WHO CONTRIBUTES TO WHOM? CROSS-ENTITY ATTRIBUTION IN PANEL TIME SERIES FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

A forecast indicates what may happen, but understanding the factors that shape the prediction can provide additional evidence for interpreting and assessing it. In panel time series, multiple entities may contain predictive information about one another. Identifying these entity-level contributions can therefore reveal which entities are relevant to a target forecast and help domain experts investigate cross-entity relationships. However, most existing panel forecasting models focus primarily on prediction and do not directly quantify how individual entities contribute to each target forecast. To our knowledge, XPanelNet is the first panel time-series forecasting architecture to integrate entity-level predictive attribution directly into the forecasting process. XPanelNet decomposes each entity's history into trend, seasonal, cyclic, and residual components to construct a base forecast, and then augments it with signed cross-entity contributions computed through a latent vector-autoregressive (VAR) coupling over a low-dimensional shared representation to obtain the final prediction. This additive formulation makes entity-level attribution an intrinsic part of the forecasting process rather than a post hoc interpretation derived from attention or similarity weights. Across nine real-world panel datasets spanning finance, energy, climate, hydrology, transportation, and public safety, XPanelNet achieves the best result of any model, including 10 deep forecasting baselines (TimeKAN, TimeMixer, iTransformer, CrossGNN, PatchTST, TimesNet, DLinear, SCINet, Non-stationary Transformer, and Autoformer), on 7 of 9 evaluated datasets by MAE and 4 of 9 by MSE, while simultaneously providing entity-level predictive attributions.

1 INTRODUCTION

Future behaviour in many high-stakes domains is uncertain, yet decisions must often be made in advance. Forecasting provides the predictions needed to support such decisions. In many practical settings, the data form a *panel*: multiple related entities observed over time (e.g., stocks, sensors, regions, or stations). Entities in a panel frequently contain predictive information about one another, so a model that can both forecast accurately *and* quantify which other entities contributed to a given prediction would be valuable to domain experts.

Existing deep learning architectures for panel time series forecasting (Wang et al., 2024; Liu et al., 2024b; Huang et al., 2023; Nie et al., 2022; Wu et al., 2022; Zeng et al., 2023; Liu et al., 2022a;b; Wu et al., 2021; Huang et al., 2025), have achieved strong forecasting accuracy by exploiting shared structure and cross-entity dependence. However, such dependence is typically captured only implicitly through attention weights, mixing coefficients, or graph structures. These quantities are not designed as predictive attributions and are not guaranteed to reflect the predictive contribution of each entity. Consequently, practitioners may obtain accurate forecasts without knowing which entities contributed to each prediction.

Post-hoc explanation methods such as SHAP and Integrated Gradients can be applied after training, but computing explanations for every forecast can be expensive, and their attributions are obtained separately from the model's predictive computation. We therefore argue that entity-level attribution should be an architectural property of the forecasting model itself.

A key challenge in panel forecasting is to provide faithful entity-level predictive contributions without sacrificing forecasting performance. Existing forecasting models primarily focus on accurately

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.

predicting future values, while providing intrinsic accounts of how individual entities contribute to a target forecast remains less explored. In particular, obtaining signed contributions that distinguish positive and negative predictive effects, without relying on post-hoc attribution, remains challenging. To address this gap, we propose XPanelNet. Our contributions are:

- We identify a limitation in existing panel forecasting models: accurate forecasts are generally not accompanied by intrinsic, entity-level attribution of how other entities contribute to each target forecast, particularly with respect to signed predictive effects.
- We propose XPanelNet, an explainable panel forecasting architecture that decomposes each forecast into a target-specific base prediction and signed cross-entity predictive contributions derived from other entities' histories.
- We introduce a low-rank latent VAR/IRF router that propagates cross-entity interactions through a shared latent space while enabling explicit source-to-target attribution and exact reconstruction of the aggregate cross-entity contribution.

The remainder of the paper is organized as follows. Section 2 reviews related work on panel forecasting, cross-entity modelling, and time-series explainability. Section 3 presents the XPanelNet architecture. Section 4 reports experimental results, including forecasting accuracy, an analysis of underperforming cases, and a faithfulness analysis of the produced attributions. Section 5 concludes and discusses limitations and future work.

2 Related Work

Long-horizon and panel forecasting. Long-horizon forecasting is challenging because prediction errors can accumulate as the forecast horizon increases. Existing studies distinguish recursive strategies, which feed previous predictions back into the model, from direct sequence-to-sequence approaches that predict the full horizon at once (Lim & Zohren, 2021; Benidis et al., 2022). Beyond the forecasting strategy, recent work has explored different ways of representing temporal structure, with patch-based, decomposition-based, and linear models showing that simpler representations can remain competitive with attention-based architectures (Nie et al., 2022; Wang et al., 2024).

The panel setting introduces an additional challenge because multiple entities may contain information relevant to one another. Global forecasting models exploit such shared information, which can be particularly useful when individual series provide limited historical observations (Benidis et al., 2022; Hewamalage et al., 2022). However, sharing information across entities can be less effective when their dynamics differ, motivating approaches that combine shared representations with entity-specific modelling (Pesaran et al., 2026). Architectures such as DeepGLO (Sen et al., 2019) and DeepDGL (Chen et al., 2021) therefore combine global and local modelling. These approaches demonstrate the value of cross-entity information, but the information from multiple entities is generally incorporated into the prediction without explicitly identifying the contribution of each source entity to a target forecast.

Cross-entity modelling. Cross-entity dependence is commonly modelled through graph-based learned edges (Corradini et al., 2025; Jin et al., 2024) or variate-centric attention; for example, iTransformer (Liu et al., 2024b) treats each series as a token and uses self-attention to model interactions between series. Such mechanisms allow a target entity to use information from other entities, but an attention weight, graph edge, or mixing coefficient represents a relationship or information flow rather than necessarily a signed predictive contribution. Once information from multiple entities is mixed, it is therefore difficult to determine how much each source entity contributed to a particular target prediction.

Explainability in time-series forecasting. Existing work on explainable forecasting can be broadly organised into three streams. *Temporal-level* methods identify influential historical time steps or attention patterns, such as Temporal Fusion Transformer (Lim et al., 2021), temporal-attention CNNs (Pantiskas et al., 2020; Lin et al., 2021), GFAN (Wu & Xiao, 2025), and TFformer (Zhao et al., 2025). *Variable-level* and variable–time methods highlight important input variables or their interactions, including IMV-LSTM (Guo et al., 2019), STAM (Gangopadhyay et al., 2021), mixture-attention LSTMs (Xu et al., 2022), PRFormer (Hu et al., 2025), DLFormer (Kim

et al., 2024), DCFA-iTimeNet (Yuan et al., 2024), and IFTT (Liu et al., 2024a). *Structural or decomposition-based* methods instead expose short-term and long-term effects or different temporal components, including Neural Additive Time-Series Models (Jo & Kim, 2023), SCNN (Deng et al., 2023), DeLELSTM (Wang et al., 2023), N-BEATS (Oreshkin et al., 2019), TSDFNet (Zhou & Gao, 2022), and PatchDecomp (Tomioka & Yoshimura, 2026). These approaches make different aspects of a forecast interpretable, but they do not directly address a distinct question that arises in panel forecasting: which source entity contributed to the forecast of a particular target entity? Recent work has begun to consider this direction. SEGCRN (García-Sigüenza et al., 2025), for example, learns context-dependent node relevance masks that indicate the relevance of nodes to a prediction. However, such relevance scores do not necessarily provide the source-to-target, horizon-specific attribution considered here. The attribution of interest is directed from source entity j to target entity i at forecast horizon h, with each source contribution forming an explicit term in the additive decomposition of the target prediction.

