 Extended Results

This section groups the remaining evidence by evaluation setting. We first report prediction-horizon-specific results for RQ1 and RQ2, followed by additional dataset and horizon tables and a longer-horizon traffic evaluation. The detailed real-world low-quality results are reported with RQ4 above.

C.10.1 RQ1 Prediction-Horizon Results

Prediction horizon 96. Table 20 isolates the shortest standard forecasting horizon and reports the CI comparison across all benchmark groups.

Table 21: RQ1 CI results at prediction horizon 192 (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.441 / 0.429	**0.425 / 0.412**	0.450 / 0.436	**0.446 / 0.432**	**0.456 / 0.440**	0.467 / 0.455	0.450 / 0.449	**0.446 / 0.436**
ETTh2	0.370 / 0.392	**0.362 / 0.378**	0.391 / 0.409	**0.390 / 0.399**	0.392 / 0.407	**0.377 / 0.404**	0.400 / 0.415	**0.372 / 0.394**
ETTm1	**0.361** / 0.379	**0.361 / 0.372**	0.399 / 0.398	**0.391 / 0.395**	0.390 / 0.397	**0.370 / 0.393**	0.402 / 0.410	**0.367 / 0.383**
ETTm2	0.239 / 0.299	**0.233 / 0.288**	0.255 / 0.315	**0.246 / 0.301**	**0.247 / 0.305**	0.258 / 0.314	0.258 / 0.318	**0.236 / 0.297**
Exchange	0.179 / 0.297	**0.172 / 0.287**	**0.184** / 0.310	0.185 / **0.306**	0.204 / 0.338	**0.142 / 0.232**	0.187 / 0.311	**0.184 / 0.301**
Weather	0.214 / 0.251	**0.203 / 0.240**	0.243 / 0.276	**0.216 / 0.253**	0.217 / **0.258**	**0.210** / 0.270	0.213 / 0.257	**0.206 / 0.249**
Solar	0.256 / 0.283	**0.222 / 0.252**	0.345 / 0.352	**0.250 / 0.282**	0.358 / 0.368	**0.221 / 0.299**	0.247 / 0.289	**0.245 / 0.286**
PEMS08	0.790 / 0.545	**0.558 / 0.443**	0.986 / 0.654	**0.581 / 0.458**	0.994 / 0.661	**0.562 / 0.435**	0.627 / 0.469	**0.489 / 0.399**
METR-LA	1.365 / 0.715	**1.323 / 0.689**	1.431 / 0.745	**1.363 / 0.732**	1.446 / 0.778	**1.257 / 0.741**	1.356 / 0.729	**1.278 / 0.699**
PEMS04	0.656 / 0.619	**0.453 / 0.521**	1.240 / 0.855	**0.569 / 0.562**	1.245 / 0.863	**0.452 / 0.501**	0.656 / 0.589	**0.434 / 0.477**
Electricity	0.186 / 0.263	**0.179 / 0.252**	0.212 / 0.292	**0.196 / 0.278**	0.222 / 0.303	**0.207 / 0.290**	0.185 / 0.263	**0.175 / 0.253**
PEMS-BAY	1.306 / 0.605	**1.155 / 0.543**	1.466 / 0.653	**0.929 / 0.466**	1.636 / 0.763	**0.715 / 0.405**	1.170 / 0.593	**1.104 / 0.564**
PEMS03	0.494 / 0.515	**0.361 / 0.421**	1.138 / 0.799	**0.436 / 0.477**	1.147 / 0.816	**0.345 / 0.425**	0.493 / 0.501	**0.359 / 0.423**
PEMS07	0.488 / 0.524	**0.247 / 0.351**	1.157 / 0.795	**0.373 / 0.443**	1.214 / 0.846	**0.313 / 0.405**	0.536 / 0.521	**0.338 / 0.414**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.452 / 0.435	**0.411 / 0.403**	0.426 / 0.440	**0.421 / 0.437**	0.536 / 0.522	**0.434 / 0.427**	**0.624** / 0.641	0.629 / **0.621**
ETTh2	**0.389 / 0.405**	0.409 / 0.414	**0.377 / 0.402**	0.385 / 0.407	0.463 / 0.485	**0.422 / 0.444**	0.554 / 0.582	**0.520 / 0.544**
ETTm1	0.387 / 0.393	**0.373 / 0.390**	0.398 / 0.413	**0.396 / 0.405**	0.481 / 0.480	**0.417 / 0.431**	0.554 / 0.591	**0.532 / 0.570**
ETTm2	**0.248** / 0.306	**0.248 / 0.303**	0.247 / 0.312	**0.241 / 0.305**	**0.301** / 0.372	0.304 / **0.370**	0.353 / 0.439	**0.336 / 0.429**
Exchange	0.184 / 0.308	**0.179 / 0.295**	0.188 / 0.308	**0.175 / 0.296**	0.222 / 0.370	**0.214 / 0.340**	0.262 / 0.437	**0.193 / 0.330**
Weather	0.213 / 0.252	**0.206 / 0.247**	0.214 / **0.276**	**0.213** / 0.280	0.304 / 0.344	**0.240 / 0.292**	**0.289 / 0.392**	0.310 / **0.392**
Solar	0.352 / 0.361	**0.241 / 0.277**	0.280 / 0.342	**0.248 / 0.338**	0.348 / 0.424	**0.247 / 0.306**	**0.334 / 0.468**	0.363 / 0.481
PEMS08	1.014 / 0.673	**0.595 / 0.459**	0.605 / 0.496	**0.437 / 0.417**	0.920 / 0.708	**0.635 / 0.555**	1.091 / 0.907	**0.517 / 0.549**
METR-LA	**1.430** / 0.739	1.433 / **0.715**	**1.187** / 0.709	1.277 / **0.702**	1.671 / 0.896	**1.402 / 0.827**	1.884 / 0.975	**1.714 / 0.913**
PEMS04	1.256 / 0.867	**0.585 / 0.570**	0.726 / 0.682	**0.457 / 0.539**	0.694 / 0.689	**0.440 / 0.523**	0.718 / 0.884	**0.377 / 0.533**
Electricity	0.212 / 0.292	**0.198 / 0.278**	**0.248 / 0.328**	0.252 / 0.330	0.291 / 0.399	**0.247 / 0.345**	0.519 / 0.549	**0.474 / 0.513**
PEMS-BAY	1.449 / 0.642	**0.904 / 0.430**	1.024 / 0.548	**0.701 / 0.406**	1.576 / 0.801	**0.973 / 0.540**	1.300 / 0.716	**1.199 / 0.657**
PEMS03	1.154 / 0.815	**0.450 / 0.486**	0.627 / 0.617	**0.372 / 0.478**	0.596 / 0.640	**0.391 / 0.484**	0.499 / 0.605	**0.361 / 0.490**
PEMS07	1.178 / 0.817	**0.424 / 0.476**	0.439 / 0.475	**0.309 / 0.383**	0.369 / 0.491	**0.244 / 0.383**	0.734 / 0.718	**0.248 / 0.387**

