Correlation-Aware Reordered Scanning Mamba for Multivariate Time Series Forecasting

Zihao Yao, Qi Zheng, and Yaying Zhang(✉)

The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China
{yaozihao,zhengqi97,yaying.zhang}@tongji.edu.cn

Abstract. The fundamental challenge in Multivariate Time Series forecasting is effectively modeling complex temporal dependencies and variable correlation. Transformer-based models achieve breakthroughs but face challenges with quadratic complexity and permutation invariant bias. A recent model, Mamba, has emerged as a competitive alternative. However, we observe that the issue of scan order sensitivity is not well concerned. In this study, we propose a novel Correlation-aware Reordered Scanning Mamba, namely CRS-Mamba, for multivariate time series forecasting. Specifically, we leverage the downsampling technique to model temporal dependencies. Then, a bidirectional Mamba layer is introduced to extract inter-variate correlations. Moreover, we propose Dimensionality Reduction Scan Algorithm to alleviate scanning sensitivity problem of Mamba. Extensive evaluations show that our approach secures superior performance in prediction accuracy on various datasets. Moreover, ablation studies demonstrate the interpretability of CRS-Mamba.

Keywords: Multivariate Time Series · Forecasting · Mamba.

1 Introduction

Multivariate time series (MTS) comprises time series with multiple variables. In recent years, MTS prediction has gained attention for its extensive use in various fields, including transportation, climate and energy systems [1].

The complex temporal dependencies and uncertain variable correlations are both pivotal for MTS. Generally, the task is challenge for two reasons: the temporal dependencies can be obscured by entangled temporal patterns [2], and the inter-variate correlation modeling may simultaneously introduce noise from irrelevant variables. Transformer-based methods demonstrated great power benefiting from self-attention mechanism [3] but suffer from quadratic complexity.

A recent model, Mamba [4], based on state space models (SSMs) [5], have shown competitive performance over Transformer in various domains [6, 7]. It employs selection mechanism to filter out irrelevant information. For efficiency, Mamba further utilizes hardware-aware algorithm with linear complexity.

Despite the effectiveness of existing models, there are still some limitations.

In temporal dimension, intricate temporal patterns present considerable uncertainty and require models to extract extensive temporal dependencies. Many

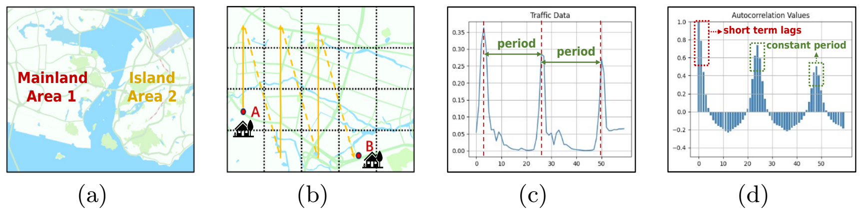


Fig. 1. Findings about MTS. (a) The correlation between routes is influenced by regions to which they belong. (b) Normal scanning mode overlook correlation between A and B. (c) The periodic characteristics of time series. (d) ACF value of period time series.

Transformer-based models focus on complex feature extraction and have not fully utilized the unique characteristic of time series. The recurrent mode of Mamba takes into account the properties of time series. But it primarily targets long-range dependencies [4], limiting its performance in short-term input tasks.

In variable dimension, the sparsity of inter-variate correlation is not well considered by current models. As shown in Figure 1(a), there will be a clear correlation between two roads within one area(eg. Area 1 or Area 2). But the correlations between roads from two different areas(eg. one road in Area 1 and another road in Area 2) may be weak. If all roads are indistinguishably treated, the existing global attention mechanism that models these variables will inevitably introduce more noise and increase unnecessary parameters.

In Mamba-based models, establishing an appropriate scanning order for variables is difficult. The sequence scanning paradigm of Mamba is only suitable for 1-D sequence. Unfolding variables into 1-D sequence will generate scan order sensitivity problem. As shown in Figure 1(b), nodes A and B have similar patterns as residential areas. With 'N'-shaped scanning order, Mamba may lead to losing crucial information due to the long scanning interval between them.