Taken together, existing work provides strong mechanisms for global forecasting, cross-entity dependence modelling, and time-series explanation. However, existing cross-entity forecasting mechanisms generally mix information from multiple entities without explicitly exposing how each source entity contributes to a particular target forecast. XPanelNet addresses this gap by making signed, directed, horizon-specific entity contributions an intrinsic part of the forecasting computation, with the individual contributions summing to the predicted target value.

3 XPanelNet

Let a panel time series contain N entities observed over T time steps. Given a lookback window of length L, the input for entity i at time t is $X_{i,t-L+1} = (x_{i,t-L+1}, \ldots, x_{i,t})$. The task is to jointly forecast the next H steps for all entities, denoted by $\hat{Y}_{i,t+1:t+H} = (\hat{y}_{i,t+1}, \ldots, \hat{y}_{i,t+H})$. We decompose each forecast into two additive terms

$$\hat{Y}_{i,t+1:t+H} = B_{i,t+1:t+H} + \sum_{j \neq i} P_{i \leftarrow j,t+1:t+H} \tag{1}$$

where $\hat{Y}_i \in \mathbb{R}^H$ denotes the forecast for entity i, $B_i \in \mathbb{R}^H$ is the *base forecast* derived from entity i's own history, and $P_{i \leftarrow j} \in \mathbb{R}^H$ denotes the signed *cross-entity contribution* of entity j to the forecast of entity i. Since the final prediction is constructed explicitly from these components, the entity-level contributions are produced directly by the model, without requiring a separate post-hoc explanation method.

3.1 Structure Overview

XPanelNet is illustrated in Figure 1. It operates in two stages. First, a *structural decomposition* stage extracts trend, seasonal, residual, and long-period cyclic components from each entity's history to construct the base forecast. Second, a *latent cross-entity coupling* stage uses the learned entity representations to model signed cross-entity contributions via a low-rank vector-autoregressive (VAR) system over a small latent space. The resulting components are summed to obtain the final forecast.

3.2 Input Representation and Normalization

For each entity i, the raw lookback window $\boldsymbol{x}_i^{\mathrm{raw}} \in \mathbb{R}^L$ is augmented with calendar features encoded using sine and cosine functions at dataset-appropriate temporal granularities. These features are mapped through a two-layer MLP to obtain a time embedding $\boldsymbol{e}^{\mathrm{time}} \in \mathbb{R}^{d_e}$. Future calendar features are constructed in the same manner for the forecast horizon and are used to provide future time information to the branch forecasts. Each entity has a learned identity embedding $\boldsymbol{e}_i \in \mathbb{R}^{d_e}$, which is used only in entity fusion and cross-entity routing. We apply a learnable gated variant of Reversible Instance Normalization (RevIN) to accommodate differences in stationarity across entities.

$$g_i = \sigma(\ell_i) \in [0, 1]. \tag{2}$$

Given the empirical mean μ_i and standard deviation σ_i of the input window, the effective normalization statistics are

$$\mu_i^{\mathrm{eff}} = g_i \mu_i, \qquad \sigma_i^{\mathrm{eff}} = g_i \sigma_i + (1 - g_i). \tag{3}$$

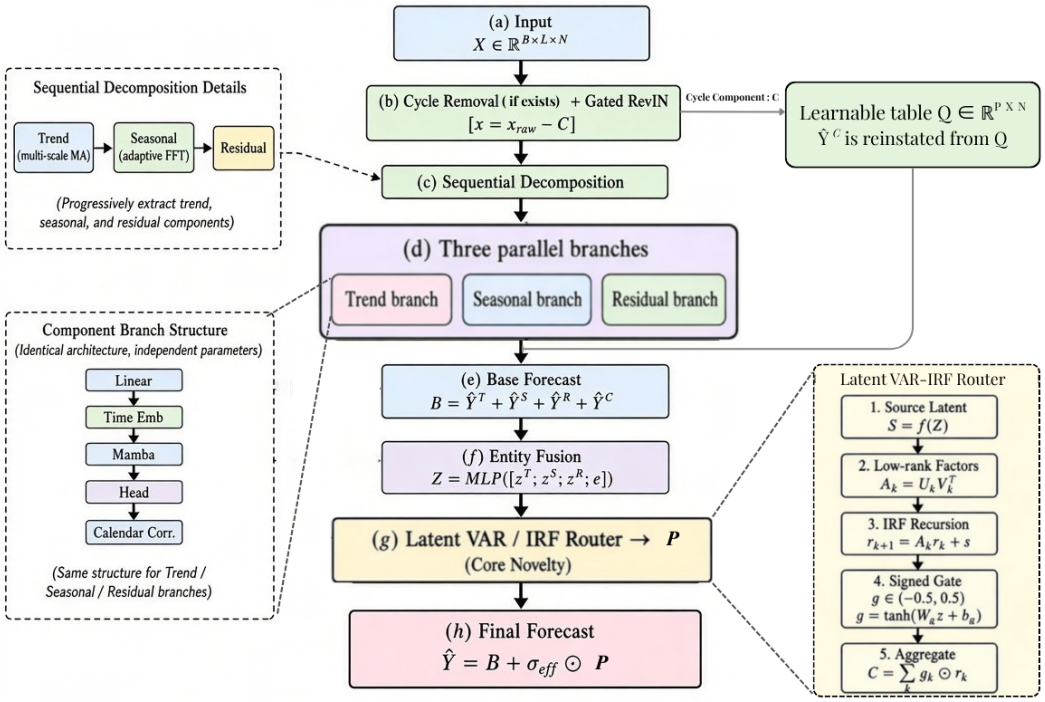

Figure 1: Overview of XPanelNet

3.3 Base Forecast

For each entity i, XPanelNet first constructs an entity-specific base forecast from its temporal dynamics before introducing cross-entity information. The observed series is decomposed into trend, seasonal, cyclic, and residual components:

$$\boldsymbol{x}_i^{\text{raw}} = T_i + S_i + C_i + R_i. \tag{4}$$

A cyclic component with a period longer than the lookback window is removed before normalization. XPanelNet addresses this with a phase-indexed periodic profile $Q \in \mathbb{R}^{P \times N}$:

$$\hat{y}^{\text{C}}_{i,t+h} = Q_{(t+h) \bmod P,\, i}. \tag{5}$$

where $P > L$ is automatically detected from the training split using the first-difference FFT spectrum and a spectral-prominence criterion. If no sufficiently clear long-period peak is detected, the cyclic branch is disabled. Then the forecast is reinstated after denormalization.

Component decomposition. The trend is obtained using a learned convex combination of moving averages, where θ_i contains entity-specific mixture parameters :

$$T_i = \sum_{k \in \{3,5,7,11\}} \alpha_{i,k}\, \text{MA}_k(\boldsymbol{x}_i), \qquad \alpha_i = \text{softmax}(\theta_i), \tag{6}$$

The seasonal component is reconstructed from the dominant spectral components of the detrended signal using the FFT, including the DC component. Then, the residual component is then defined as

$$R_i = \boldsymbol{x}_i - T_i - S_i. \tag{7}$$

Branch forecasting. The trend, seasonal, and residual components are processed by separate temporal branches. Each component is projected to d_{model}, combined with the learned time embedding, and encoded by a Mamba encoder. The final-step representation is then mapped to the forecast horizon H using a linear prediction head. Future calendar features are incorporated through a

zero-initialized time-based correction to each branch forecast. To reduce excessive long-horizon extrapolation, the trend forecast is damped toward the latest observed trend level:

$$\hat{Y}^T_{i,h} \leftarrow t_{\text{last},i} + \phi_{i,h}\left(\hat{Y}^T_{i,h} - t_{\text{last},i}\right), \tag{8}$$

where $\phi_{i,h} \in (0,1)$ is the learned damping factor, and $t\text{last}, i$ is the final value of the multi-scale trend in normalized space. The three branch forecasts, together with the forecast of the long-period cyclic component, are then summed to obtain the entity-specific base forecast:

$$B_i = \hat{Y}^T_i + \hat{Y}^S_i + \hat{Y}^R_i + \hat{Y}^{\text{C}}_i. \tag{9}$$

3.4 Entity Fusion

The final-step representations from the trend, seasonal, and residual branches, denoted by Z^T_i, Z^S_i, and Z^R_i, are concatenated with the learned entity identity embedding $\boldsymbol{e}_i$ and transformed by an MLP:

$$Z_i = \text{MLP}\left([Z^T_i; Z^S_i; Z^R_i; \boldsymbol{e}_i]\right). \tag{10}$$

The resulting fused representation $Z_i \in \mathbb{R}^{d_f}$, where d_f is the model size, is used as the input to the cross-entity router.