Prediction horizon 192. Table 21 reports the same CI comparison at a longer forecasting horizon.

Prediction horizon 336. Table 22 reports the CI results at the 336-step horizon.

Prediction horizon 720. Table 23 reports the CI results at the longest standard horizon.

C.10.2 RQ2 Prediction-Horizon Results

Prediction horizon 96. Table 24 reports the CD comparison at prediction horizon 96.

Prediction horizon 192. Table 25 reports the CD comparison at prediction horizon 192.

Prediction horizon 336. Table 26 reports the CD comparison at prediction horizon 336.

Prediction horizon 720. Table 27 reports the CD comparison at prediction horizon 720.

C.10.3 Additional Fixed-Horizon Results

Traffic Results at a Longer Horizon. Table 28 provides a supplementary comparison at a longer input–output horizon.

Table 22: RQ1 CI results at prediction horizon 336 (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.489 / 0.449	**0.466 / 0.431**	**0.493 / 0.457**	0.494 / 0.459	0.500 / **0.462**	**0.496** / 0.470	**0.532 / 0.492**	0.566 / 0.507
ETTh2	0.425 / 0.434	**0.387 / 0.399**	0.442 / 0.451	**0.416 / 0.424**	0.435 / 0.445	**0.421 / 0.443**	0.454 / 0.447	**0.434 / 0.432**
ETTm1	0.391 / 0.400	**0.389 / 0.390**	0.432 / 0.418	**0.410 / 0.410**	0.422 / 0.417	**0.397 / 0.412**	0.417 / 0.425	**0.401 / 0.411**
ETTm2	0.296 / 0.336	**0.289 / 0.324**	0.318 / 0.354	**0.303 / 0.337**	0.310 / **0.346**	**0.305** / 0.347	0.318 / 0.356	**0.299 / 0.338**
Exchange	0.323 / 0.409	**0.306 / 0.388**	0.351 / 0.433	**0.349 / 0.425**	0.383 / 0.472	**0.299 / 0.343**	0.376 / 0.452	**0.340 / 0.419**
Weather	0.266 / 0.291	**0.252 / 0.280**	0.296 / 0.313	**0.267 / 0.291**	**0.274 / 0.299**	0.275 / 0.319	0.275 / 0.303	**0.262 / 0.289**
Solar	0.281 / 0.298	**0.247 / 0.272**	0.386 / 0.367	**0.254 / 0.281**	0.400 / 0.384	**0.230 / 0.304**	0.274 / 0.306	**0.271 / 0.299**
PEMS08	0.869 / 0.585	**0.629 / 0.479**	1.099 / 0.710	**0.636 / 0.492**	1.107 / 0.722	**0.532 / 0.457**	0.680 / 0.513	**0.561 / 0.452**
METR-LA	1.408 / 0.720	**1.363 / 0.697**	1.491 / 0.764	**1.403 / 0.743**	1.510 / 0.799	**1.277 / 0.760**	1.410 / 0.767	**1.371 / 0.742**
PEMS04	0.685 / 0.574	**0.462 / 0.470**	0.924 / 0.697	**0.445 / 0.472**	0.917 / 0.697	**0.375 / 0.439**	0.526 / 0.510	**0.367 / 0.421**
Electricity	0.204 / 0.277	**0.193 / 0.264**	0.228 / 0.306	**0.211 / 0.289**	0.237 / 0.317	**0.218 / 0.299**	0.202 / 0.278	**0.189 / 0.267**