In fact, extracting and modeling complex correlations between strongly correlated data is more effective than directly using complex network. (1) As shown in Figure 1(c)1(d), ACF value is typically high in short-term lags while a distinct peak appears in constant period lags. It implies that both lag modes are beneficial for modeling current time point. (2) Graph can better reflect the correlation between variables. The graph structure is detached from the limitation of physical space. Therefore, the scan order generated from it is more versatile.

In this study, we propose the CRS-Mamba (Correlation-aware Reordered Scanning Mamba) for MTS forecasting. Specifically, we devise an Adaptive Period Identifier to obtain period of dataset. In temporal dependencies modeling, we utilize downsampling [8] to aggregate contextual semantic and model cross-period trend. In variable correlation modeling, we design a bidirectional Mamba layer to overcome the limitation of unidirectional modeling paradigm. As core contribution, we propose Dimensionality Reduction Scan Algorithm to alleviate scan order sensitivity problem. The contributions are summarized as follows:

- We design the CRS-Mamba for accurate MTS prediction. Based on the inherent characteristics of MTS in both time and variable dimension, our approach effectively extract vital information and augment correlation modeling.
- We propose Dimensionality Reduction Scan (DRS) Algorithm. It sorts relevant variables to obtain scanning order based on graph. The algorithm effectively alleviate the scanning sensitivity problem of Mamba.
- We conduct experiments on eight public benchmark datasets. Results demonstrate the superior predictive performance and interpretability of CRS-Mamba.

2 Related Work

2.1 Modeling Interaction Across Time and Variable

Modeling interaction across time aims to extract temporal dependencies. Transformer based models [2, 9, 10] that integrate the features of time series to improve attention mechanism have been widely explored. Nevertheless, the self-attention mechanism suffers from the quadratic complexity. To decrease the computational complexity, many linear-based models [11, 8] have been presented and achieve impressive results in modeling temporal dependencies.

Cross variable modeling is vital for more robust representation of variables. Graph-based methods [12] are effective in modeling message propagation and widely used in MTS forecasting. For Transformer-based methods, Crossformer [13] and iTransformer [14] embed variables into tokens with attention mechanism and achieve SOTA. Despite the efforts, we notice that these methods do not handle complex variable correlation well and result in unavoidable noise. Therefore, we leverage Mamba to address the issue with proposed scan algorithm.

2.2 Mamba and its Scanning Mode

The sequence modeling paradigm of Mamba hinders comprehensive learning process for non-sequential data. In computer vision, models focus on scanning entire region with different curves from multiple directions [7]. However, they are not suitable for MTS which cannot form a Euclidean division. To tackle the issue, MambaTS [15] presents a variable permutation training strategy, which lacks intuitive consideration of variable associations. To this end, we propose DRS Algorithm to enhance expressiveness of Mamba with high interpretability.

3 Preliminaries

In MTS forecasting, the input historical sequence across N variables is denoted by $X = \{X_1^t, X_2^t, \ldots, X_N^t\}_{t=1}^{L} \in \mathbb{R}^{L\times N}$, with each X_i^t representing the time series of the i_{th} variate at the t_{th} time step and L denoting the length of the historical observation window. The objective of MTS forecasting is to predict future values denoted by $\bar{X} = \{\bar{X}_1^t, \bar{X}_2^t, \ldots, \bar{X}_N^t\}_{t=L+1}^{L+T} \in \mathbb{R}^{T\times N}$ based on previously observed MTS data X, where T represents the prediction time steps.

