GLARE: GLOBAL AND LOCAL RECONSTRUCTION FOR TIME SERIES FORECASTING WITH EXOGENOUS VARIABLES

Anonymous authors
Paper under double-blind review

ABSTRACT

Exogenous variables can improve time series forecasting through their relationships with the endogenous variable. However, relationships estimated over the full training period may differ from those fitted to recent observations. We propose GLARE, a lightweight forecasting framework that reconstructs the endogenous variable from exogenous variables through global and local relationships. An MLP captures global relationships through a pointwise mapping learned over the full training period. Ridge regression captures local relationships by fitting coefficients to the lookback window of each instance. The global and local reconstructions are blended into a single exogenous reconstruction, and a residual forecaster predicts the remaining residual. On the 12 TSF-X datasets, GLARE achieves state-of-the-art forecasting performance with a single linear residual forecaster and only hundreds to a few thousand additional parameters. Our code is available at: `https://anonymous.4open.science/r/GLARE-6350`.

1 INTRODUCTION

In time series forecasting, exogenous information provides predictive signals that the past values of the target series alone cannot offer (Olivares et al., 2023; Wang et al., 2024b). For example, retail sales respond to scheduled prices and promotions, and electricity prices respond to grid load and wind power generation. This has motivated methods for *time series forecasting with exogenous variables* (TSF-X), which learn relationships between exogenous variables and the endogenous variable that serves as the forecasting target. TimeXer (Wang et al., 2024b) combines self- and cross-attention, while CrossLinear (Zhou et al., 2025) models cross-correlations between variables. Recent methods go beyond historical observations and incorporate future exogenous variables. ExoTST (Tayal et al., 2024) uses cross-temporal fusion, GCGNet (Li et al., 2026) models dependencies with graphs, and DAG (Qiu et al., 2026) uses a dual correlation network with attention transfer.

Beyond architectural design, a key question is whether these relationships should be estimated over the full training period or fitted to recent observations. To examine this, we compare linear regression fits estimated from the full training period (global fit) and the current lookback window (local fit). We evaluate how accurately each fit reconstructs the target from the same future exogenous variables. Figure 1 shows that the proportion of test instances where the local fit yields lower reconstruction error than the global fit ranges from 23% to 94% across datasets. This suggests that the relative predictive value of local and global relationships differs across datasets.

Motivated by this observation, we propose **GLARE** (Global and LocAl REconstruction for Time Series Forecasting with Exogenous Variables), a lightweight framework that combines relationships learned across instances with those fitted directly to recent observations. For global relationships, *Global Exogenous Mapping* learns a pointwise mapping over the full training period with an MLP. For local relationships, *Local Exogenous Regression* fits ridge coefficients to the current lookback window using a closed-form solution. Finally, the two resulting reconstructions are blended, and *Residual Forecaster* predicts the remaining residual. GLARE thus combines information from the full training period and recent observations within a single forecasting framework.

Our contributions are summarized as follows:

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.

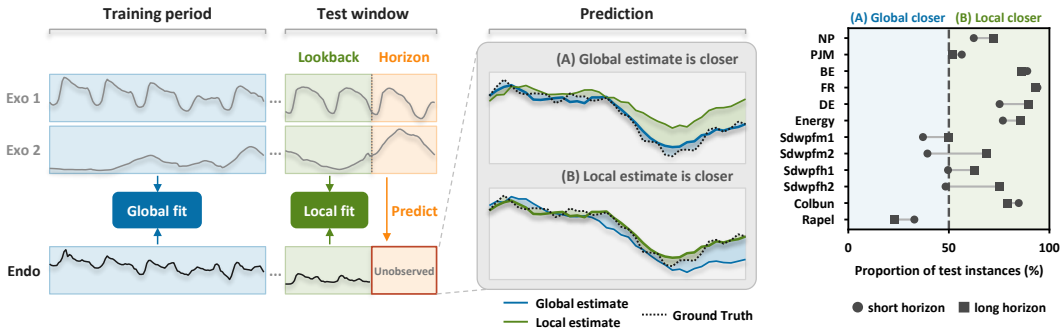

Figure 1: Left: illustration of global and local fits. Right: proportion of test instances where the local fit yields lower reconstruction error than the global fit. Details are provided in Appendix D.1.

- We propose GLARE, a lightweight TSF-X framework that blends global and local exogenous reconstructions and forecasts the remaining residual.
- With a single linear residual forecaster and only hundreds to a few thousand additional parameters, GLARE achieves state-of-the-art performance on 12 TSF-X datasets.
- GLARE also serves as a plugin that incorporates exogenous information into existing univariate backbones and improves their forecasting accuracy.

2 Related Work

The Transformer (Vaswani et al., 2017) has inspired a range of forecasting models that explore different ways to capture temporal and cross-variable relationships, including Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), PatchTST (Nie et al., 2023), iTransformer (Liu et al., 2024), and Non-stationary Transformer (Liu et al., 2022). In parallel, lightweight models such as DLinear (Zeng et al., 2023), FITS (Xu et al., 2024), and SparseTSF (Lin et al., 2024) show that models with few parameters remain competitive. TSF-X methods develop mechanisms for incorporating exogenous information into forecasting. Representative methods include TFT (Lim et al., 2021), NBEATSx (Olivares et al., 2023), TiDE (Das et al., 2023), TimeXer (Wang et al., 2024b), ExoTST (Tayal et al., 2024), CrossLinear (Zhou et al., 2025), DAG (Qiu et al., 2026), and GCGNet (Li et al., 2026). GLARE shows that competitive TSF-X performance can also be achieved with a lightweight design that combines exogenous reconstruction with a single linear residual forecaster. An extended discussion of related work is provided in Appendix A.

3 Method

Problem formulation. Time series forecasting with exogenous variables (TSF-X) distinguishes two types of variables. The *endogenous variable* is the target to be predicted, and the *exogenous variables* are external inputs that help predict it. For each input instance, we are given the endogenous variable $\mathbf{X}^{\text{endo}} \in \mathbb{R}^{L\times 1}$ and exogenous variables $\mathbf{X}^{\text{exo}} \in \mathbb{R}^{L\times C}$ over a lookback window of L steps, where C is the number of exogenous variables. The future exogenous variables $\mathbf{Y}^{\text{exo}} \in \mathbb{R}^{T\times C}$ are also given over a forecast horizon of T steps. They may represent known schedules or externally supplied forecasts, which are not prediction targets. The goal is to learn a forecasting function $\mathcal{F}$ that predicts the future endogenous variable $\mathbf{Y}^{\text{endo}} \in \mathbb{R}^{T\times 1}$:

$$\hat{\mathbf{Y}}^{\text{endo}} = \mathcal{F}\big(\mathbf{X}^{\text{endo}}, \mathbf{X}^{\text{exo}}, \mathbf{Y}^{\text{exo}}\big)\,. \tag{1}$$

The error is computed only on the endogenous variable, using the mean squared error (MSE) or the mean absolute error (MAE). For notational convenience, we concatenate the values over the lookback window and forecast horizon along the time dimension, with a total length of $N = L + T$:

$$\mathbf{Z}^{\text{exo}} = [\mathbf{X}^{\text{exo}}; \mathbf{Y}^{\text{exo}}] \in \mathbb{R}^{N\times C}, \qquad \mathbf{Z}^{\text{endo}} = [\mathbf{X}^{\text{endo}}; \mathbf{Y}^{\text{endo}}] \in \mathbb{R}^{N\times 1}. \tag{2}$$

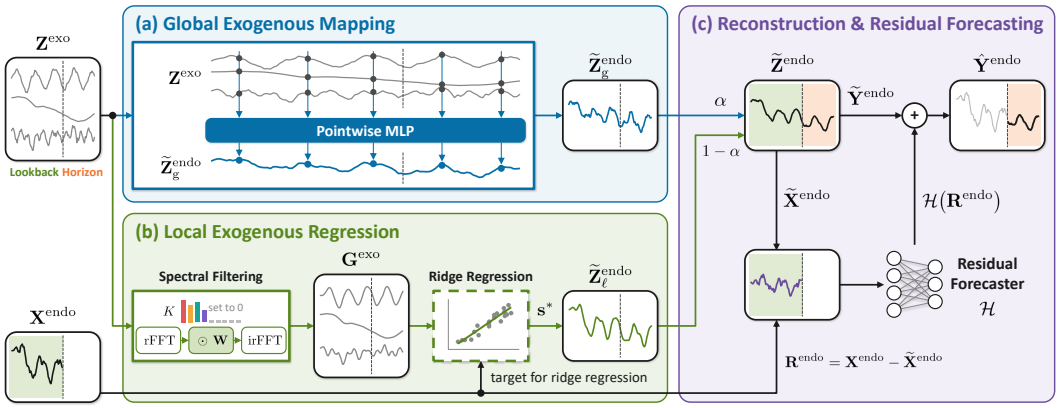

Figure 2: Overall structure of GLARE. (a) *Global Exogenous Mapping*, (b) *Local Exogenous Regression*, and (c) *Residual Forecaster* jointly produce the final prediction $\hat{\mathbf{Y}}^{\text{endo}}$.

3.1 Overview

GLARE follows a simple principle: reconstruct the contribution of exogenous variables first, then forecast what remains. As illustrated in Figure 2, the reconstruction blends two components. *Global Exogenous Mapping* (Sec. 3.2) learns a pointwise mapping from exogenous variables to the endogenous variable over the full training period. *Local Exogenous Regression* (Sec. 3.3) estimates local relationships from the current lookback window with ridge regression. Both produce a reconstruction of the endogenous variable over the lookback window and forecast horizon. We blend the two reconstructions using a weight α. *Residual Forecaster* (Sec. 3.4) then uses the difference between the endogenous history and the lookback reconstruction to predict the future residual.

3.2 Global Exogenous Mapping

To capture global relationships between exogenous and endogenous variables, *Global Exogenous Mapping* learns a mapping from exogenous values to the corresponding endogenous value at each time step. We implement this with a standard pointwise MLP (Tolstikhin et al., 2021; Chen et al., 2023) shared across all time steps and input instances. For $\mathbf{z}_t = (\mathbf{Z}^{\text{exo}}_{t,:})^\top$, the mapping is

$$\left[\widetilde{\mathbf{Z}}^{\text{endo}}_{\text{g}}\right]_t = \text{MLP}(\mathbf{z}_t), \qquad t = 1, \ldots, N. \tag{3}$$

The MLP has one hidden layer of width d with a ReLU activation and a scalar output. The resulting $\widetilde{\mathbf{Z}}^{\text{endo}}_{\text{g}} \in \mathbb{R}^{N\times 1}$ forms the global reconstruction over the lookback window and forecast horizon.

3.3 Local Exogenous Regression

Inspired by classical rolling regression (Brown et al., 1975), *Local Exogenous Regression* estimates local relationships between exogenous and endogenous variables from the current lookback window. Since this estimation is required for each input instance at inference time, we use ridge regression for its closed-form solution and controllable penalty on coefficient magnitudes.

However, ridge regression reconstructs the endogenous variable as a weighted sum of the exogenous variables, where each coefficient scales its corresponding variable by the same amount at every time step. This can lead to inaccurate reconstruction due to delayed endogenous responses or high-frequency noise in the exogenous variables. To address these limitations, we apply spectral filtering to each exogenous variable before fitting the local coefficients.

