

UNSW SYDNEY



NEURAL INFORMATION PROCESSING SYSTEMS

Robust Residual Correction via Selective Deployment for Time Series Forecasting

Jianxiang Xie · Yuncheng Hua · Mingyue Cheng

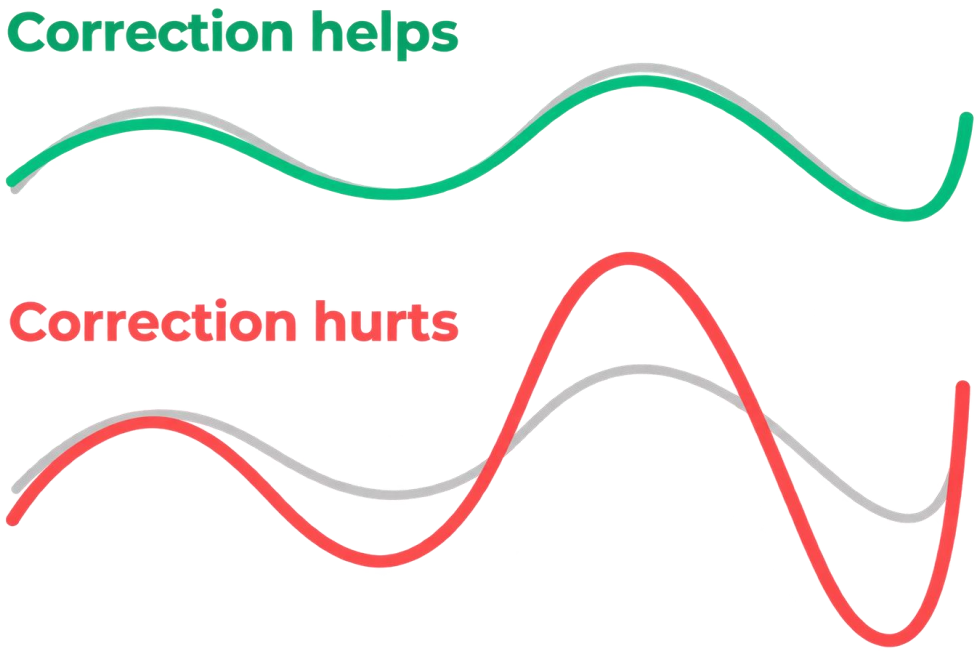
Flora D. Salim · Hao Xue

Motivation

Forecasting models often leave systematic residual errors that can be corrected to improve prediction accuracy.

However, residual patterns are not always predictable. Always-on correction may introduce additional errors and worsen forecasting performance.

Key Insight: Learning a correction is not enough — we must also decide when to deploy it.

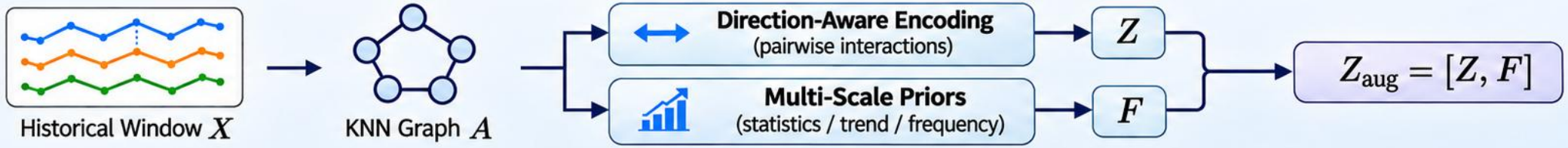


Methodology

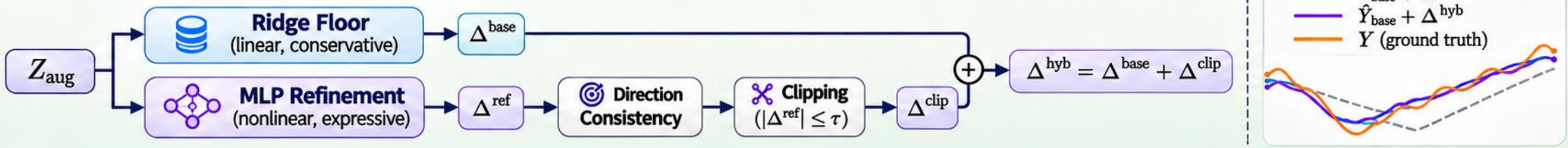
CRC: Selective Residual Correction

$\hat{Y} = \hat{Y}_{\text{base}} + \mathcal{I}(\Delta)$ | Backbone frozen; deploy only the selected correction.

1 Represent: Structured Residual Representation Z_{aug}



2 Propose: Conservative–Conditional Residual Correction



3 Select & Deploy: Validation-Calibrated Deployment



Evaluation and Results

Table 1: Horizon-averaged MSE and MAE improvements of CRC across datasets and backbones (averaged over horizons 96, 192, 336, 720). Improvements are relative to the baseline and positive values indicate gains. Full horizon-wise results are reported in Appendix Table 7.

