A Appendix

Table 6: Additional comparison results for multivariate long-term series forecasting, including Onefitsall, TimeLLM (Jin et al.), ModernTCN (Luo & Wang, 2024), UniTST (Liu et al., 2024a), and TSLANet (Eldele et al.). K refers to the number of variables in the dataset.

Models		MambaTS (Ours)		Onefitsall (2023)		TimeLLM (2024)		ModernTCN (2024)		UniTST (2024)		TSLANet (2024)	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather ($K = 21$)	96	**0.145**	**0.195**	0.162	0.212	0.147	0.201	0.149	0.200	0.156	0.202	0.148	0.197
	192	0.192	0.241	0.204	0.248	**0.189**	**0.234**	0.196	0.245	0.207	0.250	0.193	0.241
	336	0.245	0.283	0.254	0.286	0.262	0.279	**0.238**	**0.277**	0.263	0.292	0.245	0.282
	720	0.314	0.330	0.326	0.337	**0.304**	**0.316**	0.314	0.334	0.340	0.341	0.325	0.337
Electricity ($K = 321$)	96	**0.128**	**0.223**	0.139	0.238	0.131	0.224	0.129	0.226	0.139	0.235	0.136	0.229
	192	0.145	0.239	0.153	0.251	0.152	0.241	**0.143**	0.239	0.155	0.250	0.152	**0.229**
	336	0.163	0.259	0.169	0.266	**0.160**	**0.248**	0.161	0.259	0.170	0.268	0.168	0.262
	720	0.192	**0.286**	0.206	0.297	0.192	0.298	**0.191**	**0.286**	0.198	0.293	0.205	0.293
Traffic ($K = 862$)	96	**0.347**	**0.248**	0.388	0.282	0.362	**0.248**	0.368	0.253	0.402	0.255	0.372	0.261
	192	**0.358**	0.255	0.407	0.290	0.374	**0.247**	0.379	0.261	0.426	0.268	0.388	0.266
	336	**0.372**	**0.262**	0.412	0.294	0.385	0.271	0.397	0.270	0.449	0.275	0.394	0.269
	720	**0.416**	**0.284**	0.450	0.312	0.430	0.288	0.440	0.296	0.489	0.297	0.430	0.289
AVG.		**0.251**	0.259	0.273	0.276	0.257	**0.258**	0.259	0.262	0.283	0.269	0.263	0.263

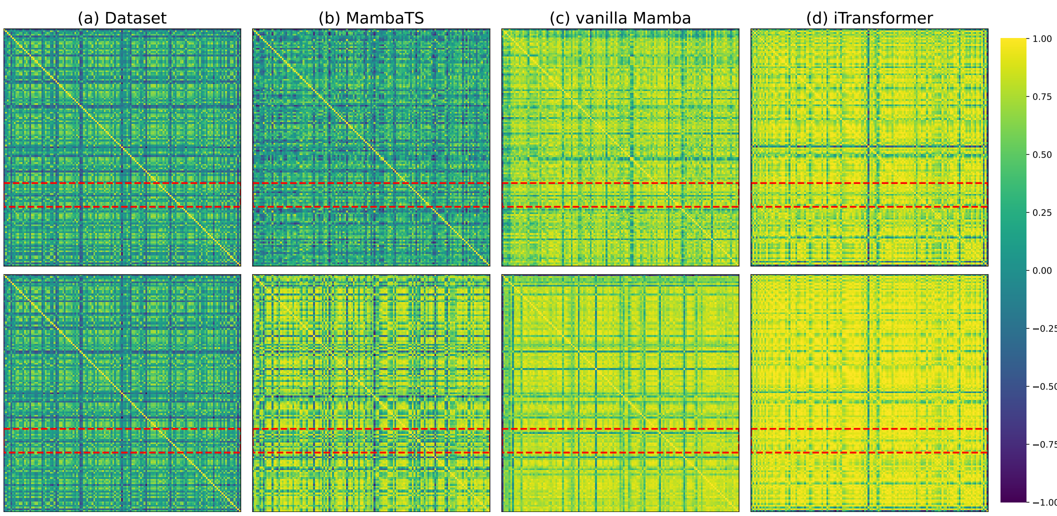

Figure 5: Visualization of variable correlations at different layers, including the first layer (top row) and the final layer (bottom row). Correlation coefficients between variables are calculated to quantify their interdependencies within the dataset (a). Notably, MambaTS (b) benefits from variable-aware scan along time (VAST), learning richer global dependencies compared to vanilla Mamba (c) (highlighted in the red-dashed regions). Additionally, while both MambaTS and iTransformer (d) capture similar variable dependencies, MambaTS learns a more intricate dependency graph, while iTransformer's graph appears smoother, particularly in the final layer.