Spectral filtering. We implement the filter in the frequency domain with variable-specific learnable complex weights. Their phases adjust the phase of each frequency component, while their magnitudes control its amplitude. Let rFFT denote the real discrete Fourier transform along the time dimension and irFFT its inverse. The rFFT produces $F = \lfloor N/2 \rfloor + 1$ complex coefficients

for each exogenous variable. With filter weights $\mathbf{W} \in \mathbb{C}^{F \times C}$, we obtain the filtered exogenous variables over both the lookback window and the forecast horizon:

$$\mathbf{G}^{\text{exo}} = \text{irFFT}(\mathbf{W} \odot \text{rFFT}(\mathbf{Z}^{\text{exo}})) \in \mathbb{R}^{N \times C}, \tag{4}$$

where $\odot$ is the elementwise product. We keep only the lowest K frequency components by fixing the remaining filter weights to zero. This reduces the number of learnable parameters while retaining most of the information in the series (Xu et al., 2024).

Fitting to the observed history. Let $\mathbf{A} = \mathbf{G}^{\text{exo}}_{1:L,:}$ denote the filtered exogenous variables over the lookback window. For each input instance, we obtain the coefficients by solving the following ridge regression problem:

$$\mathbf{s}^* = \arg\min_{\mathbf{s} \in \mathbb{R}^C} \|\mathbf{A}\mathbf{s} - \mathbf{X}^{\text{endo}}\|_2^2 + \lambda \|\mathbf{s}\|_2^2. \tag{5}$$

The fixed hyperparameter $\lambda > 0$ controls the penalty on coefficient magnitudes. The optimal coefficients are obtained in closed form[1]:

$$\mathbf{s}^* = (\mathbf{A}^\top \mathbf{A} + \lambda I)^{-1} \mathbf{A}^\top \mathbf{X}^{\text{endo}}. \tag{6}$$

Applying the fitted coefficients to the filtered exogenous variables over all N steps yields the local reconstruction:

$$\widetilde{\mathbf{Z}}^{\text{endo}}_{\ell} = \mathbf{G}^{\text{exo}} \mathbf{s}^*. \tag{7}$$

3.4 Reconstruction and Residual Forecasting

We blend the global and local reconstructions into a single exogenous reconstruction. The blending weight $\alpha \in [0, 1]$ is a dataset-specific hyperparameter that remains fixed across input instances:

$$\widetilde{\mathbf{Z}}^{\text{endo}} = \alpha\, \widetilde{\mathbf{Z}}^{\text{endo}}_{\text{g}} + (1 - \alpha)\, \widetilde{\mathbf{Z}}^{\text{endo}}_{\ell}. \tag{8}$$

Residual Forecaster predicts the part of the endogenous variable that the exogenous reconstruction does not explain. We split the exogenous reconstruction into the lookback and horizon parts as $\widetilde{\mathbf{Z}}^{\text{endo}} = [\widetilde{\mathbf{X}}^{\text{endo}}; \widetilde{\mathbf{Y}}^{\text{endo}}]$. A single linear forecaster $\mathcal{H}$ takes the lookback residual $\mathbf{R}^{\text{endo}} = \mathbf{X}^{\text{endo}} - \widetilde{\mathbf{X}}^{\text{endo}} \in \mathbb{R}^{L \times 1}$ as input and predicts its future values. Adding these predictions to $\widetilde{\mathbf{Y}}^{\text{endo}}$ yields the final forecast:

$$\hat{\mathbf{Y}}^{\text{endo}} = \widetilde{\mathbf{Y}}^{\text{endo}} + \mathcal{H}\big(\mathbf{R}^{\text{endo}}\big). \tag{9}$$

Finally, we train the pointwise MLP, the filter weights $\mathbf{W}$, and the residual forecaster $\mathcal{H}$ jointly. The loss $\mathcal{L}$ is the distance $\mathcal{D}$ (e.g., MAE) between the prediction and the future endogenous variable:

$$\mathcal{L} = \mathcal{D}\Big(\hat{\mathbf{Y}}^{\text{endo}}, \mathbf{Y}^{\text{endo}}\Big). \tag{10}$$

4 Theoretical Analysis

Refitting local coefficients for each instance raises the question of whether small changes in the endogenous history can be amplified during reconstruction and forecasting. To isolate this effect, we hold the exogenous variables fixed and derive sensitivity bounds for the lookback residual, future reconstruction, and final prediction. We express these bounds in terms of α and λ.

Throughout this section, we fix the exogenous variables, learned parameters, $\lambda > 0$, and $\alpha \in [0, 1]$. We perturb the endogenous history $\mathbf{X}^{\text{endo}}$ by $\boldsymbol{\delta} \in \mathbb{R}^{L \times 1}$ and refit the local coefficients.

Proposition 1 (Lookback residual)**.** *Let $\Delta\mathbf{R}^{\text{endo}}$ denote the resulting change in the lookback residual. Then,*

$$\alpha \|\boldsymbol{\delta}\|_2 \le \|\Delta\mathbf{R}^{\text{endo}}\|_2 \le \|\boldsymbol{\delta}\|_2. \tag{11}$$

The upper bound ensures that the lookback residual does not amplify changes in the endogenous history. When $\alpha > 0$, the lower bound prevents local fitting from eliminating these changes.

[1]When $C > L$, the equivalent dual form $\mathbf{s}^* = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top + \lambda I)^{-1}\mathbf{X}^{\text{endo}}$ reduces the system size from $C \times C$ to $L \times L$.

Proposition 2 (Future reconstruction). *Let $\mathbf{G}_{\text{fut}} = \mathbf{G}^{\text{exo}}_{L+1:N,:} \in \mathbb{R}^{T\times C}$ denote the filtered exogenous variables over the forecast horizon. Let $\Delta\widetilde{\mathbf{Y}}^{\text{endo}}$ denote the resulting change in the blended future reconstruction. Then,*

$$\|\Delta\widetilde{\mathbf{Y}}^{\text{endo}}\|_2 \leq \frac{(1-\alpha)\|\mathbf{G}_{\text{fut}}\|_2}{2\sqrt{\lambda}}\|\boldsymbol{\delta}\|_2. \tag{12}$$

The bound decreases as λ or α increases. A larger λ regularizes the fitted coefficients more strongly, while a larger α reduces the contribution of local reconstruction.

Together, these results bound the sensitivity of the final prediction for a linear residual forecaster.

Corollary 1 (Final prediction). *Let $\mathcal{H}$ be a linear residual forecaster with $\mathcal{H}(\mathbf{R}^{\text{endo}}) = \mathbf{V}\mathbf{R}^{\text{endo}} + \mathbf{b}$, where $\mathbf{V} \in \mathbb{R}^{T\times L}$. Let $\Delta\hat{\mathbf{Y}}^{\text{endo}}$ denote the resulting change in the final prediction. Then,*

$$\|\Delta\hat{\mathbf{Y}}^{\text{endo}}\|_2 \leq \left(\|\mathbf{V}\|_2 + \frac{(1-\alpha)\|\mathbf{G}_{\text{fut}}\|_2}{2\sqrt{\lambda}}\right)\|\boldsymbol{\delta}\|_2. \tag{13}$$

The final prediction therefore has bounded sensitivity to changes in $\mathbf{X}^{\text{endo}}$, even when the local coefficients are refit. Proofs of both propositions and the corollary are provided in Appendix B.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate GLARE on 12 TSF-X datasets: five electricity price datasets (NP, PJM, BE, FR, and DE) (Lago et al., 2021) and seven datasets (Energy, Colbun, Rapel, Sdwpfm1, Sdwpfm2, Sdwpfh1, and Sdwpfh2) released with DAG (Qiu et al., 2026). We also evaluate on widely used multivariate benchmarks: ETT, Weather, Electricity, Traffic, and Exchange (Zhou et al., 2021; Wu et al., 2021; Lai et al., 2018). For these benchmarks, we use the last variable as the endogenous variable and the others as exogenous variables, and report results in Appendix D.2. Dataset details are provided in Appendix C.1.

Baselines. We consider six baselines that exploit exogenous variables: DAG (Qiu et al., 2026), GCGNet (Li et al., 2026), CrossLinear (Zhou et al., 2025), TimeXer (Wang et al., 2024b), TiDE (Das et al., 2023), and TFT (Lim et al., 2021). We also include four univariate forecasting backbones to evaluate GLARE as a plugin: TimeKAN (Huang et al., 2025), xPatch (Stitsyuk & Choi, 2025), DLinear (Zeng et al., 2023), and PatchTST (Nie et al., 2023). Following the protocol of DAG and GCGNet, we use MLP fusion for methods that do not natively support future exogenous variables (all baselines except DAG, GCGNet, TFT, TiDE, and TimeXer). MLP fusion takes the concatenation of the backbone output and future exogenous variables as input and produces the final prediction.

Evaluation. We follow the evaluation protocol of TFB (Qiu et al., 2024) and report MSE and MAE. For the TSF-X datasets, we use lookback lengths of 168 and 720 with forecast horizons of 24 and 360. For the daily datasets Colbun and Rapel, we use lookback lengths of 60 and 180 with horizons of 10 and 30. For the multivariate benchmarks, we fix the lookback length at 96 and use horizons of 96, 192, 336, and 720. Implementation details are provided in Appendix C.2.

5.2 Forecasting Performance

Comparison with TSF-X methods. GLARE achieves the lowest MSE and MAE in 22 of the 24 settings in Table 1 using only a *single linear residual forecaster*. GLARE's advantage is more pronounced at longer horizons, where it ranks first on both metrics across all 12 datasets. Its average error reduction over the best baseline in each setting grows from 9.0% at the short horizon to 18.5% at the long horizon for MSE, and from 3.7% to 11.5% for MAE. These results show that combining global and local exogenous reconstruction with a simple residual forecaster can achieve strong forecasting performance.

Table 1: Forecasting performance on the TSF-X task in MSE and MAE. Best results are bold and second-best results are underlined. Standard deviations for GLARE are reported in Table 7.