PEMS-BAY	1.166 / 0.547	**1.032 / 0.504**	1.278 / 0.588	**1.080 / 0.517**	1.408 / 0.677	**0.926 / 0.513**	1.115 / 0.541	**0.953 / 0.485**
PEMS03	0.441 / 0.452	**0.330 / 0.422**	0.826 / 0.634	**0.355 / 0.408**	0.832 / 0.646	**0.314 / 0.392**	0.407 / 0.437	**0.309 / 0.378**
PEMS07	0.589 / 0.524	**0.343 / 0.401**	0.846 / 0.646	**0.325 / 0.399**	0.878 / 0.674	**0.288 / 0.377**	0.435 / 0.459	**0.290 / 0.373**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.500 / 0.461	**0.463 / 0.423**	**0.456 / 0.463**	0.468 / 0.465	0.586 / 0.545	**0.500 / 0.470**	**0.635** / 0.663	0.663 / **0.656**
ETTh2	0.436 / 0.445	**0.429 / 0.436**	0.444 / 0.459	**0.429 / 0.452**	0.515 / 0.528	**0.460 / 0.479**	0.633 / 0.644	**0.588 / 0.600**
ETTm1	**0.419 / 0.414**	0.433 / 0.420	0.417 / 0.432	**0.409 / 0.424**	0.519 / 0.505	**0.457 / 0.453**	0.585 / 0.618	**0.584 / 0.585**
ETTm2	0.311 / 0.347	**0.299 / 0.333**	0.319 / 0.355	**0.311 / 0.347**	0.375 / 0.422	**0.346 / 0.382**	0.483 / 0.503	**0.420 / 0.480**
Exchange	**0.333 / 0.421**	0.493 / 0.481	0.410 / 0.469	**0.334 / 0.416**	0.412 / 0.513	**0.385 / 0.467**	0.497 / 0.610	**0.396 / 0.480**
Weather	0.269 / 0.293	**0.265 / 0.289**	0.275 / 0.324	**0.268 / 0.321**	0.364 / 0.386	**0.301 / 0.331**	0.367 / 0.454	**0.362 / 0.437**
Solar	0.392 / 0.375	**0.265 / 0.297**	0.273 / 0.332	**0.250 / 0.324**	0.389 / 0.443	**0.243 / 0.306**	0.347 / 0.453	**0.334 / 0.408**
PEMS08	1.122 / 0.725	**0.638 / 0.495**	0.739 / 0.571	**0.547 / 0.476**	1.006 / 0.755	**0.690 / 0.575**	0.735 / 0.644	**0.677 / 0.598**
METR-LA	1.491 / 0.774	**1.407 / 0.745**	1.293 / 0.790	**1.257 / 0.753**	1.774 / 0.936	**1.481 / 0.860**	1.999 / 1.083	**1.890 / 1.023**
PEMS04	0.933 / 0.704	**0.459 / 0.479**	0.675 / 0.665	**0.464 / 0.534**	0.677 / 0.678	**0.419 / 0.501**	0.495 / 0.621	**0.447 / 0.574**
Electricity	0.227 / 0.306	**0.211 / 0.289**	**0.272 / 0.348**	0.282 / 0.355	0.309 / 0.414	**0.261 / 0.357**	0.542 / 0.564	**0.498 / 0.531**
PEMS-BAY	1.288 / 0.601	**1.131 / 0.537**	1.167 / 0.598	**1.017 / 0.546**	1.504 / 0.796	**1.089 / 0.639**	1.630 / 0.780	**1.450 / 0.750**
PEMS03	0.843 / 0.649	**0.368 / 0.416**	0.517 / 0.546	**0.384 / 0.479**	0.531 / 0.583	**0.364 / 0.457**	0.541 / 0.615	**0.351 / 0.471**
PEMS07	0.859 / 0.659	**0.353 / 0.417**	0.453 / 0.478	**0.313 / 0.403**	0.428 / 0.535	**0.295 / 0.414**	0.324 / 0.455	**0.313 / 0.449**