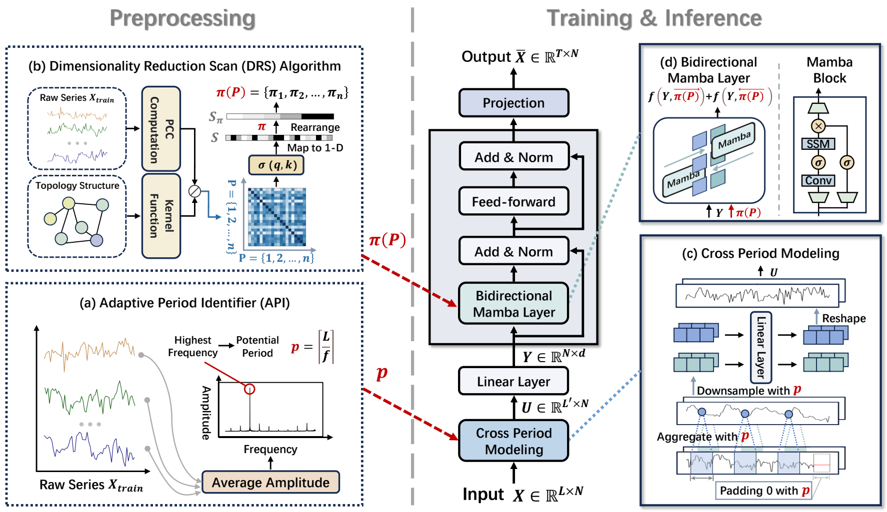


Fig. 2. The architecture of CRS-Mamba. (a) Adaptive period identifier captures period by FFT. (b) Dimensionality Reduction Scan Algorithm is applied to mitigate scan order sensitivity problem. (c) Downsampling is utilized to model cross period temporal dependencies. (d) The detailed structure of Bidirectional Mamba Layer.

4 Methodology

We divide the model into preprocessing and training phases. In preprocessing, we devise an Adaptive Period Identifier to obtain period of dataset. Then we propose the DRS Algorithm to obtain scan order of variables. During training, to model temporal dependencies, we aggregate the contextual semantic and utilize downsampling to model cross period trend. In variable correlation modeling, a bidirectional Mamba layer with scan order derived from preprocessing is introduced. The overview of proposed CRS-Mamba is illustrated in Figure 2.

4.1 Adaptive Period Identifier (API)

The API is designed to capture period of MTS. Technically, we analyze the time series in frequency domain by Fast Fourier Transform (FFT) as follows:

$$A = \mathrm{Avg}_{i=1}^{N}(\mathrm{Amp}(\mathrm{FFT}(X_{train})). \tag{1}$$

$\mathrm{Amp}(\cdot)$ denotes the calculation of amplitude values, X_{train} represents the raw MTS of training set, A represents the calculated amplitude of each frequency, which is averaged from N variables by $\mathrm{Avg}(\cdot)$. We select the most significant frequency f correspond to the top amplitude value, i.e. $f = \mathrm{argTop}(A_i)$.

The selected frequency correspond to period length $p = \left\lceil \frac{L}{f} \right\rceil$. To standardize the input X for upcoming module, we pad X to be an integer L' multiple of the period by zero. The padding length is extended to $\left\lceil \frac{L}{p} \right\rceil \times P - L$

4.2 Dimensionality Reduction Scan Algorithm

For dataset with physical distance between variables, we can construct a distance-based graph $\mathbf{G}$ using a Gaussian radial basis function [1]:

$$g_{ij} = \begin{cases} \frac{\exp\left(-\|d_{ij}\|_2\right)}{\theta}, \text{ if } d_{ij} < \epsilon \\ 0, \quad \text{otherwise} \end{cases}. \tag{2}$$

d_{ij} denotes the distance between node i and node j; θ is a hyper-parameter to control the distribution; ϵ is a predefined threshold to control the sparsity of $\mathbf{G}$.

In scenarios where distance information is absent, the adjacency matrix of a similarity-based graph $\mathbf{G}$ can be constructed by Pearson Correlation Coefficient:

$$g_{ij} = \text{abs}\left(\frac{\sum_{k=1}^{L_t}\left(x_i^k - \bar{x}_i\right)\left(x_j^k - \bar{x}_j\right)}{\sqrt{\sum_{k=1}^{L_t}\left(x_i^k - \bar{x}_i\right)^2}\sqrt{\sum_{k=1}^{L_t}\left(x_j^k - \bar{x}_j\right)^2}}\right). \tag{3}$$

x_i^k and x_j^k represents the value of node i and node j in the time series at the k_{th} time step; $\bar{x}_i$ and $\bar{x}_j$ are the mean value of time series of node i and node j in X_{train}; L_t denotes the number of samples time steps.

