Modeling Point-to-Point Dependency for High-Dimensional Long-Term Series Forecasting

Xinyu Li
22110240025@m.fudan.edu.cn
Shanghai Key Laboratory of Intelligent Information Processing
Fudan University
Shanghai, China

Kexi Chen
24210240120@m.fudan.edu.cn
Fudan University
Shanghai, China

Ying Zheng
zhengy18@fudan.edu.cn
Fudan University
Shanghai, China

Zhiyi Yao
23110720054@m.fudan.edu.cn
Fudan University
Shanghai, China

Yi Xie
18110240043@fudan.edu.cn
Fudan University
Shanghai, China

Jihan Dai
21210240003@m.fudan.edu.cn
Fudan University
Shanghai, China

Lei Bai
baisanshi@gmail.com
Shanghai AI Laboratory
Shanghai, China

Jin Zhao*
jzhao@fudan.edu.cn
Fudan University
Shanghai, China

Jiajie Shen
jiajieshen@fudan.edu.cn
Fudan University
Shanghai, China

Yunqi Cai
22210240110@m.fudan.edu.cn
Fudan University
Shanghai, China

Hong Lu
honglu@fudan.edu.cn
Fudan University
Shanghai, China

Xin Wang*
xinw@fudan.edu.cn
Shanghai Key Laboratory of Intelligent Information Processing
Fudan University
Shanghai, China

Abstract

The strategic planning and reliability of modern web services, from cloud infrastructures to e-commerce platforms, increasingly hinge on accurate long-term forecasting of high-dimensional time series. A fundamental challenge within this task is modeling the intricate point-to-point dependencies that span across both time and variable dimensions. However, many existing methods face restricted direction modeling and computational inefficiency due to their reliance on localized paradigms and Transformer architectures. To address these, we propose replacing Self-Attention with autocorrelation, achieving two key innovations: 1) We propose calculating autocorrelation across both variable and time dimensions, which is a global paradigm, to model point-to-point dependencies. 2) Our proposed Spectral Product Mechanism (SPM) optimizes traditional autocorrelation into a data-driven form suitable for deep learning. Moreover, SPM reformulates autocorrelation as spectral product and reduces the complexity from $O(N^2)$ to $O(NlogN)$, while its Hadamard product-based correlation score matrix further reduces core computation to $O(N)$ compared to Self-Attention's $O(N^2)$ matrix multiplication. We further propose a Generalized Spectral Product Mechanism (GSPM), which extends traditional autocorrelation by mapping input into distinct feature representations, enabling modeling of complex dependencies through cross-feature correlations. SPM and GSPM surpass current state-of-the-art (SOTA) methods on 14 authoritative benchmarks, collectively securing the top rank on 22 out of 28 metrics, while ranking 1st in time, 2nd in memory, and 2nd in parameter overhead. Source code is available at: https://github.com/lxy-PhD2022/SPM.

CCS Concepts

• **Computing methodologies** → **Temporal reasoning**; *Spectral methods.*

Keywords

Multivariate long-term time series forecasting; Model point-to-point dependencies; Spectral product

ACM Reference Format:
Xinyu Li, Kexi Chen, Ying Zheng, Zhiyi Yao, Yi Xie, Jihan Dai, Lei Bai, Jin Zhao, Jiajie Shen, Yunqi Cai, Hong Lu, and Xin Wang. 2026. Modeling Point-to-Point Dependency for High-Dimensional Long-Term Series Forecasting. In *Proceedings of the ACM Web Conference 2026 (WWW '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3774904.3792503

*Corresponding authors.

This work is licensed under a Creative Commons Attribution 4.0 International License.

WWW '26, Dubai, United Arab Emirates

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2307-0/2026/04
https://doi.org/10.1145/3774904.3792503

Resource Availability: An archival snapshot is available at https://doi.org/10.5281/zenodo.18371278.

1 Introduction

Accurate time series forecasting is a cornerstone of the modern World Wide Web, enabling intelligent services from resource management in cloud infrastructures [19] and e-commerce inventory planning [1] to user engagement tracking on digital platforms [15]. The data from these critical web systems is often extreme in scale, characterized by both long temporal histories and high-dimensional variables. The task, formally known as multivariate long-term series forecasting (MLTSF), aims to predict extended future trends across all these variables based on their shared history. In MLTSF, the data is best conceptualized as a 2D plane of time and variables [2]. Within this 2D plane, dependencies may exist between any two points, regardless of direction or distance. For instance, a sharp spike in CPU load on a core authentication server at a specific timestamp (point 1: $time_t, variable_A$) can trigger cascading effects, leading to an abnormal query latency on a downstream recommendation service later (point 2: $time_{t+k}, variable_B$). Therefore, accurately capturing this point-to-point (P2P) dependency that crosses both time and variable dimensions is essential for building robust web services, yet it remains a largely open challenge.

The primary dilemma in P2P modeling is preserving temporal semantic continuity. To avoid disrupting this continuity, existing methods adopt various rule-based paradigms, as illustrated in Figure 1. These paradigms, which we critique in detail in the Related Work section, fall into two main categories: 1) Axis-aligned paradigms (Figure 1(a,b,c)) are structurally biased toward horizontal and vertical interactions. An attempt to overcome this by combining the two directions [28] still fails to achieve P2P modeling, because true P2P dependencies cannot be linearly decomposed; 2) Patch-based paradigm (Figure 1(d)) allows for some oblique but inaccurate dependencies due to the loss of boundary information [8]. Although these rule-based paradigms preserve continuity, they introduce a largely overlooked yet fatal flaw: unrestricted distances but restricted directions, which prevents true P2P dependency modeling. Building upon these paradigms, many SOTA methods extensively adopt the Transformer as the backbone, resulting in high computational complexity.

To address these, we turn to 2D autocorrelation. In theory, it provides the perfect solution: its mechanism of computing correlations across all spotiotemporal lags inherently preserves semantic continuity, thus achieving true P2P modeling. At the same time, this pair-wise similarity calculation fulfills the core function of Self-Attention. However, this classical tool is rendered unusable for modern deep learning by two fatal flaws: it is a fixed, non-data-driven measure, and its own prohibitive $O(N^2)$ complexity.

In this work, we tackle these two flaws by proposing Spectral Product Mechanism (SPM), as shown in Figure 1(e). Firstly, SPM makes the process data-driven by computing autocorrelation within a latent space created by a learnable embedding layer. Secondly, SPM solves the computational bottleneck by reformulating autocorrelation as an efficient spectral product. Crucially, SPM replaces the $O(N^2)$ matrix multiplication bottleneck in Self-Attention, computing correlation scores with a highly efficient $O(N)$ Hadamard product. While the overall $O(NlogN)$ complexity is governed by FFT and IFFT, their overhead contributes only 0.019% to time, 0.306% to memory, and 0% to parameters. Thus, SPM's complexity is nearly linear in practice. Although Autoformer [25] and FEDformer [32] have respectively reduced complexity to $O(NlogN)$ and $O(N)$, their actual overhead of time, memory, and parameter number are respectively {56.12%, 176.22%}, {62.14%, 12.74%}, and {785.89%, 2769.94%} higher than SPM (details in Table 3).

In order to further model complex dependencies, we improve the traditional autocorrelation to cross-feature correlation computation. Since different feature projections of the same input still represent the same underlying signal, this operation can be regarded as a generalized form of autocorrelation. It can separate different important features from the input signal and then capture more complex dependencies from the cross-feature correlations. We call this method GSPM. Our contributions can be summarized as follows:

- We propose SPM and GSPM that model point-to-point dependencies across both time and variable dimensions, overcoming the directional limitations of existing methods.
- SPM and GSPM reinvent classical autocorrelation into a data-driven form, enabling it to act as a functional replacement for Self-Attention with only $O(N)$ core complexity and extended cross-feature correlation modeling.
- SPM and GSPM collectively secure the top rank on 22 out of 28 metrics across 14 authoritative benchmarks, while ranking 1st in time, 2nd in memory, and 2nd in parameter overhead.

2 Related work

We review prior work through the P2P dependency challenge, which requires a global modeling scope and temporal continuity. We argue that local modeling paradigms restrict existing methods to specific directions, preventing true P2P modeling.

Figure 1(a) shows the channel-independent paradigm. Methods such as PatchTST [18], DLinear [29], TimesNet [24], TimeMixer [21], TimeKAN [10], TimeBase [9], Sequence Complementor (SeqCom) [4], and WPMixer [17] adopt a channel-independent strategy that runs the model in parallel across each variable sequence for independent prediction. This strategy determines the model coefficients by summing the ACFs of all channels, which can mitigate distribution drift and improve the model's prediction accuracy [7]. However, iTransformer [14] indicates that this approach might lead to suboptimal outcomes due to the neglect of inter-variable dependencies.

Figure 1(b) shows the paradigm that embeds all variable information at each time step into independent feature vectors. Methods like Informer [31], Autoformer [25], and Liu, et al. [13, 32] process each time step's variable information independently through local segmentation and models them accordingly. This approach often results in attention maps lacking in meaningful information due to the insufficient semantic integration of variables, consequently reducing the accuracy of modeling variable relationships.

Figure 1(d) shows the paradigm that segments each variable sequence into patches and models dependencies between patches across variables. Unlike the above methods, Crossformer

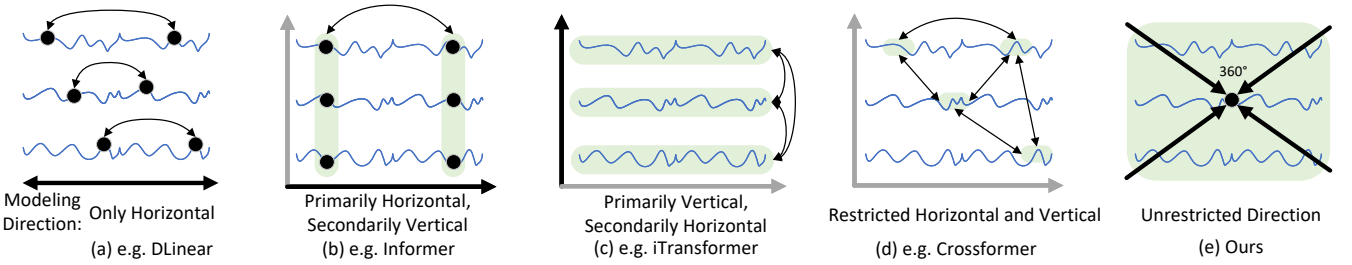

Figure 1: Existing modeling paradigms (a-d) are structurally limited to restricted directions, preventing effective point-to-point dependency modeling. To achieve it, we reformulate 2D autocorrelation (e) by overcoming its two key flaws: fixed, non-data-driven measure and $O(N^2)$ complexity.

Figure 2: The red words refer to the function that needs to maintain consistency before and after replacement. The spectral product can serve the same practical function as the Self-Attention mechanism, but our goal is not theoretical equivalence.

[30] segments each variable sequence into patches and models dependencies between patches across variables, allowing for more flexible global modeling. However, the fixed length of patch causes the loss of boundary information and semantic discontinuity of time series data [8]. Furthermore, patching increases the maximum path length at the point level [30]. Therefore, Crossformer [30] struggles to accurately model P2P dependencies. Similarly, MSGNet [3] only interacts with patches that have overlapping time ranges, learning positive and negative correlations.

Figure 1(c) shows the paradigm that independently embeds each variable series into feature vectors. To avoid the issues introduced by patching, iTransformer [14] and S-Mamba [23] embed entire variable sequences into feature vectors and model variable relationships directly. TimePro [16] further dynamically filters important time steps on this basis. This operation has improved prediction accuracy and gained widespread recognition.

The paradigm that mixes Figure 1(a) and Figure 1(c). Leddam [28] and Timexer [22] add a channel-independent learning branch on this basis to take into account the dependency modeling of inter- and intra- variables. However, the dependencies of different sampling points in any direction cannot be linearly decomposed into the superposition of horizontal and vertical directions as a whole. Therefore, this method does not accurately model P2P dependencies in an unrestricted direction. Moreover, the distinctions from Autoformer and FEDformer are in Appendix A.

3 Method

In this section, we first prove that the spectral product can achieve the function of Self-Attention (§ 3.2). Then, we present its specific design (§ 3.4) and how to extend it to the generalized form (§ 3.5).

3.1 Problem Definition

In multivariate time series forecasting, we represent the input multiple time series as $\boldsymbol{x} \in R^{M\times L}$, where M is the number of variate and L is the size of look-back window. For each single series of *i-th* variate $\boldsymbol{x}^{(i)} = (x_1^{(i)}, \ldots, x_L^{(i)}) \in R^{1\times L}$, where $i = 1, ..., M$, the goal is to forecast T future values $\boldsymbol{y}^{(i)} = (x_{L+1}^{(i)}, \ldots, x_{L+T}^{(i)}) \in R^{1\times T}$. We represent the multivariate prediction result as $\boldsymbol{y} \in R^{M\times T}$.