Dataset	Model	MSE (Baseline)	MSE (CRC)	Imp MSE (%)	MAE (Baseline)	MAE (CRC)	Imp MAE (%)
ETTh1	DLinear	0.4603	0.4582	**0.46**	0.4568	0.4554	**0.29**
	PatchTST	0.4478	0.4468	**0.22**	0.4476	0.4454	**0.50**
	TimeXer	0.4441	0.439	**1.14**	0.4451	0.444	**0.24**
	TimesNet	0.4565	0.4562	**0.06**	0.4528	0.4525	**0.06**
ETTh2	DLinear	0.5641	0.5383	**4.57**	0.5194	0.497	**4.32**
	PatchTST	0.3807	0.3794	**0.35**	0.4092	0.4078	**0.36**
	TimeXer	0.3769	0.3774	-0.13	0.4006	0.4016	-0.24
	TimesNet	0.4082	0.4082	**0.01**	0.4224	0.4219	**0.10**
ETTm1	DLinear	0.404	0.3935	**2.60**	0.4077	0.405	**0.65**
	PatchTST	0.3888	0.3863	**0.64**	0.4029	0.4019	**0.24**
	TimeXer	0.3817	0.3846	-0.76	0.3968	0.3982	-0.35
	TimesNet	0.4199	0.4142	**1.37**	0.4208	0.4195	**0.33**
ETTm2	DLinear	0.3547	0.3375	**4.86**	0.4018	0.3808	**5.21**
	PatchTST	0.2923	0.2888	**1.19**	0.3353	0.333	**0.68**
	TimeXer	0.2735	0.2688	**1.74**	0.3214	0.32	**0.44**
	TimesNet	0.2947	0.2918	**1.01**	0.3328	0.3326	**0.07**
Electricity	DLinear	0.2178	0.1801	**17.30**	0.3082	0.2775	**9.95**
	PatchTST	0.2038	0.1772	**13.05**	0.2943	0.2722	**7.49**
	TimeXer	0.1724	0.1634	**5.28**	0.2707	0.2609	**3.65**
	TimesNet	0.1967	0.1864	**5.22**	0.2963	0.2877	**2.92**
Traffic	DLinear	0.6384	0.5042	**21.03**	0.3946	0.3184	**19.31**
	PatchTST	0.4914	0.4646	**5.47**	0.3151	0.2973	**5.66**
	TimeXer	0.4705	0.4653	**1.11**	0.2875	0.2804	**2.48**
	TimesNet	0.6322	0.6039	**4.47**	0.3353	0.3138	**6.40**
Weather	DLinear	0.2654	0.2385	**10.13**	0.317	0.2833	**10.61**
	PatchTST	0.2582	0.2362	**8.52**	0.2796	0.2676	**4.33**
	TimeXer	0.2406	0.2318	**3.67**	0.2707	0.2651	**2.07**
	TimesNet	0.2631	0.2401	**8.75**	0.2888	0.2746	**4.94**

1. Consistent Improvements
CRC improves horizon-averaged MSE in 26 out of 28 dataset–backbone combinations.

2. Gains Across Diverse Settings
Evaluated on 7 datasets, 4 forecasting backbones, and 4 prediction horizons.

3. Substantial Improvements
- Traffic (DLinear): 21.03% MSE reduction
- Electricity (DLinear): 17.30% MSE reduction
- Weather (DLinear): 10.13% MSE reduction

Takeaway: CRC delivers consistent accuracy gains across diverse forecasting settings.

