TVTS: LEARNING TIME-VARYING DYNAMICS AND DIRECTED CROSS-VARIABLE DEPENDENCIES FOR MULTIVARIATE TIME SERIES FORECASTING

Anonymous authors

ABSTRACT

Multivariate Time Series Forecasting (MTSF) requires jointly modeling both temporally heterogeneous patterns and cross-variable dependencies. However, many recent designs rely on time-invariant temporal filters and symmetric similarity for variable interactions, which limits the faithful representation of position-dependent temporal dynamics and directional information propagation. To address these challenges, we propose the **T**ime-**V**arying Directed Geometry framework for Multivariate **T**ime **S**eries Forecasting (**TVTS**). For temporal modeling, we introduce a Time-Varying Convolutional Filter that learns distinct spectral responses across relative temporal positions through a structured basis expansion, thereby relaxing the shift-invariance assumption of conventional filters. For cross-variable modeling, we propose Directed Geometric Clustering, which constructs a sparse anisotropic topology to represent non-reciprocal predictive dependencies beyond symmetric similarity. Extensive experiments on diverse and challenging real-world benchmarks demonstrate that TVTS consistently achieves state-of-the-art forecasting performance with high computational efficiency across forecasting horizons [1].

1 INTRODUCTION

Multivariate Time Series Forecasting (MTSF) is central to real-world applications such as energy Guo et al. (2015); Deb et al. (2017), traffic Lippi et al. (2013); Cirstea et al. (2022), finance Cao & Tay (2003); Sezer et al. (2020), and weather Karevan & Suykens (2020). Given that these systems involve coupled variables evolving over time, accurate forecasting demands models that can capture both intra-series temporal dynamics and inter-series correlations. To capture these patterns, various approaches have been developed, ranging from Convolutional Neural Networks (CNNs) Wu et al. (2023); Luo & Wang (2024) and Recurrent Neural Networks (RNNs) Lai et al. (2018) to Graph Neural Networks (GNNs) Cai et al. (2024); Huang et al. (2023) and Multi-layer Perceptron models Zeng et al. (2023). Recently, Transformers have emerged as a prominent choice for MTSF Liu et al. (2024c); Wang et al. (2024b); Nie et al. (2023), owing to their ability to model long-range dependencies within complex temporal data.

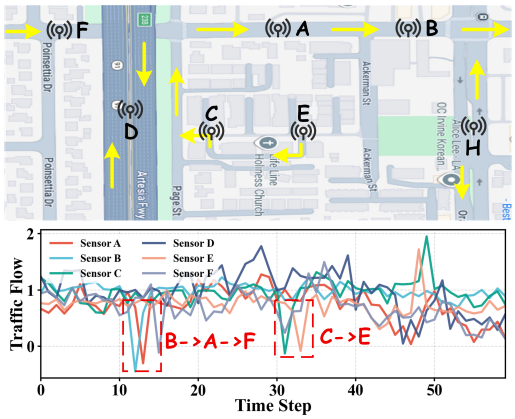


Figure 1: Example of directional propagation across interconnected traffic sensor networks. The red dashed boxes highlight the propagation paths $B \rightarrow A \rightarrow F$ and $C \rightarrow E$, demonstrating that dependencies are governed by directed connectivity rather than mutual similarity.

Despite their success, self-attention mechanisms face fundamental limitations due to their lack of inductive bias towards sequential structure. While position embeddings provide absolute temporal context, self-attention treats all positions via pairwise correlations without prioritizing local proximity Wang et al. (2022); Shin et al. (2024); He et al. (2026). This global receptive field makes it difficult

[1] The source code is anonymous online at: `https://anonymous.4open.science/r/TVTS-main`

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

to capture fine-grained, localized patterns effectively. To compensate, recent models integrate convolutional layers to extract local temporal features Wu et al. (2023); Luo & Wang (2024) or introduce frequency-domain components to emphasize periodic and high-frequency signals Yi et al. (2024); Piao et al. (2024). However, these approaches are typically time-invariant, applying a shared filtering response across all temporal positions Liu et al. (2025b); Qiu et al. (2025). When temporal statistics vary across positions, such static designs inadequately capture evolving dynamics.

Beyond temporal locality, modeling the complex interactions among variables (channels) is equally critical Nie et al. (2023); Liu et al. (2024c). In multivariate systems, different channels often exhibit heterogeneous temporal evolution Qiu et al. (2024), where the predictive accuracy for a specific channel relies on leveraging information from correlated counterparts Wang et al. (2024c). Existing approaches often assume an "*all-to-all*" structure, which introduces redundancy and noise Zhang & Yan (2023); Wang et al. (2024b). Alternative methods seek to mitigate this issue by partitioning series into disjoint clusters Liu et al. (2024a), employing metric-based soft clustering Qiu et al. (2025); Gao et al. (2025), or filtering dependencies over learned graphs Zhao et al. (2023); Hu et al. (2025).

However, these methods overlook the intrinsic topological constraints of real-world systems Seversky et al. (2016). As illustrated in Figure 1, physical dependencies are predominantly **directed** and **asymmetric** rather than mutual. For instance, consider a traffic congestion event originating at Sensor B. This disturbance propagates downstream, affecting Sensor A and subsequently Sensor F with a discernible temporal lag. Crucially, this relationship is non-reciprocal, as a fluctuation at sensor F has no retroactive effect on the upstream traffic at B and A. Current approaches typically encode relations as symmetric similarity, which cannot express directional propagation Qiu et al. (2025); Gao et al. (2025). Consequently, they often fail to capture the intrinsic geometric structure of multivariate systems, where dependencies are governed by directed connectivity rather than mutual similarity.

To address these limitations, we propose the **T**ime-**V**arying Directed Geometry framework for **T**ime **S**eries (**TVTS**). Inspired by graph signal processing, we substitute static convolutional kernels with a *Time-Varying Convolutional Filter (TVCF)* module (Section 4.2). This component learns position-specific spectral responses to model temporal heterogeneity, relaxing the shift-invariance assumption of standard convolution. Furthermore, to decode the non-reciprocal dependencies of multivariate systems, we design a *Directed Geometric Clustering (DGC)* module (Section 4.3). By leveraging direction-aware metric learning and Top-k sparsification, DGC constructs an anisotropic topology that distinguishes non-reciprocal predictive dependencies from symmetric similarity. Integrating these components, TVTS incorporates *topology-guided fusion* and *frequency-domain residual prediction* mechanisms (Section 4.4) to synergize spatio-temporal contexts and iteratively refine forecasting precision. Experiments demonstrate that TVTS achieves state-of-the-art (SOTA) performance across benchmarks in both long- and short-term forecasting. Our contributions are three-fold:

- We propose TVTS, a unified forecasting framework that jointly models position-dependent temporal dynamics and directed cross-variable dependencies, providing a principled solution to the limitations of shift-invariant temporal filtering and reciprocal variable interactions.
- We develop two core components, TVCF and DGC. TVCF learns compact position-specific spectral responses to capture temporal heterogeneity, while DGC constructs a sparse anisotropic topology through source–target decoupling to model non-reciprocal predictive dependencies.
- Comprehensive experiments across diverse real-world datasets demonstrate that TVTS consistently achieves SOTA performance in both long- and short-term forecasting. Extensive ablations, visualizations, and statistical analyses further validate the effectiveness and robustness of TVTS.

2 Related Works

2.1 Deep Architectures and Filter in MTSF

Deep learning for MTSF has evolved from complex **Transformer-based** architectures Zhou et al. (2021); Liu et al. (2024c); Wang et al. (2024b); Zhang et al. (2026) to efficient **MLP-based** models Zeng et al. (2023); Wang et al. (2024a); Tang & Zhang (2025), demonstrating that structurally minimalist linear layers can rival attention mechanisms with superior computational efficiency.