Table 23: RQ1 CI results at prediction horizon 720 (MSE / MAE).

Dataset	FITS		CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.497 / 0.472	**0.470 / 0.451**	**0.494 / 0.483**	0.512 / 0.486	**0.494 / 0.477**	0.509 / 0.507	0.588 / 0.537	**0.541 / 0.500**
ETTh2	0.429 / 0.441	**0.406 / 0.421**	0.459 / 0.468	**0.420 / 0.438**	**0.449 / 0.465**	0.457 / 0.467	0.500 / 0.490	**0.452 / 0.451**
ETTm1	**0.455** / 0.436	0.460 / **0.433**	0.493 / 0.453	**0.474 / 0.448**	0.489 / 0.453	**0.450 / 0.444**	0.473 / 0.460	**0.457 / 0.446**
ETTm2	0.393 / 0.393	**0.389 / 0.380**	0.421 / 0.412	**0.410 / 0.398**	**0.414 / 0.404**	0.465 / 0.440	0.434 / 0.425	**0.411 / 0.404**
Exchange	0.875 / 0.699	**0.763 / 0.639**	0.918 / 0.732	**0.788 / 0.665**	**0.526** / 0.409	0.652 / **0.366**	1.165 / 0.816	**1.080 / 0.764**
Weather	0.343 / 0.340	**0.329 / 0.327**	0.370 / 0.360	**0.347 / 0.342**	0.357 / **0.353**	**0.350** / 0.371	0.353 / 0.354	**0.345 / 0.343**
Solar	0.281 / 0.297	**0.243 / 0.271**	0.388 / 0.361	**0.255 / 0.276**	0.401 / 0.376	**0.222 / 0.292**	**0.268** / 0.306	0.277 / **0.302**
PEMS08	0.996 / 0.656	**0.740 / 0.546**	1.236 / 0.783	**0.714 / 0.544**	1.272 / 0.807	**0.571 / 0.480**	0.772 / 0.571	**0.627 / 0.498**
METR-LA	1.718 / 0.814	**1.600 / 0.778**	1.829 / 0.870	**1.658 / 0.813**	1.864 / 0.920	**1.447 / 0.804**	1.646 / 0.846	**1.598 / 0.818**
PEMS04	0.806 / 0.645	**0.570 / 0.539**	1.071 / 0.777	**0.505 / 0.512**	1.077 / 0.785	**0.392 / 0.452**	0.622 / 0.572	**0.439 / 0.473**
Electricity	0.246 / 0.309	**0.236 / 0.299**	0.271 / 0.337	**0.244 / 0.312**	0.280 / 0.347	**0.248 / 0.322**	0.245 / 0.311	**0.229 / 0.297**
PEMS-BAY	1.164 / 0.554	**0.854 / 0.475**	1.411 / 0.625	**1.233 / 0.564**	1.564 / 0.728	**1.036 / 0.556**	1.235 / 0.576	**1.091 / 0.534**
PEMS03	0.499 / 0.521	**0.380 / 0.441**	0.973 / 0.715	**0.419 / 0.451**	0.969 / 0.719	**0.345 / 0.414**	0.492 / 0.496	**0.376 / 0.426**
PEMS07	0.668 / 0.573	**0.422 / 0.458**	1.055 / 0.763	**0.498 / 0.511**	1.102 / 0.786	**0.454 / 0.466**	0.499 / 0.495	**0.326 / 0.397**