3.5 Cross-Entity Router: Latent VAR/IRF Coupling

XPanelNet models cross-entity dependence using a low-rank latent vector-autoregressive (VAR) system rather than explicit pairwise attention or prototype routing. Each fused entity representation Z_j is projected into a shared latent space,

$$s_j = \tanh(WZ_j) \in \mathbb{R}^m, \qquad m = 4. \tag{11}$$

The latent entity states are coupled through $p = 2$ low-rank lag matrices

$$A_k = \text{scale}_k\left(U_k V_k^\top - \text{diag}(U_k V_k^\top)\right), \qquad k = 1, \ldots, p, \tag{12}$$

where $U_k, V_k \in \mathbb{R}^{N\times r}$ and $r = N$ for $N \le 64$, otherwise $r = \lceil\sqrt{N}\rceil$. The matrices are scaled to ensure stable multi-step propagation. Starting from s, the VAR system recursively propagates cross-entity effects across the forecast horizon, with the cumulative response mapped to each horizon through a shared readout $R \in \mathbb{R}^{H\times m}$. A bounded target-specific gate then modulates the resulting signal:

$$P_{i,h} = g_{i,h}\,\text{agg}_{i,h}. \tag{13}$$

$$g_{i,h} = g_{\max}\tanh\left(\text{Linear}(Z_i)\right)_h, \qquad g_{\max} = 0.5. \tag{14}$$

where the gate is zero-initialized so that the model initially reduces to the base forecast. The router avoids materializing an $(N, N, \cdot)$ tensor during normal forecasting; its main intermediate representations scale as (B, N, H, m).Details are provided in Appendix A.5

3.6 Signed Cross-Entity Attribution

The latent VAR propagation admits an explicit impulse-response representation. Let Ψ^{cum}_h denote the cumulative VAR response at horizon h. The contribution of source entity j to target entity i is

$$P_{i\leftarrow j,h} = g_{i,h}R_h\left(\Psi^{\text{cum}}_h[i,j]\,s_j\right). \tag{15}$$

The aggregate cross-entity contribution is exactly recovered by summation:

$$P_{i,h} = \sum_{j=1}^{N} P_{i\leftarrow j,h}. \tag{16}$$

Consequently, after accounting for the target entity's denormalization scale, the forecast can be reconstructed as

$$\hat{Y}_i = B_i + \sigma^{\text{eff}}_i \odot P_i, \tag{17}$$

with no additional shift term. Thus, the pairwise terms provide signed source-to-target predictive contributions that sum exactly to the router's aggregate output. The full (B, N, N, H) attribution tensor is generated only on demand during evaluation, without gradients or retraining, so its construction does not affect the scalability of normal forecasting.

4 EXPERIMENTS

Datasets We conduct experiments on nine real-world panel datasets, including Exchange, Electricity, Traffic, Solar, SAARC Temperature, Beijing Air, USGS Streamflow, Chicago Crime, and SDWPF Wind, for long-term forecasting. Full dataset details are provided in Appendix A.1.

Baselines We compare XPanelNet against 10 well-established forecasting methods spanning different deep learning paradigms: iTransformer (Liu et al., 2024b), PatchTST (Nie et al., 2022), Non-stationary Transformer (Liu et al., 2022b), Autoformer (Wu et al., 2021), TimeMixer (Wang et al., 2024), DLinear (Zeng et al., 2023), TimeKAN (Huang et al., 2025), TimesNet (Wu et al., 2022), SCINet (Liu et al., 2022a), and CrossGNN (Huang et al., 2023).

Experimental Settings To ensure a fair comparison, we use a fixed lookback window $L = 96$ and the same four forecast horizons $H \in 96, 192, 336, 720$ for XPanelNet and all baselines. The data are split chronologically into 0.7/0.1/0.2 (train/validation/test) subsets, with a `StandardScaler` fitted on the training split only. We use Adam with a learning rate of 10^{-3}. Models are trained for up to 15 epochs with early stopping based on validation loss (patience 5). One dedicated model is trained for each forecast horizon. The remaining details provided in Appendix A.2.

4.1 FORECASTING ACCURACY

We report the forecasting results of XPanelNet and all 10 baselines, averaged across the four forecast horizons for each dataset.

Table 1: Panel data forecasting results, averaged across four prediction lengths $H \in \{96, 192, 336, 720\}$ with a fixed lookback length $L = 96$. Lower is better; **bold red** = best, underlined blue = second-best. Full per-horizon results are listed in Appendix A.3.

Dataset	**XPanelNet Ours**		TimeKAN 2025		TimeMixer 2024		iTrans. 2024		CrossGNN 2023		PatchTST 2023		TimesNet 2023		DLinear 2023		SCINet 2022		Station. 2022		Autoform. 2021	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange	**0.266**	**0.370**	0.395	0.419	0.391	0.453	0.358	0.406	0.345	0.395	0.366	0.404	0.416	0.443	0.354	0.414	0.750	0.626	0.461	0.454	0.613	0.539
Electricity	0.184	0.279	0.197	0.286	0.182	0.273	**0.178**	**0.270**	0.201	0.299	0.216	0.276	0.193	0.303	0.225	0.300	0.268	0.365	0.193	0.296	0.227	0.338
Traffic	0.537	**0.271**	0.579	0.371	0.485	0.297	**0.428**	0.282	0.583	0.323	0.522	0.286	0.620	0.336	0.625	0.383	0.804	0.509	0.624	0.340	0.628	0.379
Solar	0.243	**0.242**	0.285	0.321	**0.216**	0.280	0.233	0.262	0.249	0.312	0.236	0.333	0.402	0.374	0.330	0.401	0.282	0.375	0.350	0.390	0.593	0.589
SAARC Temp	**0.382**	**0.428**	0.924	0.722	0.814	0.669	0.904	0.729	2.006	1.125	1.099	0.790	1.006	0.784	0.623	0.582	2.230	1.203	0.498	0.528	1.702	1.043
Beijing Air	1.177	**0.700**	1.233	0.753	1.286	0.758	1.368	0.783	1.346	0.802	1.227	0.754	1.360	0.787	**1.090**	0.740	1.569	0.862	1.196	0.742	1.436	0.846
USGS Flow	**1.048**	**0.375**	1.150	0.460	1.144	0.454	1.176	0.473	1.833	0.626	1.164	0.470	1.185	0.486	1.086	0.494	1.672	0.670	1.186	0.485	1.254	0.595
Chicago	0.465	0.514	0.457	0.506	0.450	0.503	0.441	0.496	0.863	0.702	0.467	0.504	0.437	0.497	0.500	0.555	0.501	0.535	**0.432**	**0.493**	0.552	0.570
SDWPF Wind	**0.966**	**0.691**	1.032	0.744	1.034	0.738	1.070	0.746	1.642	0.885	1.070	0.750	1.163	0.789	0.996	0.803	1.148	0.784	1.056	0.790	1.239	0.850