Concurrent with these backbone evolutions, extensive research has focused on explicit feature extraction through specialized filtering mechanisms. This encompasses **CNNs** Sen et al. (2019); Wu et al. (2023); Chen et al. (2024) for local feature extraction, **GNNs** Yu et al. (2018); Lan et al.

(2022); Cai et al. (2024) for modeling inter-variable correlations, and a recent paradigm shift towards **frequency-domain learning** Yi et al. (2023; 2024); Xu et al. (2024); Piao et al. (2024); Wang et al. (2025a;c); Yue et al. (2025b); Qiu et al. (2026) to capture global periodic trends. Despite their diversity, as summarized in the "Filter Mechanism" column of Table 1, these methodologies share a critical limitation: they predominantly employ a **Fixed Filter** mechanism. Whether applying static convolutional kernels or time-invariant frequency transforms, the unified application of identical filters across the entire sequence makes it difficult to capture evolving dynamics or location-specific shifts, thereby limiting adaptability to data characterized by significant local statistical shifts.

2.2 Cross-Variable Modeling in MTSF

Effective variable correlation modeling is pivotal for MTSF. Contemporary approaches can be broadly categorized into three paradigms: First, **Channel Independence (CI)** strategy ensures robustness by modeling variates individually to mitigate spurious correlations Nie et al. (2023); Zeng et al. (2023); Xu et al. (2024); however, it inherently overlooks inter-variable dependencies, often yielding suboptimal performance in complex systems Liu et al. (2024c). Conversely, **Channel Dependence (CD)** approaches capture global interactions via joint embeddings Liu et al. (2024c); Han et al. (2024); Wang et al. (2024b). Yet, they face risks of over-smoothing and noise amplification in high-dimensional settings, as they tend to indiscriminately aggregate information from irrelevant or conflicting channels.

Table 1: Qualitative paradigm comparison. CI, CD, and CS denote Channel Independence, Channel Dependence, and Channel Selectivity, respectively. More "✔" symbols indicate better performance or efficiency.

Model	Filter Mechanism	Channel Strategy	Directed Topology	Inference Efficiency	Predictive Performance
Dlinear (2023)	✗	CI	✗	✔✔	✗
PatchTST (2023)	✗	CI	✗	✗	✗
TimesNet (2023)	Fixed	CD	✗	✗	✗
CARD (2024b)	✗	CD	✗	✗	✔
MSGNet (2024)	Fixed	CD	✗	✗	✗
FilterNet (2024)	Fixed	CD	✗	✔✔	✗
FilterTS (2025c)	Fixed	CD	✗	✔	✔
Fredfm. (2024)	Fixed	CD	✗	✗	✔✔
BIM^3 (2025)	Fixed	CS	✗	✔	✔✔
DUET (2025)	Fixed	CS	✗	✔	✔✔
TVTS (Ours)	Time-Varying	CS	✔	✔✔	✔✔

To reconcile this trade-off, **Channel Selectivity (CS)** methods introduce hybrid mechanisms that preserve individual channel independence while permitting selective influence from informative peers Zhao et al. (2023); Shang et al. (2024); Luo & Wang (2024); Wu et al. (2025). Some approaches construct learned spatial-temporal graphs to adaptively capture and filter informative inter-channel or patch-level dependencies Zhao et al. (2023); Hu et al. (2025); Wang et al. (2026c). In parallel, recent methods such as DUET Qiu et al. (2025) and BIM^3 Gao et al. (2025) employ frequency-domain metric learning for soft clustering. However, these approaches mainly determine which relationships should be retained through correlation or similarity, while overlooking interaction directionality and thus failing to distinguish directional predictive influence from symmetric association.

3 Preliminary

Problem Statement. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_L] \in \mathbb{R}^{N \times L}$ denote a multivariate historical sequence, where $\mathbf{x}_t \in \mathbb{R}^N$ contains the observations of N variables at timestamp t, and L is the look-back length. Given $\mathbf{X}$, multivariate time-series forecasting aims to predict the subsequent H timestamps, $\mathbf{Y} = [\mathbf{y}_{L+1}, \ldots, \mathbf{y}_{L+H}] \in \mathbb{R}^{N \times H}$. We parameterize the forecasting function as $\mathcal{F}_\theta : \mathbb{R}^{N \times L} \to \mathbb{R}^{N \times H}$, producing $\widehat{\mathbf{Y}} = \mathcal{F}_\theta(\mathbf{X})$. Given a training set $\mathcal{D} = \{(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)})\}_{i=1}^{M}$, the model parameters are optimized by minimizing the forecasting error across all samples, variables, and prediction steps:

$$\theta^* = \arg\min_\theta \frac{1}{MNH} \sum_{i=1}^{M} \sum_{n=1}^{N} \sum_{h=1}^{H} \ell\left(Y_{n,h}^{(i)}, \left[\mathcal{F}_\theta\left(\mathbf{X}^{(i)}\right)\right]_{n,h}\right), \tag{1}$$

where $\ell(\cdot, \cdot)$ denotes the forecasting loss, such as the Mean Squared Error (MSE).

Graph Signal Processing for Time Series. We formulate multivariate time series via Graph Signal Processing (GSP) by modeling the look-back length L as a signal supported on a Directed Cyclic Graph (DCG). The underlying topology is encoded by a circulant graph shift operator $\mathbf{S} = \mathbf{\Phi D \Phi}^{-1}$, where $\mathbf{\Phi} \in \mathbb{C}^{L \times L}$ represents the orthonormal graph Fourier basis and $\mathbf{D} = \mathrm{diag}(\lambda_0, \ldots, \lambda_{L-1})$ denotes the spectral eigenvalues. Unlike canonical shift-invariant filters Yi et al. (2024), time-varying filters accommodate non-stationary dynamics by assigning position-specific weights. Leveraging

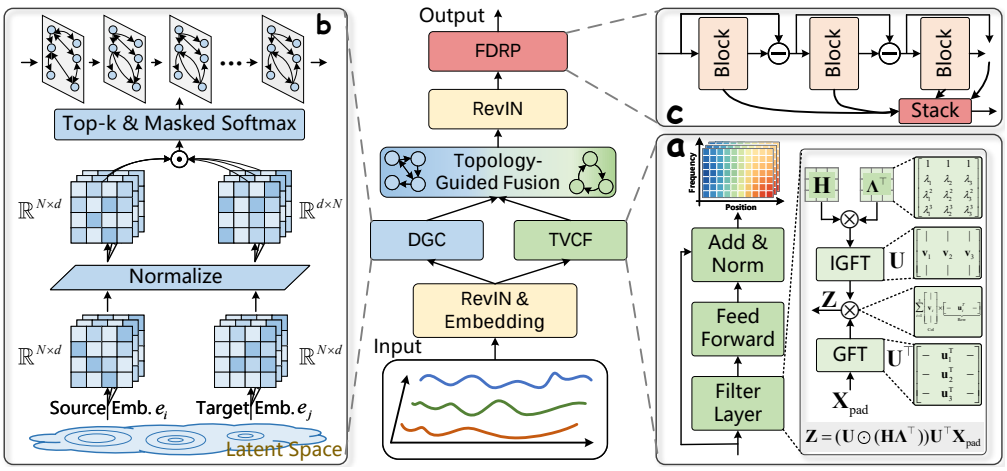


Figure 2: The framework of TVTS consists of three components: **(a) Time-Varying Convolutional Filter (TVCF)** models temporal evolution on a Directed Cyclic Graph via Graph Fourier Transform (GFT) and position-specific filter taps parameterized by Vandermonde matrix $\mathbf{\Lambda}$ to capture spectral variations; **(b) Directed Geometric Clustering (DGC)** learns an anisotropic mask from cosine similarity to build a sparse directed topology for directional dependencies; **(c) Frequency-Domain Residual Prediction (FDRP)** progressively refines forecasts through stacked spectral residual blocks.