3.2 A Functional Replacement for Self-Attention

Essentially, the Self-Attention mechanism and autocorrelation achieve the same function because they both calculate the relevance among local parts through vector inner products. As shown in Figure 2, the first step is to prove that self-convolution in signal processing can achieve the calculation of autocorrelation. In signal processing, 2D discrete self-convolution's input are two same 2D discrete signals $f[x, y]$. It is defined as:

$$(f * f)[u, v] = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f[x, y] \cdot f[u - x, v - y], \tag{1}$$

where x and y are the index, u and v are the delay, m and n are the index boundary. Flip one of the input signals and obtain:

$$(f * \tilde{f})[u, v] = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f[x, y] \cdot f[u + x, v + y], \tag{2}$$

where $\tilde{f}[x, y] = f[-x, -y]$. It is equivalent to the definition of 2D discrete autocorrelation:

$$R_f[u, v] = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} f[x, y] \cdot f[x + u, y + v]. \tag{3}$$

Therefore, self-convolving a flipped input signal can achieve the calculation of autocorrelation. According to the convolution theorem [11], the Fourier transform of a convolution is given by:

$$\mathcal{F}\{f[x, y] * g[x, y]\} = \mathcal{F}\{f[x, y]\} \cdot \mathcal{F}\{g[x, y]\}, \tag{4}$$

it proves that the Fourier transform $\mathcal{F}$ of time convolution equals the spectral product (detailed proof is shown in Appendix B.1). In this way, calculating the spectral product (Hadamard product) of the input signal and flipped input signal, then performing Inverse Fast Fourier Transform (IFFT) represents calculating their time-domain convolution, which further represents calculating autocorrelation:

$$\mathcal{F}^{-1}\{\mathcal{F}\{f[x, y]\} \cdot \mathcal{F}\{\tilde{f}[x, y]\}\} = R_f[u, v]. \tag{5}$$

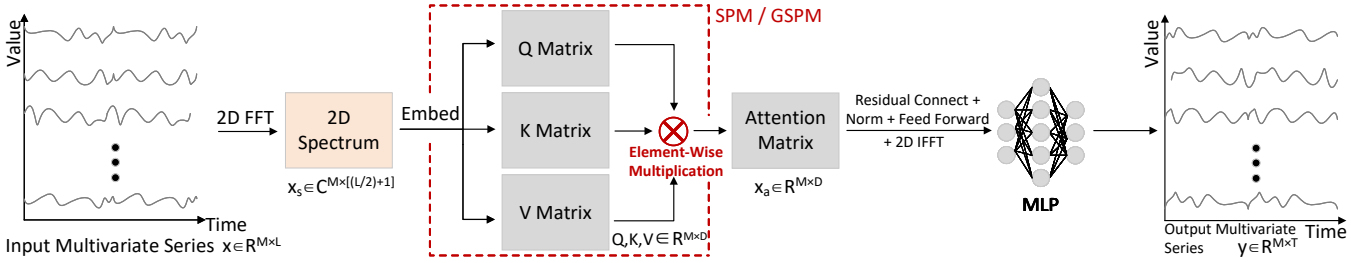

Figure 3: Framework of our method, where the SPM (GSPM) is highlighted in the dashed box. SPM employs spectral product equivalence to calculate autocorrelation across both variables and time dimensions, thereby modeling the true P2P dependencies. Meanwhile, SPM achieves the function of local relevance calculation similar to the Self-Attention mechanism. However, the overall computational complexity of SPM is lower, which is $O(NlogN)$, and through the Hadamard product, the calculation of the most core correlation score matrix is further reduced to $O(N)$.

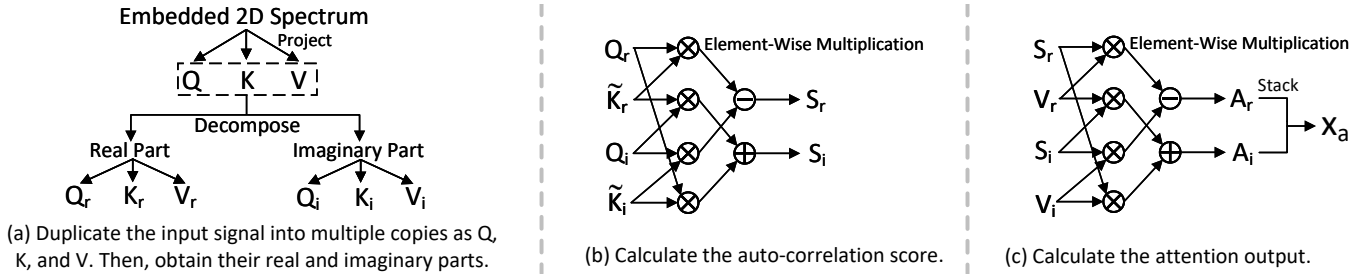

Figure 4: SPM's calculation step follows the rules of complex multiplication and Hadamard product.

Our spectral product formulation provides a dual advantage in both theoretical elegance and practical efficiency. The former, its theoretical elegance, is demonstrated in several key aspects. Firstly, by operating on the complex spectrum, it avoids the irreversible phase information loss inherent in power spectrum-based methods like Autoformer [25]. Secondly, its equivalence to global convolution [11, 27] provides a seamless mechanism to incorporate correlation scores back into the input sequence, making it a superior functional replacement for Self-Attention. Thirdly, this global convolutional nature inherently overcomes the local perception problem of CNNs [6]. The latter, its practical efficiency, is achieved by replacing the $O(N^2)$ matrix multiplication bottleneck at Self-Attention with a highly efficient $O(N)$ Hadamard product. While the overall complexity is governed by the $O(NlogN)$ FFT and IFFT, their practical overhead is negligible, contributing only 0.019% to time, 0.306% to memory, and 0% to parameters. Therefore, our complexity is nearly linear in practice.

3.3 Overview Framework

To effectively mitigate the impact of distribution drift between training and test data, we apply reversible instance normalization [5, 12] to the input, normalizing each sample point to have zero mean and unit standard deviation. This helps meet the stationarity assumption requirements of FFT. Finally, these statistics are added to the prediction results.

Figure 3 shows the framework of the proposed method. Initially, since we need to calculate the autocorrelation across variables and time dimensions and reformulate it as a spectral product, we should apply 2D instead of 1D FFT. We apply 2D FFT to transform the input into a 2D spectrum, which is then embedded into a latent space to perform data-driven autocorrelation. These are expressed as:

$$\boldsymbol{x}_s = 2DFFT(\boldsymbol{x}), \quad \boldsymbol{x}_e = Embed(\boldsymbol{x}_s), \tag{6}$$

where $\boldsymbol{x}$, $2DFFT(\cdot)$, $\boldsymbol{x}_s$, $Embed(\cdot)$, and $\boldsymbol{x}_e$ represents the input multiple time series, 2D FFT, 2D spectrum, linear mapping embedding, and embedded 2D spectrum, respectively. Furthermore, $\boldsymbol{x} \in R^{M\times L}$, $\boldsymbol{x}_s \in C^{M\times[(L/2)+1]}$, and $\boldsymbol{x}_e \in C^{M\times D}$. Here, C, M, L, and D are the complex number field, number of variables, look-back window size, and embedding dimension, respectively.

Next, we input the embedded 2D spectrum to SPM to calculate autocorrelation. The output of SPM is an attention matrix. Like in Transformer, this matrix is further processed by the feed-forward network and others. We then perform 2D IFFT to obtain the output of the encoding stage.

$$\boldsymbol{x}_a = SPM(\boldsymbol{x}_e), \tag{7}$$

$$\boldsymbol{x}_o = 2DIFFT(Compose(LN(FFN(LN(\boldsymbol{Res})) + LN(\boldsymbol{Res})))), \tag{8}$$

where $SPM(\cdot)$ and $\boldsymbol{x}_a$ represent the (Generalized) Spectral Product Mechanism and attention matrix (correlation score matrix), respectively. Let $\boldsymbol{Res} = \boldsymbol{x}_a + Decompose(\boldsymbol{x}_e)$ and it is a residue connection, while $Decompose(\cdot)$ represents the decomposition of a complex number into the real part and the imaginary part. Let $\boldsymbol{x}_o$ represents the output of the encoding stage. Furthermore, $\boldsymbol{x}_a \in R^{M\times D}$, $\boldsymbol{Res} \in R^{M\times D}$, and $\boldsymbol{x}_o \in R^{M\times[(D-1)*2]}$. Let $LN(\cdot)$, $FFN(\cdot)$, $Compose(\cdot)$, and $2DIFFT(\cdot)$ represent the layer normalization, feed-forward network, composing the real and imaginary parts stacked

in the batch size dimension into a complex number, and 2D IFFT, respectively.

Finally, we apply a multilayer perceptron (MLP) as a decoder. It maps the output of the encoding stage to the prediction result. Formally, this process is expressed as $\boldsymbol{y} = MLP(\boldsymbol{x}_o)$, where $MLP(\cdot)$ and $\boldsymbol{y}$ represent the multilayer perceptron and predicted multivariate series, respectively. Additionally, $\boldsymbol{y} \in R^{M \times T}$ and T represents the prediction length.

3.4 Spectral Product Mechanism

Figure 4 shows the details of SPM. The first step (Figure 4(a)) is to duplicate the input signal (embedded 2D spectrum) into multiple copies (Q, K, V) for calculating autocorrelation. Then we decompose to obtain the real and imaginary parts of $\boldsymbol{Q}, \boldsymbol{K}$, and $\boldsymbol{V}$. These are expressed as follows.

$$\{\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}\} = Project(\boldsymbol{x}_e), \tag{9}$$

$$\{\boldsymbol{Q}_r, \boldsymbol{Q}_i, \boldsymbol{K}_r, \boldsymbol{K}_i, \boldsymbol{V}_r, \boldsymbol{V}_i\} = Decompose(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}), \tag{10}$$

where $Project(\cdot)$, $\boldsymbol{X}_r$, and $\boldsymbol{X}_i$ represent the identity mapping (duplication), real part, and imaginary part of a complex number $\boldsymbol{X}$, respectively. Additionally, $\boldsymbol{Q}_r, \boldsymbol{K}_r, \boldsymbol{V}_r, \boldsymbol{Q}_i, \boldsymbol{K}_i, \boldsymbol{V}_i \in R^{M \times D}$.

The second step (Figure 4(b)) involves calculating autocorrelation through multiplying the two copies of the input signal $\boldsymbol{Q} = \boldsymbol{Q_r} + \boldsymbol{Q_i} j$ and $\tilde{\boldsymbol{K}} = \tilde{\boldsymbol{K}}_{\boldsymbol{r}} + \tilde{\boldsymbol{K}}_{\boldsymbol{i}} j$, where $\tilde{\boldsymbol{K}}$ represents the flipped $\boldsymbol{K}$. This is the spectral product and it requires the employment of Hadamard product instead of matrix multiplication. The calculation process can be expressed as follows.

$$\boldsymbol{Q}\tilde{\boldsymbol{K}} = (\boldsymbol{Q_r} + \boldsymbol{Q_i} j)(\tilde{\boldsymbol{K}}_{\boldsymbol{r}} + \tilde{\boldsymbol{K}}_{\boldsymbol{i}} j) = (\boldsymbol{Q_r}\tilde{\boldsymbol{K}}_{\boldsymbol{r}} - \boldsymbol{Q_i}\tilde{\boldsymbol{K}}_{\boldsymbol{i}}) + (\boldsymbol{Q_r}\tilde{\boldsymbol{K}}_{\boldsymbol{i}} + \boldsymbol{Q_i}\tilde{\boldsymbol{K}}_{\boldsymbol{r}}) j \tag{11}$$

where $(\boldsymbol{Q_r}\tilde{\boldsymbol{K}}_{\boldsymbol{r}} - \boldsymbol{Q_i}\tilde{\boldsymbol{K}}_{\boldsymbol{i}})$ and $(\boldsymbol{Q_r}\tilde{\boldsymbol{K}}_{\boldsymbol{i}} + \boldsymbol{Q_i}\tilde{\boldsymbol{K}}_{\boldsymbol{r}})$ respectively represent the real part $\boldsymbol{S_r}$ and the imaginary part $\boldsymbol{S_i}$ of the new complex number $\boldsymbol{S}$, which represents the correlation score matrix in time domain. It functions similarly to the Self-Attention score matrix. However, its element-wise multiplication (Hadamard product), compared with the matrix multiplication of Self-Attention, reduces the computational complexity from $O(N^2)$ to $O(N)$.

The third step (Figure 4(c)) is to apply the autocorrelation score to the input signal to adjust the connection between local signals according to the correlation. The process is achieved by multiplying the complex numbers $\boldsymbol{S} = \boldsymbol{S_r} + \boldsymbol{S_i} j$ and $\boldsymbol{V} = \boldsymbol{V_r} + \boldsymbol{V_i} j$. This is equivalent to performing global convolution on the input signal in the time domain using the correlation score matrix, so as to apply the score to the input. The calculation process can be expressed as follows.

$$\boldsymbol{SV} = (\boldsymbol{S_r} + \boldsymbol{S_i} j)(\boldsymbol{V_r} + \boldsymbol{V_i} j) = (\boldsymbol{S_r}\boldsymbol{V_r} - \boldsymbol{S_i}\boldsymbol{V_i}) + (\boldsymbol{S_r}\boldsymbol{V_i} + \boldsymbol{S_i}\boldsymbol{V_r}) j, \tag{12}$$

where $(\boldsymbol{S_r}\boldsymbol{V_r} - \boldsymbol{S_i}\boldsymbol{V_i})$ and $(\boldsymbol{S_r}\boldsymbol{V_i} + \boldsymbol{S_i}\boldsymbol{V_r})$ respectively represent the real part $\boldsymbol{A_r}$ and the imaginary part $\boldsymbol{A_i}$ of the new complex number $\boldsymbol{A}$, which represents the attention output matrix. Then, we stack $\boldsymbol{A_r}$ and $\boldsymbol{A_i}$ in batch size dimension, obtaining the attention output $\boldsymbol{x}_a$. This is to facilitate subsequent feed-forward network and others to directly process the real part and imaginary part respectively.

3.5 Generalized Spectral Product Mechanism

To capture more abundant and discriminative dependencies, we extend the SPM to a generalized form (GSPM). Specifically, we first set the projection in Figure 4(a) to 3 different linear mappings (SPM adopts the identity mapping) to obtain different feature representations $\boldsymbol{Q}, \boldsymbol{K}$, and $\boldsymbol{V}$ of the input signal. This process is expressed as follows.

$$\boldsymbol{Q} = Linear_1(\boldsymbol{x}_e), \quad \boldsymbol{K} = Linear_2(\boldsymbol{x}_e), \quad \boldsymbol{V} = Linear_3(\boldsymbol{x}_e), \tag{13}$$

where $Linear_1(\cdot)$, $Linear_2(\cdot)$, and $Linear_3(\cdot)$ represent 3 different linear mappings, respectively. Additionally, $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in R^{M \times D}$. Then, we remove the flipping operation of K in Figure 4(b) because the learnable linear mapping can already automatically achieve flipping. Other calculation processes of the spectral product are consistent with SPM. Their concise formal expression is $\boldsymbol{x}_a = Stack(\boldsymbol{QKV})$. The details of complex multiplication among $\boldsymbol{Q}, \boldsymbol{K}$, and $\boldsymbol{V}$, stack operations of $Stack(\cdot)$, and the attention output $\boldsymbol{x}_a$ have been introduced in the Section 3.4 above.

In this way, $\boldsymbol{Q}, \boldsymbol{K}$, and $\boldsymbol{V}$ can be driven by data to represent different important features. GSPM extends the calculation object of autocorrelation in SPM from a single fixed feature to discriminative important features, then models more complex dependencies across feature representations.