XPanelNet obtains the best average MAE among the evaluated models on 7 of 9 datasets and the best average MSE on 4 of 9 datasets. The datasets where it does not lead provide useful insight into the model's limitations. On Electricity ($N = 321$) and Traffic ($N = 862$), the two largest panels, iTransformer outperforms XPanelNet on MSE. Both datasets use the router's reduced-rank regime ($r = \lceil \sqrt{N} \rceil$), suggesting that the low-rank VAR coupling may limit the capacity to represent cross-entity interactions in very large panels. This remains a hypothesis requiring direct evaluation through rank and coupling-capacity studies. In contrast, Beijing Air ($N = 12$) is a smaller panel where XPanelNet remains competitive, winning MAE while ranking second on MSE, consistent with the difference in training objectives. Chicago Crime provides the clearest evidence that cross-entity modeling is not universally beneficial: the cross-entity ablation degrades both metrics, while the faithfulness analysis also shows zero signed occlusion agreement. These results indicate that the usefulness of cross-entity predictive structure depends on the characteristics of the underlying panel dynamics.

4.2 FAITHFULNESS OF ENTITY-LEVEL ATTRIBUTIONS

The reconstruction property ensures that pairwise attributions sum to the cross-entity output, but does not establish their faithfulness to source-entity sensitivity. We therefore evaluate faithfulness using decomposition, deletion, router randomization, and full-model occlusion (Table 2).

Exact decomposition. The sum of pairwise contributions reproduces the router's cross-entity output with a maximum absolute error of 8.3×10^{-7}, empirically confirming the reconstruction property.

Table 2: Faithfulness diagnostics across datasets. Exactness is lower-is-better; higher occlusion agreement, higher Top-k/lower Bottom-k deletion ratios, and near-zero randomized-router correlation indicate stronger faithfulness.

Case	Exactness (max abs err)	Occlusion Spearman		Deletion (ratio to random)		Randomized-router corr.
		Signed	Magnitude	Top-k	Bottom-k	
SAARC Temp	2.4e-7	0.18	0.34	1.39	0.52	0.10
Exchange	1.2e-7	0.06	0.27	1.40	0.55	0.14
USGS Flow	4.8e-7	0.44	0.26	1.93	0.07	0.03
Chicago	7.5e-8	0.00	0.60	1.54	0.09	0.22
Beijing Air	7.5e-8	0.52	0.49	1.89	0.33	-0.04
SDWPF Wind	1.5e-7	0.67	0.33	4.73	0.03	0.01
Solar	2.4e-7	0.45	0.27	3.66	0.02	0.03
Mean (range)	$\leq$8.3e-7	0.33 (0.00–0.67)	0.37 (0.26–0.60)	2.36 (1.39–4.73)	0.23 (0.02–0.55)	0.07 ($-$0.04–0.22)

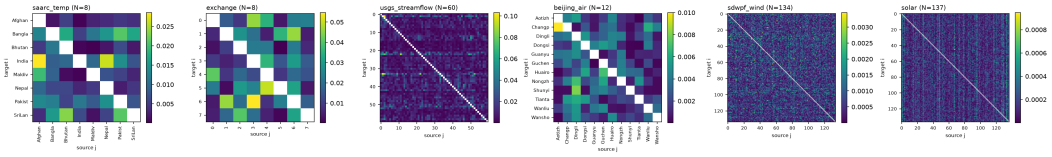

Figure 2: Contribution heatmaps (mean $|C_{i\leftarrow j}|$) across the faithfulness test cases in Table 2.

Deletion ordering Removing the highest-ranked source entities changes the forecast by 1.4–$4.7\times$ more than removing the same number of random entities, whereas removing the lowest-ranked sources produces only 0.02–$0.55\times$ the change caused by random deletion. Thus, the attribution ranking consistently distinguishes more influential from less influential sources across all cases.

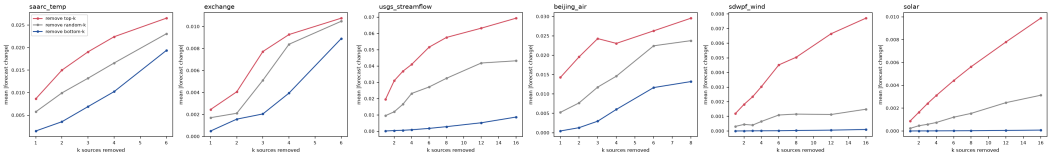

Figure 3: Deletion curves (top-k / random-k / bottom-k) for each faithfulness test case, corresponding to the deletion columns of Table 2.

Training dependence Re-initializing the router produces explanations with near-zero agreement with those of the trained model (mean Spearman $\rho = 0.07$, range -0.04–0.22), indicating that the attributions depend on learned parameters rather than architecture alone, consistent with the sanity-check principle of Adebayo et al. (2018).

Full-model occlusion. Agreement with full-model occlusion varies across panels. Signed agreement is stronger for SDWPF Wind, Beijing Air, Solar, and USGS Streamflow (0.44–0.67), where cross-entity physical coupling is plausible, and weaker for SAARC Temperature, Exchange, and Chicago Crime (0.00–0.18). Notably, Chicago Crime has zero signed agreement but 0.60 magnitude agreement, indicating that the attribution identifies relevant source entities while not consistently matching the direction of their effects in the full model.

Limitations of the faithfulness evaluation. The pairwise contributions represent linear counterfactual effects within the router, whereas perturbing an input entity in the full model can induce nonlinear changes throughout the forecasting pipeline. This mismatch may partly explain the weaker signed occlusion agreement observed on panels with limited cross-entity coupling. In some datasets, such as Solar and Chicago Crime, the cross-entity component also accounts for only a small fraction of the overall forecast, so even a faithful ranking may correspond to a small absolute effect. Finally, all evaluations except Exchange use a single seed.

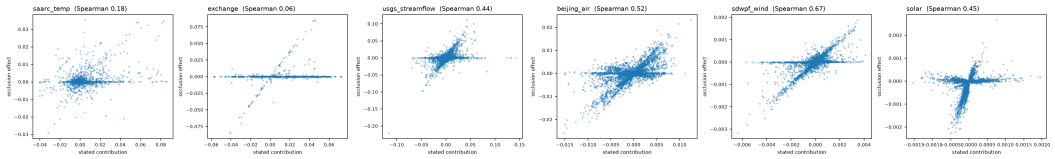

Figure 4: Occlusion scatter plots (stated vs. actual effect) for each faithfulness test case, corresponding to the occlusion Spearman columns of Table 2.

Interpretation. These attributions should therefore be interpreted as correlational rather than causal: they identify which entities the trained router relies on and provide evidence that this reliance is meaningful through deletion and training-dependent through randomization, but they do not establish that changes in one entity causally produce changes in another.

4.3 Ablation Studies

Cross-entity contribution term on vs. off. We compare the base forecast $\hat{Y}_i = B_i$ with the full model $\hat{Y}_i = B_i + C_i$ across all 9 datasets. The cross-entity term improves both MSE and MAE

Table 3: Ablation: XPanelNet with vs. without the cross-entity contribution term C_i, averaged across four horizons and all nine datasets. Lower is better; **bold** denotes the better variant. Positive Δ (w/o. − full) indicates an improvement from including the cross-entity term.