node-variant graph filters Segarra et al. (2017); Gama et al. (2022), we allocate unique filter coefficients to individual nodes to model non-stationarity. Formally, we define a filter tap matrix $\mathbf{W} \in \mathbb{R}^{L\times(K+1)}$, where the k-th column $\mathbf{w}_k$ governs the k-th order shift response. Leveraging a structured Vandermonde matrix $\mathbf{\Lambda} \in \mathbb{C}^{L\times(K+1)}$ (with $\mathbf{\Lambda}_{i,k} = \lambda_i^k$) derived from graph frequencies, the filtering operation admits a computationally efficient spectral representation. The mapping from input $\mathbf{x}$ to output $\mathbf{y}$ is formulated as follows (Please refer to Appendix A for more details):

$$\mathbf{y} = \sum_{k=0}^{K} \text{diag}(\mathbf{w}_k)\mathbf{S}^k\mathbf{x} \iff \mathbf{y} = \left(\mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)\right)\tilde{\mathbf{x}}, \tag{2}$$

where $\tilde{\mathbf{x}} = \mathbf{\Phi}^{-1}\mathbf{x}$ denotes the Graph Fourier Transform (GFT) of the input. The term $\mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)$ constitutes a learnable time-varying convolution kernel that assigns distinct spectral responses across temporal positions, effectively capturing localized temporal variations.

4 Proposed Method

4.1 Series Representation

Following standard practices to address distribution discrepancies, we first normalize the input using Reversible Instance Normalization (RevIN) Kim et al. (2021). And then, to adapt linear sequences to the DCG topology and suppress boundary artifacts, we generate an extended representation $\mathbf{X}_{\text{pad}}$ by appending a zero-tensor as follows (see Appendix B for further theoretical justification):

$$\mathbf{X}_{\text{pad}} = \text{Concat}(\mathbf{X}, \mathbf{P}) \in \mathbb{R}^{L_{pad}\times N}, \tag{3}$$

where $\mathbf{X}$ denotes the normalized input, $\mathbf{P} = \mathbb{R}^{(K-1)\times N}$ represents the zero-padding tensor, and $L_{pad} = L + K - 1$ is the extended sequence length determined by the filter order K. This padding strategy isolates cyclic shift artifacts, thereby aligning spectral filtering with temporal directionality.

4.2 Time-Varying Convolutional Filtering

Prevailing MTSF models typically employ time-invariant convolutional kernels Wu et al. (2023) or fixed frequency-domain transforms Yi et al. (2024). However, these static designs apply a shared response across temporal positions, limiting their ability to represent position-dependent spectral

variations. To overcome this rigidity, we propose the **Time-Varying Convolutional Filter (TVCF)**, which models temporal heterogeneity by learning position-specific spectral representations.

As illustrated in Figure 2 (a), TVCF performs position-varying spectral filtering on the extended sequence. Inspired by Eq. (2), the GFT step and the reconstruction can be formulated as:

$$\mathbf{Z} = \mathcal{T}\Big(\big[\mathbf{U} \odot \big(\mathbf{H}\mathbf{\Lambda}^\top\big)\big]\, \widetilde{\mathbf{X}}_{\text{pad}}\Big) = \mathcal{T}\big(\big[\mathbf{U} \odot \big(\mathbf{H}\mathbf{\Lambda}^\top\big)\big]\, \mathbf{U}^\top \mathbf{X}_{\text{pad}}\big) \in \mathbb{R}^{L\times N}, \tag{4}$$

where $\mathbf{U}^\top$ denotes the GFT basis matrix derived from the cyclic shift operator, while $\mathbf{U}$ provides the corresponding reconstruction basis. $\mathbf{\Lambda}$ represents the corresponding eigenvalues, and $\mathbf{H}$ contains the position-specific filter coefficients. Their product $\mathbf{H}\mathbf{\Lambda}^\top$ therefore defines the time-varying frequency response. Finally, $\mathcal{T}(\cdot)$ removes the padding buffer and restores the original sequence length L.

Notably, to model position-dependent filters without introducing an unstructured parameter matrix, we factorize the learnable filter $\mathbf{H}$ through a structured basis expansion:

$$\overline{\mathbf{B}}_{m,:} = \frac{\mathbf{B}_{m,:}}{\|\mathbf{B}_{m,:}\|_2 + \varepsilon}, \qquad \mathbf{H} = \mathbf{C}\overline{\mathbf{B}} \in \mathbb{C}^{L_{\text{pad}}\times(K+1)}, \tag{5}$$

where $\mathbf{B} \in \mathbb{C}^{M\times(K+1)}$ denotes the shared complex-valued basis and $\mathbf{C} \in \mathbb{R}^{L_{\text{pad}}\times M}$ contains learnable position-specific coefficients. We normalize each complex basis vector independently as $\overline{\mathbf{B}}_{m,:} = \mathbf{B}_{m,:}/(\|\mathbf{B}_{m,:}\|_2 + \varepsilon)$, and construct the position-dependent filter matrix as $\mathbf{H} = \mathbf{C}\overline{\mathbf{B}}$. Finally, feed-forward-based refinement further improves the learned temporal representations.

4.3 Directed Geometric Clustering

Many real-world multivariate systems exhibit non-reciprocal predictive dependencies, such that the influence of one variable on another can be asymmetric. However, symmetric similarity-based methods are limited in explicitly representing such anisotropic information flow Qiu et al. (2025); Gao et al. (2025). To address this, we propose **Directed Geometric Clustering (DGC)**, which learns a sparse, asymmetric topology as a predictive inductive bias for modeling directed cross-variable interactions within the topology-guided attention mechanism (as illustrated in Figure 2 (b)).

Specifically, we quantify the directed dependency between variables by computing an asymmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ via a dual-embedding interaction mechanism (more on Appendix C):

$$\mathbf{A}_{ij} = \tanh\big(\alpha \mathbf{E}_s \mathbf{E}_t^\top\big) = \tanh\left(\alpha \cdot \frac{\mathbf{e}_{i,s}(\mathbf{e}_{j,t})^\top}{\|\mathbf{e}_{i,s}\|_2 \|\mathbf{e}_{j,t}\|_2}\right), \tag{6}$$

where $\mathbf{e}_{i,s}, \mathbf{e}_{i,t} \in \mathbb{R}^d$ represent the source ($\mathbf{E}_s$) and target ($\mathbf{E}_t$) embeddings for variable i, respectively, and α is a scaling factor controlling saturation. Notably, this source-target separation breaks the symmetry assumption (i.e., $\mathbf{A}_{ij} \neq \mathbf{A}_{ji}$), enabling the precise modeling of directional flows.

To further eliminate noisy correlations and ensure computational efficiency, we refine $\mathbf{A}$ into a sparse topological mask $\mathbf{M}$ using a hard-thresholding strategy:

$$\mathbf{M}_{ij} = \begin{cases} \mathbf{A}_{ij} & \text{if } \mathbf{A}_{ij} \in \text{Top-}k(\mathbf{A}_{i,:}) \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $k (k \leq N)$ determines the sparsity budget by retaining only the most significant neighbors. The resulting mask $\mathbf{M}$ is subsequently normalized via Masked $\mathrm{Softmax}(\cdot)$ and injected into the decoder as a structural bias, effectively functioning as a spatial geometric clustering operator.

4.4 Fusion Attention and Prediction Layer

To synthesize the dynamic features $\mathbf{Z}$ extracted by TVCF (Section 4.2) and the spatial topology $\mathbf{M}$ derived from DGC (Section 4.3), we introduce the **Topology-Guided Fusion (TGF)** attention mechanism (Appendix Figure 7a). Instead of relying on dense all-to-all attention, TGF utilizes the sparse adjacency matrix $\mathbf{M}$ as a structural prior to regulate cross-variable information flow.