Methods		GLARE (ours)		DAG (ICML'26)		GCGNet (ICLR'26)		CrossLinear (KDD'25)		TimeXer (NeurIPS'24)		TiDE (TMLR'23)		TFT (IJF'21)	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.180**	**0.219**	0.202	0.237	0.208	0.237	0.210	0.266	0.236	0.266	0.284	0.301	0.219	0.249
	360	**0.380**	**0.388**	0.521	0.451	0.531	0.459	0.531	0.508	0.600	0.475	0.601	0.498	0.539	0.501
PJM	24	0.062	0.152	**0.057**	**0.143**	0.060	0.150	0.088	0.191	0.075	0.166	0.106	0.214	0.095	0.195
	360	**0.121**	**0.214**	0.130	0.218	0.129	0.223	0.135	0.254	0.140	0.231	0.177	0.279	0.133	0.219
BE	24	**0.328**	**0.225**	0.361	0.229	0.350	0.248	0.391	0.259	0.392	0.253	0.426	0.285	0.426	0.272
	360	**0.460**	**0.294**	0.485	0.330	0.511	0.340	0.568	0.416	0.512	0.327	0.571	0.364	0.482	0.310
FR	24	**0.322**	**0.170**	0.355	0.171	0.347	0.188	0.390	0.226	0.366	0.208	0.418	0.255	0.543	0.253
	360	**0.453**	**0.252**	0.473	0.268	0.482	0.279	0.575	0.370	0.489	0.273	0.551	0.308	0.465	0.261
DE	24	**0.273**	**0.318**	0.277	0.322	0.280	0.331	0.387	0.396	0.339	0.362	0.367	0.383	0.380	0.383
	360	**0.382**	**0.389**	0.462	0.418	0.523	0.447	0.583	0.507	0.610	0.474	0.630	0.511	0.599	0.509
Energy	24	**0.051**	**0.174**	0.079	0.215	0.081	0.221	0.241	0.418	0.122	0.273	0.103	0.248	0.093	0.235
	360	**0.091**	**0.236**	0.169	0.320	0.182	0.334	0.237	0.385	0.204	0.357	0.202	0.355	0.167	0.331
Colbun	10	**0.058**	**0.093**	0.061	0.094	0.065	0.108	0.071	0.102	0.113	0.172	0.089	0.131	0.092	0.135
	30	**0.127**	**0.201**	0.135	0.215	0.149	0.243	0.182	0.288	0.176	0.299	0.240	0.322	0.383	0.460
Rapel	10	0.179	0.223	**0.151**	**0.203**	0.211	0.230	0.163	0.209	0.301	0.308	0.228	0.271	0.201	0.253
	30	**0.304**	**0.340**	0.309	0.408	0.401	0.384	0.340	0.417	0.387	0.416	0.411	0.432	0.409	0.414
Sdwpfm1	24	**0.301**	**0.373**	0.351	0.400	0.376	0.415	0.355	0.473	0.558	0.533	0.474	0.488	0.366	0.421
	360	**0.331**	**0.412**	0.495	0.522	0.472	0.499	0.497	0.532	0.845	0.684	0.492	0.526	0.597	0.528
Sdwpfm2	24	**0.339**	**0.406**	0.372	0.414	0.421	0.441	0.477	0.536	0.627	0.570	0.461	0.492	0.411	0.458
	360	**0.369**	**0.442**	0.583	0.556	0.529	0.531	0.589	0.611	0.978	0.736	0.511	0.540	0.541	0.519
Sdwpfh1	24	**0.296**	**0.381**	0.408	0.438	0.435	0.486	0.548	0.585	0.651	0.587	0.434	0.489	0.401	0.460
	360	**0.355**	**0.433**	0.489	0.534	0.465	0.514	0.566	0.601	0.841	0.700	0.472	0.527	0.557	0.523
Sdwpfh2	24	**0.353**	**0.429**	0.438	0.465	0.473	0.506	0.468	0.540	0.820	0.677	0.579	0.553	0.474	0.493
	360	**0.399**	**0.470**	0.608	0.595	0.566	0.565	0.608	0.609	0.962	0.761	0.619	0.614	0.657	0.549
1st count		**22**	**22**	2	2	0	0	0	0	0	0	0	0	0	0

GLARE as a plugin. As shown in Table 2, GLARE improves performance across all evaluated datasets and horizons when each of the four backbones serves as the residual forecaster $\mathcal{H}$. Relative to the backbone alone, it reduces MSE and MAE by 44.1% and 31.1% on average, while MLP fusion reduces them by 14.5% and 3.2%. These results support GLARE as an effective plugin for incorporating exogenous information into univariate forecasting models.

Local fitting and forecasting gains. Our linear regression comparison in the introduction shows that the relative accuracy of global and local fits varies across datasets and horizons. We examine how local win rates relate to GLARE's forecasting gains over DAG on the 12 TSF-X datasets. Figure 3 shows that higher local win rates tend to accompany larger gains within the same dataset type. This trend is clearest for wind power, while Energy exhibits the largest gains over DAG at both horizons. These observations suggest an association between the relative advantage of local fitting and GLARE's forecasting gains.

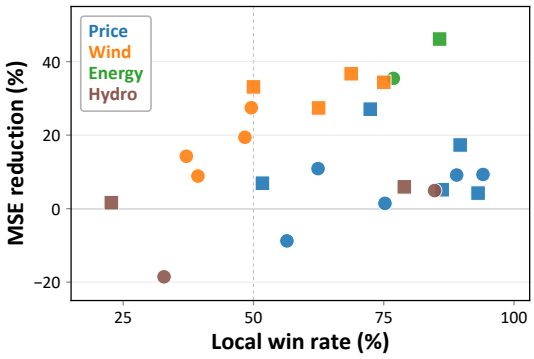

Figure 3: Local win rate versus relative MSE reduction over DAG. Circles and squares indicate short and long horizons, respectively.

Table 2: Forecasting performance of univariate backbones in MSE and MAE. +M and +G denote MLP fusion and GLARE, respectively. Best results for each backbone are bold.

Backbone		TimeKAN (ICLR'25)						xPatch (AAAI'25)						DLinear (AAAI'23)						PatchTST (ICLR'23)					
		MSE			MAE			MSE			MAE			MSE			MAE			MSE			MAE		
Dataset	T	Base	+M	+G	Base	+M	+G	Base	+M	+G	Base	+M	+G	Base	+M	+G	Base	+M	+G	Base	+M	+G	Base	+M	+G
NP	24	0.336	0.273	**0.189**	0.344	0.310	**0.216**	0.260	0.234	**0.187**	0.277	0.278	**0.224**	0.294	0.282	**0.181**	0.308	0.337	**0.225**	0.283	0.249	**0.186**	0.296	0.294	**0.213**
	360	0.632	0.538	**0.362**	0.527	0.529	**0.371**	0.717	0.523	**0.440**	0.528	0.461	**0.424**	0.647	0.601	**0.394**	0.537	0.552	**0.397**	0.630	0.530	**0.446**	0.505	0.498	**0.400**
PJM	24	0.130	0.115	**0.064**	0.236	0.244	**0.152**	0.085	0.077	**0.063**	0.178	0.168	**0.153**	0.109	0.096	**0.062**	0.215	0.208	**0.150**	0.106	0.116	**0.063**	0.212	0.239	**0.151**
	360	0.219	0.162	**0.113**	0.304	0.281	**0.209**	0.193	0.131	**0.120**	0.267	0.220	**0.209**	0.199	0.164	**0.119**	0.285	0.278	**0.212**	0.189	0.149	**0.115**	0.273	0.279	**0.211**
BE	24	0.424	0.451	**0.342**	0.280	0.319	**0.218**	0.404	0.411	**0.352**	0.258	0.256	**0.219**	0.453	0.451	**0.323**	0.308	0.310	**0.226**	0.405	0.452	**0.350**	0.264	0.326	**0.212**
	360	0.552	0.645	**0.445**	0.350	0.495	**0.290**	0.562	0.579	**0.456**	0.346	0.410	**0.302**	0.645	0.600	**0.470**	0.454	0.442	**0.310**	0.564	0.702	**0.450**	0.354	0.538	**0.284**
FR	24	0.448	0.454	**0.324**	0.250	0.296	**0.173**	0.402	0.424	**0.341**	0.208	0.235	**0.169**	0.412	0.422	**0.317**	0.253	0.265	**0.174**	0.397	0.518	**0.325**	0.223	0.368	**0.164**
	360	0.534	0.641	**0.426**	0.303	0.452	**0.221**	0.553	0.556	**0.460**	0.306	0.350	**0.243**	0.576	0.572	**0.450**	0.352	0.362	**0.246**	0.542	0.658	**0.443**	0.306	0.452	**0.233**
DE	24	0.504	0.399	**0.271**	0.462	0.412	**0.311**	0.455	0.399	**0.288**	0.425	0.394	**0.324**	0.511	0.423	**0.271**	0.458	0.426	**0.315**	0.503	0.413	**0.268**	0.450	0.412	**0.307**
	360	0.834	0.547	**0.388**	0.590	0.479	**0.378**	0.810	0.550	**0.429**	0.575	0.472	**0.405**	0.868	0.573	**0.387**	0.600	0.490	**0.381**	0.889	0.589	**0.407**	0.604	0.498	**0.393**
Energy	24	0.109	0.135	**0.052**	0.254	0.298	**0.178**	0.107	0.092	**0.040**	0.252	0.240	**0.153**	0.106	0.085	**0.050**	0.248	0.230	**0.174**	0.108	0.239	**0.045**	0.254	0.390	**0.164**
	360	0.186	0.302	**0.098**	0.338	0.464	**0.246**	0.232	0.204	**0.124**	0.377	0.367	**0.280**	0.214	0.263	**0.092**	0.367	0.416	**0.237**	0.298	0.214	**0.128**	0.428	0.363	**0.288**
Colbun	10	0.091	0.061	**0.056**	0.129	0.101	**0.087**	0.071	0.090	**0.053**	0.092	0.104	**0.091**	0.077	0.069	**0.058**	0.135	0.151	**0.091**	0.083	0.062	**0.058**	0.125	0.122	**0.107**
	30	0.311	0.195	**0.144**	0.377	0.249	**0.225**	0.405	0.217	**0.140**	0.379	0.315	**0.224**	0.318	0.245	**0.110**	0.359	0.285	**0.187**	0.217	0.417	**0.126**	0.316	0.496	**0.237**
Rapel	10	0.226	0.174	**0.151**	0.238	0.231	**0.225**	0.421	0.161	**0.146**	0.421	0.216	**0.202**	0.223	0.184	**0.163**	0.253	0.239	**0.221**	0.215	0.163	**0.143**	0.232	0.218	**0.214**
	30	0.458	0.325	**0.296**	0.437	0.390	**0.356**	0.307	0.505	**0.235**	0.343	0.546	**0.325**	0.414	0.316	**0.314**	0.423	0.410	**0.382**	0.436	0.374	**0.232**	0.436	0.445	**0.313**
Sdwpfm1	24	0.598	0.418	**0.299**	0.579	0.503	**0.373**	0.564	0.669	**0.318**	0.509	0.580	**0.388**	0.547	0.417	**0.331**	0.589	0.501	**0.380**	0.557	0.374	**0.344**	0.557	0.454	**0.384**
	360	0.853	0.476	**0.301**	0.764	0.566	**0.388**	0.910	0.592	**0.316**	0.672	0.556	**0.398**	0.839	0.480	**0.332**	0.773	0.562	**0.398**	0.910	0.497	**0.343**	0.745	0.550	**0.424**
Sdwpfm2	24	0.684	0.474	**0.343**	0.621	0.538	**0.402**	0.641	0.508	**0.375**	0.539	0.478	**0.422**	0.631	0.459	**0.368**	0.631	0.530	**0.402**	0.648	0.418	**0.343**	0.610	0.492	**0.408**
	360	0.987	0.520	**0.343**	0.780	0.589	**0.419**	1.058	0.523	**0.375**	0.726	0.513	**0.436**	0.980	0.589	**0.364**	0.824	0.630	**0.418**	1.020	0.602	**0.381**	0.819	0.603	**0.460**
Sdwpfh1	24	0.755	0.511	**0.294**	0.679	0.582	**0.386**	0.701	0.524	**0.294**	0.579	0.505	**0.385**	0.736	0.557	**0.317**	0.721	0.613	**0.379**	0.669	0.422	**0.312**	0.629	0.505	**0.388**
	360	0.853	0.643	**0.295**	0.762	0.694	**0.385**	0.882	0.472	**0.309**	0.680	0.504	**0.389**	0.829	0.600	**0.342**	0.781	0.635	**0.429**	0.939	0.513	**0.382**	0.804	0.548	**0.455**
Sdwpfh2	24	0.892	0.580	**0.328**	0.740	0.614	**0.393**	0.829	0.640	**0.336**	0.632	0.554	**0.408**	0.838	0.633	**0.351**	0.759	0.653	**0.408**	0.769	0.457	**0.349**	0.680	0.527	**0.423**
	360	0.990	0.713	**0.313**	0.825	0.729	**0.405**	1.053	0.579	**0.342**	0.742	0.547	**0.430**	0.989	0.704	**0.416**	0.837	0.674	**0.477**	1.280	0.641	**0.347**	0.925	0.632	**0.430**

Table 3: Ablation of GLARE components. The left group compares global and local reconstruction, and the right group exogenous reconstruction and residual forecasting. Best results are bold.