Then, an improved activation function $\sigma(q,k)$ defined with a domain and range $[0,1]$ is applied to control the element values of graph $\mathbf{G}$:

$$\sigma = \left(\frac{\frac{1}{1+e^{-kx+\frac{1}{2}k}} - \frac{1}{1+e^{\frac{1}{2}k}}}{\frac{1}{1+e^{-\frac{1}{2}k}} - \frac{1}{1+e^{\frac{1}{2}k}}}\right)^q, \tag{4}$$

where q controls the attenuation rate and k controls the degree of neglect of low values. Specially, with appropriate parameter tuning, it can be regarded as none activation function or treat $\mathbf{G}$ as unweighted topology-based graph.

The core idea is that the related variables should be adjacent during scanning, while unrelated variables should be kept apart. Assuming that each variable corresponds to a real value z, then the absolute value of the difference between variables i and j should be close to the distance g_{ij} between variables i and j. We aim to minimize the overall disparity across all variables. Therefore, the problem can be translated into mathematical forms to solve the objective function:

$$\min \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|z_i - z_j\|^2 g_{ij}. \tag{5}$$

z_i corresponds to the mapped value of variable i, g_{ij} denotes the element of $\mathbf{G}$.

The original objective function can be simplified to $\text{tr}\left(Z^T\mathbf{L}Z\right)$, where $Z = (z_1, z_2, \ldots, z_N)^T \in \mathbb{R}^N$ represents the mapping vector of variables, $\mathbf{L}$ denotes Laplacian matrix of $\mathbf{G}$, $\text{tr}(\cdot)$ denotes trace of matrix. To ensure that the mapped nodes are distributed throughout the mapping space rather than clustering together, additional constraint $Z^T\mathbf{D}Z = \mathbf{I}$ is introduced. We can then use Lagrange multipliers to solve the objective function:

$$\begin{array}{l} f(Z) = \text{tr}\left(Z^T\mathbf{L}Z\right) + \text{tr}\left[\mathbf{\Lambda}\left(Z^T\mathbf{D}Z - \mathbf{I}\right)\right], \frac{\partial f(Z)}{\partial Z} = 0 \\ \Rightarrow \mathbf{L}Z = -\mathbf{D}Z\mathbf{\Lambda} \end{array}, \tag{6}$$

where $\mathbf{\Lambda}$ is a diagonal matrix. The above optimization function is transformed into a generalized eigenvalue problem. By finding the eigenvector corresponding to the smallest non-zero eigenvalue, the desired vector Z can be obtained.

Sort the elements of the obtained one-dimensional vector Z. Since the mapping values of the related variables are numerically close, the corresponding order after sorting will also be adjacent, thus yielding the desired scanning sequence.

4.3 Cross Period Modeling

From the analysis of auto-correlation, the variation of time points are similar with adjacent area and the same phase among periods. Here, we utilize downsampling [8] to separate the periodic components and focus on trend variations.

To incorporate local semantic and mitigate the impact of outliers, we perform a sliding aggregation in terms of temporal dimension. Each aggregated data point incorporates information from other points within its surrounding period. Technically, this sliding aggregation can be implemented using a 1D convolution and a kernel size of $2 \times \left\lfloor \frac{p}{2} \right\rfloor + 1$. It can be formulated as follows:

$$x^i_{t-L+1:t} = x^i_{t-L+1:t} + \text{Conv1D}(x^i_{t-L+1:t}) \tag{7}$$

Then, downsample the series into p subsequences of length $\left\lfloor \frac{L'}{p} \right\rfloor$. A fully connected layer with parameter sharing is then applied to these subsequences. After modeling, p subsequences are reshaped back to the sequence of original length.

4.4 Inter-Variate Correlations Modeling

The scanning process of Mamba is recursive pattern which exhibits unidirectionality, resulting in only incorporating preceding variables information. However, there is no sequential recursion pattern between different variables. In order to focus on the global context and better model the correlations among effective variables, we use two Mamba modules to form a bidirectional Mamba layer:

$$Y = f(Y; \overleftarrow{\pi(P)}) + f(Y; \overrightarrow{\pi(P)}) + Y, \tag{8}$$

where Y denotes the input, $\overrightarrow{\pi(P)}$ represents the forward order obtained from DRS algorithm, $\overleftarrow{\pi(P)}$ represents the inverse order, f is the Mamba module.