We project the temporal features $\mathbf{Z}$ into a latent representation $\mathbf{Z}' \in \mathbb{R}^{N\times D}$. The learned topology is then introduced as an additive structural mask in the attention computation:

$$\mathbf{Z}_{\text{TGF}} = \text{Att}_{\mathbf{M}}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{(\mathbf{Z}'\mathbf{W}_Q)(\mathbf{Z}'\mathbf{W}_K)^\top}{\sqrt{D}} + \psi(\mathbf{M})\right)(\mathbf{Z}'\mathbf{W}_V) \in \mathbb{R}^{N\times D}, \tag{8}$$

where $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V \in \mathbb{R}^{D\times D}$ are learnable projection matrices, yielding $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{N\times D}$. The masking function $\psi(\mathbf{M})$ assigns zero to connected pairs and a large negative value to disconnected pairs, thereby restricting cross-variable information flow to the learned directed topology.

Following Transformer-style refinement Vaswani et al. (2017), the topology-aware representation is processed by a position-wise FFN with residual connection and Layer Normalization (LNorm):

$$\hat{\mathbf{Z}} = \mathrm{LNorm}(\mathbf{Z}' + \mathrm{FFN}(\mathbf{Z}_{\mathrm{TGF}})) = \mathrm{LNorm}(\mathbf{Z}' + \mathrm{FFN}(\mathrm{Att}_{\mathbf{M}}(\mathbf{Q}, \mathbf{K}, \mathbf{V}))) \in \mathbb{R}^{N\times D}, \quad (9)$$

Frequency-Domain Residual Prediction (FDRP). To further refine forecasting precision, we employ I stacked residual blocks in the frequency domain (Appendix Figure 7b). Let $\mathbf{R}^{(0)} = \hat{\mathbf{Z}}$ denote the initial residual derived from the TGF output (Eq. 9). Each block bridges the length mismatch between look-back window L and horizon H using a complex upsampler ϕ and spectral modulator $\mathcal{W}$:

$$\tilde{\mathbf{O}}^{(i)} = \mathcal{F}^{-1}\left[\phi\Big(\mathcal{F}(\mathbf{R}^{(i-1)})\Big) \odot \mathcal{W}\Big(\mathcal{F}(\mathbf{R}^{(i-1)})\Big)\right], \quad (10)$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fast Fourier Transform (FFT) and its inverse, respectively. The output $\tilde{\mathbf{O}}^{(i)}$ is split into reconstruction $\hat{\mathbf{X}}^{(i)}$ and prediction $\hat{\mathbf{Y}}^{(i)}$ to recursively update the residual and accumulate the final forecast $\hat{\mathbf{Y}} = \sum_{i=1}^{I} \hat{\mathbf{Y}}^{(i)} \in \mathbb{R}^{N\times H}$. Further details are provided in Appendix E, and the overall pseudocode of TVTS is presented in Algorithm 1.

4.5 Learning Objective

The overall objective function combines orthogonal regularization and time-frequency loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{pred}} + \lambda \mathcal{L}_{\mathrm{ortho}}, \quad \lambda \in [0, 1]. \quad (11)$$

Orthogonal Regularization ($\mathcal{L}_{\mathrm{ortho}}$). To prevent mode collapse and encourage diverse frequency representations in the spectral basis $\mathbf{B}$ (Eq. 5), we enforce orthogonality on its real and imaginary parts. Let $\tilde{\mathbf{B}}$ denote the row-normalized matrix ($\tilde{\mathbf{B}}_p = \mathbf{B}_p / \|\mathbf{B}_p\|_2$). The regularization term is:

$$\mathcal{L}_{\mathrm{ortho}} = \sum_{p\in\{\mathrm{real,imag}\}} \|\tilde{\mathbf{B}}_p \tilde{\mathbf{B}}_p^{\top} - \mathbf{I}\|^2. \quad (12)$$

Here, the subscript p iterates over the real and imaginary parts of $\mathbf{B}$, while $\mathbf{I}$ is the identity matrix.

Time-Frequency Consistency Loss ($\mathcal{L}_{\mathrm{pred}}$). To ensure accuracy in both domains, we define

$$\mathcal{L}_{\mathrm{pred}} = (1-\beta)\|\mathbf{Y} - \hat{\mathbf{Y}}\|^2 + \beta\|\mathcal{F}(\mathbf{Y}) - \mathcal{F}(\hat{\mathbf{Y}})\|_1, \quad \beta \in [0, 1]. \quad (13)$$

5 Experiment

5.1 Experimental Setups

Datasets. We evaluate TVTS on 17 real-world multivariate time series datasets covering energy, traffic, meteorology, and finance. For standard long-term and short-term forecasting, we use widely adopted benchmarks, including **ETT (ETTh1/h2/m1/m2)**, **Weather**, **Electricity (ECL)**, **Traffic**, **Exchange**, and **PEMS (PEMS03/04/07/08)**. To further assess *DGC* and *TVCF* under higher dimensionality and stronger dynamics, we additionally include five more challenging datasets: **DowJones**, **Wiki**, **METR-LA**, **COVID-19**, and **NASDAQ**. Detailed statistics are reported in Appendix G.1.

Baselines. We compare TVTS with representative state-of-the-art (SOTA) methods across four distinct categories: (i) **Linear-based models:** GTR Cao et al. (2026), TimePro Ma et al. (2025), TFPS Sun et al. (2025), TimeMixer++ Wang et al. (2025b), TimeMixer Wang et al. (2024a), TimesNet Wu et al. (2023), DLinear Zeng et al. (2023); (ii) **Transformer-based models:** DistDF Wang et al. (2026a), CARD Wang et al. (2024b), iTransformer Liu et al. (2024c), Leddam Yu et al. (2024), PatchTST Nie et al. (2023), Crossformer Zhang & Yan (2023); (iii) **Frequency-domain models:** FACT Wang et al. (2026b), FreDF Wang et al. (2025a), FilterTS Wang et al. (2025c), FilterNet Yi et al. (2024), FITS Xu et al. (2024), Fredformer Piao et al. (2024), FreTS Yi et al. (2023). (iv) **Graph-based models:** MSGNet Cai et al. (2024), CrossGNN Huang et al. (2023).

Evaluation Metrics. We evaluate forecasting accuracy using five standard metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (R^2), Pearson Correlation Coefficient (r), and Mean Absolute Scaled Error (MASE). More details are reported in Appendix G.2.

Table 2: Aggregated long-term MTSF performance averaged over four horizons $H \in \{96, 192, 336, 720\}$ with a fixed look-back window of $L = 96$. **Bold** and underlined values indicate the best and second-best results, respectively. Full results are provided in Table 8.