Variants		Global vs. Local				Recon. vs. Residual					
		Global only		Local only		Recon only		$\mathcal{H}$ only		Full GLARE	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	0.190	0.226	0.206	0.237	0.255	0.286	0.292	0.295	**0.180**	**0.219**
	360	0.392	0.394	0.478	0.427	0.582	0.542	0.623	0.507	**0.380**	**0.388**
PJM	24	0.066	0.160	0.066	0.153	0.096	0.191	0.090	0.189	**0.062**	**0.152**
	360	0.132	0.225	0.140	0.225	0.132	0.230	0.198	0.278	**0.121**	**0.214**
DE	24	0.424	0.433	**0.253**	**0.312**	0.397	0.396	0.523	0.463	0.273	0.318
	360	0.502	0.444	0.393	0.392	0.454	0.433	0.899	0.613	**0.382**	**0.389**
Colbun	10	0.071	0.105	0.061	0.103	0.313	0.375	0.077	0.098	**0.058**	**0.093**
	30	0.219	0.251	0.181	0.253	0.712	0.575	0.225	0.257	**0.127**	**0.201**
Sdwpfm1	24	0.310	0.379	0.436	0.447	0.369	0.418	0.549	0.493	**0.301**	**0.373**
	360	**0.331**	0.412	0.566	0.569	0.334	**0.400**	0.925	0.699	**0.331**	0.412

5.3 Ablation Study

We evaluate the contributions of GLARE's three components: *Global Exogenous Mapping*, *Local Exogenous Regression*, and *Residual Forecaster*. We also compare alternatives to spectral filtering and the pointwise MLP. These experiments address four questions: (1) Does blending global and local reconstruction improve upon either component alone? (2) Do exogenous reconstruction and residual forecasting each contribute to the final performance? (3) Is spectral filtering a better transformation of the exogenous variables for local fitting than the alternatives? (4) How does the pointwise MLP compare with alternative global mappings? Details of the alternatives are provided in Appendix C.2.

(1) Global and local reconstruction. As described in Appendix C.2, full GLARE blends both reconstructions using α selected from $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. We compare this blend with global-only ($\alpha = 1$) and local-only ($\alpha = 0$) variants in Table 3. Neither component alone is consistently preferable across datasets and horizons, while blending outperforms both standalone variants in most settings. This suggests that relationships learned over the full training period and those fitted to the current lookback window provide complementary information.

(2) Exogenous reconstruction and residual forecasting. We compare GLARE with reconstruction-only and $\mathcal{H}$-only variants in Table 3, where $\mathcal{H}$ only applies the residual forecaster directly to the endogenous history. Although the stronger standalone component varies across datasets, using both together improves forecasting accuracy in most settings. This suggests that the exogenous reconstruction effectively incorporates exogenous information, and the residual forecaster complements it by capturing the remaining temporal structure.

(3) Filtering design. We replace the spectral filter with an identity mapping, a temporal linear layer for each exogenous variable, or a depthwise 1D convolution. Table 4 shows that spectral filtering generally improves upon these alternatives with only $\mathcal{O}(CK)$ additional learnable parameters. These results support spectral filtering as a compact and effective transformation of the full exogenous sequence before local regression.

Table 4: Comparison of filtering methods.

Variants		Identity		Linear		Conv		Spectral (ours)	
Params		0		$\mathcal{O}(CN^2)$		$\mathcal{O}(Ck)$		$\mathcal{O}(CK)$	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	0.194	0.241	0.184	0.225	0.185	0.225	**0.180**	**0.219**
	360	0.398	0.400	0.467	0.460	0.394	0.403	**0.380**	**0.388**
PJM	24	0.062	0.155	**0.061**	0.153	0.062	0.154	0.062	**0.152**
	360	0.132	0.224	0.135	0.223	0.133	0.224	**0.121**	**0.214**
DE	24	0.285	0.328	0.298	0.339	0.290	0.332	**0.273**	**0.318**
	360	0.406	0.400	0.578	0.497	0.426	0.406	**0.382**	**0.389**
Colbun	10	0.092	0.115	0.075	0.114	0.123	0.127	**0.058**	**0.093**
	30	0.458	0.351	0.183	0.234	0.230	0.266	**0.127**	**0.201**
Sdwpfm1	24	0.314	0.379	0.390	0.448	0.314	0.379	**0.301**	**0.373**
	360	0.361	0.428	0.392	0.472	0.363	0.427	**0.331**	**0.412**

(4) Global mapping design. We compare the pointwise MLP in *Global Exogenous Mapping* with linear, convolutional, and attention-based alternatives. Table 5 shows that the MLP performs best in most settings with $\mathcal{O}(Cd)$ parameters. Attention performs better in some settings but requires more parameters and computation. These results support the pointwise MLP as a simple and effective choice for global exogenous mapping.

Table 5: Comparison of global mapping designs.

Variants		Linear		Attention		Conv		MLP (ours)	
Params		$\mathcal{O}(C)$		$\mathcal{O}(Cd + d^2)$		$\mathcal{O}(Ck)$		$\mathcal{O}(Cd)$	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	0.208	0.233	0.182	0.223	0.208	0.232	**0.180**	**0.219**
	360	0.626	0.538	0.395	0.393	0.463	0.418	**0.380**	**0.388**
PJM	24	0.065	0.154	0.063	0.154	0.065	0.154	**0.062**	**0.152**
	360	0.129	0.216	0.123	0.217	0.128	0.217	**0.121**	**0.214**
DE	24	0.276	0.320	**0.271**	**0.316**	0.352	0.389	0.273	0.318
	360	0.386	0.390	0.384	0.393	0.387	0.391	**0.382**	**0.389**
Colbun	10	0.068	0.094	0.063	0.097	0.060	0.097	**0.058**	**0.093**
	30	0.134	0.219	0.145	0.217	0.133	0.207	**0.127**	**0.201**
Sdwpfm1	24	0.381	0.421	0.367	0.397	0.379	0.415	**0.301**	**0.373**
	360	0.411	0.466	0.335	**0.402**	0.406	0.466	**0.331**	0.412

k: kernel size.

5.4 Efficiency Analysis

Figure 4 compares forecasting accuracy against inference time on four settings, with bubble area indicating the parameter count. GLARE lies in the lower left corner in every setting, achieving the lowest error with the fewest parameters and the shortest inference time. Its parameter count is $\mathcal{O}(LT + C(d + K))$, where the linear residual forecaster accounts for the LT term and the global mapping and spectral filter add only $C(d + K)$. The $C(d + K)$ term corresponds to only hundreds to a few thousand additional parameters on the TSF-X datasets.

Even on multivariate benchmarks with hundreds of variables, such as Electricity and Traffic, GLARE remains highly efficient by using the dual form of ridge regression (Sec. 3.3). GLARE also has the lowest estimated FLOPs and peak GPU memory across diverse settings. Full results are reported in Appendix D.4.

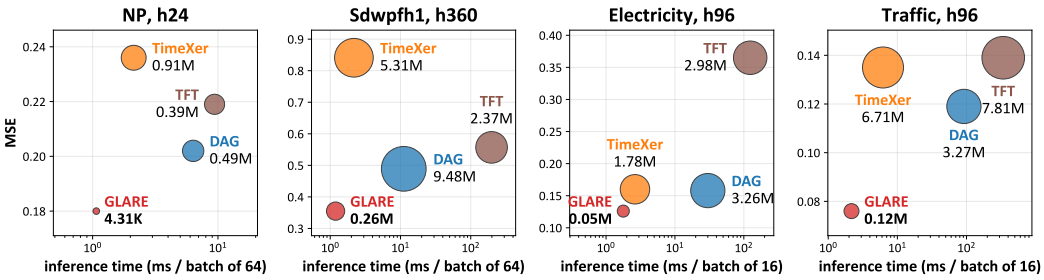

Figure 4: Forecasting accuracy versus inference time. Larger bubbles indicate more parameters.

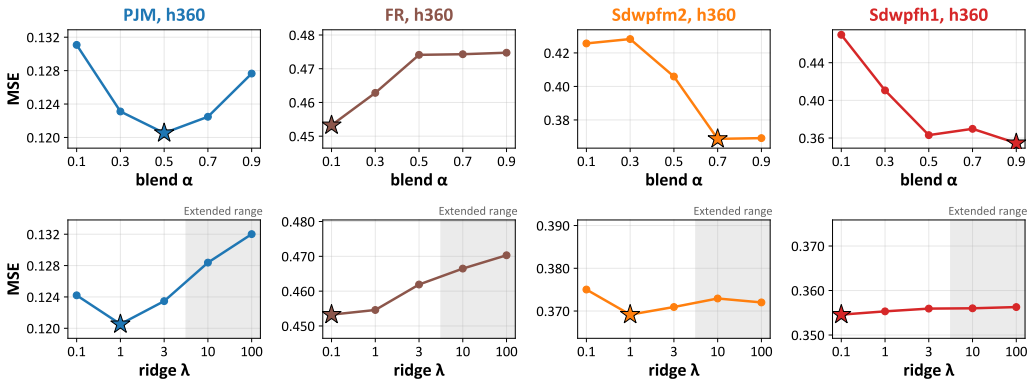

Figure 5: Sensitivity to the blending weight α and ridge strength λ at longer horizons. Shaded regions indicate larger regularization strengths evaluated beyond the original search range.

5.5 Hyperparameter Analysis

We examine GLARE's sensitivity to α and λ in Figure 5. The preferred α differs across datasets, reflecting the varying usefulness of global and local reconstruction. We extend λ beyond the original search range to examine the effect of stronger regularization on local fitting. The increased MSE on PJM and FR at larger λ suggests that excessive regularization can weaken the benefit of local fitting. In contrast, Sdwpfm2 and Sdwpfh1 show smaller changes even in the extended range, consistent with the larger global weights used on these datasets.

6 Limitations

GLARE uses a dataset-specific blending weight α selected from a coarse grid of five values, without fine-grained tuning. However, the relative importance of global and local reconstruction can vary across windows within a dataset, which a fixed weight cannot reflect. GLARE is also designed for settings with available future exogenous variables, and its gains are more limited when these are unavailable (Appendix D.3). Addressing these limitations through adaptive blending and more effective forecasting without future exogenous variables is a direction for future work.