4 EXPERIMENTS

4.1 Experimental Setup

Due to space limitations, the detailed experimental setup is provided in Appendix C.1.

4.2 Main Results

Compare with the latest SOTA methods. In this section, we focus on discussing the differences from other methods. Since GSPM and SPM essentially represent the same concept, that is, spectral product, they are collectively referred to as our method and discussed together. Their differences are discussed separately in the next section. As shown in Table 1, considering the overall MSE and MAE, our method achieves 1st place in 14 and 8 metrics, respectively, significantly surpassing the latest models published in 2025, including recent Mamba- and KAN-based architectures [10, 16, 23]. Specifically, on the most authoritative Traffic dataset, our method achieves the lowest average MSE and MAE, surpassing not only the latest SOTA methods but also outperforming well-known methods like iTransformer [14] and PatchTST [18], with MSE 3.51% and 22.01% lower, respectively. On the highly authoritative Electricity dataset, our method also achieves the lowest average MSE and the second lowest average MAE. Specifically, our MSE is 2.96% to 16.75% lower than other methods, while MAE is 0.763% to 12.16% lower. On the other 12 datasets with 24 metrics, our method ranks 1st in 19 metrics and ranks 2nd in 16 metrics.

Furthermore, it can be observed that on the two datasets with the highest dimensional variables (Traffic, Electricity), the methods of adopting the paradigm of modeling variable relationships (as shown in Figure 1(c), such as S-Mamba [23]) are more effective than the other methods without adoption (as shown in Figure 1(a), such as TimeKAN [10]). However, on the datasets with lower-dimensional variables (ETT), the situation is reversed. This indicates that modeling dependencies in only one direction (either the variable or the time dimension) cannot adapt to a broader range of scenarios. To address this, our global modeling paradigm (shown in Figure 1(e))

Table 1: Quantitative results averaged from all prediction length $T \in \{96, 192, 336, 720\}$ of ours and latest SOTA methods published in 2025. The full results and results of earlier methods published before 2025 can be found in the Appendix Table 6 and Table 7, respectively. Red and blue: 1st and 2nd. "-": unreported in its paper, unmeasurable as its code not open-sourced.

Models	GSPM		SPM		S-Mamba [23]		TimeKAN [10]		TimePro [16]		FreDF [20]		TimeBase [9]		WPMixer [17]		Seq-Com [4]	
Venue	-		-		NC 25		ICLR 25		ICML 25		ICLR 25		ICML 25		AAAI 25		AAAI 25	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	**0.413**	**0.269**	0.416	0.270	0.414	0.276	0.572	0.370	0.439	0.286	0.421	0.279	0.682	0.375	0.489	0.297	0.472	0.301
Electricity	0.166	0.261	**0.164**	0.260	0.170	0.265	0.197	0.286	0.169	0.262	0.170	**0.259**	0.227	0.296	0.177	0.267	0.192	0.282
Solar	**0.229**	0.263	0.233	0.265	0.240	0.273	0.276	0.310	0.233	0.267	0.240	**0.255**	0.621	0.488	0.237	0.260	-	-
Weather	**0.239**	0.271	**0.239**	0.271	0.251	0.276	0.243	0.272	0.252	0.276	0.254	0.274	0.252	0.279	0.243	**0.269**	0.243	0.273
Exchange	**0.324**	**0.377**	0.372	0.400	0.367	0.408	0.385	0.415	0.352	0.399	0.399	0.432	0.476	0.483	0.391	0.418	0.356	0.400
ILI	**1.040**	**0.652**	1.397	0.754	2.817	1.126	2.174	0.959	2.966	1.172	4.095	1.475	5.786	1.731	2.093	0.900	-	-
ETTh1	**0.408**	0.424	0.432	0.438	0.455	0.450	0.418	0.427	0.438	0.438	0.437	0.435	0.463	0.429	0.422	**0.423**	0.429	0.435
ETTh2	**0.311**	**0.365**	0.327	0.376	0.381	0.405	0.383	0.404	0.377	0.404	0.372	0.396	0.410	0.426	0.355	0.387	0.368	0.398
ETTm1	**0.376**	0.394	0.380	0.397	0.398	0.405	0.377	0.395	0.391	0.401	0.392	0.399	0.431	0.420	**0.376**	**0.388**	0.381	0.398
ETTm2	**0.269**	**0.317**	0.277	0.323	0.288	0.332	0.277	0.323	0.281	0.326	0.278	0.319	0.290	0.332	0.271	**0.317**	0.275	0.323
PEMS03	0.239	0.325	**0.222**	**0.314**	0.264	0.346	0.527	0.509	0.376	0.422	0.309	0.378	1.290	0.875	0.432	0.452	-	-
PEMS04	0.196	0.300	0.164	0.273	**0.162**	**0.266**	0.516	0.514	0.297	0.379	0.340	0.400	1.379	0.917	0.449	0.474	-	-
PEMS07	0.159	0.257	**0.144**	**0.244**	0.161	0.256	0.546	0.522	0.378	0.414	0.262	0.339	1.322	0.894	0.427	0.443	-	-
PEMS08	0.351	0.322	**0.308**	**0.303**	0.374	0.352	0.731	0.544	0.616	0.494	0.482	0.412	1.502	0.926	0.641	0.504	-	-
1st Count:	**9**	**5**	5	3	1	1	0	0	0	0	0	2	0	0	1	4	0	0

Table 2: Component ablation study results. GSPM is the optimal one overall. The full results are in the Appendix Table 5. We set the training epochs for Traffic to 300 to speed up the experiment.

Model	w/o Vari FFT		w/o Time FFT		w/o FFT		SPM		SPM w/o Emb		w/ Nonl-QKV		w/ Sigm-Scor		w/o Nonl-FFN		w/o Nonl-MLP		w/ Mult		w/ Add		w/o FFT; w/ Add		GSPM	
Dataset	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	0.456	0.286	0.418	0.272	0.460	0.288	0.416	**0.270**	0.437	0.282	**0.415**	0.271	0.425	0.273	0.424	0.275	0.513	0.337	0.422	0.273	0.425	0.273	0.463	0.291	0.419	0.272
Electricity	0.177	0.269	0.166	0.263	0.177	0.269	**0.164**	**0.260**	0.171	0.267	0.166	0.263	0.168	0.264	0.168	0.264	0.199	0.287	0.177	0.272	0.166	0.261	0.178	0.270	0.166	0.261
Solar	0.239	0.269	0.235	0.269	0.242	0.273	0.233	0.265	0.235	0.268	**0.229**	**0.263**	0.234	0.265	0.234	0.268	0.275	0.291	0.242	0.278	0.232	0.264	0.247	0.278	**0.229**	**0.263**
Weather	**0.239**	0.272	0.241	0.273	0.240	0.272	**0.239**	**0.271**	0.240	0.272	0.240	0.272	0.240	0.272	**0.239**	**0.271**	0.248	0.278	0.242	0.273	0.242	0.274	0.240	0.272	**0.239**	**0.271**
Exchange	0.317	0.375	0.368	0.392	0.314	0.373	0.372	0.400	0.344	0.385	0.373	0.394	0.327	0.380	0.358	0.387	0.339	0.382	0.334	0.380	0.347	0.388	**0.312**	**0.372**	0.324	0.377
ILI	1.483	0.767	1.298	0.719	1.409	0.761	1.397	0.754	1.475	0.775	1.306	0.727	1.560	0.794	1.375	0.754	1.431	0.774	1.423	0.762	1.325	0.727	1.336	0.731	**1.040**	**0.652**
ETTh1	0.435	0.436	0.435	0.441	0.422	0.433	0.432	0.438	0.431	0.437	0.430	0.439	0.431	0.436	0.429	0.437	0.430	0.437	0.433	0.437	0.430	0.436	0.420	0.430	**0.408**	**0.424**
ETTh2	0.327	0.374	0.329	0.376	0.334	0.377	0.327	0.376	0.331	0.377	0.327	0.375	0.329	0.377	0.328	0.376	0.329	0.375	0.324	0.373	0.330	0.377	0.333	0.376	**0.311**	**0.365**
ETTm1	0.382	0.397	0.386	0.399	0.382	0.397	0.380	0.397	0.380	0.398	0.381	0.397	0.381	0.397	0.382	0.398	0.395	0.402	0.382	0.398	0.383	0.399	0.383	0.398	**0.376**	**0.394**
ETTm2	0.274	0.322	0.276	0.323	0.273	0.321	0.277	0.323	0.280	0.325	0.275	0.322	0.275	0.322	0.274	0.322	0.280	0.325	0.278	0.324	0.278	0.325	0.273	0.321	**0.269**	**0.317**
AVG	0.433	0.377	0.415	0.373	0.425	0.377	0.424	0.375	0.432	0.379	0.414	0.372	0.437	0.378	0.421	0.375	0.444	0.389	0.426	0.377	0.416	0.372	0.419	0.374	**0.378**	**0.360**
1st Count:	1	0	0	0	0	0	2	3	0	0	2	1	0	0	1	1	0	0	0	0	0	0	1	1	**8**	**8**

models the dependencies across variables and the time dimension. It treats the time and variable dimensions equally to model the dependencies with unrestricted direction. In this way, our method can overcome this limitation and achieve the lowest average MSE and MAE in the greatest number of cases, significantly surpassing the latest SOTA methods.

Comparison between GSPM and SPM. Considering the overall MSE and MAE, GSPM achieves 175% more 1st-place rankings than SPM, with its MSE and MAE decreasing by up to 25.55% and 13.53%, respectively. This verifies the effectiveness of extending the traditional autocorrelation by mapping the input into distinct feature representations, which enables the modeling of complex dependencies through cross-feature correlations. This approach can improve the adaptability to generalized scenarios, capturing complex patterns.

4.3 Ablation Study

Component ablation. We explore this from four aspects and report the results in Table 2. The first part is to investigate the role of global modeling across both variable and time dimensions. We remove the FFT of variables, time, and all dimensions respectively, as shown in the left 1st to 3rd columns. Their overall average MSEs are 14.6%, 9.8%, and 12.4% higher than that of the 2D FFT of GSPM, verifying the effectiveness of global modeling.

The second part investigates whether optimizing the 2D autocorrelation into a data-driven form is effective. We remove this optimization from the pure SPM, and the results are shown in the left 4th to 5th columns. The data-driven SPM ranks first in five cases, whereas the original 2D autocorrelation not only drops to zero but also increases the MSE by up to 5.58%. This demonstrates that the data-driven optimization in SPM is highly effective.

The third part is to explore the source of GSPM's non-linear modeling ability. We conduct the following studies respectively: adding non-linear mapping to $\boldsymbol{Q}$, $\boldsymbol{K}$, and $\boldsymbol{V}$ (left 6th) increases the average MSE by 9.52%; adding a sigmoid non-linear activation function (because GSPM does not employ matrix multiplication, Softmax is not applicable) to the correlation score matrix (left 7th) increases the average MSE by 15.61%. These two studies indicate that non-linear mapping is not needed when capturing the discriminative features of the input and is not needed for the score matrix either. Then, we remove the non-linear activation function from the feed-forward network (left 8th), only two indicators are on par with GSPM, but the average MSE increase by 11.38%; removing the non-linear activation function from the MLP (left 9th), the average MSE increase

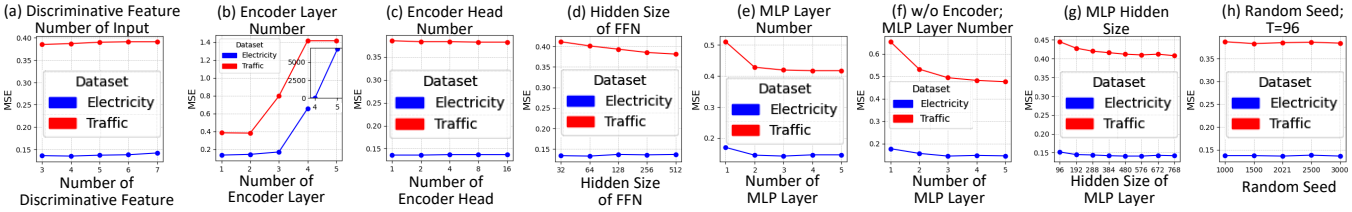

Figure 5: Hyperparameter sensitivity analysis on largest-scaled datasets Traffic and Electricity. For the encoder and decoder, smaller and larger depths are respectively more optimal. The random seed has no effect.

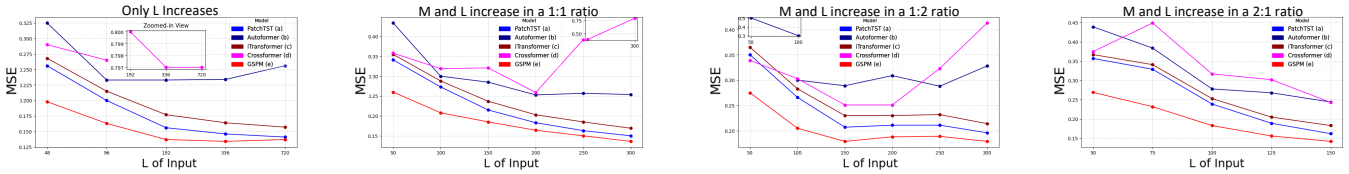

Figure 6: Influence of look-back window size on Electricity dataset with T=96 and unified hyper-parameters. GSPM achieves consistently decreased and lower MSE than baselines under all settings.

by 17.46%. The last two studies indicate that the non-linear modeling ability of GSPM mainly comes from the MLP, followed by the FFN.

The fourth part is to study the effectiveness of the Hadamard product. We replace it with matrix multiplication (left 10th) and addition (left 11th), respectively. Moreover, to exclude the influence of FFT, we construct a time-domain addition (left 12th). Their average MSE increases by 12.70%, 10.05%, and 10.85% , respectively, indicating that the Hadamard product is more suitable for the spectral product mechanism both in theory and in practice.

Hyperparameter sensitivity analysis. We conduct this study in four parts. Firstly, we study the impact of the number of discriminative features (like $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$), as shown in Figure 5(a). Taking three as the basic version, the MSE is already the lowest. This indicates that three feature matrices can already accurately calculate the autocorrelation and assign values. The second part is to study the impact of the number of parameters in the encoder. Figures 5(b) and 5(c) respectively show that increasing the number of layers and heads of the encoder has no benefits and too many layers significantly increases MSE. These results are because the autocorrelation scores will be distorted due to repeated calculations. In fact, we only use one layer and one head, confirming that we rely on an efficient theory, not parameter piling. Moreover, Figure 5(d) shows that expand feed-forward network decreases MSE in the partial dataset.