Then a normalization layer integrating feed-forward network (FFN) is employed to enhance convergence and training stability in deep networks by standardizing all variates to a Gaussian distribution [14]. With the stacking of blocks, CRS-Mamba can encode the observed time series and decode the representations for future series using dense non-linear connections.

5 Experiments

Datasets and Baseline: We verify the performance of CRS-Mamba on eight real-world datasets, including ECL, Traffic, ETTh1, ETTh2[2] and PEMS03,

Table 1. Accuracy comparison over different MTS forecasting models, the lower the better. A look-back window size L is fixed at 720 for PEMS datasets and 96 for the remaining datasets. The best results are in **bold** and the second best are underlined.

Method	T	**CRS-Mamba**		S-Mamba		SparseTSF		DLinear		iTransformer		PatchTST		Crossformer		FEDFormer		AutoFormer	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.384	0.404	0.388	0.406	0.385	**0.391**	0.383	0.396	0.387	0.405	0.379	0.399	0.406	0.437	**0.375**	0.415	0.533	0.490
	192	0.433	0.434	0.445	0.441	0.435	**0.420**	0.433	0.426	0.441	0.436	**0.425**	0.431	0.458	0.457	0.427	0.448	0.514	0.487
	336	0.470	0.453	0.490	0.465	0.476	**0.440**	0.491	0.467	0.491	0.462	0.470	0.457	0.611	0.574	**0.458**	0.465	0.507	0.493
	720	0.492	0.485	0.506	0.497	**0.461**	**0.454**	0.528	0.519	0.509	0.494	0.523	0.507	0.731	0.646	0.484	0.496	0.547	0.522
ETTh2	96	**0.293**	**0.346**	0.297	0.349	0.303	0.347	0.329	0.380	0.301	0.350	0.294	0.347	0.713	0.593	0.341	0.385	0.377	0.411
	192	0.375	0.398	0.378	0.399	0.385	0.399	0.431	0.443	0.380	0.399	**0.372**	**0.397**	2.861	1.316	0.433	0.441	0.443	0.451
	336	0.423	0.434	0.425	0.435	**0.421**	**0.428**	0.459	0.462	0.424	0.434	0.426	0.440	2.048	1.145	0.504	0.495	0.501	0.496
	720	0.431	0.451	0.432	0.448	**0.420**	**0.437**	0.774	0.631	0.432	0.447	0.435	0.456	2.835	1.449	0.479	0.486	0.501	0.502
ECL	96	**0.138**	**0.234**	0.139	0.236	0.210	0.280	0.195	0.277	0.148	0.239	0.180	0.273	0.153	0.254	0.188	0.304	0.229	0.338
	192	**0.157**	**0.252**	0.162	0.259	0.206	0.282	0.194	0.280	0.167	0.258	0.187	0.280	0.175	0.274	0.196	0.311	0.221	0.330
	336	**0.176**	**0.272**	0.178	0.274	0.219	0.295	0.207	0.296	0.179	**0.272**	0.204	0.296	0.222	0.315	0.212	0.327	0.265	0.363
	720	**0.201**	**0.295**	0.204	0.299	0.260	0.328	0.243	0.328	0.208	0.297	0.246	0.328	0.243	0.329	0.244	0.352	0.321	0.395