Dataset	MSE		MAE		Δ MSE	Δ MAE
	Full (with C_i)	w/o C_i	Full (with C_i)	w/o C_i		
Exchange	**0.266**	0.270	0.370	**0.366**	+0.004	-0.004
Electricity	**0.184**	0.192	**0.279**	0.284	+0.008	+0.005
Traffic	0.537	0.537	**0.271**	0.278	+0.000	+0.007
Solar	**0.243**	0.259	**0.242**	0.246	+0.015	+0.004
SAARC Temp	**0.382**	0.406	**0.428**	0.447	+0.024	+0.018
Beijing Air	1.177	**1.164**	0.700	**0.699**	-0.013	-0.001
USGS Flow	**1.048**	1.070	**0.375**	0.386	+0.022	+0.011
Chicago	0.465	**0.457**	0.514	**0.510**	-0.008	-0.004
SDWPF Wind	**0.966**	1.001	**0.691**	0.699	+0.035	+0.008
# datasets helped	6 / 9		6 / 9			

on 6 of 9 datasets, with the largest MSE gain on SDWPF Wind (-0.036), followed by SAARC Temperature (-0.025) and USGS Streamflow (-0.022). The largest corresponding MAE gains occur on SAARC Temperature (-0.018) and USGS Streamflow (-0.011). Performance decreases on both metrics for Beijing Air and Chicago Crime, and on MAE for Exchange Rate, indicating that the benefit of cross-entity information varies across datasets.

Adaptive Fourier frequency selection and future calendar correction. We evaluate adaptive per-dataset frequency selection against a fixed 8-frequency decomposition, followed by the addition of future calendar correction, across all 9 datasets.

Adaptive frequency selection improves MAE and MSE on 4 of 9 datasets each. Its largest gain occurs on Exchange (MAE -0.015, MSE -0.015), while Traffic shows a small MAE regression ($+0.009$) despite a larger MSE improvement (-0.059). Adding future calendar correction yields more consistent gains, improving both MAE and MSE on 7 of 9 datasets, with the largest improvements on Traffic (MAE -0.037, MSE -0.082) and SDWPF Wind (MAE -0.016).

5 Conclusion

We presented XPanelNet, an explainable panel time series forecasting framework that jointly models each entity's temporal dynamics and cross-entity predictive interactions. Its latent VAR/IRF router

Table 4: Ablation of adaptive Fourier frequency selection and future calendar correction. Each cell reports MAE/MSE; **bold** indicates improvement over the preceding stage.

Dataset	Fixed 8 frequencies	+ Adaptive freqs	+ Adaptive + Calendar
Exchange	0.230 / 0.101	**0.215 / 0.086**	**0.209 / 0.084**
Electricity	0.261 / 0.161	0.262 / 0.162	0.266 / 0.168
Traffic	0.312 / 0.627	0.321 / **0.568**	**0.284 / 0.486**
Solar	0.244 / 0.219	**0.236 / 0.208**	**0.233** / 0.217
SAARC Temp	0.413 / 0.349	**0.410** / 0.357	**0.408 / 0.347**
Beijing Air	0.657 / 1.077	0.657 / **1.071**	**0.656 / 1.062**
USGS Flow	0.360 / 1.080	**0.354** / 1.080	**0.350 / 1.074**
Chicago	0.482 / 0.420	0.482 / 0.421	0.482 / **0.418**
SDWPF Wind	0.708 / 1.085	0.716 / 1.094	**0.700** / 1.093

produces signed source-to-target contributions that are explicitly decomposable and exactly reconstruct the aggregate cross-entity output, providing a reconstructable account of how other entities contribute to each forecast. Across nine diverse panel datasets, XPanelNet achieves competitive forecasting accuracy while providing entity-level attribution without constructing dense pairwise tensors during normal inference. Empirical faithfulness evaluations further show that the learned attributions depend on the trained router and distinguish higher- and lower-ranked source entities under deletion-based evaluation. Overall, XPanelNet demonstrates that forecasting and interpretable cross-entity predictive attribution can be integrated within a scalable panel forecasting framework.

The current framework has several limitations. The sequential VAR propagation increases computational cost for longer forecast horizons and large panels. In addition, the router uses a fixed lag order and low-rank coupling, which impose a specific inductive bias on the structure of cross-entity interactions. Finally, the attribution represents predictive dependence within the learned model and should not be interpreted as causal influence.

Future work will investigate more efficient long-horizon propagation and adaptive lag structures while preserving explicit source-to-target decomposition. More flexible dynamic coupling mechanisms could further capture heterogeneous cross-entity dependencies across domains. Additional intervention-based and domain-informed evaluations could also provide stronger evidence for the faithfulness and practical utility of the learned predictive attributions.

AI Use Statement

Generative AI tools were used to assist with manuscript editing, language refinement, and image generation. All research ideas, model design, experiments, analyses, and reported results were produced by the authors. AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript and its contents.

Ethics Statement

This work focuses on panel time series forecasting using publicly available, aggregate, non-identifying datasets. None of the datasets used in our experiments contain individual-level personal data. We do not identify, infer, or make decisions about individuals. Dataset sources and licensing information are provided in Appendix A.1.

Reproducibility Statement

We formally describe the XPanelNet architecture and its components in the main text. Detailed implementation, dataset, training, and experimental configurations are provided in the Appendix.Code and scripts to reproduce the experiments are available at an anonymized repository: `[https://anonymous.4open.science/r/XPanelNet-CC2F]`.

REFERENCES

Julius Adebayo, Justin Gilmer, Michael Muelly, Ian J. Goodfellow, Moritz Hardt, and Been Kim. Sanity checks for saliency maps. In *Neural Information Processing Systems*, 2018. URL https://api.semanticscholar.org/CorpusID:52938797.

Konstantinos Benidis, Syama Sundar Rangapuram, Valentin Flunkert, Yuyang Wang, Danielle Maddix, Caner Turkmen, Jan Gasthaus, Michael Bohlke-Schneider, David Salinas, Lorenzo Stella, et al. Deep learning for time series forecasting: Tutorial and literature survey. *ACM Computing Surveys*, 55(6):1–36, 2022.

Ling Chen, Weiqi Chen, Binqing Wu, Youdong Zhang, Bo Wen, and Chenghu Yang. Learning from multiple time series: A deep disentangled approach to diversified time series forecasting. *arXiv preprint arXiv:2111.04942*, 2021.

Flavio Corradini, Flavio Gerosa, Marco Gori, Carlo Lucheroni, Marco Piangerelli, and Martina Zannotti. A systematic literature review of spatio-temporal graph neural network models for time series forecasting and classification. *Neural Networks*, pp. 108269, 2025.

Jinliang Deng, Xiusi Chen, Renhe Jiang, Du Yin, Yezhou Yang, Xuan Song, and Ivor Wai-Hung Tsang. Disentangling structured components: Towards adaptive, interpretable and scalable time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36:3783–3800, 2023. URL https://api.semanticscholar.org/CorpusID:258832704.

Tryambak Gangopadhyay, Sin Yong Tan, Zhanhong Jiang, Rui Meng, and Soumik Sarkar. Spatiotemporal attention for multivariate time series prediction and interpretation. In *ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 3560–3564. IEEE, 2021.

Javier García-Sigüenza, Manuel Curado, Faraon Llorens-Largo, and Jose F Vicent. Self explainable graph convolutional recurrent network for spatio-temporal forecasting. *Machine Learning*, 114 (1):2, 2025.

Tian Guo, Tao Lin, and Nino Antulov-Fantulin. Exploring interpretable lstm neural networks over multi-variable data. In *International conference on machine learning*, pp. 2494–2504. PMLR, 2019.

Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. Global models for time series forecasting: A simulation study. *Pattern Recognition*, 124:108441, 2022.