Dataset	RLinear		SegRNN		TCN		GRU-CI	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.497 / 0.478	**0.475 / 0.448**	**0.456 / 0.477**	0.481 / 0.498	**0.578 / 0.569**	0.710 / 0.584	**0.731 / 0.729**	0.766 / 0.744
ETTh2	0.446 / 0.458	**0.438 / 0.448**	0.467 / 0.488	**0.441 / 0.479**	0.525 / 0.553	**0.506 / 0.517**	0.655 / 0.652	**0.644 / 0.647**
ETTm1	0.486 / 0.449	**0.462 / 0.445**	**0.470** / 0.460	**0.470 / 0.458**	0.596 / 0.548	**0.526 / 0.501**	0.654 / 0.644	**0.653 / 0.632**
ETTm2	**0.414** / 0.406	0.419 / **0.400**	**0.420** / 0.437	0.424 / **0.418**	0.496 / 0.492	**0.459 / 0.465**	0.620 / 0.596	**0.548 / 0.554**
Exchange	**0.864 / 0.706**	0.922 / 0.708	**1.079 / 0.793**	1.189 / 0.828	**1.108** / 0.882	1.198 / **0.865**	**1.451** / 1.092	1.747 / **1.051**
Weather	0.350 / 0.346	**0.343 / 0.339**	0.364 / 0.376	**0.322 / 0.359**	0.452 / 0.436	**0.387 / 0.394**	0.480 / 0.516	**0.474 / 0.503**
Solar	0.393 / 0.368	**0.253 / 0.286**	**0.267 / 0.325**	0.276 / 0.352	0.392 / 0.438	**0.235 / 0.309**	0.332 / 0.445	**0.331 / 0.424**
PEMS08	1.272 / 0.804	**0.718 / 0.548**	0.946 / 0.694	**0.670 / 0.545**	1.122 / 0.821	**0.797 / 0.638**	0.861 / 0.680	**0.809 / 0.665**
METR-LA	1.832 / 0.887	**1.671 / 0.822**	1.456 / 0.841	**1.416 / 0.794**	2.150 / 1.053	**1.710 / 0.940**	2.147 / 1.132	**2.121 / 1.066**
PEMS04	1.084 / 0.787	**0.521 / 0.520**	0.844 / 0.771	**0.566 / 0.606**	0.752 / 0.724	**0.492 / 0.549**	0.559 / 0.671	**0.499 / 0.613**
Electricity	0.270 / 0.336	**0.246 / 0.313**	**0.316 / 0.381**	0.337 / 0.395	0.366 / 0.455	**0.306 / 0.389**	0.584 / 0.590	**0.540 / 0.549**
PEMS-BAY	1.422 / 0.642	**1.213 / 0.562**	1.284 / 0.640	**1.140 / 0.599**	1.712 / 0.810	**1.394 / 0.732**	2.598 / 1.025	**1.707 / 0.762**
PEMS03	0.984 / 0.726	**0.437 / 0.462**	0.612 / 0.615	**0.424 / 0.508**	0.638 / 0.650	**0.448 / 0.513**	0.611 / 0.660	**0.414 / 0.526**
PEMS07	1.077 / 0.769	**0.530 / 0.512**	0.466 / 0.483	**0.321 / 0.381**	0.519 / 0.606	**0.343 / 0.470**	0.826 / 0.819	**0.344 / 0.460**