The third part is to study the impact of the number of parameters in the decoder. Figure 5(e) shows that increasing the number of layers of the MLP can gradually reduce the MSE. To verify whether the effectiveness comes from our method (located in the encoder), we further conduct experiments by removing the encoder, as shown in Figure 5(f). At this time, although the trend also remains downward, the lowest MSE of traffic and electricity increases by 14.149% and 2.11%, respectively (note that the scales of the vertical axes are different). Moreover, it requires two more layers to converge, leading to a larger number of parameters and higher costs but also a larger error. These verify that the effectiveness comes from our method. In addition, we also studied the impact of the hidden size of the MLP, as shown in Figure 5(g). A larger model generally yields higher accuracy, as the encoder captures large and complex P2P dependencies requiring stronger nonlinear decoding. Increasing the decoder's layers and hidden size enhances this ability, but once sufficient (3 layers, 576 size), further growth brings no gain. The fourth part focuses on robustness, as shown in Figure 5(h). Changing the random seed over a large range almost does not change the MSE, which verifies our robustness.

Influence of look-back window size. We specifically study the impact of the common input time length (L) and the input size at unrestricted direction. The latter refers to the situation where the variable (M) and time (L) dimensions change simultaneously. We increase M and L at multiple ratios and compare with four most representative local paradigm methods shown in Figure 1(a), (b), (c), and (d). The first sub-figure of Figure 6 shows that GSPM is optimal in terms of the conventional utilization of time length input. Its MSE not only gradually decreases but is also the lowest throughout the process. The other three sub-figures illustrate that GSPM performs the same when the input size at unrestricted direction with multiple ratios increases. These jointly verify that GSPM fully leverages more input by modeling true P2P dependencies, resulting in more accurate high-dimension long-term forecasting.

More results. Figure 8 and Figure 9 show that GSPM achieves lower MSE and MAE than the baselines globally, verifying its strong zero-shot forecasting ability and anti-disturbance robustness.

4.4 Resource Overhead Analysis

Since one of our goals is to reduce the overhead of Transformer, our baselines are the Transformer-based methods and the latest Mamba-based method, as shown in Table 3. GSPM has the second-lowest computational complexity, outperforming most square-level methods. It trails only FEDformer and S-Mamba's linear-level complexity. In fact, the autocorrelation score calculation in GSPM and

Table 3: Comparison of training resource overhead to Transformer- and Mamba-based models on the largest-scale dataset Traffic. We adopt a unified parameter setting (e.g., batch size) to ensure a fair comparison of computational overhead, and adopt the prediction errors reported in publicly published papers to enable an optimal comparison of predictive accuracy. Red and blue: 1st and 2nd. We use "N" to denote the number of varying time steps or variables, and simplify the complexity expression by keeping only the highest-order term. GSPM is optimal overall.

Model	Complexity	Time (s/epoch)					Memory (GB)					Parameter (M)					AVG
		96	192	336	720	AVG	96	192	336	720	AVG	96	192	336	720	AVG	MSE/MAE
Autoformer [25]	$O(NlogN)$	17.564	25.063	29.063	42.284	28.494	1.406	2.282	3.134	5.440	3.066	1.369	1.468	1.468	1.468	1.444	0.628/0.379
FEDformer [32]	$O(N)$	41.203	45.659	49.687	65.099	50.412	1.286	1.572	2.160	3.510	2.132	4.515	4.515	4.515	4.515	4.515	0.610/0.376
Crossformer [30]	$O(N^2)$	63.143	61.121	99.538	109.653	83.364	15.546	26.348	42.114	85.598	42.402	2.332	2.999	3.999	6.665	3.999	0.550/0.304
PatchTST [18]	$O(N^2)$	24.632	26.813	31.090	39.568	30.526	6.522	6.254	6.962	7.542	6.820	0.249	0.397	0.618	1.209	0.618	0.529/0.341
iTransformer [14]	$O(N^2)$	11.455	14.808	19.731	31.556	19.388	1.776	2.088	2.342	2.880	2.272	**0.125**	**0.137**	**0.156**	0.205	**0.156**	0.428/0.282
Leddam [28]	$O(N^2)$	118.715	166.338	170.349	179.943	158.836	2.130	1.686	1.776	2.528	2.030	0.305	3.293	3.367	3.565	2.633	0.467/0.294
S-Mamba [23]	$O(N)$	11.661	15.988	21.254	32.552	20.364	**1.278**	**1.432**	**1.706**	2.806	**1.806**	0.189	0.202	0.220	0.270	0.220	0.414/0.276
GSPM	$O(NlogN)$	**10.154**	**14.016**	**18.639**	**30.195**	**18.251**	1.582	1.594	1.998	**2.390**	1.891	0.140	0.149	0.163	**0.200**	0.163	**0.413/0.269**

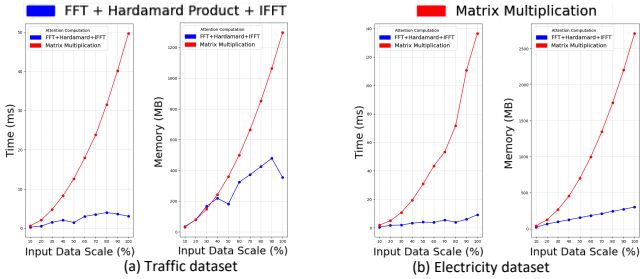

Figure 7: For the Self-Attention, a comparison of the spatiotemporal overhead between our "(I)FFT + Hadamard product" and matrix multiplication on the large-scale datasets. Ours grows linearly, while matrix multiplication grows quadratically. This verifies the validity of our SPM theory.

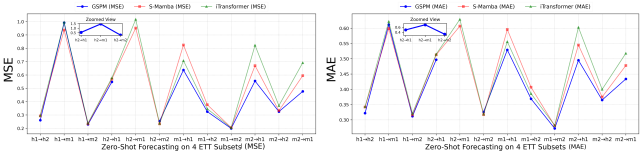

Figure 8: Zero-Shot forecasting results on 4 ETT subsets. From a global perspective, the two blue lines of GSPM are at the bottom of their group, achieving the global optimum.

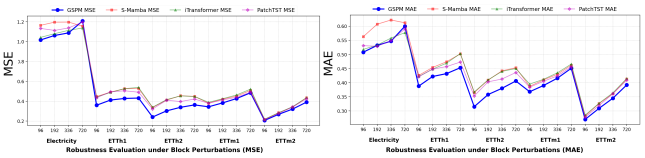

Figure 9: Results of the robustness verification based on 5% block perturbation under multiple electricity scenarios. From a global perspective, the two blue lines of GSPM are at the bottom of their group, verified its excellent robustness.

SPM has $O(N)$ complexity. Although FFT adds $O(LlogL)$, its impact on time, memory, and parameters is minimal (4.38%, 1.61%, 0% respectively), making GSPM and SPM almost $O(N)$ overall.

To verify this, we will independently compare the above computational form with the matrix multiplication of Self-Attention. On the two largest datasets as shown in Figure 7, our spatiotemporal overhead (blue line) increases linearly with the size of the input data, while that of matrix multiplication (red line) increases quadratically, which strongly verifies that GSPM and SPM are almost $O(N)$ complexity in general. Crucially, under respective hyperparameters, GSPM's overall average MSE/MAE is 44.01%/38.37% and 38.82%/16.80% lower than FEDformer and S-Mamba, respectively, giving higher practical value.

In time overhead, GSPM ranks 1st across all prediction lengths. Notably, FEDformer, despite its lowest complexity, has an average time overhead 1.77 times and 1.65 times larger than Autoformer and PatchTST respectively, proving lower complexity does not always translate to better practicality. For memory and parameter overhead, GSPM ranks 2nd in the average prediction length and 1st in the longest prediction. This shows that the advantage of GSPM lies in ultra-long-term prediction, which is exactly the problem scenario we are targeting. By comprehensively comparing the four indicators, it can be verified that GSPM has achieved our goal. While achieving the lowest MSE by modeling true P2P dependencies, GSPM effectively reduces the overhead of Transformer.

5 Conclusion

To achieve more accurate high-dimensional long-term time series forecasting, we propose a new global modeling paradigm. It models true P2P dependencies across variable and time dimensions. To reduce Transformer's cost, we replace the Self-Attention with autocorrelation and reformulate it into Spectral Product Mechanism (SPM), reducing complexity from $O(N^2)$ to $O(NlogN)$. Moreover, its Hadamard product-based correlation score matrix further reduces core computation to $O(N)$ compared to Self-Attention's $O(N^2)$ matrix multiplication. For more complex patterns, we introduce GSPM, which maps and models cross-feature correlations. Our method surpasses the latest SOTA methods on 14 authoritative benchmarks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62472101) and the Songshan Laboratory Research Fund (Grant No. ZZK202402010).

References

[1] Aakanksha Ankit Gandhi, Sivaramakrishnan Kaveri, and Vineet Chaoji. [n. d.]. Spatio-temporal Multi-graph Networks for Demand Forecasting in Online Marketplaces. ([n. d.]).

[2] Ali Behrouz, Michele Santacatterina, and Ramin Zabih. 2024. Chimera: Effectively modeling multivariate time series with 2-dimensional state space models. *Advances in Neural Information Processing Systems* 37 (2024), 119886–119918.

[3] Wanlin Cai, Yuxuan Liang, Xianggen Liu, Jianshuai Feng, and Yuankai Wu. 2024. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 11141–11149.

[4] Xiwen Chen, Peijie Qiu, Wenhui Zhu, Huayu Li, Hao Wang, Aristeidis Sotiras, Yalin Wang, and Abolfazl Razi. 2025. Sequence Complementor: Complementing Transformers For Time Series Forecasting with Learnable Sequences. 39 (2025).

[5] Wei Fan, Pengyang Wang, Dongkun Wang, Dongjie Wang, Yuanchun Zhou, and Yanjie Fu. 2023. Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37. 7522–7529.

[6] Shanghua Gao, Zhong-Yu Li, Qi Han, Ming-Ming Cheng, and Liang Wang. 2022. Rf-next: Efficient receptive field search for convolutional neural networks. *IEEE transactions on pattern analysis and machine intelligence* 45, 3 (2022), 2984–3002.

[7] Lu Han, Han-Jia Ye, and De-Chuan Zhan. 2024. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering* (2024).

[8] Qihe Huang, Lei Shen, Ruixin Zhang, Jiahuan Cheng, Shouhong Ding, Zhengyang Zhou, and Yang Wang. 2024. Hdmixer: Hierarchical dependency with extendable patch for multivariate time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 12608–12616.

[9] Qihe Huang, Zhengyang Zhou, Kuo Yang, Zhongchao Yi, Xu Wang, and Yang Wang. [n. d.]. TimeBase: The Power of Minimalism in Efficient Long-term Time Series Forecasting. In *Forty-second International Conference on Machine Learning*.

[10] Songtao Huang, Zhen Zhao, Can Li, and Lei Bai. 2025. Timekan: Kan-based frequency decomposition learning architecture for long-term time series forecasting. *arXiv preprint arXiv:2502.06910* (2025).

[11] B Hunt. 1971. A matrix theory proof of the discrete convolution theorem. *IEEE Transactions on Audio and Electroacoustics* 19, 4 (1971), 285–288.

[12] Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. 2021. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International Conference on Learning Representations*.

[13] Shizhan Liu, Hang Yu, Cong Liao, Jianguo Li, Weiyao Lin, Alex X Liu, and Schahram Dustdar. 2021. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. In *International conference on learning representations*.

[14] Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. 2023. itransformer: Inverted transformers are effective for time series forecasting. *arXiv preprint arXiv:2310.06625* (2023).

[15] Yozen Liu, Xiaolin Shi, Lucas Pierce, and Xiang Ren. 2019. Characterizing and forecasting user engagement with in-app action graph: A case study of snapchat. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2023–2031.

[16] Xiaowen Ma, Zhenliang Ni, Shuai Xiao, and Xinghao Chen. 2025. TimePro: Efficient Multivariate Long-term Time Series Forecasting with Variable-and Time-Aware Hyper-state. *arXiv preprint arXiv:2505.20774* (2025).

[17] Md Mahmuddun Nabi Murad, Mehmet Aktukmak, and Yasin Yilmaz. 2025. WPMixer: Efficient multi-resolution mixing for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 19581–19588.

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

[19] Arnak V. Poghosyan, Ashot N. Harutyunyan, Naira Grigoryan, Clement Pang, George Oganesyan, Sirak Ghazaryan, and Narek A. Hovhannisyan. 2021. An Enterprise Time Series Forecasting System for Cloud Applications Using Transfer Learning †. *Sensors (Basel, Switzerland)* 21 (2021). https://api.semanticscholar.org/CorpusID:232128975

[20] Hao Wang, Licheng Pan, Zhichao Chen, Degui Yang, Sen Zhang, Yifei Yang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. 2024. Fredf: Learning to forecast in the frequency domain. *arXiv preprint arXiv:2402.02399* (2024).

[21] Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and Jun Zhou. 2024. Timemixer: Decomposable multiscale mixing for time series forecasting. *arXiv preprint arXiv:2405.14616* (2024).

[22] Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Guo Qin, Haoran Zhang, Yong Liu, Yunzhong Qiu, Jianmin Wang, and Mingsheng Long. 2024. Timexer: Empowering transformers for time series forecasting with exogenous variables. *Advances in Neural Information Processing Systems* 37 (2024), 469–498.

[23] Zihan Wang, Fanheng Kong, Shi Feng, Ming Wang, Xiaocui Yang, Han Zhao, Daling Wang, and Yifei Zhang. 2024. Is mamba effective for time series forecasting? *Neurocomputing* (2024), 129178.

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

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

[26] Weiwei Ye, Songgaojun Deng, Qiaosha Zou, and Ning Gui. 2024. Frequency Adaptive Normalization For Non-stationary Time Series Forecasting. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

[27] Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. 2024. Frequency-domain MLPs are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems* 36 (2024).