Full Proof of Proposition 2

Proposition 2. *Given a causal graph $G = (V, E)$ with unknown relationships among nodes $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K\}$, if the total cost of a random walk without return is known, then the causal relationships can be estimated through infinite random walks without return.*

Proof. Let $G = (V, E)$ be the causal graph, with nodes $\mathbf{V} = \{\mathbf{V}_1, \ldots, \mathbf{V}_K\}$ and unknown causal relationships. The goal is to estimate these causal relationships based on random walks without return.

Define the cost matrix $\boldsymbol{P}$, where $p_{i,j}$ represents the cost $C_{i,j}$ of transitioning from node $\mathbf{V}_i$ to $\mathbf{V}_j$. For a random walk without return involving K nodes, there are $K-1$ transitions. Since the individual contribution of each transition to the total cost C is unknown, we assume that the cost is evenly distributed across these $K-1$ transitions. Therefore, we focus on the question of whether the total cost C reflects the actual transition costs.

Transitions are classified into three types:

- **Positive transition (PT).** A transition from node $\mathbf{V}_i$ to node $\mathbf{V}_j$ is positive if there exists a direct or indirect causal relationship from $\mathbf{V}_i$ to $\mathbf{V}_j$ in G, i.e., there exists a directed path from $\mathbf{V}_i$ to $\mathbf{V}_j$ in the graph. This transition incurs a positive cost, $C_{i,j} > 0$, reflecting the causal influence of $\mathbf{V}_i$ on $\mathbf{V}_j$, either directly or through intermediate nodes.
- **Negative transition (NT).** A transition from node $\mathbf{V}_i$ to node $\mathbf{V}_j$ is negative if $\mathbf{V}_i$ is causally influenced by $\mathbf{V}_j$, i.e., there exists a directed path from $\mathbf{V}_j$ to $\mathbf{V}_i$ in G, either directly or through intermediate nodes. This transition incurs a negative cost, $C_{i,j} < 0$, with magnitude equal to the positive transition in the reverse direction.
- **Independent node transition (IN).** A transition from $\mathbf{V}_i$ to $\mathbf{V}_j$ is independent if there is no direct causal relationship between them, i.e., $(\mathbf{V}_i, \mathbf{V}_j) \notin E$ and $(\mathbf{V}_j, \mathbf{V}_i) \notin E$. Such transitions are assigned a positive cost $0 < C_{i,j} < C_{\text{max}}$, reflecting valid movement within the graph without disrupting causal structure. [1]

A key observation is that, due to the symmetry of causal relationships in the graph, the number of positive transitions #PT equals the number of negative transitions #NT. Consequently, for any given random walk without return, the proportion of transitions that contribute correctly to the cost update is at least $\frac{\#\text{PT}+\#\text{IN}}{\#\text{PT}+\#\text{NT}+\#\text{IN}} \geq \frac{1}{2}$. This bound is tight and equality holds if and only if there are no independent node transitions, i.e., $\#\text{IN} = 0$.

As the number of random walks $N \to \infty$, the expected cost for each transition converges to a positive value, reflecting the underlying causal relationships:

$$\mathbb{E}[C_{i,j}^{(n)}] > 0, \quad \text{for all } i, j. \tag{8}$$

After N random walks, the average cost $p_{i,j}$ for the transition from node $\mathbf{V}_i$ to node $\mathbf{V}_j$ is given by:

$$p_{i,j} = \frac{1}{N} \sum_{n=1}^{N} C_{i,j}^{(n)}. \tag{9}$$

Finally, the strength of the causal relationship between $\mathbf{V}_i$ and $\mathbf{V}_j$ is estimated as:

$$\hat{R}_{i,j} = \frac{p_{i,j}}{\sum_{k \in V} p_{i,k}}, \tag{10}$$

where the denominator normalizes the cost relative to all other outgoing transitions from $\mathbf{V}_i$, providing a measure of the relative strength of the causal relationship.

Thus, as the number of random walks approaches infinity, the causal relationships among nodes in the graph can be accurately estimated through random walks. □

[1] Our experimental results, as shown in Table 3, confirm that handling independent node transitions this way contributes effectively to the overall performance of the VAST method, demonstrating that even transitions between causally independent nodes can provide valuable information within the causal inference framework.