Traffic	96	**0.375**	**0.259**	0.382	0.261	0.663	0.393	0.650	0.397	0.393	0.268	0.459	0.298	0.519	0.275	0.575	0.357	0.692	0.418
	192	**0.391**	**0.266**	0.396	0.267	0.611	0.367	0.600	0.372	0.413	0.277	0.469	0.301	0.550	0.288	0.613	0.381	0.627	0.398
	336	**0.416**	**0.277**	0.417	**0.277**	0.617	0.367	0.606	0.374	0.424	0.283	0.483	0.307	0.565	0.298	0.621	0.380	0.636	0.403
	720	0.463	**0.297**	0.461	0.298	0.655	0.389	0.646	0.396	**0.458**	0.300	0.517	0.326	0.597	0.334	0.630	0.383	0.647	0.399
PEMS03	12	**0.062**	**0.164**	0.065	0.168	0.103	0.211	0.076	0.183	0.064	0.169	0.064	0.173	0.071	0.172	0.192	0.315	0.527	0.562
	24	**0.078**	**0.184**	0.081	0.185	0.134	0.237	0.135	0.258	0.091	0.198	**0.078**	**0.184**	0.086	0.187	0.203	0.329	0.393	0.490
	48	**0.106**	**0.211**	0.114	0.213	0.175	0.264	0.166	0.272	0.114	0.217	**0.106**	0.214	0.110	0.217	0.234	0.359	0.643	0.633
	96	0.135	0.234	0.147	0.240	0.207	0.286	0.210	0.313	**0.132**	**0.233**	0.135	0.236	0.163	0.259	0.301	0.406	0.833	0.739
PEMS04	12	**0.072**	**0.173**	0.073	0.174	0.122	0.234	0.092	0.197	0.075	0.179	0.081	0.194	**0.072**	0.174	0.173	0.298	0.431	0.502
	24	0.082	**0.184**	0.083	0.186	0.151	0.258	0.126	0.234	0.088	0.194	0.101	0.224	**0.080**	**0.184**	0.182	0.309	0.325	0.440
	48	**0.095**	**0.197**	0.097	0.198	0.192	0.287	0.174	0.270	0.104	0.209	0.113	0.217	0.119	0.220	0.210	0.336	0.499	0.560
	96	**0.107**	**0.208**	0.111	0.215	0.226	0.306	0.211	0.304	0.123	0.229	0.133	0.238	0.148	0.245	0.251	0.370	0.829	0.740
PEMS07	12	**0.050**	**0.144**	0.075	0.171	0.098	0.213	0.073	0.180	0.064	0.160	0.055	0.156	0.057	0.146	0.154	0.280	0.306	0.306
	24	**0.057**	**0.151**	0.115	0.215	0.130	0.243	0.108	0.223	0.075	0.174	0.065	0.166	0.071	0.157	0.166	0.294	0.590	0.615
	48	**0.067**	**0.164**	0.135	0.230	0.171	0.277	0.161	0.269	1.031	0.868	0.081	0.184	0.100	0.182	0.182	0.308	0.776	0.735
	96	**0.076**	**0.173**	0.117	0.204	0.204	0.299	0.213	0.303	0.926	0.747	0.109	0.223	0.124	0.198	0.216	0.339	1.024	0.838
PEMS08	12	**0.075**	**0.172**	0.078	0.178	0.131	0.233	0.093	0.196	0.084	0.181	0.084	0.197	0.174	0.190	0.327	0.350	0.737	0.637
	24	**0.092**	**0.183**	0.104	0.196	0.178	0.262	0.142	0.236	0.114	0.205	0.101	0.207	0.192	0.200	0.340	0.365	0.661	0.595
	48	**0.142**	**0.206**	0.155	0.233	0.264	0.301	0.229	0.285	0.390	0.396	**0.142**	0.215	0.222	0.215	0.365	0.382	0.922	0.736
	96	**0.169**	**0.209**	0.183	0.225	0.334	0.322	0.313	0.319	0.336	0.351	0.196	0.240	0.242	0.237	0.448	0.448	1.214	0.903