[28] Guoqi Yu, Jing Zou, Xiaowei Hu, Angelica I Aviles-Rivero, Jing Qin, and Shujun Wang. 2024. Revitalizing Multivariate Time Series Forecasting: Learnable Decomposition with Inter-Series Dependencies and Intra-Series Variations Modeling. In *Forty-first International Conference on Machine Learning*. https://openreview.net/forum?id=87CYNyCGOo

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

[30] Yunhao Zhang and Junchi Yan. 2023. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In *The eleventh international conference on learning representations*.

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

[32] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*. PMLR, 27268–27286.

A Related work

A.1 Existing MLTSF Modeling Paradigms

The paradigm that mixes Figure 1(a) and Figure 1(c). Leddam [28] and Timexer [22] add a channel-independent learning branch on this basis to take into account the dependency modeling of inter- and intra- variables. However, the dependencies of different sampling points in any direction cannot be linearly decomposed into the superposition of horizontal and vertical directions as a whole. Therefore, this method does not accurately model P2P dependencies in an unrestricted direction.

A.2 Distinctions from Autoformer and FEDformer

It is crucial to distinguish our method from prior works that also leverage similar signal processing concepts, in order to precisely delineate our core innovations. The common and most important distinction is that they follow a local paradigm (Figure 1(b)), while ours follows a global modeling paradigm (Figure 1(e)). Specific distinctions are as follows.

Autoformer [25]: The primary distinction is not a simple extension from 1D to 2D, but a fundamental difference in goal and mechanism. Autoformer [25] employs classical, fixed autocorrelation as a pre-processing step to identify and aggregate the most relevant periodic sub-series. In sharp contrast, our work's central thesis is to transform autocorrelation from this fixed measure into a fully learnable, data-driven operator (SPM/GSPM) that operates in a latent space to model the entire P2P dependency structure, not just periodicity.

FEDformer [32]: This method does not use autocorrelation. While it operates in the frequency domain, it relies on the standard,

computationally expensive Self-Attention mechanism to model interactions between different frequency components. Our approach is fundamentally different: we use the frequency domain as a vehicle to efficiently implement autocorrelation, leveraging a highly efficient Hadamard product to compute correlations within the same frequency components.

B Method

B.1 A Functional Replacement for Self-Attention

Given 1D continuous signals $f(x)$ and $g(x)$, their convolution is equivalent to frequency-domain multiplication. This conclusion also holds in the 2D discrete case. Formally,

$$f * g = \mathcal{F}^{-1}\{\mathcal{F}\{f\} \cdot \mathcal{F}\{g\}\},$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and the inverse Fourier transform, respectively.

Proof. The definition of the convolution is defined as:

$$(f * g)(\tau) = \int_{-\infty}^{\infty} f(x)g(\tau - x)\, dx,$$

where τ is the time delay.

The definition of the Fourier transform of a signal is defined as:

$$\mathcal{F}\{f(\tau)\}(\omega) = \int_{-\infty}^{\infty} f(\tau)e^{-i\omega\tau}\, d\tau,$$

where ω is angular frequency and $\mathcal{F}$ is the Fourier transform.

Substitute the definition of convolution into the formula of Fourier transform. The formula is defined as:

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \mathcal{F}\left\{\int_{-\infty}^{\infty} f(x)g(\tau - x)\, dx\right\}(\omega)$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x)g(\tau - x)\, dx\right)e^{-i\omega\tau}\, d\tau,$$

according to the linear property of Fourier transform, the order of integration can be exchanged.

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} g(\tau - x)e^{-i\omega\tau}\, d\tau\right) dx,$$

let $u = \tau - x$, then $\tau = u + x$ and $d\tau = du$.

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} g(u)e^{-i\omega(u+x)}\, du\right) dx,$$

decompose the exponential term.

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\left(\int_{-\infty}^{\infty} g(u)e^{-i\omega u}\, du\right) dx,$$

identify the internal integral as the Fourier transform of g.

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x}\, dx \cdot \mathcal{F}\{g\}(\omega),$$

identify the external integral as the Fourier transform of f.

$$\mathcal{F}\{(f * g)(\tau)\}(\omega) = \mathcal{F}\{f\}(\omega) \cdot \mathcal{F}\{g\}(\omega),$$

obtain the product form of the spectra of two signals.

Proved. □

C Experiments

C.1 Experimental Setup

Datasets. We employ 14 real-world benchmark datasets to evaluate models. All these datasets can be obtained from [14, 23, 25]. The statistic information of these benchmarks are shown in Table 4. According to [26], these datasets are non-stationary with evolving trend and seasonal patterns, and can be used to evaluate the model's ability to handle non-stationary time series. According to FreTS [27], the details of these datasets are as follows:

Traffic[1]: This dataset records hourly traffic flow data for 963 freeway lanes in San Francisco. It supports long-term forecasting using 862 lanes. The data collection began on January 1, 2015, with an hourly sampling interval.

Electricity[2]: This dataset captures electricity consumption patterns of 370 clients for short-term forecasting and 321 clients for long-term predictions. The data spans from January 1, 2011, with a 15-minute sampling interval.

Solar[3]: This dataset consists of solar power measurements collected by the National Renewable Energy Laboratory. It includes data points from various power plants in Florida, comprising 593 records, spanning from January 1, 2006, to December 31, 2016, with an hourly sampling interval.

PEMS[4]: This dataset contains the public traffic network data in California collected by 5-minute windows. It includes four subsets (PEMS03, PEMS04, PEMS07, PEMS08).

Weather[5]: This dataset collects meteorological indicators, such as humidity and air temperature, from the Max Planck Biogeochemistry Institute's Weather Station in Germany. It includes data from 2020, with a 10-minute sampling frequency.

Exchange[6]: This dataset collects features of daily exchange rates from eight foreign currencies (including those from Australia, Canada, the UK, Switzerland, China, Japan, New Zealand, and Singapore). This dataset spans from 1990 to 2016, with a 1-day sampling interval.

ILI[7]: This dataset collects the number of patients and influenza-like illness ratio in a weekly frequency.

ETT[8]: This dataset includes four subsets (ETTh1, ETTh2, ETTm1, and ETTm2). It measurements from two distinct electric transformers, labeled ETTh1 and ETTm1, representing different temporal resolutions (hourly and 15-minute intervals, respectively). These datasets serve as benchmarks for long-term forecasting.

Baselines. Latest SOTA methods published in 2025 are reported as follows. The corresponding relationships between their modeling paradigms and Figure 1 are as follows: Figure 1(a): TimeKAN [10], TimeBase [9], WPMixer [17], Seq-Com [4]; Figure 1(c): S-Mamba [23], TimePro [16], FreDF [20] based on iTransformer [14].

Earlier published methods are reported as follows. (1) Transformer-based methods: iTransformer [14], Leddam [28], PatchTST [18], Crossformer [30], FEDformer [32], Autoformer [25], Informer [31];

[1] https://pems.dot.ca.gov/
[2] https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014
[3] https://www.nrel.gov/grid/solar-power-data.html
[4] https://pems.dot.ca.gov/
[5] https://www.bgc-jena.mpg.de/wetter/
[6] https://github.com/laiguokun/multivariate-time-series-data
[7] https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html
[8] https://github.com/zhouhaoyi/ETDataset

Table 4: The statistic information of benchmarks.

Datasets	Traffic	Electricity	Solar	PEMS03	PEMS04	PEMS07	PEMS08	Weather	Exchange	ILI	ETTh1	ETTh2	ETTm1	ETTm2
Variates	862	321	137	358	307	883	170	21	8	7	7	7	7	7
Timesteps	17544	26304	52560	26209	16992	28224	17856	52696	7588	966	17420	17420	69680	69680
Granularity	1hour	1hour	10min	5min	5min	5min	5min	10min	1day	1week	1hour	1hour	5min	5min

Table 5: The full component ablation study results. GSPM is the optimal one overall.