7 Conclusion

We presented GLARE, a framework that combines global exogenous mapping with local exogenous regression and forecasts the remaining residual. The core idea is that relationships learned across instances and those fitted within each instance can complement each other in forecasting. Guided by this design principle, our work demonstrates that a simple architecture can achieve state-of-the-art forecasting performance at low computational cost. An important direction for future work is to explore whether scaling models built on this design principle can improve forecasting accuracy.

AI USE STATEMENT

We used generative AI tools to polish the English writing and to help find related work. Beyond this assistance, all the scientific work was conducted by the authors.

REPRODUCIBILITY STATEMENT

We implement GLARE using the DAG[2] codebase, which is built on TFB[3]. Our code and selected hyperparameters are available at `https://anonymous.4open.science/r/GLARE-6350`.

REFERENCES

R. L. Brown, J. Durbin, and J. M. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society: Series B (Methodological)*, 37(2): 149–163, 1975. doi: 10.1111/j.2517-6161.1975.tb01532.x.

Si-An Chen, Chun-Liang Li, Nathanael Christian Yoder, Sercan O Arik, and Tomas Pfister. TSMixer: An all-MLP architecture for time series forecasting. *Transactions on Machine Learning Research*, 2023.

Abhimanyu Das, Weihao Kong, Andrew Leach, Shaan K Mathur, Rajat Sen, and Rose Yu. Long-term forecasting with TiDE: Time-series dense encoder. *Transactions on Machine Learning Research*, 2023.

Seungha Hong, Sukang Chae, Suyeon Kim, Sanghwan Jang, and Hwanjo Yu. Dynamic multi-period experts for online time series forecasting. In *Proceedings of the ACM Web Conference 2026*, pp. 7669–7678, 2026. doi: 10.1145/3774904.3792731.

Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. TimeKAN: KAN-based frequency decomposition learning architecture for long-term time series forecasting. In *The Thirteenth International Conference on Learning Representations*, 2025.

HyunGi Kim, Siwon Kim, Jisoo Mok, and Sungroh Yoon. Battling the non-stationarity in time series forecasting via test-time adaptation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, 2025. doi: 10.1609/aaai.v39i17.33965.

Jesus Lago, Grzegorz Marcjasz, Bart De Schutter, and Rafał Weron. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy*, 293:116983, 2021. doi: 10.1016/j.apenergy.2021.116983.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 95–104, 2018. doi: 10.1145/3209978.3210006.

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.

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*, 2026.

Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4): 1748–1764, 2021. doi: 10.1016/j.ijforecast.2021.03.012.

[2] `https://github.com/decisionintelligence/DAG`
[3] `https://github.com/decisionintelligence/TFB`

Shengsheng Lin, Weiwei Lin, Wentai Wu, Haojun Chen, and Junjie Yang. SparseTSF: Modeling long-term time series forecasting with 1k parameters. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30211–30226. PMLR, 2024.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *Advances in Neural Information Processing Systems*, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. iTransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Donghao Luo and Xue Wang. ModernTCN: A modern pure convolution structure for general time series analysis. In *The Twelfth International Conference on Learning Representations*, 2024.

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

Kin G. Olivares, Cristian Challu, Grzegorz Marcjasz, Rafał Weron, and Artur Dubrawski. Neural basis expansion analysis with exogenous variables: Forecasting electricity prices with nbeatsx. *International Journal of Forecasting*, 39(2):884–900, 2023. doi: 10.1016/j.ijforecast.2022.03.001.

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. *Proc. VLDB Endow.*, 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 V.1*, pp. 1185–1196, 2025. doi: 10.1145/3690624.3709325.

Xiangfei Qiu, Yuhan Zhu, Zhengyu Li, Xingjian Wu, Bin Yang, and Jilin Hu. DAG: A dual correlation network for time series forecasting with exogenous variables. In *Forty-third International Conference on Machine Learning*, 2026.

Artyom Stitsyuk and Jaesik Choi. xpatch: Dual-stream time series forecasting with exponential seasonal-trend decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20601–20609, 2025.

Kshitij Tayal, Arvind Renganathan, Xiaowei Jia, Vipin Kumar, and Dan Lu. ExoTST: Exogenous-aware temporal sequence transformer for time series prediction. In *2024 IEEE International Conference on Data Mining (ICDM)*, pp. 857–862, 2024. doi: 10.1109/ICDM59182.2024.00105.

Ilya Tolstikhin, Neil Houlsby, Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Thomas Unterthiner, Jessica Yung, Andreas Peter Steiner, Daniel Keysers, Jakob Uszkoreit, Mario Lucic, and Alexey Dosovitskiy. MLP-mixer: An all-MLP architecture for vision. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

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*, 2024a.

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. Timexer: Empowering transformers for time series forecasting with exogenous variables. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b.

Zihan Wang, Fanheng Kong, Shi Feng, Ming Wang, Xiaocui Yang, Han Zhao, Daling Wang, and Yifei Zhang. Is Mamba effective for time series forecasting? *Neurocomputing*, 619:129178, 2025. doi: 10.1016/j.neucom.2024.129178.

Billy M Williams. Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record*, 1776(1):194–200, 2001. doi: 10.3141/1776-25.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023.

Zhijian Xu, Ailing Zeng, and Qiang Xu. FITS: Modeling time series with $10k$ parameters. In *The Twelfth International Conference on Learning Representations*, 2024.

Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. Frequency adaptive normalization for non-stationary time series forecasting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Ying yee Ava Lau, Zhiwen Shao, and Dit-Yan Yeung. Fast and slow streams for online time series forecasting without information leakage. In *The Thirteenth International Conference on Learning Representations*, 2025.

Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, and Zhendong Niu. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023a.

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain MLPs are more effective learners in time series forecasting. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b.

Ailing Zeng, Muxi Chen, Lei Zhang, and Qiang Xu. Are transformers effective for time series forecasting? In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, 2023. doi: 10.1609/aaai.v37i9.26317.

YiFan Zhang, Qingsong Wen, Xue Wang, Weiqi Chen, Liang Sun, Zhang Zhang, Liang Wang, Rong Jin, and Tieniu Tan. OneNet: Enhancing time series forecasting models under concept drift by online ensembling. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Yunhao Zhang and Junchi Yan. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The Eleventh International Conference on Learning Representations*, 2023.

Lifan Zhao and Yanyan Shen. Proactive model adaptation against concept drift for online time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, pp. 2020–2031, 2025. doi: 10.1145/3690624.3709210.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Pengfei Zhou, Yunlong Liu, Junli Liang, Qi Song, and Xiangyang Li. Crosslinear: Plug-and-play cross-correlation embedding for time series forecasting with exogenous variables. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, pp. 4120–4131, 2025. doi: 10.1145/3711896.3736899.

Tian Zhou, Ziqing Ma, Xue Wang, Qingsong Wen, Liang Sun, Tao Yao, Wotao Yin, and Rong Jin. FiLM: Frequency improved legendre memory model for long-term time series forecasting. In *Advances in Neural Information Processing Systems*, 2022a.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 27268–27286. PMLR, 2022b.

A Extended Related Work

A.1 Time Series Forecasting

Transformer-based (Vaswani et al., 2017) forecasting methods explore different ways to represent temporal and cross-variable relationships. Informer (Zhou et al., 2021) reduces attention cost through sparse attention, Autoformer (Wu et al., 2021) combines series decomposition with an auto-correlation mechanism, and Crossformer (Zhang & Yan, 2023) models dependencies across time and variables. PatchTST (Nie et al., 2023) embeds temporal patches of each variable as tokens, whereas iTransformer (Liu et al., 2024) embeds each variable's lookback sequence as a single token. Non-stationary Transformer (Liu et al., 2022) addresses non-stationarity through series normalization and de-stationary attention.

Other architectures offer alternatives to attention-based forecasting. TimesNet (Wu et al., 2023) and ModernTCN (Luo & Wang, 2024) use convolutional architectures. TSMixer (Chen et al., 2023) mixes information across time and variables with MLPs, while TimeMixer (Wang et al., 2024a) combines patterns across multiple temporal scales. Lightweight methods further demonstrate the effectiveness of compact forecasting models. DLinear (Zeng et al., 2023) and RLinear (Li et al., 2023) combine linear mappings with decomposition and normalization, respectively. SparseTSF (Lin et al., 2024) exploits periodicity by forecasting downsampled subsequences.

Frequency-domain methods provide another approach to representing time series. FEDformer (Zhou et al., 2022b) and FiLM (Zhou et al., 2022a) incorporate frequency-domain processing into their forecasting architectures. FreTS (Yi et al., 2023b) learns MLPs in the frequency domain, while FourierGNN (Yi et al., 2023a) uses learnable Fourier graph operators. FITS (Xu et al., 2024) uses complex-valued linear interpolation over retained low-frequency components, while TimeKAN (Huang et al., 2025) learns representations of decomposed frequency bands. FAN (Ye et al., 2024) uses frequency-adaptive normalization to address non-stationarity, and DUET (Qiu et al., 2025) learns channel relationships through metric learning in the frequency domain.

GLARE draws on both lightweight forecasting and frequency-domain methods. It uses a single linear layer for residual forecasting and a spectral filter to transform exogenous variables before local regression.

A.2 Time Series Forecasting with Exogenous Variables

Exogenous variables have long been incorporated into statistical forecasting models such as ARIMAX (Williams, 2001), which combines autoregressive modeling with exogenous inputs. Deep learning models incorporate exogenous information through various architectures. TFT (Lim et al., 2021) combines variable selection and temporal attention to process static covariates, historical observations, and known future inputs. NBEATSx (Olivares et al., 2023) incorporates exogenous information into a basis expansion architecture, while TiDE (Das et al., 2023) uses an MLP-based encoder-decoder to process historical observations and covariates. TimeXer (Wang et al., 2024b) combines self-attention over endogenous patches with cross-attention to exogenous variable tokens, while ExoTST (Tayal et al., 2024) integrates historical and future exogenous information through cross-temporal fusion. CrossLinear (Zhou et al., 2025) uses 1D convolution to model

cross-correlation between endogenous and exogenous variables. GCGNet (Li et al., 2026) uses graph-based correlation consistency to guide prediction generation and refinement, while DAG (Qiu et al., 2026) models temporal and cross-variable correlations through a dual network with attention transfer. GLARE combines global and locally fitted exogenous reconstructions with a single linear residual forecaster.

A.3 Online Forecasting and Test-Time Adaptation

Fitting coefficients to the current lookback window connects GLARE to online forecasting and test-time adaptation. Methods such as OneNet (Zhang et al., 2023), PROCEED (Zhao & Shen, 2025), DSOF (yee Ava Lau et al., 2025), TAFAS (Kim et al., 2025), and DynaME (Hong et al., 2026) adapt forecasters using newly available observations. GLARE keeps its learned parameters fixed and computes local regression coefficients independently for each input instance. This requires no additional trainable adaptation module or sequential processing of test windows. We therefore focus our comparisons on TSF-X methods evaluated under the same offline protocol.

B Proofs

Proof of Proposition 1. The filtered lookback matrix $\mathbf{A}$ and the global reconstruction remain unchanged when the endogenous history is perturbed.