PEMS04, PEMS07, PEMS08 [14]. To demonstrate the effectiveness of CRS-Mamba, we fairly compare it with eight representative SOTA forecasting models, including Transformer-based models: Autoformer [2], FEDformer [9], Crossformer [13], PatchTST [10], iTransformer [14], Linear-based models: DLinear [11], SparesTSF [8] and Mamba-based model: S-Mamba [6]. We evaluate the models using Mean Squared Error (MSE) and Mean Absolute Error (MAE).

Experimental Setting: We split datasets into training, validation and test set by the ratio of 6 : 2 : 2 for ETT and PEMS, and 7 : 1 : 2 for ECL and Traffic. To ensure that input length covers appropriate periods, all models employ a look-back window $L = 720$ for PEMS and $L = 96$ for the remaining datasets. Parameter θ is set to 1 and ϵ is set to 0. All experiments are conducted on a machine with NVIDIA V100 GPU and 32GB memory using the Adam optimizer.

5.1 Performance Analysis with Baselines

The quantitative results of MTS forecasting is shown in Table 1. As we can observe from the table, CRS-Mamba achieves outstanding performance on most datasets across various prediction length settings, obtaining 46 first-place and 10 second-place rankings in total 64 settings. Specially in complex scenarios

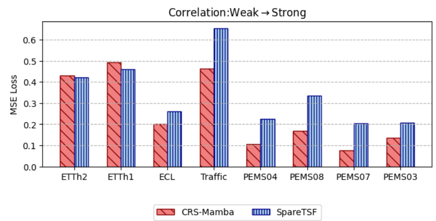


Fig. 3. Results comparison for varying correlation with datasets.

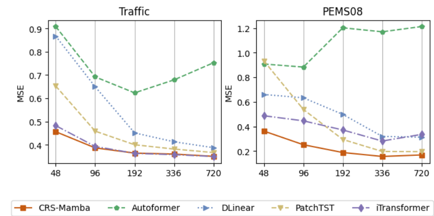


Fig. 4. Forecasting performance with the varing lookback window size.

Table 2. Ablations Studies on components. CPM: Cross period modeling. Mamba: Bidirectional Mamba layer. DRS: Dimensionality Reduction Scan Algorithm. Lookback window $L = 720$ for PEMS08, $L = 96$ for remaining datasets. Prediction length $T = 96$.

CPM	Mamba	DRS	ETTh1		ECL		Traffic		PEMS08	
			MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
●	○	○	0.392	0.407	0.197	0.274	0.645	0.383	0.308	0.309
○	●	●	0.386	**0.404**	0.142	0.237	0.388	0.262	0.177	0.210
●	●	○	0.386	0.405	**0.137**	**0.234**	0.382	0.261	0.198	0.249
●	●	●	**0.384**	**0.404**	0.138	**0.234**	**0.375**	**0.259**	**0.169**	**0.209**

with a large number of relevant variables, while other models struggle to model effectively, CRS-Mamba demonstrates significant forecasting performance.

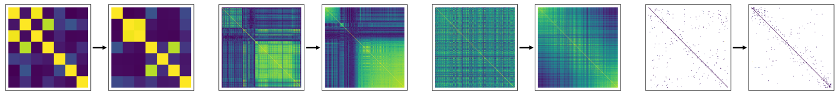
Channel Independent Strategy: CRS-Mamba does not achieve the best performance on ETT compared with SparseTSF, which assumes variable independence. However, its performance deteriorates on PEMS. It likely occurs as CRS-Mamba is beneficial for capturing complex variate correlations while inadvertently introduce noise into datasets with weak channel dependencies.

The results of datasets correlation are shown in Figuer 3. When the dataset correlation is not pronounced, CRS-Mamba that considers correlation modeling may not outperform models with channel independent strategy.

Lookback Window Size: Although a growing lookback window offers more information, previous studies have shown that Transformer based methods may not necessarily benefit from it [11]. We experiment CRS-Mamba and several baselines in this context. The results are shown in Figure 4. The prediction performance of CRS-Mamba indeed improves with increasing input length, which is attributed to the cross period modeling technique that extracts periodic features over longer horizon. DLinear, PatchTST and iTransformer also show this characteristic, but the overall curve of CRS-Mamba is the lowest.

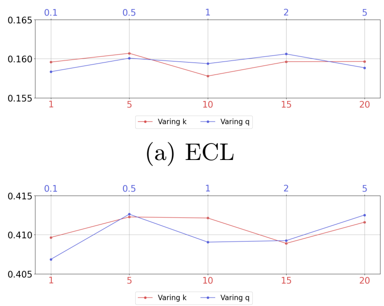
5.2 Ablation Studies and Analysis

Table 2 presents ablation study results. CRS-Mamba with complete structure achieves the best results, validating the effectiveness of proposed components. The unimproved performance of certain components may be related to the characteristics of datasets. CPM and DRS do not significantly improve the performance on ETTh1 possibly due to the unobvious correlation between variables



(a) ETTh1 (b) ECL (c) Traffic (d) PEMS08

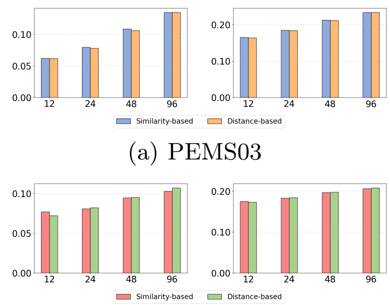
Fig. 5. Visualization results of DRS algorithm.