Dataset		w/o Vari FFT		w/o Time FFT		w/o FFT		SPM		SPM w/o Emb		w/ Nonl-QKV		w/ Sigm-Scor		w/o Nonl-FFN		w/o Nonl-MLP		w/ Mult		w/ Add		w/o FFT; w/ Add		GSPM	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	0.432	0.275	0.385	0.255	0.437	0.277	0.384	**0.253**	0.410	0.270	**0.380**	**0.253**	0.393	0.257	0.389	0.257	0.502	0.339	0.391	0.258	0.393	0.257	0.440	0.280	0.385	0.254
	192	0.445	0.279	0.406	0.265	0.446	0.280	**0.403**	**0.263**	0.426	0.277	0.404	0.264	0.414	0.266	0.413	0.268	0.498	0.327	0.411	0.266	0.414	0.266	0.449	0.282	0.408	0.266
	336	0.456	0.284	0.422	0.274	0.461	0.289	**0.420**	**0.271**	0.439	0.282	**0.420**	0.272	0.429	0.274	0.428	0.276	0.508	0.330	0.426	0.274	0.429	0.276	0.464	0.290	0.424	0.274
	720	0.490	0.304	0.458	0.295	0.495	0.309	**0.457**	**0.294**	0.472	0.301	0.458	0.295	0.463	**0.294**	0.467	0.299	0.545	0.351	0.461	**0.294**	0.462	**0.294**	0.497	0.310	0.459	**0.294**
	AVG	0.456	0.286	0.418	0.272	0.460	0.288	0.416	**0.270**	0.437	0.282	**0.415**	0.271	0.425	0.273	0.424	0.275	0.513	0.337	0.422	0.273	0.425	0.273	0.463	0.291	0.419	0.272
Electricity	96	0.145	0.244	0.135	0.238	0.145	0.244	**0.132**	0.236	0.135	0.238	0.135	0.238	0.135	0.237	0.139	0.242	0.167	0.265	0.136	0.241	0.133	**0.235**	0.148	0.247	0.136	0.238
	192	0.164	0.255	**0.155**	**0.249**	0.164	0.254	0.156	**0.249**	0.160	0.254	0.157	0.252	0.156	**0.249**	0.158	0.251	0.183	0.271	0.170	0.263	0.156	**0.249**	0.164	0.255	0.157	**0.249**
	336	0.181	0.273	**0.171**	0.269	0.180	0.271	**0.171**	**0.266**	0.173	0.269	0.173	0.270	**0.171**	**0.266**	0.174	0.268	0.198	0.288	0.189	0.280	**0.171**	**0.266**	0.179	0.271	**0.171**	0.267
	720	0.218	0.306	0.201	0.296	0.221	0.306	**0.196**	**0.291**	0.214	0.307	0.198	0.292	0.208	0.302	0.201	0.294	0.247	0.323	0.215	0.306	0.203	0.295	0.220	0.305	0.198	**0.291**
	AVG	0.177	0.269	0.166	0.263	0.177	0.269	**0.164**	**0.260**	0.171	0.267	0.166	0.263	0.168	0.264	0.168	0.264	0.199	0.287	0.177	0.272	0.166	0.261	0.178	0.270	0.166	0.261
Solar	96	0.205	0.244	0.203	0.245	0.211	0.251	0.194	0.235	0.193	0.239	0.193	0.233	0.198	0.236	0.194	0.238	0.227	0.261	0.217	0.264	0.195	0.233	0.227	0.266	**0.190**	**0.231**
	192	0.244	0.275	0.241	0.273	0.234	0.268	0.232	0.266	0.232	0.265	0.229	**0.263**	0.232	0.265	0.230	0.266	0.260	0.283	0.245	0.280	0.232	0.266	0.235	0.268	**0.227**	**0.263**
	336	0.256	0.282	0.247	**0.276**	0.267	0.288	0.254	0.282	0.253	0.280	0.246	**0.276**	0.255	0.283	0.251	0.281	0.291	0.303	0.251	0.281	0.252	0.281	0.268	0.292	**0.244**	0.279
	720	**0.249**	**0.276**	0.251	0.282	0.256	0.285	0.250	**0.276**	0.261	0.288	0.250	0.281	**0.249**	**0.276**	0.262	0.289	0.321	0.315	0.256	0.286	0.250	**0.276**	0.258	0.286	0.253	0.280
	AVG	0.239	0.269	0.235	0.269	0.242	0.273	0.233	0.265	0.235	0.268	**0.229**	**0.263**	0.234	0.265	0.234	0.268	0.275	0.291	0.242	0.278	0.232	0.264	0.247	0.278	**0.229**	**0.263**
Weather	96	0.164	0.209	0.165	0.209	0.165	0.210	0.162	0.207	0.164	0.209	0.162	0.207	0.162	0.207	0.163	0.208	0.172	0.218	0.162	**0.206**	0.164	0.209	0.165	0.210	**0.161**	**0.206**
	192	0.211	0.251	0.213	0.252	0.211	0.251	0.209	0.250	0.210	0.251	0.210	0.251	0.209	0.250	0.211	0.251	0.223	0.261	0.210	0.251	0.212	0.252	0.211	0.251	**0.208**	**0.249**
	336	**0.254**	**0.288**	0.258	0.292	0.255	0.290	0.256	0.290	0.258	0.291	0.261	0.293	0.259	0.292	0.255	0.289	0.262	0.295	0.265	0.297	0.261	0.294	0.255	0.289	0.258	0.291
	720	**0.327**	0.338	0.331	0.340	0.329	0.338	0.328	0.338	**0.327**	**0.336**	0.329	0.339	0.328	0.338	0.328	0.338	0.333	0.340	0.330	0.340	0.330	0.340	0.328	0.337	0.329	0.339
	AVG	**0.239**	0.272	0.241	0.273	0.240	0.272	**0.239**	**0.271**	0.240	0.272	0.240	0.272	0.240	0.272	**0.239**	**0.271**	0.248	0.278	0.242	0.273	0.242	0.274	0.240	0.272	**0.239**	**0.271**
Exchange	96	0.082	0.198	0.081	0.199	0.087	0.203	0.103	0.224	0.086	0.204	0.083	0.199	0.085	0.203	0.087	0.205	0.084	0.201	0.088	0.205	0.091	0.208	0.082	0.197	**0.079**	**0.196**
	192	0.179	0.300	0.172	0.295	0.169	0.293	0.184	0.305	0.181	0.305	0.184	0.304	0.199	0.315	0.193	0.311	0.179	0.298	0.192	0.309	0.198	0.316	0.183	0.304	**0.164**	**0.289**
	336	0.330	0.415	0.333	0.418	0.325	0.412	0.399	0.456	0.341	0.422	0.337	0.420	**0.323**	**0.410**	0.333	0.417	0.336	0.417	0.324	**0.410**	0.327	0.413	**0.323**	**0.410**	0.378	0.445
	720	0.678	0.585	0.884	0.655	0.676	0.585	0.803	0.617	0.766	0.609	0.889	0.655	0.702	0.590	0.821	0.615	0.756	0.612	0.731	0.597	0.773	0.617	**0.661**	**0.575**	0.676	0.578
	AVG	0.317	0.375	0.368	0.392	0.314	0.373	0.372	0.400	0.344	0.385	0.373	0.394	0.327	0.380	0.358	0.387	0.339	0.382	0.334	0.380	0.347	0.388	**0.312**	**0.372**	0.324	0.377
ILI	24	1.557	0.766	1.302	0.734	1.340	0.721	1.348	0.712	1.610	0.762	1.112	0.645	1.322	0.722	1.302	0.693	1.370	0.749	1.106	0.671	1.126	0.651	1.447	0.740	**0.928**	**0.609**
	36	1.579	0.776	1.211	0.674	1.247	0.714	1.509	0.790	1.345	0.747	1.462	0.776	1.338	0.708	1.516	0.802	1.629	0.837	1.552	0.792	1.435	0.749	1.188	0.693	**1.110**	**0.673**
	48	1.286	0.727	1.299	0.728	1.651	0.837	1.375	0.748	1.429	0.774	1.288	0.730	1.743	0.849	1.296	0.735	1.288	0.738	1.466	0.776	1.439	0.759	1.358	0.737	**1.007**	**0.645**
	60	1.509	0.799	1.379	0.741	1.397	0.773	1.355	0.766	1.514	0.817	1.364	0.756	1.837	0.897	1.385	0.788	1.437	0.773	1.570	0.811	1.298	0.747	1.353	0.754	**1.114**	**0.683**
	AVG	1.483	0.767	1.298	0.719	1.409	0.761	1.397	0.754	1.475	0.775	1.306	0.727	1.560	0.794	1.375	0.754	1.431	0.774	1.423	0.762	1.325	0.727	1.336	0.731	**1.040**	**0.652**
ETTh1	96	0.360	**0.387**	0.362	0.390	**0.359**	**0.387**	0.362	0.389	0.365	0.392	0.361	0.389	0.362	0.389	0.361	0.388	0.374	0.391	0.361	0.389	0.362	0.390	**0.359**	**0.387**	0.360	0.388
	192	0.440	0.436	0.443	0.446	0.400	0.420	0.432	0.430	0.420	0.429	0.427	0.435	0.424	0.427	0.424	0.428	0.428	0.432	0.433	0.433	0.425	0.430	**0.399**	**0.417**	0.412	0.422
	336	0.453	0.450	0.456	0.449	0.446	0.445	0.454	0.450	0.464	0.453	0.454	0.450	0.457	0.448	0.453	0.450	0.452	0.450	0.454	0.449	0.456	0.449	0.434	0.437	**0.427**	**0.431**
	720	0.489	0.471	0.477	0.480	0.482	0.479	0.479	0.482	0.475	0.475	0.479	0.482	0.482	0.478	0.478	0.482	0.465	0.476	0.484	0.479	0.475	0.476	0.490	0.478	**0.432**	**0.453**
	AVG	0.435	0.436	0.435	0.441	0.422	0.433	0.432	0.438	0.431	0.437	0.430	0.439	0.431	0.436	0.429	0.437	0.430	0.437	0.433	0.437	0.430	0.436	0.420	0.430	**0.408**	**0.424**
ETTh2	96	0.257	0.329	0.254	0.325	0.255	0.327	0.256	0.329	0.250	0.322	0.256	0.328	0.261	0.331	0.260	0.331	0.259	0.326	0.245	0.317	0.258	0.329	0.258	0.327	**0.241**	**0.315**
	192	0.313	0.363	0.320	0.364	0.324	0.367	0.315	0.364	0.314	0.364	0.314	0.363	0.315	0.364	0.314	0.363	0.323	0.369	0.319	0.367	0.315	0.364	0.323	0.366	**0.302**	**0.357**
	336	0.357	0.388	0.355	0.394	0.369	0.395	0.354	0.392	0.355	0.392	0.354	0.392	0.354	0.391	0.355	0.392	0.344	0.384	0.355	0.393	0.357	0.393	0.366	0.391	**0.339**	**0.380**
	720	0.379	0.416	0.388	0.421	0.387	0.421	0.384	0.419	0.404	0.432	0.384	0.419	0.388	0.421	0.383	0.419	0.392	0.423	0.379	0.416	0.389	0.421	0.386	0.421	**0.363**	**0.406**
	AVG	0.327	0.374	0.329	0.376	0.334	0.377	0.327	0.376	0.331	0.377	0.327	0.375	0.329	0.377	0.328	0.376	0.329	0.375	0.324	0.373	0.330	0.377	0.333	0.376	**0.311**	**0.365**
ETTm1	96	0.316	0.362	0.323	0.365	0.313	0.359	0.315	0.361	0.317	0.363	0.315	0.361	0.314	0.361	0.317	0.362	0.334	0.372	0.320	0.364	0.319	0.365	0.317	0.361	**0.308**	**0.355**
	192	0.352	0.380	0.354	0.382	0.355	0.382	0.351	0.380	0.352	0.382	0.351	0.379	0.351	0.380	0.352	0.380	0.367	0.386	0.351	0.381	0.352	0.380	0.356	0.382	**0.349**	**0.378**
	336	0.397	0.406	0.398	0.406	0.398	0.407	0.397	0.406	0.396	0.405	0.396	0.406	0.398	0.407	0.397	0.406	0.411	0.410	0.396	0.406	0.398	0.407	0.399	0.406	**0.391**	**0.403**
	720	0.462	0.442	0.468	0.445	0.460	0.441	0.460	0.442	0.454	0.441	0.462	0.442	0.459	0.441	0.462	0.442	0.469	**0.440**	0.459	0.443	0.462	0.443	0.461	0.441	**0.455**	0.441
	AVG	0.382	0.397	0.386	0.399	0.382	0.397	0.380	0.397	0.380	0.398	0.381	0.397	0.381	0.397	0.382	0.398	0.395	0.402	0.382	0.398	0.383	0.399	0.383	0.398	**0.376**	**0.394**
ETTm2	96	0.177	0.259	0.178	0.261	0.179	0.261	0.178	0.261	0.179	0.260	0.178	0.261	0.178	0.259	0.178	0.261	0.181	0.263	0.175	0.258	0.178	0.259	0.177	0.259	**0.173**	**0.254**
	192	0.240	0.300	0.239	0.299	0.240	0.299	0.241	0.301	0.242	0.301	0.240	0.300	0.242	0.302	0.241	0.301	0.245	0.304	0.243	0.302	0.243	0.303	0.239	0.299	**0.236**	**0.296**
	336	0.296	0.336	0.300	0.339	0.296	0.336	0.295	0.335	0.296	0.336	0.295	0.336	0.298	0.338	0.295	0.335	0.304	0.341	0.306	0.343	0.300	0.339	0.297	0.336	**0.292**	**0.333**
	720	0.382	0.391	0.386	0.394	0.380	0.389	0.393	0.395	0.403	0.402	0.385	0.392	0.383	0.391	0.384	0.391	0.390	0.394	0.387	0.392	0.391	0.398	0.381	0.390	**0.375**	**0.386**
	AVG	0.274	0.322	0.276	0.323	0.273	0.321	0.277	0.323	0.280	0.325	0.275	0.322	0.275	0.322	0.274	0.322	0.280	0.325	0.278	0.324	0.278	0.325	0.273	0.321	**0.269**	**0.317**

Table 6: Full quantitative results of ours and latest SOTA methods published in 2025. The results of earlier methods published before 2025 can be found in the Table 7. Red and blue: 1st and 2nd. "-": unreported in its paper, unmeasurable as its code not open-sourced.