NDR-Stability

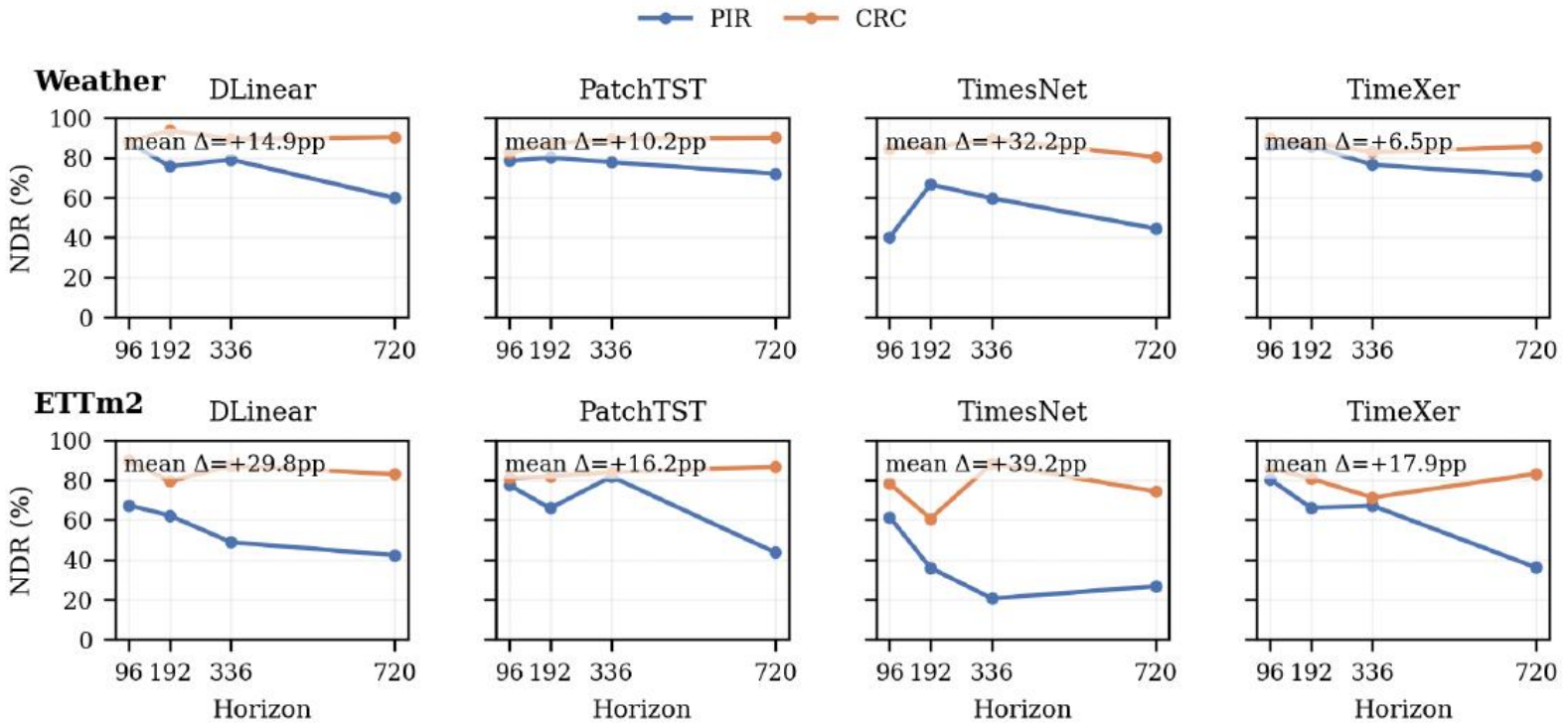


Figure 2: **NDR stability curves.** Non-degradation rate (NDR, %) of PIR vs. CRC across horizons on **Weather** and **ETTm2** (higher is better). Each panel corresponds to one backbone; annotations report mean ΔNDR in percentage points (pp). Exact values are provided in Appendix Table 5.

1. Higher Non-Degradation Rate
CRC consistently achieves higher NDR than PIR across the evaluated datasets, backbones, and horizons.

2. Substantial Stability Gains
•Weather (TimesNet): +32.2 pp
•ETTm2 (TimesNet): +39.2 pp

3. Better Performance at Long Horizons
CRC maintains higher NDR than PIR even at longer forecasting horizons, including H = 720.

Takeaway: Selective deployment helps reduce the risk of harmful residual corrections.

Accuracy–Stability Trade-off

Table 3: Controlled deployment ladder (G0–G5) on B1–B3 (seed 2021). B1/B2/B3: ETTh1+DLinear at pred_len 96/336 and Weather+TimesNet at 96. **Not** horizon-averaged like Table 1. G0 baseline, G1 always-deploy (blend search off; τ/direction gate retained), G2 ridge+validation gate only (no MLP), G3 w/o direction gate, G4 deploy-only encoder, G5 full CRC. Lower MAE/MSE and higher NDR (%) are better.

Config	B1 MAE	B1 MSE	B1 NDR	B2 MAE	B2 MSE	B2 NDR	B3 MAE	B3 MSE	B3 NDR
G0 Baseline	0.4108	0.3962	—	0.4654	0.4874	—	0.2184	0.1685	—
G1 Always-on	0.4738	0.4665	8.2	0.5689	0.6328	10.6	0.2326	0.1803	33.9
G2 Ridge+val gate	0.4097	0.3941	97.9	0.4693	0.4920	97.1	0.2111	0.1626	90.0
G3 No dir gate	0.4079	0.3865	81.8	0.4684	0.4786	78.5	0.2098	0.1611	81.5
G4 Deploy-only	0.4097	0.3941	62.1	0.4592	0.4838	77.8	0.2081	0.1584	79.4
G5 Full CRC	**0.3914**	**0.3780**	**80.3**	**0.4392**	**0.4612**	**79.7**	**0.1903**	**0.1408**	**78.5**

- Always-on (G1): Often worsens accuracy, with low NDR.
- Ridge-only (G2): Highest NDR, but smaller accuracy gains.
- Full CRC (G5): Lowest MAE/MSE across all three settings, while retaining relatively high NDR.

Takeaway: Full CRC improves accuracy without sacrificing too much stability.

Conclusion

•Selective correction matters: Residual correction can help, but always-on deployment may worsen forecasts.

•CRC: Combines structured residual features, conservative correction, and validation-based selection.

•Results: Improves MSE in 26/28 settings and shows higher NDR than PIR in the evaluated comparisons.

•Takeaway: The key is not only how to correct residuals, but also when to apply corrections.