Hanwen Hu, Zhangchi Han, Shiyou Qian, Dingyu Yang, Jian Cao, and Guangtao Xue. Pattern-oriented attention mechanism for multivariate time series forecasting. *ACM Transactions on Knowledge Discovery from Data*, 19(2):1–26, 2025.

Qihe Huang, Lei Shen, Ruixin Zhang, Shouhong Ding, Binwu Wang, Zhengyang Zhou, and Yang Wang. Crossgnn: Confronting noisy multivariate time series via cross interaction refinement. *Advances in Neural Information Processing Systems*, 36:46885–46902, 2023.

Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. Timekan: Kan-based frequency decomposition learning architecture for long-term time series forecasting. In *International Conference on Learning Representations*, volume 2025, pp. 28514–28529, 2025.

Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE transactions on pattern analysis and machine intelligence*, 46(12):10466–10485, 2024.

Wonkeun Jo and Dongil Kim. Neural additive time-series models: Explainable deep learning for multivariate time-series prediction. *Expert Syst. Appl.*, 228:120307, 2023. URL https://api.semanticscholar.org/CorpusID:258633019.

Younghwi Kim, Dohee Kim, Joong-Kyun Kim, and Sunghyun Sim. Distributed lag transformer based on time-variable-aware learning for explainable multivariate time series forecasting. *IEEE Access*, 14:115642–115664, 2024. URL https://api.semanticscholar.org/CorpusID:272310433.

Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A*, 379(2194):20200209, 2021.

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.

Yang Lin, Irena Koprinska, and Mashud Rana. Temporal convolutional attention neural networks for time series forecasting. In *2021 International joint conference on neural networks (IJCNN)*, pp. 1–8. IEEE, 2021.

Lei Liu, Xinyu Wang, Xue Dong, Kang Chen, Qiu-Ju Chen, and Bin Li. Interpretable feature-temporal transformer for short-term wind power forecasting with multivariate time series. *Applied Energy*, 2024a. URL https://api.semanticscholar.org/CorpusID:271578167.

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

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. *Advances in neural information processing systems*, 35: 9881–9893, 2022b.

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 *International conference on learning representations*, volume 2024, pp. 11116–11140, 2024b.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*, 2022.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. *ArXiv*, abs/1905.10437, 2019. URL https://api.semanticscholar.org/CorpusID:166228758.

Leonardos Pantiskas, Kees Verstoep, and Henri Bal. Interpretable multivariate time series forecasting with temporal attention convolutional neural networks. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1687–1694. IEEE, 2020.

M Hashem Pesaran, Andreas Pick, and Allan Timmermann. Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity. *Quantitative Economics*, 17(2):342–393, 2026.

Rajat Sen, Hsiang-Fu Yu, and Inderjit S Dhillon. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. *Advances in neural information processing systems*, 32, 2019.

Hiroki Tomioka and Genta Yoshimura. Patchdecomp: Interpretable patch-based time series forecasting. *ArXiv*, abs/2603.03902, 2026. URL https://api.semanticscholar.org/CorpusID:286238279.

Chaoqun Wang, Yijun Li, Xiangqian Sun, Qi Wu, Dongdong Wang, and Zhixiang Huang. Delelstm: Decomposition-based linear explainable lstm to capture instantaneous and long-term effects in time series. *arXiv preprint arXiv:2308.13797*, 2023.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International conference on learning representations*, volume 2024, pp. 38626–38652, 2024.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34:22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*, 2022.

Qilong Wu and Jian Xiao. An explainable fully attentional network for multivariate time-series forecasting. *Knowledge-Based Systems*, 324:113780, 2025.

Chongchong Xu, Chaojie Li, and Xiaojun Zhou. Interpretable lstm based on mixture attention mechanism for multi-step residential load forecasting. *Electronics*, 11(14):2189, 2022.

Jianjun Yuan, Fujun Wu, Luo-Ming Zhao, Dongbo Pan, and Xinyue Yu. Dcfa-itimenet: Dynamic cross-fusion attention network for interpretable time series prediction. *Applied Intelligence*, 55, 2024. URL https://api.semanticscholar.org/CorpusID:274918682.

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

Tianlong Zhao, Lexin Fang, Xiang Ma, Xuemei Li, and Caiming Zhang. Tfformer: A time–frequency domain bidirectional sequence-level attention based transformer for interpretable long-term sequence forecasting. *Pattern Recognition*, 158:110994, 2025.

Liwang Zhou and Jing Gao. Temporal spatial decomposition and fusion network for time series forecasting*. *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–10, 2022. URL https://api.semanticscholar.org/CorpusID:252762469.

A Appendix

A.1 Additional Dataset Details

Table 5 lists the public source and license/terms for each of the nine datasets used in this paper (Section 4), plus the eleventh candidate dataset (Commodities) excluded because it only supports horizons up to 336, not the full $\{96, 192, 336, 720\}$ used elsewhere.

Table 5: Dataset sources and licensing terms. Entries marked "verify" should be double-checked against the current source page before submission.

Dataset	Source	Public?	License / terms
Electricity (321 clients)	UCI ElectricityLoadDiagrams20112014 (via standard LTSF version, Autoformer/Time-Series-Library)	Yes	CC BY 4.0 (verify)
Traffic (862 sensors)	Caltrans PeMS (via LTSF version)	Yes (free registration for raw PeMS)	Caltrans terms of use
Exchange Rate (8 countries)	Lai et al. 2018 (LSTNet)	Yes	Repo code MIT; data has no separate stated license
Solar-Energy (137 plants, Alabama)	NREL Solar Power Data for Integrated Grid Studies (via LTSF version)	Yes	NREL open data
Beijing Air	UCI #501 Beijing Multi-Site Air Quality	Yes	CC BY 4.0 (verify)
SDWPF Wind	Baidu KDD Cup 2022 SDWPF, HuggingFace `AI4Climate/WindFarm_raw`	Yes	Verify on HuggingFace page and associated papers
Chicago Crime (77 areas)	Chicago Data Portal, "Crimes – 2001 to Present"	Yes	City open-data terms
USGS Streamflow (60 gauges)	USGS NWIS Water Services	Yes	US government public-domain data
Commodities (11 futures, excluded)	Yahoo Finance via `yfinance`, 2000–2025	Accessible, not open-licensed	Yahoo terms: personal use; redistribution restricted
SAARC Temperature (8 capitals)	Open-Meteo Historical Archive API	Yes	Attribution required (CC BY 4.0, verify); underlying reanalysis has its own terms

Preprocessing. Missing values are forward- and back-filled; Chicago Crime additionally fills missing days with 0. SDWPF Wind is downsampled from 10-minute to 30-minute means. All datasets use a chronological 0.7/0.1/0.2 split with the scaler fit on the training split only (Section 4).

A.2 Training Configuration Differences

Table 6: Training configuration differences between XPanelNet and the rerun baselines. Both share the same data protocol (Section 4: $L = 96$, four horizons, chronological 0.7/0.1/0.2 split, train-only `StandardScaler`).

	Baselines (Time-Series-Library)	XPanelNet
Runs	1 run each (`--itr 1`)	1 run (seed 1)
Epochs	10 (most models); 15 (TimeMixer); 20 (DLinear)	up to 15
Early stopping patience	3 (library default)	5
Learning rate	10^{-4} default	10^{-3}
Model width (d_{model})	512 (transformers), 32 (TimesNet), 16 (TimeMixer)	32–64
Batch size	8–512, varies by model/dataset	8–32
Hyperparameter search	None (defaults, some model-specific exceptions)	None, but architecture iterated over many design rounds
Loss function	MSE (library default)	MAE

A.3 Additional Results

Per-dataset, per-horizon breakdown of the main results table (Section 4, Table 1), plus additional ablations and extended faithfulness evaluation results once available.