Define the lookback reconstruction matrix as

$$\mathbf{P}_\lambda = \mathbf{A}(\mathbf{A}^\top\mathbf{A} + \lambda I)^{-1}\mathbf{A}^\top. \tag{14}$$

By Eq. (6), the residual formed from the blended reconstruction is

$$\mathbf{R}^{\text{endo}} = \big[I - (1-\alpha)\mathbf{P}_\lambda\big]\mathbf{X}^{\text{endo}} - \alpha\widetilde{\mathbf{X}}_{\text{g}}^{\text{endo}}. \tag{15}$$

Here, $\widetilde{\mathbf{X}}_{\text{g}}^{\text{endo}}$ denotes the global reconstruction over the lookback window. The global term cancels when comparing the original and perturbed residuals. Therefore,

$$\Delta\mathbf{R}^{\text{endo}} = \big[I - (1-\alpha)\mathbf{P}_\lambda\big]\boldsymbol{\delta}. \tag{16}$$

The matrix $\mathbf{P}_\lambda$ is symmetric. By the singular value decomposition of $\mathbf{A}$, its nonzero eigenvalues are $\sigma_i^2/(\sigma_i^2+\lambda)$, where σ_i are the nonzero singular values of $\mathbf{A}$. Its remaining eigenvalues are zero. Since $\lambda > 0$, all eigenvalues of $\mathbf{P}_\lambda$ lie in $[0, 1)$. Consequently, the symmetric matrix $I - (1-\alpha)\mathbf{P}_\lambda$ has all eigenvalues in $[\alpha, 1]$. Applying these bounds to Eq. (16) yields

$$\alpha\|\boldsymbol{\delta}\|_2 \le \|\Delta\mathbf{R}^{\text{endo}}\|_2 \le \|\boldsymbol{\delta}\|_2. \tag{17}$$

□

Proof of Proposition 2. Define the ridge operator as

$$\mathbf{Q}_\lambda = (\mathbf{A}^\top\mathbf{A} + \lambda I)^{-1}\mathbf{A}^\top. \tag{18}$$

Refitting the local coefficients after a perturbation $\boldsymbol{\delta}$ gives $\Delta\mathbf{s}^* = \mathbf{Q}_\lambda\boldsymbol{\delta}$.

Since the global reconstruction is fixed, the change in the blended future reconstruction is

$$\Delta\widetilde{\mathbf{Y}}^{\text{endo}} = (1-\alpha)\mathbf{G}_{\text{fut}}\Delta\mathbf{s}^* = (1-\alpha)\mathbf{G}_{\text{fut}}\mathbf{Q}_\lambda\boldsymbol{\delta}. \tag{19}$$

The singular values of $\mathbf{Q}_\lambda$ have the form $\sigma_i/(\sigma_i^2+\lambda)$, where σ_i are the singular values of $\mathbf{A}$. For every $\sigma_i \ge 0$,

$$\frac{\sigma_i}{\sigma_i^2+\lambda} \le \frac{1}{2\sqrt{\lambda}}, \tag{20}$$

because $\sigma_i^2 + \lambda \ge 2\sigma_i\sqrt{\lambda}$. Thus, $\|\mathbf{Q}_\lambda\|_2 \le 1/(2\sqrt{\lambda})$. Taking norms in Eq. (19) gives

$$\|\Delta\widetilde{\mathbf{Y}}^{\text{endo}}\|_2 \le (1-\alpha)\|\mathbf{G}_{\text{fut}}\|_2\|\mathbf{Q}_\lambda\|_2\|\boldsymbol{\delta}\|_2 \le \frac{(1-\alpha)\|\mathbf{G}_{\text{fut}}\|_2}{2\sqrt{\lambda}}\|\boldsymbol{\delta}\|_2. \tag{21}$$

□

Proof of Corollary 1. The bias of the residual forecaster cancels when comparing the original and perturbed predictions. By Eq. (9),

$$\Delta\hat{\mathbf{Y}}^{\text{endo}} = \Delta\widetilde{\mathbf{Y}}^{\text{endo}} + \mathbf{V}\Delta\mathbf{R}^{\text{endo}}. \tag{22}$$

The triangle inequality and Propositions 1 and 2 give

$$\begin{aligned}\|\Delta\hat{\mathbf{Y}}^{\text{endo}}\|_2 &\le \|\mathbf{V}\|_2\|\Delta\mathbf{R}^{\text{endo}}\|_2 + \|\Delta\widetilde{\mathbf{Y}}^{\text{endo}}\|_2 \\ &\le \left(\|\mathbf{V}\|_2 + \frac{(1-\alpha)\|\mathbf{G}_{\text{fut}}\|_2}{2\sqrt{\lambda}}\right)\|\boldsymbol{\delta}\|_2.\end{aligned} \tag{23}$$

□

C Experimental Details

C.1 Dataset Details

We follow the dataset setup of previous works (Li et al., 2026; Qiu et al., 2026). Table 6 summarizes their statistics, where C denotes the number of exogenous variables. Datasets are split chronologically into training, validation, and test sets with a ratio of 7:1:2, except for ETT with 6:2:2.

C.2 Implementation Details

All experiments are conducted following TFB (Qiu et al., 2024). We set `drop_last=False` in all experiments. We use baseline results reported by DAG and results from our own reruns under the same evaluation protocol. We report GLARE's mean results across three random seeds in Table 1 and standard deviations in Table 7. All experiments run on a single NVIDIA TITAN Xp GPU.

We train GLARE with MAE loss and a batch size of 64, selecting the learning rate from {1e-4, 1e-3, 1e-2, 3e-2}. We select the blending weight α from the coarse grid $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ without fine-grained tuning. The ridge strength λ is selected from $\{0.1, 1, 3\}$. We fix the MLP hidden dimension d in *Global Exogenous Mapping* to 32. We set K to retain the lowest 50% of frequency bins in the spectral filter. For the daily datasets Colbun and Rapel, we retain all frequency bins.

For the filtering comparison in Table 4, the identity mapping leaves the exogenous sequence unchanged. The temporal linear alternative applies a separate $N \times N$ matrix to each exogenous variable. The convolutional alternative uses a depthwise 1D convolution with kernel size $k = 3$. For the global mapping comparison in Table 5, all alternatives map raw exogenous variables to one scalar per time step. The linear alternative uses a single linear layer. The convolutional alternative uses a 1D convolution with kernel size $k = 3$. The attention alternative embeds each variable into a d-dimensional token and applies multi-head attention with four heads.

D Additional Experiments

D.1 Global and Local Fitting Comparison

Figure 1 compares global and local reconstructions obtained by ordinary least squares (OLS) on the twelve TSF-X datasets used in our main experiments. The global fit estimates coefficients once using the full training split. The local fit estimates coefficients separately using the L observations immediately preceding each test horizon. The right panel reports the percentage of test windows in which the local reconstruction has lower MSE than the global reconstruction.

D.2 Results on Multivariate Benchmarks

Table 8 reports results on the multivariate benchmarks. We use the last variable as the endogenous variable and the others as exogenous variables, with a lookback length of 96 and horizons of 96, 192, 336, and 720. On Weather, Table 9 compares GLARE with DAG, the second-best method, at five decimal places. GLARE uses a simple linear residual forecaster and shows weaker performance on ETT, where weak cross-variable dependencies can limit the benefits of cross-variable modeling (Chen et al., 2023; Wang et al., 2025).

Table 6: Dataset statistics.

Dataset	C	Exogenous Variables	Endogenous Variable	Frequency	Length
NP	2	Load, wind power	Electricity price	1 Hour	52,416
PJM	2	Power load	Electricity price	1 Hour	52,416
BE	2	Load, generation	Electricity price	1 Hour	52,416
FR	2	Load, generation	Electricity price	1 Hour	52,416
DE	2	Load, wind and solar power	Electricity price	1 Hour	52,416
Energy	5	Generation and storage	Thermal power generation	1 Hour	13,064
Colbun	2	Precipitation, inflow	Water level	1 Day	2,958
Rapel	2	Precipitation, inflow	Water level	1 Day	3,366
Sdwpfh	6	Meteorological variables	Active power	1 Hour	14,641
Sdwpfm	6	Meteorological variables	Active power	30 Min	29,281
ETTh1, ETTh2	6	Power load	Oil temperature	1 Hour	14,400
ETTm1, ETTm2	6	Power load	Oil temperature	15 Min	57,600
Weather	20	Meteorological variables	CO_2 concentration	10 Min	52,696
Electricity	320	Electricity consumption	Electricity consumption	1 Hour	26,304
Traffic	861	Road occupancy rates	Road occupancy rates	1 Hour	17,544
Exchange	7	Exchange rate	Exchange rate	1 Day	7,588

Table 7: GLARE results with future exogenous variables on the 12 TSF-X datasets, reported as mean $\pm$ standard deviation across three random seeds.

	Short horizon			Long horizon		
Dataset	T	MSE	MAE	T	MSE	MAE
NP	24	0.180 ± 0.000	0.219 ± 0.002	360	0.380 ± 0.002	0.388 ± 0.001
PJM	24	0.062 ± 0.000	0.152 ± 0.001	360	0.121 ± 0.003	0.214 ± 0.002
BE	24	0.328 ± 0.001	0.225 ± 0.001	360	0.460 ± 0.002	0.294 ± 0.001
FR	24	0.322 ± 0.003	0.170 ± 0.003	360	0.453 ± 0.003	0.252 ± 0.002
DE	24	0.273 ± 0.009	0.318 ± 0.004	360	0.382 ± 0.002	0.389 ± 0.002
Energy	24	0.051 ± 0.003	0.174 ± 0.004	360	0.091 ± 0.002	0.236 ± 0.003
Colbun	10	0.058 ± 0.002	0.093 ± 0.002	30	0.127 ± 0.007	0.201 ± 0.008
Rapel	10	0.179 ± 0.013	0.223 ± 0.007	30	0.304 ± 0.017	0.340 ± 0.010
Sdwpfm1	24	0.301 ± 0.009	0.373 ± 0.004	360	0.331 ± 0.014	0.412 ± 0.007
Sdwpfm2	24	0.339 ± 0.008	0.406 ± 0.005	360	0.369 ± 0.010	0.442 ± 0.007
Sdwpfh1	24	0.296 ± 0.015	0.381 ± 0.008	360	0.355 ± 0.016	0.433 ± 0.009
Sdwpfh2	24	0.353 ± 0.019	0.429 ± 0.009	360	0.399 ± 0.007	0.470 ± 0.004

D.3 Forecasting without Future Exogenous Variables

GLARE is designed for settings where future exogenous variables are available. When they are unavailable, we train a forecaster as a warmup and use its forecasts as pseudo future exogenous variables, adding approximately 0.2M to 1M parameters. The forecaster follows the patch embedding of CrossLinear (Zhou et al., 2025), with a patch length of 8, an embedding dimension of 64, and an MLP hidden width of 256. It is shared across exogenous variables and frozen during GLARE training. Table 10 shows that GLARE remains comparable to DAG, although its gains are smaller than with future exogenous variables.

D.4 Efficiency Measurements

Table 11 reports the efficiency measurements behind Figure 4. Peak GPU memory and inference time are measured per batch of 64 samples on NP and Sdwpfh1 and per batch of 16 samples on Electricity and Traffic. Measurements are averaged over 100 iterations after 20 warmup iterations.

Table 8: Forecasting performance on multivariate benchmarks in MSE and MAE. Best results are bold and second-best results are underlined.