Table 24: RQ2 CD results at prediction horizon 96 (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.401 / 0.421	**0.366 / 0.402**	**0.396 / 0.405**	0.426 / 0.451	0.495 / 0.610	**0.375 / 0.458**	**0.377 / 0.415**	0.474 / 0.476
ETTh2	**0.318 / 0.376**	0.366 / 0.402	**0.300 / 0.350**	0.426 / 0.451	0.495 / 0.610	**0.375 / 0.458**	**0.335 / 0.374**	0.474 / 0.476
ETTm1	**0.321 / 0.372**	0.366 / 0.402	**0.331 / 0.368**	0.426 / 0.451	0.495 / 0.610	**0.375 / 0.458**	**0.357 / 0.385**	0.474 / 0.476
ETTm2	**0.185 / 0.276**	0.366 / 0.402	**0.173 / 0.257**	0.426 / 0.451	0.495 / 0.610	**0.375 / 0.458**	**0.186 / 0.269**	0.474 / 0.476
Exchange	**0.096 / 0.220**	0.366 / 0.402	**0.090 / 0.209**	0.426 / 0.451	0.495 / 0.610	**0.338 / 0.413**	**0.102 / 0.231**	0.474 / 0.476
Weather	**0.166 / 0.244**	0.366 / 0.402	**0.161 / 0.208**	0.426 / 0.451	0.495 / 0.610	**0.375 / 0.458**	**0.174 / 0.215**	0.474 / 0.476
Solar	**0.191 / 0.266**	0.366 / 0.402	**0.202 / 0.249**	0.426 / 0.451	0.545 / 0.671	**0.413 / 0.504**	**0.207 / 0.252**	0.474 / 0.476
PEMS08	0.408 / 0.436	**0.366 / 0.402**	0.493 / 0.496	**0.426 / 0.451**	0.495 / 0.610	**0.375 / 0.458**	0.501 / 0.498	**0.474 / 0.476**
METR-LA	1.167 / 0.674	**1.092 / 0.639**	1.188 / 0.639	**1.123 / 0.618**	1.507 / 0.807	**1.379 / 0.776**	1.203 / 0.666	**1.158 / 0.647**
PEMS04	0.171 / 0.306	**0.163 / 0.291**	0.359 / 0.449	**0.330 / 0.420**	0.490 / 0.590	**0.300 / 0.453**	0.427 / 0.478	**0.410 / 0.461**
Electricity	0.193 / 0.296	**0.191 / 0.290**	0.176 / 0.262	**0.170 / 0.254**	0.456 / 0.508	**0.434 / 0.495**	0.183 / 0.263	**0.176 / 0.255**
PEMS-BAY	0.915 / 0.511	**0.819 / 0.469**	0.928 / 0.487	**0.805 / 0.452**	1.466 / 0.679	**1.210 / 0.624**	1.190 / 0.544	**1.119 / 0.519**
PEMS03	**0.228 / 0.338**	0.366 / 0.402	**0.378 / 0.436**	0.426 / 0.451	0.545 / 0.671	**0.413 / 0.504**	**0.376 / 0.446**	0.474 / 0.476
PEMS07	0.186 / 0.307	**0.175 / 0.291**	0.272 / 0.378	**0.236 / 0.347**	0.384 / 0.474	**0.213 / 0.344**	0.322 / 0.412	**0.303 / 0.391**

Table 25: RQ2 CD results at prediction horizon 192 (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	**0.446** / 0.451	0.465 / **0.437**	**0.445** / **0.431**	0.585 / 0.503	0.740 / 0.695	**0.526** / **0.525**	**0.441** / **0.445**	0.639 / 0.529
ETTh2	**0.420** / **0.436**	0.465 / 0.437	**0.386** / **0.401**	0.585 / 0.503	0.740 / 0.695	**0.526** / **0.525**	**0.405** / **0.415**	0.639 / 0.529
ETTm1	**0.366** / **0.397**	0.465 / 0.437	**0.393** / **0.401**	0.585 / 0.503	0.740 / 0.695	**0.526** / **0.525**	**0.389** / **0.403**	0.639 / 0.529
ETTm2	**0.259** / **0.327**	0.465 / 0.437	**0.239** / **0.300**	0.585 / 0.503	0.740 / 0.695	**0.526** / **0.525**	**0.257** / **0.317**	0.639 / 0.529
Exchange	**0.192** / **0.316**	0.465 / 0.437	**0.184** / **0.307**	0.585 / 0.503	0.740 / 0.695	**0.473** / **0.473**	**0.202** / **0.328**	0.639 / 0.529
Weather	**0.226** / **0.299**	0.465 / 0.437	**0.209** / **0.253**	0.585 / 0.503	0.740 / 0.695	**0.526** / **0.525**	**0.227** / **0.262**	0.639 / 0.529
Solar	**0.229** / **0.280**	0.465 / 0.437	**0.251** / **0.284**	0.585 / 0.503	0.814 / 0.765	**0.579** / **0.578**	**0.245** / **0.282**	0.639 / 0.529
PEMS08	0.517 / 0.474	**0.465** / **0.437**	0.680 / 0.548	**0.585** / **0.503**	0.740 / 0.695	**0.526** / **0.525**	0.662 / 0.550	**0.639** / **0.529**
METR-LA	1.285 / 0.707	**1.226** / **0.672**	1.449 / 0.719	**1.334** / **0.710**	1.722 / 0.893	**1.601** / **0.829**	1.400 / 0.753	**1.382** / **0.718**
PEMS04	0.215 / 0.341	**0.199** / **0.324**	0.439 / 0.495	**0.397** / **0.461**	0.541 / 0.695	**0.384** / **0.521**	0.499 / 0.523	**0.468** / **0.498**
Electricity	0.209 / 0.309	**0.200** / **0.297**	0.184 / 0.270	**0.177** / **0.261**	0.471 / 0.518	**0.450** / **0.506**	0.191 / 0.272	**0.185** / **0.264**
PEMS-BAY	0.952 / 0.559	**0.847** / **0.482**	1.056 / 0.532	**0.907** / **0.457**	1.161 / 0.648	**1.113** / **0.607**	1.120 / 0.527	**1.070** / **0.513**
PEMS03	**0.279** / **0.376**	0.465 / 0.437	**0.432** / **0.474**	0.585 / 0.503	0.814 / 0.765	**0.579** / **0.578**	**0.436** / **0.479**	0.639 / 0.529
PEMS07	0.195 / 0.313	**0.181** / **0.296**	0.347 / 0.427	**0.296** / **0.387**	0.465 / 0.500	**0.268** / **0.392**	0.385 / 0.456	**0.363** / **0.433**