Model	TVTS (Ours)		FACT (2026b)		DistDF (2026a)		GTR (2026)		TimePro (2025)		TFPS (2025)		FreDF (2025a)		FilterTS (2025c)		FilterNet (2024)		iTrans. (2024c)		MSGNet (2024)		CrossGNN (2023)		PatchTST (2023)		DLinear (2023)	
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	MSE	MAE	MSE	MAE
ETTm1	**0.375**	**0.386**	0.385	0.395	0.378	0.394	0.378	0.394	0.391	0.400	0.395	0.406	0.392	0.399	0.385	0.396	0.392	0.401	0.407	0.410	0.398	0.411	0.393	0.404	0.387	0.400	0.403	0.407
ETTm2	**0.270**	**0.315**	0.281	0.323	0.277	0.321	0.277	0.321	0.281	0.326	0.276	0.321	0.278	0.319	0.276	0.321	0.285	0.328	0.288	0.332	0.288	0.330	0.282	0.330	0.281	0.326	0.350	0.401
ETTh1	**0.428**	**0.426**	0.438	0.436	0.430	0.429	0.430	0.429	0.438	0.438	0.448	0.443	0.437	0.435	0.433	0.430	0.441	0.439	0.454	0.447	0.452	0.452	0.437	0.434	0.469	0.454	0.456	0.452
ETTh2	**0.363**	**0.390**	0.382	0.406	0.367	0.393	0.372	0.400	0.377	0.403	0.380	0.403	0.371	0.396	0.372	0.396	0.383	0.407	0.383	0.407	0.396	0.417	0.393	0.418	0.384	0.405	0.559	0.515
ECL	**0.157**	**0.249**	0.162	0.257	0.172	0.267	0.166	0.260	0.169	0.262	0.183	0.280	0.170	0.259	0.180	0.271	0.173	0.268	0.178	0.270	0.194	0.300	0.201	0.300	0.208	0.295	0.212	0.300
Exchange	**0.334**	**0.390**	0.348	0.404	–	–	–	–	0.352	0.399	0.395	0.414	0.387	0.421	0.352	0.397	0.388	0.419	0.360	0.403	0.399	0.430	0.345	0.395	0.367	0.404	0.354	0.414
Traffic	0.428	**0.277**	0.429	0.278	**0.417**	0.279	0.470	0.280	–	–	–	–	0.421	0.279	0.471	0.315	0.463	0.310	0.428	0.282	0.641	0.370	0.583	0.323	0.531	0.343	0.625	0.383
Weather	**0.240**	**0.267**	0.248	0.275	0.248	0.275	0.248	0.275	0.251	0.276	0.241	0.271	0.254	0.274	0.244	0.274	0.245	0.272	0.258	0.279	0.249	0.278	0.247	0.289	0.259	0.281	0.265	0.317
1st count	**7**	**8**	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3: Short-term forecasting performance evaluation on PEMS datasets. The look-back window is fixed at $L = 96$, and metrics are averaged across four prediction horizons $H \in \{12, 24, 48, 96\}$. Other details are consistent with Table 2. See Table 10 for full results.

Dataset	TVTS (Ours)		TimeMix.++ (2025b)		FilterNet (2024)		FITS (2024)		CARD (2024b)		Fredfm. (2024)		TimeMix. (2024a)		PatchTST (2023)		TimesNet (2023)		Crossfm. (2023)		Dlinear (2023)	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS03	**0.120**	**0.226**	0.165	0.263	0.145	0.251	0.489	0.465	0.174	0.275	0.135	0.243	0.167	0.267	0.180	0.291	0.147	0.248	0.169	0.282	0.278	0.375
PEMS04	**0.112**	**0.215**	0.136	0.251	0.146	0.258	0.531	0.489	0.206	0.299	0.162	0.261	0.185	0.287	0.195	0.307	0.129	0.241	0.209	0.314	0.295	0.388
PEMS07	**0.103**	**0.203**	0.152	0.258	0.123	0.229	0.500	0.472	0.149	0.247	0.121	0.222	0.181	0.271	0.211	0.303	0.124	0.225	0.235	0.315	0.329	0.396
PEMS08	**0.145**	**0.229**	0.200	0.279	0.172	0.260	0.534	0.487	0.201	0.280	0.161	0.250	0.226	0.299	0.280	0.321	0.193	0.271	0.268	0.307	0.379	0.416

5.2 Overall Performance

Long-term Forecasting. Table 2 compares TVTS on eight datasets, with detailed results reported in Table 8. TVTS achieves the best performance on **15 of 16** aggregated metrics, ranking first for both MSE and MAE on seven datasets and obtaining the lowest MAE on Traffic. Its advantages over frequency-domain methods, including TFPS, FreDF, FilterTS, and FilterNet, indicate that TVCF better represents position-dependent spectral variations than static frequency modeling. Meanwhile, the gains over variable-interaction methods such as FACT, iTransformer, MSGNet, and CrossGNN suggest that DGC effectively captures directed cross-variable dependencies. Together, these results demonstrate robust forecasting across heterogeneous temporal patterns and variable interactions.

Short-term Forecasting. Table 3 reports performance on high-dimensional traffic datasets (PEMS). TVTS ranks first on all **8 metrics** (detailed statistics in Table 10), demonstrating robust performance under dense sensor settings. Compared with the second-best model (Fredformer), TVTS yields a large gain, with average MSE reductions of 14.9% on PEMS07. These consistent gains at short horizons suggest that TVTS is effective not only for long-range trend extrapolation, but also for short-term forecasting where performance hinges on capturing rapid local changes.

We further provide comprehensive comparisons on **additional datasets** (Appendix H.1.3), **univariate** forecasting (Appendix H.1.4), **LLM-based** comparisons (Appendix H.2), and **few-shot** evaluations using **10%** and **5%** training data (Appendix H.3), further validating the effectiveness of TVTS.

5.3 Ablation Studies

We evaluate nine variants to isolate each component of TVTS: (i) **w/o DGC/TVCF/FDRP** removes the corresponding module; (ii) **Full Attention** replaces topology-guided attention with dense all-to-all attention; (iii) **Random-M** replaces the learned topology with a random matrix; (iv) **Based-H** replaces the time-varying filter with a static filter; (v) **Dual-A** replaces the directed topology with its bidirectional counterpart; and (vi) **w/o $\mathcal{L}_F$/$\mathcal{L}_O$** removes the frequency-domain or orthogonal loss.

Table 4: Average ablation results with $L = 96$ across the forecast horizons $H \in \{96, 192, 336, 720\}$ on seven datasets. Please refer to Appendix I for more detailed experimental results.

Model	TVTS		w/o FDRP		w/o TVCF		w/o DGC		Full attention		Random-$\mathbf{M}$		Based-$\mathbf{H}$		Dual-$\mathbf{A}$		w/o $\mathcal{L}_{\mathcal{F}}$		w/o $\mathcal{L}_{\mathcal{O}}$	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.375**	**0.386**	0.378	0.389	0.380	0.393	0.378	0.393	0.379	0.392	0.379	0.393	0.377	0.388	0.376	0.391	0.379	0.395	0.379	0.393
ETTm2	**0.270**	**0.315**	0.272	0.317	0.276	0.319	0.274	0.319	0.273	0.317	0.272	0.316	0.272	0.316	0.270	0.316	0.276	0.320	0.273	0.316
ETTh1	**0.428**	**0.426**	0.436	0.431	0.441	0.435	0.439	0.432	0.440	0.433	0.461	0.442	0.433	0.427	0.429	0.427	0.445	0.436	0.442	0.434
ETTh2	**0.363**	**0.390**	0.367	0.392	0.371	0.396	0.373	0.398	0.367	0.395	0.377	0.401	0.366	0.392	0.366	0.391	0.376	0.401	0.375	0.400
ECL	**0.157**	**0.249**	0.160	0.254	0.172	0.263	0.184	0.269	0.175	0.263	0.178	0.269	0.161	0.254	0.162	0.253	0.169	0.262	0.162	0.253
Exchange	**0.334**	**0.390**	0.349	0.401	0.364	0.408	0.359	0.403	0.357	0.404	0.363	0.407	0.352	0.401	0.355	0.399	0.357	0.405	0.347	0.393
Weather	**0.240**	**0.267**	0.244	0.269	0.248	0.274	0.246	0.270	0.249	0.272	0.247	0.273	0.244	0.270	0.241	0.269	0.248	0.274	0.246	0.270

Temporal Modeling. As shown in Table 4, removing TVCF (w/o TVCF) consistently increases forecasting errors, particularly on ETTh1 and Exchange. Replacing its time-varying filter with Based-$\mathbf{H}$ also degrades performance, confirming that the gains arise from position-varying responses rather than spectral filtering alone. The consistent decline without FDRP (w/o FDRP) further demonstrates the benefit of iterative frequency-domain residual refinement.