Methods		GLARE (ours)		DAG (ICML'26)		CrossLinear (KDD'25)		TimeXer (NeurIPS'24)		TiDE (TMLR'23)		TFT (IJF'21)	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.059	0.188	**0.054**	**0.176**	0.152	0.321	0.056	0.179	0.056	0.179	0.081	0.214
	192	0.079	0.219	**0.072**	0.208	0.199	0.374	0.073	**0.207**	0.075	0.210	0.073	0.213
	336	0.098	0.246	**0.079**	**0.219**	0.218	0.383	0.089	0.234	0.090	0.236	0.099	0.246
	720	0.110	0.264	**0.079**	**0.221**	0.295	0.475	0.104	0.254	0.096	0.244	0.124	0.282
ETTh2	96	**0.123**	**0.269**	0.127	0.273	0.135	0.289	0.141	0.289	0.127	0.271	0.217	0.364
	192	0.171	**0.319**	0.177	0.327	**0.154**	0.323	0.194	0.346	0.179	0.327	0.256	0.405
	336	**0.172**	**0.331**	0.215	0.369	0.179	0.343	0.226	0.380	0.223	0.375	0.287	0.435
	720	0.259	0.409	**0.214**	**0.370**	0.319	0.463	0.234	0.385	0.253	0.405	0.283	0.431
ETTm1	96	**0.029**	**0.127**	**0.029**	**0.127**	0.087	0.247	0.030	0.129	0.030	0.130	0.030	0.131
	192	0.046	0.164	**0.044**	**0.159**	0.164	0.351	**0.044**	0.161	0.049	0.168	0.048	0.168
	336	0.060	0.190	**0.057**	**0.184**	0.239	0.426	0.060	0.189	0.069	0.203	0.068	0.200
	720	0.081	0.221	0.079	0.216	0.286	0.462	0.082	0.222	**0.073**	**0.211**	0.095	0.240
ETTm2	96	0.072	0.199	**0.063**	**0.183**	0.156	0.332	0.072	0.195	0.070	0.190	0.117	0.253
	192	0.107	0.243	**0.098**	**0.230**	0.203	0.366	0.110	0.249	0.101	0.235	0.179	0.324
	336	**0.127**	0.273	**0.127**	**0.270**	0.136	0.292	0.139	0.284	0.128	0.272	0.206	0.356
	720	0.166	0.318	0.179	0.328	**0.156**	**0.316**	0.190	0.340	0.179	0.328	0.271	0.417
Weather	96	**0.001**	**0.017**	0.001	0.021	0.012	0.093	0.001	0.023	0.001	0.026	0.001	0.028
	192	**0.001**	**0.018**	0.001	0.023	0.005	0.056	0.001	0.026	0.001	0.028	0.002	0.030
	336	**0.001**	**0.020**	0.001	0.025	0.009	0.080	0.001	0.027	0.002	0.031	0.002	0.030
	720	**0.001**	**0.023**	0.001	0.028	0.007	0.067	0.002	0.032	0.002	0.036	0.002	0.035
Electricity	96	**0.126**	**0.270**	0.158	0.298	0.281	0.416	0.160	0.296	0.312	0.414	0.365	0.456
	192	**0.130**	**0.274**	0.156	0.295	0.248	0.384	0.186	0.320	0.274	0.391	0.339	0.448
	336	**0.144**	**0.289**	0.217	0.353	0.260	0.396	0.218	0.346	0.313	0.417	0.367	0.471
	720	**0.192**	**0.333**	0.246	0.370	0.268	0.401	0.267	0.386	0.353	0.445	0.429	0.510
Traffic	96	**0.076**	**0.148**	0.119	0.187	0.261	0.322	0.135	0.233	0.149	0.251	0.139	0.221
	192	**0.076**	**0.146**	0.121	0.191	0.232	0.289	0.135	0.230	0.149	0.246	0.116	0.212
	336	**0.080**	**0.150**	0.125	0.200	0.223	0.280	0.143	0.237	0.150	0.251	0.134	0.228
	720	**0.091**	**0.164**	0.152	0.232	0.213	0.272	0.156	0.250	0.169	0.271	0.138	0.220
Exchange	96	**0.044**	**0.156**	0.087	0.229	0.234	0.425	0.109	0.245	0.096	0.230	0.105	0.258
	192	**0.092**	**0.243**	0.147	0.298	0.231	0.430	0.197	0.350	0.205	0.340	0.258	0.393
	336	**0.133**	**0.301**	0.274	0.413	0.394	0.563	0.331	0.444	0.453	0.519	0.422	0.524
	720	**0.182**	**0.344**	0.338	0.470	0.390	0.558	0.362	0.482	0.419	0.502	0.552	0.577
1st count		**20**	**20**	11	10	2	1	1	1	1	1	0	0

Table 9: Comparison of GLARE and DAG on Weather in MSE and MAE. Best results are bold.

	GLARE (ours)		DAG (re-run)	
T	MSE	MAE	MSE	MAE
96	**0.00059**	**0.01662**	0.00091	0.02145
192	**0.00066**	**0.01801**	0.00103	0.02293
336	**0.00081**	**0.02033**	0.00123	0.02546
720	**0.00102**	**0.02334**	0.00147	0.02803

D.5 Effect of the Frequency Cutoff

We compare five cutoff ratios $K/F \in \{1, 0.75, 0.5, 0.25, 0.125\}$ in Figure 6. Reducing the ratio to 0.5 generally maintains or improves forecasting accuracy while reducing the number of learnable filter parameters. More truncation can remove useful information and increase MSE.

Table 10: Forecasting performance without future exogenous variables in MSE and MAE. Best results are bold and second-best results are underlined.

Methods		GLARE (ours)		DAG (ICML'26)		GCGNet (ICLR'26)		CrossLinear (KDD'25)		TimeXer (NeurIPS'24)		TiDE (TMLR'23)		TFT (IJF'21)	
Dataset	T	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
NP	24	**0.226**	**0.257**	0.237	0.259	0.240	0.260	0.246	0.282	0.284	0.300	0.297	0.309	0.330	0.321
	360	**0.591**	**0.493**	0.601	0.500	0.620	0.502	0.655	0.505	0.663	0.509	0.638	0.523	0.964	0.655
PJM	24	0.074	0.168	**0.072**	**0.162**	**0.072**	0.164	0.096	0.198	0.101	0.197	0.105	0.213	0.129	0.231
	360	0.184	0.262	**0.180**	**0.259**	0.191	0.269	0.197	0.283	0.182	0.264	0.210	0.292	0.270	0.310
BE	24	**0.348**	**0.242**	0.376	0.244	0.365	0.259	0.389	0.253	0.428	0.267	0.492	0.322	0.400	0.267
	360	**0.540**	0.354	0.547	0.350	0.553	0.354	0.564	**0.346**	0.576	0.352	0.602	0.375	0.727	0.442
FR	24	**0.343**	0.193	0.345	**0.190**	0.352	0.199	0.397	0.204	0.395	0.225	0.415	0.254	0.445	0.231
	360	0.527	0.312	0.526	0.307	**0.519**	**0.293**	0.555	0.310	0.536	0.297	0.574	0.325	0.624	0.345
DE	24	0.437	0.415	0.422	0.411	**0.413**	**0.402**	0.438	0.423	0.493	0.441	0.465	0.433	0.576	0.452
	360	0.795	0.579	**0.784**	**0.563**	0.816	0.600	0.832	0.592	0.857	0.585	0.823	0.604	0.792	0.578
Energy	24	0.104	0.250	0.112	0.257	0.112	0.259	**0.102**	**0.246**	0.135	0.288	0.117	0.265	0.352	0.463
	360	0.208	0.362	**0.189**	**0.342**	0.195	0.351	0.299	0.425	0.221	0.372	0.207	0.358	0.399	0.497
Colbun	10	**0.065**	**0.088**	0.072	0.103	0.070	0.094	0.068	0.101	0.098	0.158	0.095	0.140	0.449	0.279
	30	0.230	0.230	**0.161**	**0.208**	0.168	0.236	0.190	0.289	0.230	0.320	0.282	0.341	0.663	0.494
Rapel	10	0.197	**0.214**	**0.195**	0.215	0.217	0.221	0.201	0.224	0.284	0.305	0.238	0.270	0.238	0.274
	30	0.391	0.388	0.327	0.347	0.313	**0.346**	**0.278**	0.374	0.475	0.449	0.409	0.444	0.377	0.394
Sdwpfm1	24	**0.523**	**0.486**	0.532	0.512	0.537	0.509	0.549	0.553	0.628	0.561	0.561	0.540	0.596	0.531
	360	0.900	0.684	**0.845**	0.695	0.885	**0.682**	1.070	0.834	0.867	0.710	0.864	0.725	1.372	0.925
Sdwpfm2	24	**0.606**	**0.528**	0.609	0.548	0.630	0.547	0.635	0.586	0.739	0.613	0.636	0.582	0.749	0.657
	360	1.015	**0.736**	**0.962**	0.766	0.983	0.739	0.965	0.788	1.008	0.756	0.995	0.783	1.340	0.877
Sdwpfh1	24	0.681	0.583	**0.647**	0.592	0.651	**0.578**	0.687	0.652	0.735	0.637	0.719	0.641	0.754	0.670
	360	0.911	0.717	**0.820**	0.695	0.830	**0.677**	0.849	0.752	0.830	0.712	0.897	0.757	0.831	0.727
Sdwpfh2	24	0.782	**0.622**	**0.739**	0.635	0.798	0.645	0.810	0.718	0.875	0.701	0.783	0.674	0.811	0.714
	360	1.072	0.780	1.000	**0.755**	**0.974**	0.757	1.102	0.830	0.975	0.763	1.056	0.829	1.042	0.779
1st count		8	**9**	**11**	7	4	6	2	2	0	0	0	0	0	0

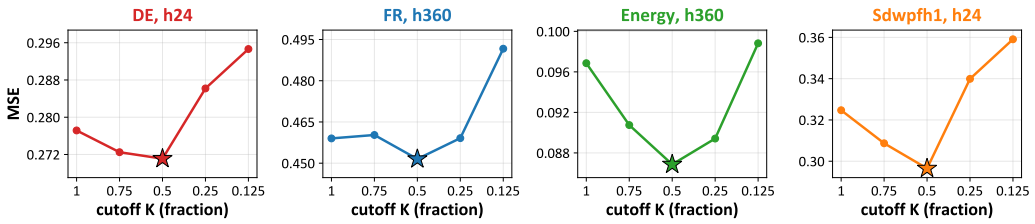

Figure 6: Test MSE for different fractions of frequency components retained by the spectral filter.

D.6 Effect of the Lookback Length

We vary the lookback length over $\{48, 96, 192, 336, 720\}$ with a fixed forecast horizon of 360 on NP, DE, Energy, and Sdwpfm2. Figure 7 shows that GLARE and DAG generally benefit from longer lookback windows, whereas TimeXer and CrossLinear improve less consistently. GLARE's MSE decreases monotonically on all four datasets and is the lowest in 19 of the 20 settings.

D.7 Case Study

Figures 8 and 9 show forecasting examples of GLARE and the baselines on the 12 TSF-X datasets.

Table 11: Parameter counts, estimated FLOPs, peak GPU memory, and inference time. FLOPs are reported per sample. Best results in each setting are bold.