Models		GSPM		SPM		S-Mamba		TimeKAN		TimePro		FreDF		TimeBase		WPMixer		Sequence-Com	
Venue		-		-		NC 25		ICLR 25		ICML 25		ICLR 25		ICML 25		AAAI 25		AAAI 25	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	**0.378**	**0.249**	0.384	0.253	0.382	0.261	0.570	0.376	0.406	0.269	0.391	0.265	0.713	0.385	0.465	0.286	0.455	0.291
	192	0.400	**0.262**	0.403	0.263	**0.396**	0.267	0.553	0.362	0.424	0.276	0.410	0.273	0.652	0.362	0.475	0.290	0.460	0.291
	336	**0.417**	0.272	0.420	**0.271**	**0.417**	0.276	0.557	0.363	0.444	0.287	0.424	0.280	0.660	0.365	0.489	0.296	0.473	0.302
	720	**0.457**	0.295	**0.457**	**0.294**	0.460	0.300	0.606	0.379	0.483	0.312	0.460	0.298	0.702	0.386	0.527	0.318	0.500	0.321
	AVG	**0.413**	**0.269**	0.416	0.270	0.414	0.276	0.572	0.370	0.439	0.286	0.421	0.279	0.682	0.375	0.489	0.297	0.472	0.301
Electricity	96	0.136	0.238	**0.132**	0.236	0.139	0.235	0.174	0.266	0.139	**0.234**	0.144	0.233	0.212	0.279	0.150	0.241	0.156	0.252
	192	0.157	0.249	**0.156**	0.249	0.159	0.255	0.182	0.273	**0.156**	0.249	0.159	**0.247**	0.209	0.281	0.162	0.252	0.175	0.269
	336	**0.171**	0.267	**0.171**	0.266	0.176	0.272	0.197	0.286	0.172	0.267	0.172	**0.263**	0.223	0.295	0.179	0.270	0.190	0.284
	720	0.198	**0.291**	**0.196**	**0.291**	0.204	0.298	0.236	0.320	0.209	0.299	0.204	0.294	0.264	0.327	0.217	0.304	0.246	0.324
	AVG	0.166	0.261	**0.164**	0.260	0.170	0.265	0.197	0.286	0.169	0.262	0.170	**0.259**	0.227	0.296	0.177	0.267	0.192	0.282
Solar	96	**0.190**	0.231	0.194	0.235	0.205	0.244	0.232	0.288	0.196	0.237	0.207	**0.221**	0.619	0.492	0.205	0.239	-	-
	192	**0.227**	0.263	0.232	0.266	0.237	0.270	0.279	0.311	0.231	0.263	0.244	**0.254**	0.642	0.505	0.234	0.257	-	-
	336	**0.244**	0.279	0.254	0.282	0.258	0.288	0.297	0.324	0.250	0.281	0.253	**0.269**	0.638	0.496	0.254	0.270	-	-
	720	0.253	0.280	**0.250**	0.276	0.260	0.288	0.294	0.318	0.253	0.285	0.256	0.274	0.584	0.457	0.253	**0.272**	-	-
	AVG	**0.229**	0.263	0.233	0.265	0.240	0.273	0.276	0.310	0.233	0.267	0.240	**0.255**	0.621	0.488	0.237	0.260	-	-
Weather	96	0.161	0.206	0.162	0.207	0.165	0.210	0.162	0.208	0.166	0.207	0.164	**0.202**	0.170	0.215	0.162	0.204	**0.159**	0.206
	192	0.208	0.249	0.209	0.250	0.214	0.252	0.207	0.249	0.216	0.254	0.220	0.253	0.216	0.256	0.209	**0.246**	**0.205**	0.249
	336	0.258	0.291	**0.256**	0.290	0.274	0.297	0.263	0.290	0.273	0.296	0.275	0.294	0.271	0.297	0.263	**0.287**	0.263	0.291
	720	0.329	0.339	**0.328**	**0.338**	0.350	0.345	0.338	0.340	0.351	0.346	0.356	0.347	0.351	0.348	0.339	**0.338**	0.344	0.345
	AVG	**0.239**	0.271	**0.239**	0.271	0.251	0.276	0.243	0.272	0.252	0.276	0.254	0.274	0.252	0.279	0.243	**0.269**	0.243	0.273
Exchange	96	**0.079**	**0.196**	0.103	0.224	0.086	0.207	0.088	0.207	0.085	0.204	0.102	0.228	0.111	0.238	0.094	0.211	0.082	0.200
	192	**0.164**	**0.289**	0.184	0.305	0.182	0.304	0.192	0.310	0.178	0.299	0.204	0.326	0.280	0.399	0.187	0.306	0.175	0.297
	336	0.378	0.445	0.399	0.456	0.332	0.418	0.357	0.433	0.328	0.414	0.360	0.438	0.499	0.522	0.360	0.432	**0.325**	**0.413**
	720	**0.676**	**0.578**	0.803	0.617	0.867	0.703	0.901	0.711	0.817	0.679	0.928	0.734	1.015	0.771	0.923	0.724	0.840	0.690
	AVG	**0.324**	**0.377**	0.372	0.400	0.367	0.408	0.385	0.415	0.352	0.399	0.399	0.432	0.476	0.483	0.391	0.418	0.356	0.400
ILI	24	**0.928**	**0.609**	1.348	0.712	2.908	1.139	2.265	1.029	2.822	1.125	4.250	1.528	6.804	1.902	2.107	0.909	-	-
	36	**1.110**	**0.673**	1.509	0.790	2.778	1.115	2.105	0.955	2.747	1.119	4.102	1.472	6.166	1.794	1.652	0.805	-	-
	48	**1.007**	**0.645**	1.375	0.748	2.760	1.112	2.409	0.955	2.784	1.131	3.992	1.452	5.282	1.634	1.982	0.851	-	-
	60	**1.114**	**0.683**	1.355	0.766	2.821	1.136	1.917	0.896	3.511	1.314	4.037	1.449	4.893	1.592	2.630	1.035	-	-
	AVG	**1.040**	**0.652**	1.397	0.754	2.817	1.126	2.174	0.959	2.966	1.172	4.095	1.475	5.786	1.731	2.093	0.900	-	-
ETTh1	96	**0.360**	0.388	0.362	0.389	0.386	0.405	0.367	0.395	0.375	0.398	0.382	0.400	0.399	**0.392**	0.368	0.394	0.375	0.394
	192	**0.412**	0.422	0.432	0.430	0.443	0.437	0.414	0.420	0.427	0.429	0.430	0.427	0.455	0.423	0.420	**0.418**	0.423	0.425
	336	**0.427**	**0.431**	0.454	0.450	0.489	0.468	0.445	0.434	0.472	0.450	0.474	0.451	0.501	0.443	0.452	0.433	0.457	0.448
	720	**0.432**	0.453	0.479	0.482	0.502	0.489	0.444	0.459	0.476	0.474	0.463	0.462	0.498	0.458	0.449	**0.449**	0.462	0.472
	AVG	**0.408**	0.424	0.432	0.438	0.455	0.450	0.418	0.427	0.438	0.438	0.437	0.435	0.463	0.429	0.422	**0.423**	0.429	0.435
ETTh2	96	**0.241**	**0.315**	0.256	0.329	0.296	0.348	0.290	0.340	0.293	0.345	0.289	0.337	0.338	0.376	0.282	0.334	0.283	0.337
	192	**0.302**	**0.357**	0.315	0.364	0.376	0.396	0.375	0.392	0.367	0.394	0.363	0.385	0.402	0.405	0.359	0.385	0.362	0.388
	336	**0.339**	**0.380**	0.354	0.392	0.424	0.431	0.423	0.435	0.419	0.431	0.419	0.426	0.439	0.444	0.374	0.404	0.408	0.425
	720	**0.363**	**0.406**	0.384	0.419	0.426	0.444	0.443	0.449	0.427	0.445	0.415	0.437	0.460	0.477	0.405	0.427	0.419	0.441
	AVG	**0.311**	**0.365**	0.327	0.376	0.381	0.405	0.383	0.404	0.377	0.404	0.372	0.396	0.410	0.426	0.355	0.387	0.368	0.398
ETTm1	96	**0.308**	0.355	0.315	0.361	0.333	0.368	0.322	0.361	0.326	0.364	0.324	0.362	0.373	0.388	0.314	**0.350**	0.321	0.359
	192	**0.349**	0.378	0.351	0.380	0.376	0.390	0.357	0.383	0.367	0.383	0.373	0.385	0.411	0.409	0.358	**0.375**	0.362	0.386
	336	0.391	0.403	0.397	0.406	0.408	0.413	**0.382**	0.401	0.402	0.409	0.402	0.404	0.436	0.421	0.384	**0.395**	0.393	0.406
	720	0.455	0.441	0.460	0.442	0.475	0.448	**0.445**	0.435	0.469	0.446	0.469	0.444	0.503	0.461	0.448	**0.432**	0.450	0.442
	AVG	**0.376**	0.394	0.380	0.397	0.398	0.405	0.377	0.395	0.391	0.401	0.392	0.399	0.431	0.420	**0.376**	**0.388**	0.381	0.398
ETTm2	96	0.173	0.254	0.178	0.261	0.179	0.263	0.174	0.255	0.178	0.260	0.173	**0.252**	0.188	0.271	**0.171**	0.253	0.172	0.256
	192	0.236	0.296	0.241	0.301	0.250	0.309	0.239	0.299	0.242	0.303	0.241	0.298	0.251	0.309	**0.234**	**0.294**	0.236	0.299
	336	**0.292**	**0.333**	0.295	0.335	0.312	0.349	0.301	0.340	0.303	0.342	0.298	0.334	0.311	0.346	**0.292**	**0.333**	0.298	0.339
	720	**0.375**	**0.386**	0.393	0.395	0.411	0.406	0.395	0.396	0.400	0.399	0.398	0.393	0.411	0.401	0.387	0.390	0.395	0.397
	AVG	**0.269**	**0.317**	0.277	0.323	0.288	0.332	0.277	0.323	0.281	0.326	0.278	0.319	0.290	0.332	0.271	**0.317**	0.275	0.323
PEMS03	96	0.186	0.291	**0.179**	**0.285**	0.209	0.313	0.504	0.507	0.653	0.597	0.281	0.368	1.436	0.967	0.360	0.410	-	-
	192	0.222	0.318	**0.208**	**0.309**	0.242	0.335	0.586	0.546	0.266	0.359	0.305	0.380	1.472	0.965	0.489	0.491	-	-
	336	0.241	0.323	**0.223**	**0.310**	0.286	0.363	0.460	0.462	0.252	0.339	0.284	0.356	1.048	0.749	0.436	0.449	-	-
	720	0.308	0.369	**0.277**	**0.350**	0.317	0.373	0.558	0.521	0.332	0.393	0.365	0.408	1.203	0.820	0.443	0.459	-	-
	AVG	0.239	0.325	**0.222**	**0.314**	0.264	0.346	0.527	0.509	0.376	0.422	0.309	0.378	1.290	0.875	0.432	0.452	-	-
PEMS04	96	0.164	0.275	0.142	0.254	**0.127**	**0.237**	0.486	0.505	0.248	0.346	0.319	0.389	1.534	1.000	0.377	0.431	-	-
	192	0.191	0.299	0.157	0.268	**0.151**	**0.256**	0.569	0.550	0.298	0.383	0.363	0.417	1.552	0.999	0.543	0.534	-	-
	336	0.199	0.297	**0.165**	**0.270**	0.169	**0.270**	0.462	0.473	0.300	0.374	0.302	0.369	1.119	0.787	0.423	0.453	-	-
	720	0.230	0.330	**0.193**	0.300	0.201	**0.299**	0.545	0.526	0.343	0.411	0.374	0.423	1.312	0.883	0.453	0.478	-	-
	AVG	0.196	0.300	0.164	0.273	**0.162**	**0.266**	0.516	0.514	0.297	0.379	0.340	0.400	1.379	0.917	0.449	0.474	-	-
PEMS07	96	0.129	0.237	**0.118**	**0.222**	0.132	0.235	0.512	0.513	0.790	0.659	0.239	0.324	1.473	0.983	0.379	0.405	-	-
	192	0.163	0.256	**0.142**	**0.242**	0.167	0.267	0.617	0.566	0.240	0.335	0.278	0.352	1.508	0.985	0.495	0.487	-	-
	336	0.154	0.253	**0.139**	**0.239**	0.159	0.251	0.481	0.476	0.210	0.306	0.232	0.315	1.074	0.767	0.415	0.439	-	-
	720	0.190	0.280	**0.178**	0.273	0.187	**0.272**	0.572	0.532	0.271	0.357	0.297	0.366	1.231	0.841	0.418	0.442	-	-
	AVG	0.159	0.257	**0.144**	**0.244**	0.161	0.256	0.546	0.522	0.378	0.414	0.262	0.339	1.322	0.894	0.427	0.443	-	-
PEMS08	96	0.232	0.289	**0.209**	**0.276**	0.284	0.332	0.596	0.515	0.986	0.725	0.384	0.396	1.579	1.007	0.487	0.453	-	-
	192	0.349	0.325	**0.284**	**0.298**	0.370	0.364	0.830	0.580	0.599	0.510	0.490	0.421	1.672	1.008	0.711	0.542	-	-
	336	0.374	0.314	**0.341**	**0.298**	0.382	0.331	0.697	0.510	0.405	0.343	0.478	0.381	1.301	0.809	0.711	0.517	-	-
	720	0.450	0.361	**0.399**	**0.339**	0.458	0.380	0.799	0.569	0.475	0.397	0.577	0.450	1.455	0.881	0.655	0.504	-	-
	AVG	0.351	0.322	**0.308**	**0.303**	0.374	0.352	0.731	0.544	0.616	0.494	0.482	0.412	1.502	0.926	0.641	0.504	-	-

(2) Linear-based methods: TimeMixer [21], DLinear [29]; (3) CNN-based methods: TimesNet [24].

Implementation Details. Our method is implemented using PyTorch with Adam optimizer on a single RTX 3090 GPU. Full hyperparameter settings are reported in the scripts of the source code https://github.com/lxy-PhD2022/SPM. In addition, in Table 3 of Section 4.4 (Resource Overhead Analysis), we unify the loss function of all models to MSE and keep the encoder hyper-parameters consistent. Since the baselines mainly employ a single-layer linear decoder while ours uses a multi-layer MLP, we set the MLP to two layers to reduce potential unfairness caused by non-core factors.

For Loss Function, we compute a frequency domain loss for Traffic dataset and MSE loss for others. The formalized representations of the frequency domain and MSE loss are as follows, respectively.

$$\mathcal{L}_{fre} = \mathbb{E}_{\boldsymbol{x}} \left\| FFT(\boldsymbol{y}_{L+1:L+T}) - FFT(\boldsymbol{x}_{L+1:L+T}) \right\|_1, \tag{14}$$

$$\mathcal{L}_{mse} = \mathbb{E}_{\boldsymbol{x}} \left\| \boldsymbol{y}_{L+1:L+T} - \boldsymbol{x}_{L+1:L+T} \right\|_2^2 \tag{15}$$

where $\mathcal{L}_{fre}$, $\mathcal{L}_{mse}$, $\boldsymbol{x}_{L+1:L+T}$, and $\boldsymbol{y}_{L+1:L+T}$ denote the frequency domain loss, MSE loss, true future sequences, and predicted future sequences of length T across M variables, respectively. Let $FFT(\cdot)$ denotes the FFT transformation along the temporal dimension.

We set the training epoch and patience of the Traffic dataset to {1000, 1000}, and set those of other datasets to {300, 20}.

C.2 Main Results

Full quantitative results of ours and 2025 latest methods are shown in Table 6. The results of earlier methods published before 2025 can be found in Table 7. It can be seen that SPM and GSPM achieve Top-2 performance in the vast majority of metrics. Our method is capable of modeling dependencies across both variables and time while treating these two dimensions on an equal footing, enabling unrestricted directional dependency modeling. As a result, it overcomes the P2P modeling limitation and attains the lowest MSE and MAE in the largest number of cases, significantly outperforming the vast number of methods published from 2021 to 2025.

C.3 Ablation Study

Component ablation. Table 5 shows the full component ablation study results: (i) Removing variable/time/all FFT components increases the average MSE by 14.6%/9.8%/12.4%, confirming the value of global modeling across both dimensions. (ii) Disabling the data-driven optimization raises 5.58% MSE, validating its effectiveness. (iii) Adding nonlinearities to $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}$ or the score matrix increases MSE by 9.52%-15.61%, while removing activations from FFN and MLP increases MSE by 11.38% and 17.46%. (iv) Replacing the Hadamard product with multiplication/addition (including a

Table 7: The results of earlier methods published before 2025. Red and blue: 1st and 2nd.