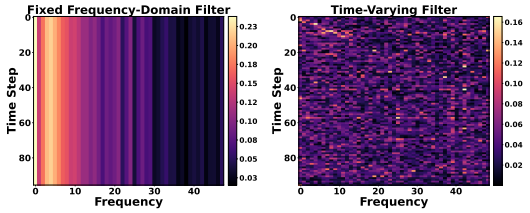
Directed Topology. Removing DGC (w/o DGC) causes substantial degradation, including a 17.2% MSE increase on ECL. Full Attention and Random-$\mathbf{M}$ perform worse than TVTS, showing that dense or unstructured interactions introduce irrelevant cross-variable information. Moreover, Dual-$\mathbf{A}$ remains inferior to the directed formulation, especially on ECL and Exchange, validating the importance of modeling asymmetric dependencies rather than merely preserving graph connectivity.

Auxiliary Objectives. Removing either $\mathcal{L}_{\mathcal{F}}$ or $\mathcal{L}_{\mathcal{O}}$ consistently reduces accuracy, supporting their roles in maintaining time-frequency consistency and spectral basis diversity.

5.4 Visualization Analysis

Time-Varying Spectral Analysis.

Figure 3 compares the **Fixed** Yi et al. (2024) and **Time-Varying (TV)** filters learned on Weather. The fixed filter in Figure 3 (a) exhibits nearly uniform vertical patterns across timestamps, reflecting a time-invariant spectral response. In contrast, the TV filter in Figure 3 (b) shows clear position-dependent variations, with its spectral emphasis continuously evolving over time. This comparison provides intuitive evidence that TVCF models position-dependent spectral variations beyond static spectral filtering.



(a) Fixed Filter (b) TV Filter

Figure 3: Learned graph spectral filters on the Weather dataset. The horizontal and vertical axes denote spectral bands and time steps, respectively. Additional results are provided in Appendix J.1.

Cross-Variable Dependencies. Figure 4 contrasts the static cosine similarity in sensor space with the learned DGC mask attention. The similarity matrices (left) are dense and strictly symmetric, which makes dependencies appear inherently mutual (e.g., DUET and BIM[3]). In contrast, the DGC mask (right) is sparse and clearly asymmetric, indicating that the model does not simply mirror undirected similarity but instead learns an anisotropic dependency structure. For example, sensors C251 and C256 exhibit high symmetric similarity (0.9) in the sensor space, yet DGC assigns a much stronger weight from C251 to C256 (0.3) than in the reverse direction (0.1). This discrepancy suggests that DGC can disentangle non-reciprocal interactions and translate static correlation patterns into directed dependencies that are more suitable for modeling asymmetric cross-sensor influence in real-world systems.

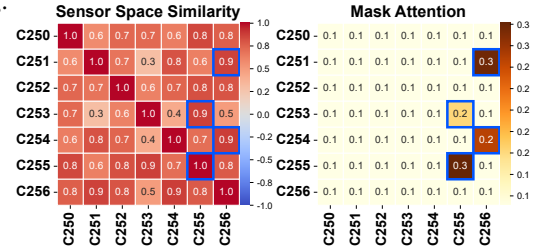
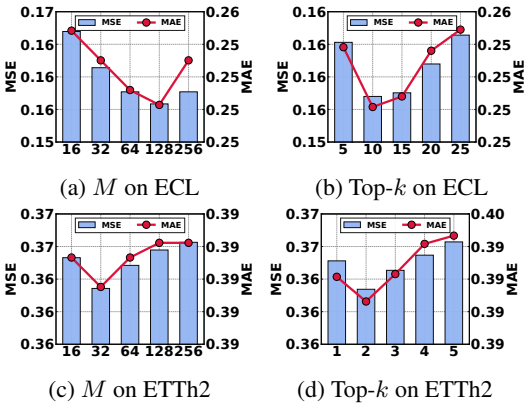


Figure 4: Topological visualization of learned dependencies on PEMS07. The left panel shows static sensor-space similarity, while the right panel presents the learned Geometric Mask $\mathbf{M}$. Additional results are provided in Appendix J.2.

Parameter Sensitivity. Figure 5 analyzes the sensitivity to spectral basis M and sparsity k.

- **Impact of Spectral Basis Size M.** Performance follows a convex trajectory (Figure 5 (a, c)). Increasing M enhances resolution to more effectively capture evolving temporal patterns, yet excessive values (e.g., $M > 128$) introduce spectral noise. This confirms that a balanced basis count is critical for modeling temporal variations without overfitting.
- **Impact of Sparsity Top-k Neighbors.** Optimal k correlates positively with dimensionality (Figure 5 (b, d)). While high-dimensional data (ECL) requires broader receptive fields ($k = 10$), low-dimensional ones (ETTh2) favor strict sparsity ($k = 2$). This validates that DGC effectively prunes the redundancy of symmetric "all-to-all" assumptions, constructing a precise anisotropic topology tailored to the dataset's intrinsic structural complexity.



(a) M on ECL (b) Top-k on ECL

(c) M on ETTh2 (d) Top-k on ETTh2

Figure 5: Sensitivity analysis of basis size M and sparsity Top-k. Results are averaged over $H \in \{96, 192, 336, 720\}$ with look-back length $L = 96$. Additional results are provided in Appendix M.

5.5 Complexity and Efficiency Analysis

Figure 6 benchmarks computational viability. Attributable to its streamlined spectral architecture, TVTS demonstrates an optimal memory profile (**0.44G**) and rapid convergence (**33ms/iter**). It significantly outperforms computationally intensive Transformers (e.g., Crossformer) while matching the efficiency of lightweight linear baselines (e.g., FilterTS). Crucially, this efficiency does not compromise fidelity as TVTS simultaneously attains the lowest MSE, validating that our framework strikes a superior balance between computational cost and predictive performance for resource-constrained scenarios. Detailed complexity analysis and metrics are provided in Appendix F and Table 23.

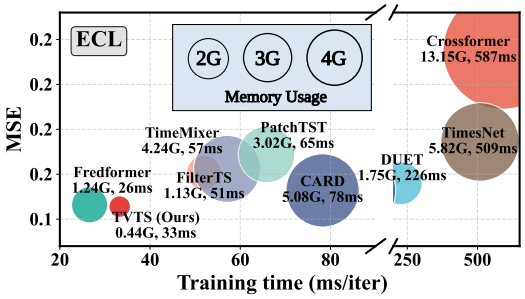


Figure 6: Model efficiency comparison on ECL. The horizontal axis is training time, while bubble size represents training GPU memory.

6 Limitation

TVTS introduces moderate training overhead due to repeated spectral transformations and position-specific filter generation, although its inference efficiency remains high, as shown in Table 23. Moreover, TVCF generates filters based on temporal positions rather than individual sequence content, which may limit its adaptability to highly irregular instance-specific dynamics. Finally, the asymmetric topology learned by DGC provides a useful approximation of directional dependencies, while a rigorous causal interpretation may require additional assumptions and interventional validation.

7 Conclusion

In this work, we proposed TVTS to address the challenges of evolving dynamics and directed dependencies in multivariate time series forecasting. Unlike existing methods relying on static filters and symmetric attention, TVTS integrates a Time-Varying Convolutional Filter to capture heterogeneous temporal patterns and a Directed Geometric Clustering module to model anisotropic variable interactions via sparse topology learning. Complementing these core designs, the framework further incorporates topology-guided fusion and frequency-domain residual prediction to rigorously harmonize spatial-temporal representations and iterative forecasting. Comprehensive experiments confirm that TVTS significantly outperforms SOTA baselines across various benchmarks. By effectively unifying time-varying filtering with directed geometric modeling, our approach offers a precise and efficient paradigm for capturing the intrinsic structure of complex multivariate systems.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language polishing, restructuring selected passages, and improving the visual presentation of figures and tables. We did not use generative AI to generate experimental data, fabricate results, conduct statistical analyses, or determine the scientific conclusions of this work. The model design, implementation, experiments, and analyses were developed and verified by the authors. All AI-assisted text and visual materials were manually reviewed against the original technical content, while all equations, citations, tables, and reported results were independently checked for correctness. We take full responsibility for the final content of this work, including all text, claims, results, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies multivariate time-series forecasting using publicly available benchmark datasets and does not involve human subjects, personally identifiable information, or the collection of sensitive data. We follow the licenses and intended research uses of all datasets and baseline implementations. Although forecasting models may be applied in high-impact domains, such as energy, transportation, and finance, the results reported in this work are intended for methodological research and should not be used as the sole basis for safety-critical or consequential decisions without appropriate domain-specific validation.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the TVTS architecture, mathematical formulations, learning objectives, and overall pseudocode in the main paper and appendix. The experimental section reports dataset configurations, forecasting horizons, evaluation metrics, baseline settings, and implementation details. The appendix further includes complete results for all prediction horizons, hyperparameter analyses, ablation studies, computational-efficiency comparisons, repeated experiments over five random seeds, and statistical significance tests. These materials are intended to support independent verification and reproduction of the reported results.