	NP, $L = 168, T = 24$				NP, $L = 720, T = 360$			
	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)
TFT	0.39M	56.33M	202.66	9.40	0.39M	565.82M	4804.10	127.45
TiDE	0.66M	1.77M	35.34	1.21	2.62M	8.56M	158.01	4.00
TimeXer	0.91M	12.86M	22.36	2.13	5.31M	61.14M	62.44	2.14
GCGNet	2.34M	48.87M	46.39	5.93	2.18M	27.13M	39.73	5.36
DAG	0.49M	2.29M	14.92	6.34	6.06M	17.87M	41.62	6.22
GLARE	**4.31K**	**0.06M**	**12.12**	**1.07**	**0.26M**	**0.86M**	**31.18**	**1.18**
	Sdwpfh1, $L = 168, T = 24$				Sdwpfh1, $L = 720, T = 360$			
	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)
TFT	2.37M	357.97M	322.68	21.75	2.37M	2518.46M	5289.59	197.92
TiDE	0.91M	2.71M	36.87	1.18	3.57M	13.40M	165.40	4.10
TimeXer	0.91M	14.31M	23.09	1.75	5.31M	64.41M	65.92	2.17
GCGNet	9.55M	166.94M	101.58	11.06	5.36M	81.68M	67.38	5.62
DAG	0.51M	12.09M	31.99	6.12	9.48M	94.83M	233.90	11.13
GLARE	**4.70K**	**0.17M**	**13.14**	**1.08**	**0.26M**	**1.47M**	**36.81**	**1.19**
	Electricity, $L = 96, T = 96$				Traffic, $L = 96, T = 96$			
	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)	Params	FLOPs	Peak Mem. (MiB)	Inf. Time (ms)
TFT	2.98M	0.41G	477.35	123.68	7.81M	1.07G	1258.75	339.88
TiDE	16.57M	15.83G	90.56	276.76	21.89M	55.78G	137.24	1286.97
TimeXer	1.78M	0.20G	49.45	2.63	6.71M	1.96G	176.49	6.20
GCGNet	161.42M	3.76G	1251.86	124.46	1.15G	19.73G	8554.51	523.28
DAG	3.26M	5.43G	436.53	30.01	3.27M	16.00G	1114.90	91.76
GLARE	**0.05M**	**0.02G**	**45.24**	**1.78**	**0.12M**	**0.05G**	**107.78**	**2.17**

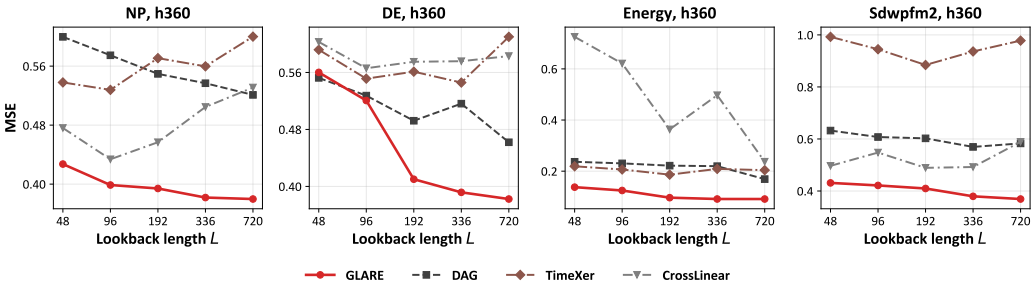

Figure 7: Test MSE for different lookback lengths at a forecast horizon of 360.

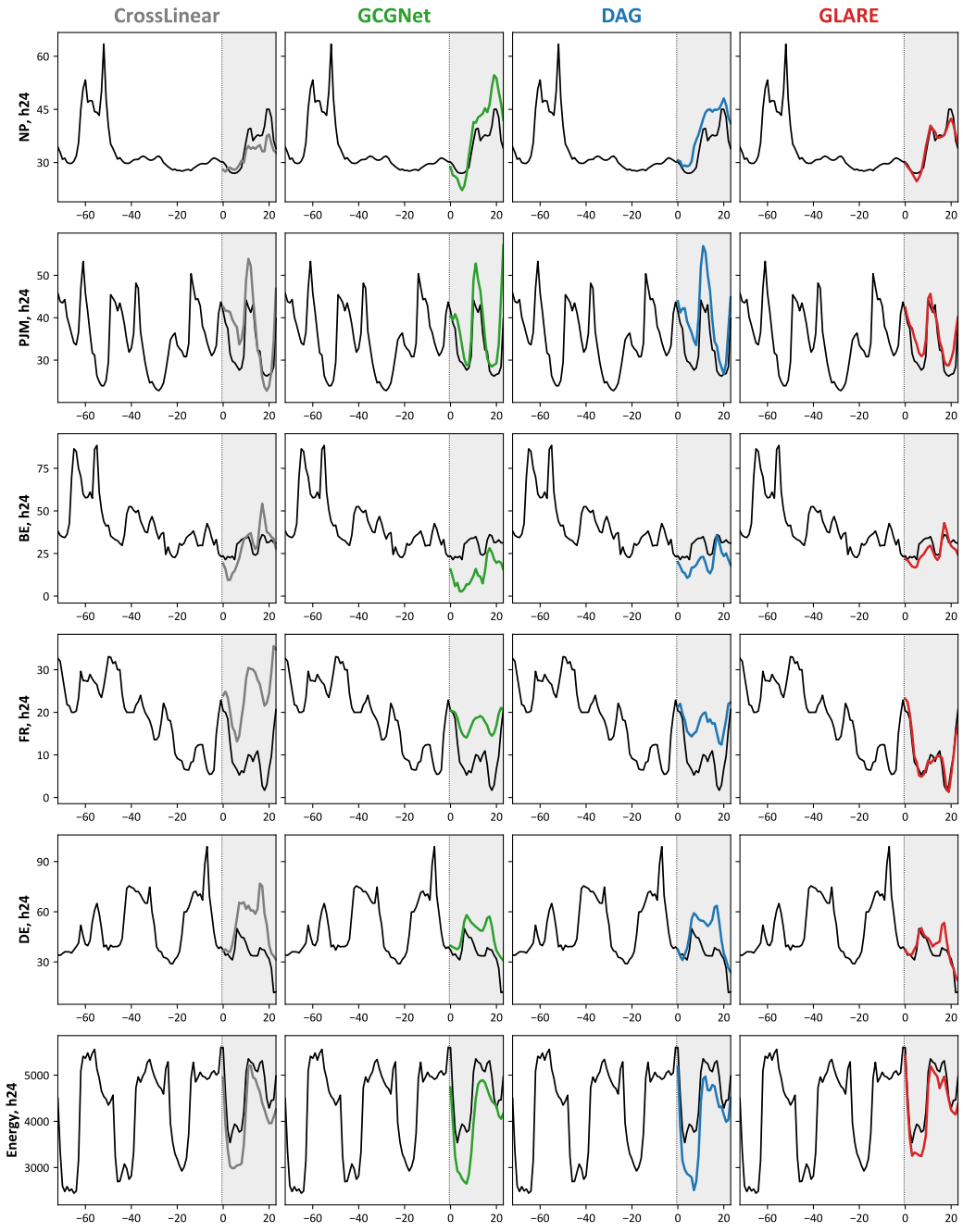

Figure 8: Forecasting examples on NP, PJM, BE, FR, DE, and Energy. Black lines indicate the ground truth, and shaded regions indicate the forecast horizon.

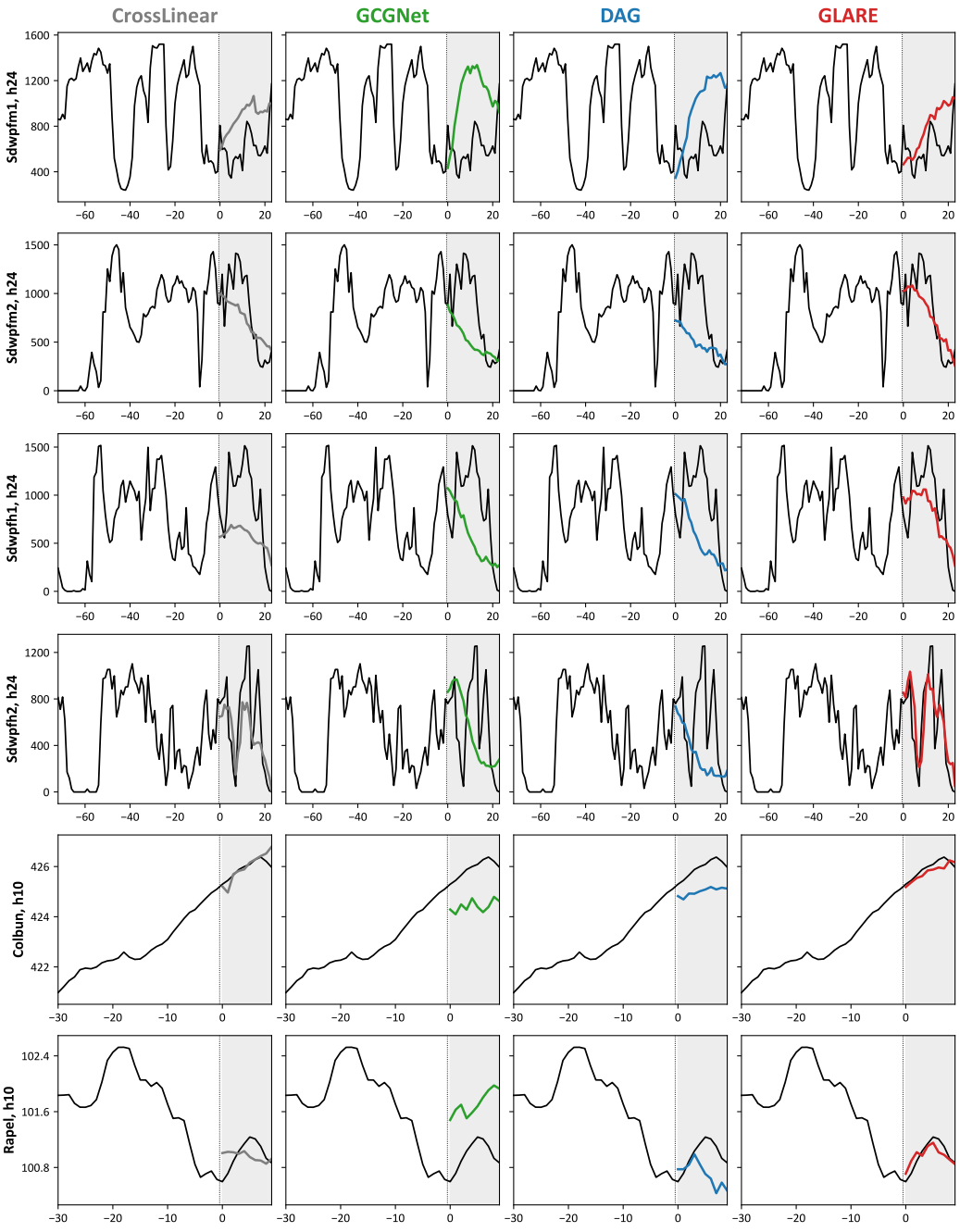

Figure 9: Forecasting examples on Sdwpfm1, Sdwpfm2, Sdwpfh1, Sdwpfh2, Colbun, and Rapel. Black lines indicate the ground truth, and shaded regions indicate the forecast horizon.