(a) ECL

(b) Traffic

Fig. 6. MSE results of CRS-Mamba on ECL and Traffic. The input length is set as 96 and output length is set as 192.



(a) PEMS03

(b) PEMS04

Fig. 7. MSE (left) and MAE (right) results of CRS-Mamba on PEMS03 and PEMS04 with different graph type.

and the weak periodic patterns. As the visualize of DRS algorithm in Figure 5, the variable order of ECL almost unchanged after performing DRS algorithm, inferring the original variable distribution aligns with the result of algorithm.

5.3 Dimensionality Reduction Scan Algorithm Analysis

Activate Function Parameters k, q: We use an improved activation function $\sigma(k, q)$ with hyperparameters k, q to control the attention span of algorithm. Here, we fix one of the parameters while varying the other and conducting experiments. Figure 6 illustrate the results. We observe that CRS-Mamba is not obviously sensitive to q and k, which suggests prioritizing the relative strength of correlations over their specific numerical differences.

Distance-based Graph and Similarity-based Graph: We experiment distance-based and similarity-based graph on PEMS03 and PEMS04. As shown in Figure 7, there are slight differences in performance between different graph. Considering the computation cost and competitive performance, distance-based graph can be the preferred method. However, if additional variable distribution is unavailable, similarity-based graph serves as the sole selection.

6 Conclusion and Future Work

In this work, we propose CRS-Mamba to address the challenges of modeling temporal dependencies and variable correlation. An Adaptive Period Identifier is leveraged to obtain main period. In temporal dependencies modeling, we utilize downsampling to aggregate contextual semantic and predict cross-period trend. Then, we design a bidirectional Mamba layer to capture variable correlation. To mitigate scan order sensitivity problem, we propose DRS algorithm to interact between correlated variables. Extensive experiments demonstrate the superior performance of CRS-Mamba over existing SOTA methods.

Acknowledgments. This work was partly supported by the National Key Research and Development Program of China under Grant 2022YFB4501704, National Natural Science Foundation of China under Grant 72342026, and Fundamental Research Funds for the Central Universities under Grant 2024-6-ZD-02.

References

1. Jin, G., Liang, Y., Fang, Y., Shao, Z., Huang, J., Zhang, J., Zheng, Y.: Spatio-temporal graph neural networks for predictive learning in urban computing: A survey. IEEE Transactions on Knowledge and Data Engineering (2023)
2. Wu, H., Xu, J., Wang, J., Long, M.: Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Advances in neural information processing systems **34**, 22419–22430 (2021)
3. Vaswani, A.: Attention is all you need. Advances in Neural Information Processing Systems (2017)
4. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
5. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
6. Wang, Z., Kong, F., Feng, S., Wang, M., Zhao, H., Wang, D., Zhang, Y.: Is mamba effective for time series forecasting? arXiv preprint arXiv:2403.11144 (2024)
7. Qu, et al.: A survey of mamba. arXiv preprint arXiv:2408.01129 (2024)
8. Lin, S., Lin, W., Wu, W., Chen, H., Yang, J.: Sparsetsf: Modeling long-term time series forecasting with 1k parameters. arXiv preprint arXiv:2405.00946 (2024)
9. Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., Jin, R.: Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In: International conference on machine learning. pp. 27268–27286. PMLR (2022)
10. Nie, et al.: A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730 (2022)
11. Zeng, A., Chen, M., Zhang, L., Xu, Q.: Are transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 11121–11128 (2023)
12. Huang, Q., Shen, L., Zhang, R., Ding, S., Wang, B., Zhou, Z., Wang, Y.: Crossgnn: Confronting noisy multivariate time series via cross interaction refinement. Advances in Neural Information Processing Systems **36**, 46885–46902 (2023)
13. Zhang, Y., Yan, J.: Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: The eleventh international conference on learning representations (2023)
14. Liu, et al.: itransformer: Inverted transformers are effective for time series forecasting. arXiv preprint arXiv:2310.06625 (2023)
15. Cai, X., Zhu, Y., Wang, X., Yao, Y.: Mambats: Improved selective state space models for long-term time series forecasting. arXiv preprint arXiv:2405.16440 (2024)