A.4 Forecast Visualizations

Figure 5 shows example forecasts against ground truth for each of the nine datasets used in this paper (Section 4).

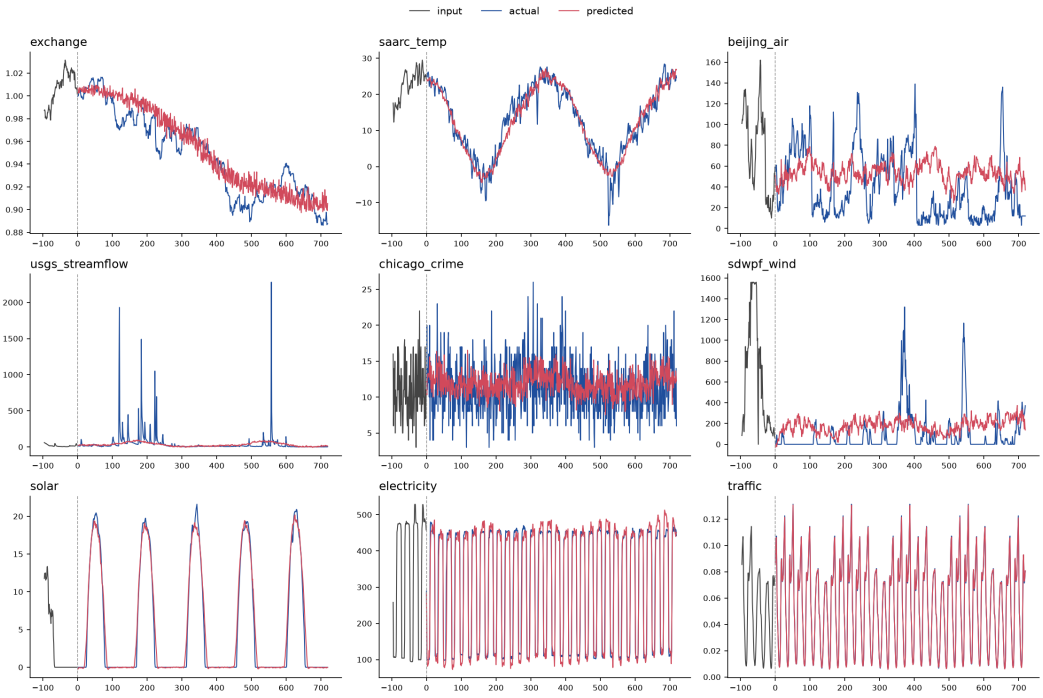

Figure 5: Example forecasts vs. ground truth for one representative entity per dataset (left to right, top to bottom: Exchange, Electricity, Traffic, Solar, SAARC Temperature, Beijing Air, USGS Streamflow, Chicago Crime, SDWPF Wind).

A.5 Implementation Details

VAR router stability bound (Section 3.5). Each lag's coupling matrix is constructed as $A_k = \text{scale}_k \cdot (U_k V_k^\top - \text{diag}(U_k V_k^\top))$, with scale_k chosen so that $\|U_k V_k^\top\|_2 = \rho/p$ before diagonal removal ($\rho = 0.45$, $p = 2$ lags, so each pre-diagonal-removal factor has spectral norm 0.225). Removing the diagonal of a matrix can increase its spectral norm by at most a factor of 2 in the worst case, so each $\|A_k\|_2 \le 2 \times 0.225 = 0.45$ after diagonal removal, giving $\sum_{k=1}^{p} \|A_k\|_2 \le 2 \times 0.45 = 0.9 < 1$. This keeps the VAR recursion $x_\tau = \sum_k A_k x_{\tau-k}$ stable (bounded, non-exploding) for any H. [NOTE: this bound is derived, not directly asserted in the codebase; the code's own comment states only the pre-diagonal-removal bound. Confirm this derivation independently before relying on it as a formal stability guarantee in the paper, since the factor-of-2 worst case for diagonal removal is a general matrix-norm bound, not one specific to this construction, and may be loose in practice.]

Low-rank factor initialization and rank rule. $U_k, V_k \in \mathbb{R}^{N \times r}$ are initialized i.i.d. $\mathcal{N}(0, 1/r)$. The rank r is set to N for panels with $N \le 64$ entities (effectively full-rank coupling) and to $\lceil \sqrt{N} \rceil$ for larger panels, trading off expressiveness against parameter count and compute as N grows; Traffic ($N = 862$) uses $r = \lceil \sqrt{862} \rceil = 30$.

Table 7: Full results (MSE, MAE) across all nine datasets and four horizons, XPanelNet against all 10 baselines, ordered chronologically (newest first). Bold marks the best value per metric within each row, all models rerun under a matched data protocol (Section 4). See Table 6 (Appendix A.2) for training-configuration differences, including the MSE/MAE loss asymmetry between XPanelNet and the baselines.