REFERENCES

Wanlin Cai, Yuxuan Liang, Xianggen Liu, Jianshuai Feng, and Yuankai Wu. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11141–11149, 2024.

Defu Cao, Furong Jia, Sercan O Arik, Tomas Pfister, Yixiang Zheng, Wen Ye, and Yan Liu. TEMPO: Prompt-based generative pre-trained transformer for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Fanpu Cao, Lu Dai, Jindong Han, and Hui Xiong. Enhancing multivariate time series forecasting with global temporal retrieval. In *The Fourteenth International Conference on Learning Representations*, 2026.

Li-Juan Cao and Francis Eng Hock Tay. Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on neural networks*, 14(6):1506–1518, 2003.

Cristian Challu, Kin G Olivares, Boris N Oreshkin, Federico Garza Ramirez, Max Mergenthaler Canseco, and Artur Dubrawski. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 6989–6997, 2023.

Ching Chang, Wei-Yao Wang, Wen-Chih Peng, and Tien-Fu Chen. Llm4ts: Aligning pre-trained llms as data-efficient time-series forecasters. *ACM Transactions on Intelligent Systems and Technology*, 16(3):1–20, 2025.

Ji Chen, Zhihao Ying, Chonghui Zhang, and Tomas Balezentis. Forecasting tourism demand with search engine data: A hybrid cnn-bilstm model based on boruta feature selection. *Information Processing & Management*, 61(3):103699, 2024.

Yuzhou Chen, Ignacio Segovia-Dominguez, Baris Coskunuzer, and Yulia Gel. Tamp-s2gcnets: coupling time-aware multipersistence knowledge representation with spatio-supra graph convolutional networks for time-series forecasting. In *International conference on learning representations*, 2022.

Razvan-Gabriel Cirstea, Bin Yang, Chenjuan Guo, Tung Kieu, and Shirui Pan. Towards spatio-temporal aware traffic time series forecasting. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pp. 2900–2913. IEEE, 2022.

Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017.

Fernando Gama, Brendon G Anderson, and Somayeh Sojoudi. Node-variant graph filters in graph neural networks. In *2022 IEEE Data Science and Learning Workshop (DSLW)*, pp. 1–6. IEEE, 2022.

Yifan Gao, Boming Zhao, Haocheng Peng, Hujun Bao, Jiashu Zhao, and Zhaopeng Cui. Bidirectional temporal-aware modeling with multi-scale mixture-of-experts for multivariate time series forecasting. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 696–706, 2025.

Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pp. 424–438, 1969.

Chenjuan Guo, Bin Yang, Ove Andersen, Christian S Jensen, and Kristian Torp. Ecomark 2.0: empowering eco-routing with vehicular environmental models and actual vehicle fuel consumption data. *GeoInformatica*, 19(3):567–599, 2015.

Lu Han, Han-Jia Ye, and De-Chuan Zhan. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7129–7142, 2024.

Peng He, Yao Liu, Yanglei Gan, Run Lin, Yuxiang Cai, and Qiao Liu. Fait: Frequency-aware inverted transformer for multivariate time series forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 1614–1625, 2026.

Yifan Hu, Guibin Zhang, Peiyuan Liu, Disen Lan, Naiqi Li, Dawei Cheng, Tao Dai, Shu-Tao Xia, and Shirui Pan. Timefilter: Patch-specific spatial-temporal graph filtration for time series forecasting. In *International Conference on Machine Learning*, pp. 24893–24911. PMLR, 2025.

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

Elvin Isufi, Fernando Gama, David I Shuman, and Santiago Segarra. Graph filters for signal processing and machine learning on graphs. *IEEE Transactions on Signal Processing*, 72:4745–4781, 2024.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-fang Li, Shirui Pan, et al. Time-llm: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations*, 2024a.

Ming Jin, Shiyu Wang, Lintao Ma, Zhixuan Chu, James Y Zhang, Xiaoming Shi, Pin-Yu Chen, Yuxuan Liang, Yuan-Fang Li, Shirui Pan, and Qingsong Wen. Time-LLM: Time series forecasting by reprogramming large language models. In *International Conference on Learning Representations (ICLR)*, 2024b.

Zahra Karevan and Johan AK Suykens. Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9, 2020.

Taesung Kim, Jinhee Kim, Yunwon Tae, Cheonbok Park, Jang-Ho Choi, and Jaegul Choo. Reversible instance normalization for accurate time-series forecasting against distribution shift. In *International conference on learning representations*, 2021.

Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In *International Conference on Learning Representations*, 2020.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pp. 95–104, 2018.

Shiyong Lan, Yitong Ma, Weikang Huang, Wenwu Wang, Hongyu Yang, and Pyang Li. Dstagnn: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting. In *International conference on machine learning*, pp. 11906–11917. PMLR, 2022.

Marco Lippi, Matteo Bertini, and Paolo Frasconi. Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):871–882, 2013.

Chenxi Liu, Qianxiong Xu, Hao Miao, Sun Yang, Lingzheng Zhang, Cheng Long, Ziyue Li, and Rui Zhao. Timecma: Towards llm-empowered multivariate time series forecasting via cross-modality alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18780–18788, 2025a.

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

Peiyuan Liu, Beiliang Wu, Yifan Hu, Naiqi Li, Tao Dai, Jigang Bao, and Shu-Tao Xia. Timebridge: Non-stationarity matters for long-term time series forecasting. *International Conference on Machine Learning*, 2025b.

Qinshuo Liu, Yanwen Fang, Pengtao Jiang, and Guodong Li. Dgcformer: Deep graph clustering transformer for multivariate time series forecasting. *arXiv preprint arXiv:2405.08440*, 2024a.

Xu Liu, Junfeng Hu, Yuan Li, Shizhe Diao, Yuxuan Liang, Bryan Hooi, and Roger Zimmermann. Unitime: A language-empowered unified model for cross-domain time series forecasting. In *Proceedings of the ACM Web Conference 2024*, pp. 4095–4106, 2024b.

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

Yong Liu, Chenyu Li, Jianmin Wang, and Mingsheng Long. Koopa: Learning non-stationary time series dynamics with koopman predictors. *Advances in neural information processing systems*, 36: 12271–12290, 2023.

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

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

Xiaowen Ma, Zhen-Liang Ni, Shuai Xiao, and Xinghao Chen. Timepro: Efficient multivariate long-term time series forecasting with variable- and time-aware hyper-state. In *Forty-second International Conference on Machine Learning*, 2025.

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

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-beats: Neural basis expansion analysis for interpretable time series forecasting. In *International Conference on Learning Representations*, 2020.

Xihao Piao, Zheng Chen, Taichi Murayama, Yasuko Matsubara, and Yasushi Sakurai. Fredformer: Frequency debiased transformer for time series forecasting. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 2400–2410, 2024.