Table 26: RQ2 CD results at prediction horizon 336 (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.495 / 0.475	**0.480** / **0.472**	0.495 / 0.459	**0.485** / **0.451**	**0.569** / **0.579**	0.598 / 0.586	**0.470** / 0.473	0.477 / **0.454**
ETTh2	0.450 / 0.465	**0.434** / **0.454**	0.425 / 0.438	**0.405** / **0.418**	0.567 / 0.564	**0.514** / **0.541**	0.434 / 0.441	**0.429** / **0.430**
ETTm1	0.394 / 0.423	**0.389** / **0.410**	0.440 / 0.428	**0.403** / **0.402**	0.529 / 0.536	**0.514** / **0.517**	**0.441** / 0.435	0.446 / **0.426**
ETTm2	0.322 / 0.366	**0.315** / **0.358**	0.303 / 0.341	**0.301** / **0.333**	0.425 / 0.450	**0.368** / **0.425**	0.320 / 0.357	**0.305** / **0.341**
Exchange	0.370 / 0.445	**0.344** / **0.426**	0.363 / 0.438	**0.283** / **0.383**	0.434 / 0.536	**0.376** / **0.445**	**0.334** / **0.423**	0.359 / 0.433
Weather	0.294 / 0.352	**0.277** / **0.334**	0.269 / 0.299	**0.261** / **0.287**	0.333 / 0.404	**0.325** / **0.383**	0.288 / 0.307	**0.275** / **0.294**
Solar	0.261 / 0.339	**0.244** / **0.314**	0.251 / 0.285	**0.247** / **0.274**	0.318 / 0.413	**0.261** / **0.313**	0.248 / 0.284	**0.243** / **0.274**
PEMS08	0.537 / 0.464	**0.483** / **0.423**	0.672 / 0.526	**0.573** / **0.456**	0.679 / 0.581	**0.609** / **0.541**	0.650 / 0.506	**0.618** / **0.483**
METR-LA	1.382 / 0.766	**1.337** / **0.738**	1.459 / 0.736	**1.397** / **0.731**	1.807 / 0.995	**1.709** / **0.942**	1.452 / 0.756	**1.393** / **0.728**
PEMS04	0.227 / 0.347	**0.219** / **0.336**	0.390 / 0.447	**0.357** / **0.419**	0.453 / 0.554	**0.405** / **0.537**	0.419 / 0.461	**0.391** / **0.437**
Electricity	0.229 / 0.325	**0.216** / **0.311**	0.204 / 0.288	**0.197** / **0.280**	0.493 / 0.533	**0.458** / **0.513**	0.211 / 0.290	**0.206** / **0.283**
PEMS-BAY	0.965 / 0.592	**0.856** / **0.507**	1.074 / 0.527	**0.924** / **0.482**	1.515 / 0.717	**1.328** / **0.697**	1.090 / 0.529	**1.086** / **0.514**
PEMS03	0.284 / 0.382	**0.248** / **0.351**	0.371 / 0.421	**0.335** / **0.394**	0.424 / 0.523	**0.371** / **0.490**	0.371 / 0.425	**0.349** / **0.404**
PEMS07	0.203 / 0.322	**0.189** / **0.305**	0.324 / 0.401	**0.275** / **0.365**	0.299 / 0.429	**0.294** / **0.422**	0.340 / 0.414	**0.324** / **0.397**

Table 27: RQ2 CD results at prediction horizon 720 (MSE / MAE).