Models	H	**XPanelNet (Ours)**		TimeKAN		TimeMixer		iTransformer		CrossGNN		PatchTST		TimesNet		DLinear		SCINet		Stationary		Autoformer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Exchange Rate	96	**0.082**	0.206	0.087	0.206	0.090	0.235	0.088	0.209	0.084	**0.203**	0.088	0.205	0.107	0.234	0.088	0.218	0.267	0.396	0.111	0.237	0.197	0.323
	192	**0.161**	**0.290**	0.191	0.310	0.187	0.343	0.181	0.304	0.171	0.294	0.176	0.299	0.226	0.344	0.176	0.315	0.351	0.459	0.219	0.335	0.300	0.369
	336	**0.292**	0.414	0.347	0.425	0.353	0.473	0.334	0.419	0.319	0.407	0.301	**0.397**	0.367	0.448	0.313	0.427	1.324	0.853	0.421	0.476	0.509	0.524
	720	**0.528**	**0.570**	0.954	0.736	0.934	0.761	0.829	0.691	0.805	0.677	0.901	0.714	0.964	0.746	0.839	0.695	1.058	0.797	1.092	0.769	1.447	0.941
Electricity	96	**0.146**	**0.240**	0.174	0.266	0.153	0.247	0.148	0.240	0.173	0.275	0.190	0.251	0.168	0.272	0.210	0.282	0.247	0.345	0.169	0.273	0.201	0.317
	192	0.179	0.274	0.182	0.273	0.166	0.256	**0.162**	**0.253**	0.195	0.288	0.199	0.262	0.184	0.322	0.210	0.285	0.257	0.355	0.182	0.286	0.222	0.334
	336	0.190	0.286	0.197	0.286	0.185	0.277	**0.178**	**0.269**	0.206	0.300	0.217	0.279	0.198	0.300	0.223	0.301	0.269	0.369	0.200	0.304	0.231	0.338
	720	0.221	0.318	0.236	0.320	0.225	**0.310**	0.225	0.319	0.231	0.335	0.258	0.313	**0.220**	0.320	0.258	0.333	0.299	0.390	0.222	0.321	0.254	0.361
Traffic	96	0.511	**0.260**	0.589	0.380	0.462	0.285	**0.395**	0.268	0.570	0.310	0.526	0.272	0.593	0.321	0.650	0.396	0.788	0.499	0.612	0.338	0.613	0.388
	192	0.514	**0.265**	0.560	0.362	0.473	0.296	**0.417**	0.276	0.577	0.321	0.522	0.289	0.617	0.336	0.598	0.370	0.789	0.505	0.613	0.340	0.616	0.382
	336	0.542	**0.270**	0.562	0.363	0.498	0.296	**0.433**	0.283	0.588	0.324	0.517	0.282	0.629	0.336	0.605	0.373	0.797	0.508	0.618	0.328	0.622	0.337
	720	0.582	**0.288**	0.603	0.381	0.506	0.313	**0.467**	0.302	0.597	0.337	0.522	0.301	0.640	0.350	0.645	0.394	0.841	0.523	0.653	0.355	0.660	0.408
Solar-Energy	96	0.217	**0.230**	0.253	0.300	**0.189**	0.259	0.203	0.237	0.222	0.301	0.205	0.323	0.373	0.358	0.290	0.378	0.237	0.344	0.321	0.380	0.456	0.446
	192	0.237	**0.240**	0.279	0.319	**0.222**	0.283	0.233	0.261	0.246	0.307	0.237	0.332	0.397	0.376	0.320	0.398	0.280	0.380	0.346	0.369	0.588	0.588
	336	0.252	**0.246**	0.307	0.336	**0.231**	0.292	0.248	0.273	0.263	0.324	0.250	0.339	0.420	0.380	0.353	0.415	0.304	0.389	0.357	0.387	0.595	0.590
	720	0.268	**0.254**	0.300	0.327	**0.223**	0.285	0.249	0.275	0.265	0.318	0.252	0.336	0.420	0.381	0.357	0.413	0.308	0.388	0.375	0.424	0.733	0.733
SAARC Temperature	96	**0.357**	**0.413**	0.750	0.656	0.671	0.612	0.739	0.652	1.221	0.850	0.890	0.716	0.938	0.761	0.570	0.562	2.075	1.165	0.437	0.493	1.056	0.822
	192	**0.377**	**0.430**	1.139	0.823	0.968	0.747	1.094	0.829	2.188	1.185	1.333	0.900	1.135	0.847	0.673	0.607	2.662	1.356	0.529	0.552	2.143	1.210
	336	**0.391**	**0.434**	0.912	0.713	0.758	0.636	0.880	0.713	2.346	1.250	1.069	0.767	0.902	0.733	0.594	0.553	2.105	1.153	0.477	0.511	1.709	1.035
	720	**0.401**	**0.436**	0.896	0.697	0.860	0.681	0.902	0.721	2.270	1.217	1.106	0.777	1.049	0.796	0.654	0.606	2.078	1.137	0.551	0.554	1.898	1.106
Beijing Air	96	1.059	**0.656**	1.127	0.715	1.189	0.715	1.243	0.738	1.192	0.749	1.103	0.704	1.241	0.745	**0.990**	0.697	1.474	0.838	1.105	0.698	1.333	0.822
	192	1.160	**0.700**	1.243	0.753	1.257	0.752	1.379	0.784	1.352	0.802	1.210	0.749	1.370	0.791	**1.081**	0.738	1.585	0.866	1.173	0.750	1.453	0.865
	336	1.213	**0.712**	1.222	0.765	1.336	0.781	1.400	0.797	1.379	0.817	1.266	0.773	1.377	0.795	**1.118**	0.754	1.576	0.865	1.205	0.754	1.444	0.841
	720	1.276	**0.730**	1.340	0.778	1.360	0.784	1.449	0.815	1.460	0.842	1.330	0.789	1.453	0.819	**1.169**	0.769	1.640	0.880	1.302	0.768	1.513	0.856
USGS Streamflow	96	**1.144**	**0.377**	1.211	0.443	1.205	0.435	1.230	0.454	1.694	0.566	1.223	0.453	1.215	0.462	1.174	0.486	1.686	0.645	1.214	0.461	1.283	0.584
	192	**1.144**	**0.383**	1.274	0.480	1.265	0.475	1.300	0.490	1.908	0.638	1.291	0.492	1.269	0.492	1.199	0.511	1.775	0.687	1.299	0.502	1.330	0.614
	336	**1.011**	**0.379**	1.112	0.464	1.107	0.462	1.138	0.472	1.834	0.640	1.124	0.467	1.178	0.493	1.048	0.493	1.616	0.663	1.176	0.491	1.229	0.587
	720	**0.892**	**0.361**	1.003	0.452	0.998	0.444	1.037	0.476	1.897	0.659	1.019	0.469	1.080	0.499	0.923	0.487	1.610	0.684	1.057	0.486	1.173	0.596
Chicago Crime	96	0.423	0.486	0.425	0.489	0.424	0.490	0.416	**0.482**	0.845	0.691	0.425	0.488	**0.414**	0.484	0.431	0.500	0.478	0.521	0.414	0.482	0.435	0.500
	192	0.450	0.504	0.445	0.498	0.433	0.496	0.427	0.491	0.860	0.700	0.442	0.499	0.429	0.493	0.448	0.514	0.497	0.534	**0.423**	**0.488**	0.450	0.509
	336	0.469	0.517	0.451	0.503	0.443	0.500	0.449	0.498	0.864	0.703	0.456	0.504	0.440	0.500	0.502	0.563	0.501	0.536	**0.433**	**0.494**	0.488	0.538
	720	0.516	0.550	0.506	0.532	0.501	0.526	0.472	0.513	0.884	0.713	0.546	0.524	0.466	0.513	0.619	0.642	0.528	0.550	**0.460**	**0.507**	0.834	0.732
SDWPF Wind	96	1.072	**0.704**	1.088	0.781	1.108	0.739	1.148	0.763	1.482	0.818	1.132	0.775	1.200	0.788	**0.968**	0.771	1.219	0.799	1.050	0.752	1.287	0.898
	192	**0.955**	**0.701**	1.078	0.739	1.023	0.745	1.050	0.720	1.611	0.870	1.069	0.746	1.153	0.771	1.043	0.823	1.180	0.783	1.029	0.775	1.194	0.803
	336	**0.945**	**0.697**	1.023	0.737	1.035	0.726	1.072	0.738	1.684	0.896	1.091	0.733	1.162	0.784	1.025	0.818	1.179	0.800	1.029	0.775	1.224	0.830
	720	**0.891**	**0.663**	0.937	0.717	0.971	0.744	1.011	0.761	1.792	0.956	0.989	0.748	1.139	0.812	0.948	0.802	1.014	0.753	1.116	0.859	1.252	0.868

Gate cap. The router's readout gate $g_{i,h} = g_{\max} \tanh(\cdot)$ uses $g_{\max} = 0.5$, the same cap used in the earlier prototype-based router design; it was not independently re-tuned for the VAR router and this should be flagged as a design choice inherited rather than re-validated for the current architecture.

Router scalability. The VAR router never forms an $(N, N, \cdot)$ tensor in the normal forward pass; the largest intermediate tensors are (B, N, H, m) with $m = 4$. Per-forward cost is $O(H \cdot p \cdot B \cdot N \cdot r \cdot m)$ for the recursion plus $O(p \cdot N \cdot r^2)$ for the per-lag spectral-norm computation (via QR/small SVD on the two $N \times r$ factors), giving $O(N^2)$ when $r = N$ ($N \leq 64$) and $O(N^{1.5})$ when $r = \lceil \sqrt{N} \rceil$. The H-step recursion is a sequential loop, which made horizon-720 training runs on Electricity roughly $3\times$ slower than under the earlier prototype-based router (approximately 25 vs. 9 minutes per epoch); this cost should be reported alongside the accuracy results as a practical limitation.

Code state at time of writing. The reported architecture (VARRouter, in the module referred to internally as `XPanelNet_trdvar`) coexists in the released codebase with an earlier, unused prototype-based router implementation (`CrossEntityRouter`), which remains present but is not invoked by the final model. Before any code release, the unused prototype-router path should be removed or clearly marked as deprecated/superseded, so that released code matches the architecture described in this paper exactly and does not mislead a reader auditing the repository.

[Full hyperparameters, architecture dimensions, training details, and compute resources used.]