Models		GSPM		SPM		iTransformer		Leddam		MSGNet		PatchTST		Crossformer		TimesNet		DLinear		FEDformer		Autoformer		Informer	
Venue		-		-		ICLR 24		ICML 24		AAAI 24		ICLR 23		ICLR 23		ICLR 23		AAAI 23		ICML 22		NeurIPS 21		AAAI 21	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	**0.378**	**0.249**	0.384	0.253	0.395	0.268	0.426	0.276	0.609	0.349	0.526	0.347	0.522	0.290	0.593	0.321	0.650	0.396	0.587	0.366	0.613	0.388	0.719	0.391
	192	**0.400**	**0.262**	0.403	0.263	0.417	0.276	0.458	0.289	0.636	0.371	0.522	0.332	0.530	0.293	0.617	0.336	0.598	0.370	0.604	0.373	0.616	0.382	0.696	0.379
	336	**0.417**	0.272	0.420	**0.271**	0.433	0.283	0.486	0.297	0.671	0.393	0.517	0.334	0.558	0.305	0.629	0.336	0.605	0.373	0.621	0.383	0.622	0.337	0.777	0.420
	720	**0.457**	0.295	**0.457**	**0.294**	0.467	0.302	0.498	0.313	0.727	0.422	0.552	0.352	0.589	0.328	0.640	0.350	0.645	0.394	0.626	0.382	0.660	0.408	0.864	0.472
	AVG	**0.413**	**0.269**	0.416	0.270	0.428	0.282	0.467	0.294	0.661	0.384	0.529	0.341	0.550	0.304	0.620	0.336	0.625	0.383	0.610	0.376	0.628	0.379	0.764	0.416
Electricity	96	0.136	0.238	**0.132**	0.236	0.148	0.240	0.141	**0.235**	0.165	0.274	0.190	0.296	0.219	0.314	0.168	0.272	0.197	0.282	0.193	0.308	0.201	0.317	0.274	0.368
	192	0.157	**0.249**	**0.156**	**0.249**	0.162	0.253	0.159	0.252	0.184	0.292	0.199	0.304	0.231	0.322	0.184	0.289	0.196	0.285	0.201	0.315	0.222	0.334	0.296	0.386
	336	**0.171**	0.267	**0.171**	**0.266**	0.178	0.269	0.173	0.268	0.195	0.302	0.217	0.319	0.246	0.337	0.198	0.300	0.209	0.301	0.214	0.329	0.231	0.338	0.300	0.394
	720	0.198	**0.291**	**0.196**	**0.291**	0.225	0.317	0.201	0.295	0.231	0.332	0.258	0.352	0.280	0.363	0.220	0.320	0.245	0.333	0.246	0.355	0.254	0.361	0.373	0.439
	AVG	0.166	0.261	**0.164**	**0.260**	0.178	0.270	0.169	0.263	0.194	0.300	0.216	0.318	0.244	0.334	0.193	0.295	0.212	0.300	0.214	0.327	0.227	0.338	0.311	0.397
Solar	96	**0.190**	**0.231**	0.194	0.235	0.203	0.237	0.197	0.241	0.225	0.263	0.234	0.286	0.310	0.331	0.250	0.292	0.290	0.378	0.242	0.342	0.884	0.711	0.212	0.260
	192	**0.227**	0.263	0.232	0.266	0.233	0.261	0.231	0.264	0.265	0.285	0.267	0.310	0.734	0.725	0.296	0.318	0.320	0.398	0.285	0.380	0.834	0.692	0.243	**0.259**
	336	0.244	0.279	0.254	0.282	0.248	0.273	**0.241**	**0.268**	0.293	0.315	0.290	0.315	0.750	0.735	0.319	0.330	0.353	0.415	0.282	0.376	0.941	0.723	0.270	0.276
	720	0.253	0.280	0.250	0.276	0.249	**0.275**	0.250	0.281	0.291	0.315	0.289	0.317	0.769	0.765	0.338	0.337	0.356	0.413	0.357	0.427	0.882	0.717	**0.241**	0.300
	AVG	**0.229**	0.263	0.233	0.265	0.233	**0.262**	0.230	0.264	0.269	0.295	0.270	0.307	0.641	0.639	0.301	0.319	0.330	0.401	0.292	0.381	0.885	0.711	0.242	0.274
Weather	96	0.161	0.206	0.162	0.207	0.174	0.214	**0.156**	**0.202**	0.163	0.212	0.186	0.227	0.158	0.230	0.172	0.220	0.196	0.255	0.217	0.296	0.266	0.336	0.300	0.384
	192	0.208	**0.249**	0.209	0.250	0.221	0.254	0.207	0.250	0.212	0.254	0.234	0.265	**0.206**	0.277	0.219	0.261	0.237	0.296	0.276	0.336	0.307	0.367	0.598	0.544
	336	0.258	0.291	**0.256**	**0.290**	0.278	0.296	0.262	0.291	0.272	0.299	0.284	0.301	0.272	0.335	0.280	0.306	0.283	0.335	0.339	0.380	0.359	0.395	0.578	0.523
	720	0.329	0.339	**0.328**	**0.338**	0.358	0.347	0.343	0.343	0.350	0.348	0.356	0.349	0.398	0.418	0.365	0.359	0.345	0.381	0.403	0.428	0.419	0.428	1.059	0.741
	AVG	**0.239**	**0.271**	**0.239**	**0.271**	0.258	0.278	0.242	0.272	0.249	0.278	0.265	0.286	0.259	0.315	0.259	0.287	0.265	0.317	0.309	0.360	0.338	0.382	0.634	0.548
Exchange	96	**0.079**	**0.196**	0.103	0.224	0.086	0.206	0.084	0.204	0.102	0.230	0.081	0.198	0.256	0.367	0.107	0.234	0.088	0.219	0.148	0.278	0.197	0.323	0.847	0.752
	192	**0.164**	**0.289**	0.184	0.305	0.177	0.299	0.189	0.312	0.195	0.317	0.171	0.293	0.470	0.509	0.226	0.344	0.176	0.315	0.271	0.380	0.300	0.369	1.204	0.895
	336	0.378	0.445	0.399	0.456	0.331	0.417	0.338	0.421	0.359	0.436	0.319	**0.407**	1.268	0.883	0.367	0.448	**0.313**	0.427	0.460	0.500	0.509	0.524	1.672	1.036
	720	**0.676**	**0.578**	0.803	0.617	0.847	0.691	1.072	0.771	0.940	0.738	0.836	0.688	1.767	1.068	0.964	0.746	0.839	0.695	1.195	0.841	1.447	0.941	2.478	1.310
	AVG	**0.324**	**0.377**	0.372	0.400	0.360	0.403	0.421	0.427	0.399	0.430	0.352	0.397	0.940	0.707	0.416	0.443	0.354	0.414	0.519	0.500	0.613	0.539	1.550	0.998
ILI	24	**0.928**	**0.609**	1.348	0.712	1.832	0.861	1.815	0.838	3.993	1.411	1.581	0.787	3.442	1.240	2.317	0.934	2.398	1.040	3.228	1.260	3.483	1.287	5.764	1.677
	36	**1.110**	**0.673**	1.509	0.790	1.638	0.832	2.077	0.930	3.728	1.310	1.254	0.733	3.463	1.257	1.972	0.920	2.646	1.088	2.679	1.080	3.103	1.148	4.755	1.467
	48	**1.007**	**0.645**	1.375	0.748	2.117	0.914	2.084	0.887	3.379	1.259	1.999	0.854	3.640	1.296	2.238	0.940	2.614	1.086	2.622	1.078	2.669	1.085	4.763	1.469
	60	**1.114**	**0.683**	1.355	0.766	1.838	0.885	1.997	0.896	4.055	1.430	1.697	0.829	3.928	1.352	2.027	0.928	2.804	1.146	2.857	1.157	2.770	1.125	5.264	1.564
	AVG	**1.040**	**0.652**	1.397	0.754	1.856	0.873	1.993	0.888	3.789	1.353	1.633	0.801	3.618	1.286	2.139	0.931	2.616	1.090	2.847	1.144	3.006	1.161	5.137	1.544
ETTh1	96	**0.360**	**0.388**	0.362	0.389	0.386	0.405	0.377	0.394	0.39	0.411	0.460	0.447	0.423	0.448	0.384	0.402	0.386	0.400	0.376	0.419	0.449	0.459	0.865	0.713
	192	**0.412**	**0.422**	0.432	0.430	0.441	0.436	0.424	**0.422**	0.442	0.442	0.512	0.477	0.471	0.474	0.436	0.429	0.437	0.432	0.420	0.448	0.500	0.482	1.008	0.792
	336	**0.427**	**0.431**	0.454	0.450	0.487	0.458	0.459	0.442	0.480	0.468	0.546	0.496	0.570	0.546	0.491	0.469	0.481	0.459	0.459	0.465	0.521	0.496	1.107	0.809
	720	**0.432**	**0.453**	0.479	0.482	0.503	0.491	0.463	0.459	0.494	0.488	0.544	0.517	0.653	0.621	0.521	0.500	0.519	0.516	0.506	0.507	0.514	0.512	1.181	0.865
	AVG	**0.408**	**0.424**	0.432	0.438	0.454	0.448	0.431	0.429	0.452	0.452	0.516	0.484	0.529	0.522	0.458	0.450	0.456	0.452	0.440	0.460	0.496	0.487	1.040	0.795
ETTh2	96	**0.241**	**0.315**	0.256	0.329	0.297	0.349	0.292	0.343	0.328	0.371	0.308	0.355	0.745	0.584	0.340	0.374	0.333	0.387	0.358	0.397	0.346	0.388	3.755	1.525
	192	**0.302**	**0.357**	0.315	0.364	0.380	0.400	0.367	0.389	0.402	0.414	0.393	0.405	0.877	0.656	0.402	0.414	0.477	0.476	0.429	0.439	0.456	0.452	5.602	1.931
	336	**0.339**	**0.380**	0.354	0.392	0.428	0.432	0.412	0.424	0.435	0.443	0.427	0.436	1.043	0.731	0.452	0.452	0.594	0.541	0.496	0.487	0.482	0.486	4.721	1.835
	720	**0.363**	**0.406**	0.384	0.419	0.427	0.445	0.419	0.438	0.417	0.441	0.436	0.450	1.104	0.763	0.462	0.468	0.831	0.657	0.463	0.474	0.515	0.511	3.647	1.625
	AVG	**0.311**	**0.365**	0.327	0.376	0.383	0.407	0.373	0.399	0.396	0.417	0.391	0.412	0.942	0.684	0.414	0.427	0.559	0.515	0.437	0.449	0.450	0.459	4.431	1.729
ETTm1	96	**0.308**	**0.355**	0.315	0.361	0.334	0.368	0.319	0.359	0.319	0.366	0.352	0.374	0.404	0.426	0.338	0.375	0.345	0.372	0.379	0.419	0.505	0.475	0.672	0.571
	192	**0.349**	**0.378**	0.351	0.380	0.377	0.391	0.369	0.383	0.376	0.397	0.390	0.393	0.450	0.451	0.374	0.387	0.380	0.389	0.426	0.441	0.553	0.496	0.795	0.669
	336	**0.391**	0.403	0.397	0.406	0.426	0.420	0.394	**0.402**	0.417	0.422	0.421	0.414	0.532	0.515	0.410	0.411	0.413	0.413	0.445	0.459	0.621	0.537	1.212	0.871
	720	**0.455**	**0.441**	0.460	0.442	0.491	0.459	0.460	0.442	0.481	0.458	0.462	0.449	0.666	0.589	0.478	0.450	0.474	0.453	0.543	0.490	0.671	0.561	1.166	0.823
	AVG	**0.376**	**0.394**	0.380	0.397	0.407	0.410	0.386	0.397	0.398	0.411	0.406	0.408	0.513	0.495	0.400	0.406	0.403	0.407	0.448	0.452	0.588	0.517	0.961	0.734
ETTm2	96	**0.173**	**0.254**	0.178	0.261	0.180	0.264	0.176	0.257	0.177	0.262	0.183	0.270	0.287	0.366	0.187	0.267	0.193	0.292	0.203	0.287	0.255	0.339	0.365	0.453
	192	**0.236**	**0.296**	0.241	0.301	0.250	0.309	0.243	0.303	0.247	0.307	0.255	0.314	0.414	0.492	0.249	0.309	0.248	0.362	0.269	0.328	0.281	0.340	0.533	0.563
	336	**0.292**	**0.333**	0.295	0.335	0.311	0.348	0.303	0.402	0.312	0.346	0.309	0.347	0.597	0.542	0.321	0.351	0.369	0.427	0.325	0.366	0.339	0.372	1.363	0.887
	720	**0.375**	**0.386**	0.393	0.395	0.412	0.407	0.400	0.442	0.414	0.403	0.412	0.404	1.730	1.042	0.408	0.403	0.554	0.522	0.421	0.415	0.433	0.432	3.379	1.338
	AVG	**0.269**	**0.317**	0.277	0.323	0.288	0.332	0.281	0.351	0.288	0.330	0.290	0.334	0.757	0.611	0.291	0.333	0.341	0.401	0.305	0.349	0.327	0.371	1.410	0.810
PEMS03	96	0.186	0.291	**0.179**	**0.285**	0.276	0.369	0.252	0.347	0.267	0.341	0.505	0.517	0.262	0.367	0.228	0.317	0.457	0.515	0.348	0.434	1.031	0.796	0.225	0.316
	192	0.222	0.318	0.208	0.309	0.325	0.400	0.286	0.370	0.425	0.610	0.549	0.545	**0.173**	**0.279**	0.290	0.364	0.477	0.527	0.486	0.535	2.893	1.401	0.204	0.304
	336	0.241	0.323	0.223	0.310	0.306	0.373	0.273	0.354	0.412	0.605	0.431	0.459	0.221	0.303	0.264	0.342	0.396	0.456	0.579	0.564	1.654	0.999	**0.206**	**0.301**
	720	0.308	0.369	0.277	0.350	0.371	0.419	0.331	0.395	0.471	0.675	0.540	0.528	0.252	**0.320**	0.335	0.391	0.439	0.492	0.710	0.631	1.772	1.078	**0.245**	0.332
	AVG	0.239	0.325	0.222	0.314	0.320	0.390	0.286	0.367	0.394	0.558	0.506	0.512	0.227	0.317	0.279	0.354	0.442	0.498	0.531	0.541	1.838	1.069	**0.220**	**0.313**
PEMS04	96	0.164	0.275	0.142	0.254	0.314	0.392	0.170	0.285	0.194	0.314	0.644	0.594	0.402	0.457	0.190	0.303	0.452	0.504	0.341	0.427	0.912	0.748	**0.128**	**0.243**
	192	0.191	0.299	0.157	0.268	0.354	0.419	0.213	0.319	0.395	0.550	0.689	0.616	0.199	0.292	0.223	0.332	0.470	0.516	0.599	0.595	1.651	1.005	**0.151**	**0.260**
	336	0.199	0.297	0.165	**0.270**	0.314	0.383	0.207	0.305	0.394	0.558	0.536	0.517	0.188	0.289	0.215	0.321	0.399	0.459	0.637	0.598	1.412	0.926	**0.161**	0.700
	720	0.230	0.330	**0.193**	0.300	0.366	0.423	0.215	0.318	0.461	0.637	0.629	0.578	0.210	**0.293**	0.262	0.362	0.462	0.501	0.753	0.659	1.707	1.060	0.219	0.310
	AVG	0.196	0.300	**0.164**	**0.273**	0.337	0.404	0.201	0.307	0.361	0.515	0.625	0.576	0.250	0.333	0.223	0.330	0.446	0.495	0.583	0.570	1.421	0.935	0.165	0.378
PEMS07	96	0.129	0.237	**0.118**	**0.222**	0.240	0.336	0.162	0.267	0.187	0.289	0.492	0.507	0.396	0.442	0.181	0.279	0.594	0.553	0.262	0.376	0.554	0.578	0.201	0.274
	192	0.163	0.256	**0.142**	**0.242**	0.284	0.368	0.239	0.336	0.359	0.545	0.545	0.538	0.196	0.274	0.223	0.315	0.621	0.567	0.419	0.489	1.257	0.897	0.206	0.272
	336	0.154	0.253	**0.139**	**0.239**	0.238	0.329	0.186	0.288	0.344	0.526	0.432	0.463	0.192	0.265	0.214	0.312	0.491	0.490	0.445	0.494	1.421	0.898	0.221	0.284
	720	0.190	0.280	**0.178**	**0.273**	0.316	0.386	0.246	0.339	0.263	-	0.509	0.513	0.233	0.291	0.271	0.353	0.561	0.532	0.482	0.519	1.325	0.923	0.306	0.345
	AVG	0.159	0.257	**0.144**	**0.244**	0.270	0.355	0.208	0.308	0.288	0.453	0.495	0.505	0.254	0.318	0.222	0.315	0.567	0.536	0.402	0.470	1.139	0.824	0.234	0.294
PEMS08	96	0.232	0.289	**0.209**	**0.276**	0.384	0.402	0.253	0.320	0.370	0.400	0.682	0.591	0.377	0.397	0.320	0.351	0.674	0.565	0.442	0.465	1.385	0.915	0.455	0.421
	192	0.349	0.325	0.284	0.298	0.506	0.441	0.365	0.352	0.486	0.782	0.778	0.599	**0.251**	**0.274**	0.496	0.445	0.727	0.58	0.743	0.621	1.754	1.016	0.425	0.390
	336	0.374	0.314	0.341	**0.298**	0.495	0.400	0.392	0.340	0.462	0.790	0.711	0.540	**0.292**	0.304	0.488	0.401	0.65	0.518	1.047	0.724	0.938	0.702	0.427	0.383
	720	0.450	0.361	0.399	0.339	0.587	0.460	0.640	0.500	0.523	0.876	0.807	0.599	**0.343**	**0.317**	0.588	0.448	0.713	0.563	1.050	0.739	1.537	0.964	0.425	0.387
	AVG	0.351	0.322	**0.308**	**0.303**	0.493	0.426	0.413	0.378	0.460	0.712	0.745	0.582	0.316	0.323	0.473	0.411	0.691	0.557	0.821	0.637	1.404	0.899	0.433	0.395

time-domain variant) increases MSE by 10.05%–12.70%, indicating it best matches the spectral product mechanism.