Dataset	TSMixer		XLinear		GRU-CD		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE	MSE / MAE
ETTh1	0.524 / 0.509	**0.501** / **0.504**	0.506 / 0.488	**0.487** / **0.468**	**0.652** / **0.658**	0.673 / 0.660	0.488 / 0.491	**0.487** / **0.478**
ETTh2	0.481 / 0.494	**0.467** / **0.477**	0.450 / 0.461	**0.433** / **0.444**	0.581 / 0.590	**0.567** / **0.568**	0.458 / 0.464	**0.427** / **0.439**
ETTm1	0.456 / 0.461	**0.443** / **0.445**	0.504 / 0.470	**0.483** / **0.445**	**0.573** / 0.576	0.586 / **0.559**	0.501 / 0.469	**0.486** / **0.455**
ETTm2	**0.422** / **0.423**	0.441 / 0.433	**0.405** / 0.405	0.414 / **0.398**	0.542 / 0.523	**0.498** / **0.495**	0.425 / 0.416	**0.412** / **0.401**
Exchange	1.075 / 0.797	**1.057** / **0.779**	0.932 / 0.731	**0.771** / **0.657**	**1.258** / 0.990	1.605 / **0.989**	1.050 / 0.779	**0.936** / **0.719**
Weather	0.382 / 0.411	**0.354** / **0.389**	0.350 / 0.352	**0.342** / **0.341**	0.434 / 0.469	**0.417** / **0.451**	0.366 / 0.358	**0.350** / **0.342**
Solar	0.244 / 0.320	**0.222** / **0.290**	0.257 / 0.288	**0.249** / **0.272**	0.311 / 0.402	**0.260** / **0.334**	0.258 / 0.287	**0.250** / **0.278**
PEMS08	0.631 / 0.524	**0.570** / **0.484**	0.792 / 0.593	**0.662** / **0.508**	0.789 / 0.633	**0.730** / **0.611**	0.752 / 0.571	**0.720** / **0.549**
METR-LA	1.549 / 0.819	**1.486** / **0.797**	1.730 / 0.818	**1.655** / **0.815**	1.917 / 1.017	**1.910** / **0.985**	1.740 / 0.846	**1.670** / **0.820**
PEMS04	0.299 / 0.406	**0.284** / **0.388**	0.484 / 0.510	**0.464** / **0.491**	0.521 / 0.615	**0.444** / **0.564**	0.515 / 0.524	**0.478** / **0.495**
Electricity	0.260 / 0.351	**0.251** / **0.339**	0.249 / 0.321	**0.242** / **0.317**	0.534 / 0.556	**0.501** / **0.541**	0.257 / 0.326	**0.250** / **0.317**
PEMS-BAY	1.633 / 0.789	**1.461** / **0.728**	1.305 / 0.598	**1.133** / **0.552**	1.901 / 0.833	**1.603** / **0.771**	1.315 / 0.592	**1.258** / **0.569**
PEMS03	0.335 / 0.416	**0.314** / **0.396**	0.464 / 0.486	**0.394** / **0.437**	0.485 / 0.577	**0.429** / **0.526**	0.456 / 0.483	**0.431** / **0.459**
PEMS07	0.265 / 0.371	**0.244** / **0.352**	0.387 / 0.449	**0.334** / **0.410**	0.563 / 0.555	**0.342** / **0.448**	0.415 / 0.466	**0.399** / **0.452**

Table 28: Results on the CA, GBA, GLA, and SD traffic datasets with the 12-to-12 setting for CI models (RMSE / MAPE (%); lower is better). Each Origin/+Snow pair uses the same backbone, and bold indicates the lower value.

Dataset	CycleNet		NLinear		PatchTST	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)
CA	80.32 / 34.51	**67.85 / 30.01**	80.85 / 35.00	**59.26** / 35.05	76.60 / 34.30	**66.82 / 29.40**
GBA	81.53 / 42.39	**68.37 / 36.03**	81.97 / 42.13	**62.74 / 39.03**	77.78 / 41.28	**67.05 / 34.88**
GLA	87.92 / 31.89	**65.47 / 25.84**	88.09 / 32.36	**63.64** / 34.89	82.82 / 31.61	**72.74 / 26.49**
SD	91.90 / 34.66	**79.44 / 29.59**	89.52 / 33.71	**71.50 / 31.94**	84.55 / 32.12	**74.16 / 27.14**

Dataset	RLinear		SegRNN		iTransformer	
Variant	Origin	+Snow	Origin	+Snow	Origin	+Snow
Metric	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)	RMSE / MAPE (%)
CA	80.09 / 34.38	**70.90 / 29.23**	69.86 / 34.00	**53.36 / 29.26**	63.40 / 28.73	**59.24 / 26.52**
GBA	81.34 / 42.14	**72.48 / 36.42**	73.14 / 40.67	**59.89 / 36.42**	66.99 / 35.95	**63.20 / 32.47**
GLA	87.93 / 31.69	**70.73 / 26.78**	79.44 / 31.61	**58.60 / 27.67**	67.79 / 26.40	**63.63 / 24.48**
SD	91.78 / 34.57	**81.64 / 29.84**	82.36 / 32.55	**69.30** / 33.26	74.47 / 27.20	**65.30 / 25.10**