Xiangfei Qiu, Jilin Hu, Lekui Zhou, Xingjian Wu, Junyang Du, Buang Zhang, Chenjuan Guo, Aoying Zhou, Christian S Jensen, Zhenli Sheng, et al. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *Proceedings of the VLDB Endowment*, 17(9):2363–2377, 2024.

Xiangfei Qiu, Xingjian Wu, Yan Lin, Chenjuan Guo, Jilin Hu, and Bin Yang. Duet: Dual clustering enhanced multivariate time series forecasting. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp. 1185–1196, 2025.

Xiangfei Qiu, Kangjia Yan, Xvyuan Liu, Xingjian Wu, and Jilin Hu. Bridging time and frequency: A joint modeling framework for irregular multivariate time series forecasting. In *Forty-third International Conference on Machine Learning*, 2026.

Santiago Segarra, Antonio G Marques, and Alejandro Ribeiro. Optimal graph-filter design and applications to distributed linear network operators. *IEEE Transactions on Signal Processing*, 65 (15):4117–4131, 2017.

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

Lee M Seversky, Shelby Davis, and Matthew Berger. On time-series topological data analysis: New data and opportunities. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 59–67, 2016.

Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing*, 90:106181, 2020.

Zongjiang Shang, Ling Chen, Binqing Wu, and Dongliang Cui. Ada-mshyper: adaptive multi-scale hypergraph transformer for time series forecasting. *Advances in Neural Information Processing Systems*, 37:33310–33337, 2024.

Yehjin Shin, Jeongwhan Choi, Hyowon Wi, and Noseong Park. An attentive inductive bias for sequential recommendation beyond the self-attention. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8984–8992, 2024.

Yehjin Shin, Jeongwhan Choi, Seojin Kim, and Noseong Park. TV-rec: Time-variant convolutional filter for sequential recommendation. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Ljubiša Stanković, Miloš Daković, Ali Bagheri Bardi, Miloš Brajović, and Isidora Stanković. Fourier analysis of signals on directed acyclic graphs (dag) using graph zero-padding. *Digital Signal Processing*, 159:104995, 2025.

Yanru Sun, Zongxia Xie, Emadeldeen Eldele, Dongyue Chen, Qinghua Hu, and Min Wu. Learning pattern-specific experts for time series forecasting under patch-level distribution shift. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Peiwang Tang and Weitai Zhang. Unlocking the power of patch: Patch-based mlp for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12640–12648, 2025.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010, 2017.

Hao Wang, Lichen Pan, Yuan Shen, Zhichao Chen, Degui Yang, Yifei Yang, Sen Zhang, Xinggao Liu, Haoxuan Li, and Dacheng Tao. FreDF: Learning to forecast in the frequency domain. In *The Thirteenth International Conference on Learning Representations*, 2025a.

Hao Wang, Licheng Pan, Yuan Lu, Zhixuan Chu, Xiaoxi Li, Shuting He, Zi Chan, Qingsong Wen, Haoxuan Li, and Zhouchen Lin. Distdf: Time-series forecasting needs joint-distribution wasserstein alignment. In *International Conference on Learning Representations*, volume 2026, pp. 131512–131535, 2026a.

Huiqiang Wang, Jieming Shi, and Li Qing. FACT: Fine-grained across-variable convolution for multivariate time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026b.

Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. In *International Conference on Learning Representations*, 2022.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y Zhang, and JUN ZHOU. Timemixer: Decomposable multiscale mixing for time series forecasting. In *International Conference on Learning Representations (ICLR)*, 2024a.

Shiyu Wang, Jiawei LI, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Ju Shengtong, Zhixuan Chu, and Ming Jin. Timemixer++: A general time series pattern machine for universal predictive analysis. In *The Thirteenth International Conference on Learning Representations*, 2025b.

Xue Wang, Tian Zhou, Qingsong Wen, Jinyang Gao, Bolin Ding, and Rong Jin. CARD: Channel aligned robust blend transformer for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024b.

Yulong Wang, Yushuo Liu, Xiaoyi Duan, and Kai Wang. Filterts: Comprehensive frequency filtering for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21375–21383, 2025c.

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

Yuxuan Wang, Haixu Wu, Jiaxiang Dong, Yong Liu, Chen Wang, Mingsheng Long, and Jianmin Wang. Deep time series models: A comprehensive survey and benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026c.

Bernard L Welch. The generalization of 'student's'problem when several different population varlances are involved. *Biometrika*, 34(1-2):28–35, 1947.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. ETSformer: Exponential smoothing transformers for time-series forecasting, 2023.

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

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

Xingjian Wu, Xiangfei Qiu, Zhengyu Li, Yihang Wang, Jilin Hu, Chenjuan Guo, Hui Xiong, and Bin Yang. Catch: Channel-aware multivariate time series anomaly detection via frequency patching. In *International conference on learning representations*, volume 2025, pp. 17017–17045, 2025.

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

Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Lian, Longbing Cao, and Zhendong Niu. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36:76656–76679, 2023.

Kun Yi, Jingru Fei, Qi Zhang, Hui He, Shufeng Hao, Defu Lian, and Wei Fan. Filternet: Harnessing frequency filters for time series forecasting. *Advances in Neural Information Processing Systems*, 37:55115–55140, 2024.

Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pp. 3634–3640. International Joint Conferences on Artificial Intelligence Organization, 2018.

Guoqi Yu, Jing Zou, Xiaowei Hu, Angelica I Aviles-Rivero, Jing Qin, and Shujun Wang. Revitalizing multivariate time series forecasting: Learnable decomposition with inter-series dependencies and intra-series variations modeling. In *International Conference on Machine Learning*, pp. 57818–57841. PMLR, 2024.

Wenzhen Yue, Yong Liu, Hao Wang, Haoxuan Li, Xianghua Ying, Ruohao Guo, Bowei Xing, and Ji Shi. OLinear: A linear model for time series forecasting in orthogonally transformed domain. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a.

Wenzhen Yue, Yong Liu, Xianghua Ying, Bowei Xing, Ruohao Guo, and Ji Shi. Freeformer: frequency enhanced transformer for multivariate time series forecasting. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, IJCAI '25, 2025b. ISBN 978-1-956792-06-5. doi: 10.24963/ijcai.2025/401.

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

T. Zhang, Yizhuo Zhang, Wei Cao, J. Bian, Xiaohan Yi, Shun Zheng, and Jian Li. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. *ArXiv*, abs/2207.01186, 2022.

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

Zhiwei Zhang, Xinyi Du, Xuanchi Guo, Weihao Wang, and Wenjuan Han. Emaformer: enhancing transformer through embedding armor for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28636–28644, 2026.

Kai Zhao, Chenjuan Guo, Yunyao Cheng, Peng Han, Miao Zhang, and Bin Yang. Multiple time series forecasting with dynamic graph modeling. *Proceedings of the VLDB Endowment*, 17(4): 753–765, 2023.

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

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

Tian Zhou, Peisong Niu, Liang Sun, Rong Jin, et al. One fits all: Power general time series analysis by pretrained lm. *Advances in neural information processing systems*, 36:43322–43355, 2023.

A Proof of Spectral Representation for Time-Varying Convolutional Filters

In this section, we provide a detailed proof of the Spectral Representation for Time-Varying Convolutional Filters (TVCF), corresponding to Eq. 2 of the main text. Our derivation references Gama et al. (2022); Isufi et al. (2024).

Theorem 1. *Given a graph shift operator $\mathbf{S} \in \mathbb{R}^{L\times L}$ with eigendecomposition $\mathbf{S} = \mathbf{\Phi}\mathbf{D}\mathbf{\Phi}^{-1}$, a Time-Varying Convolutional filter defined by the filter tap matrix $\mathbf{W} \in \mathbb{R}^{L\times(K+1)}$ operates on an input signal $\mathbf{x}$ as:*

$$\mathbf{y} = \sum_{k=0}^{K} \text{diag}(\mathbf{w}_k)\mathbf{S}^k\mathbf{x}. \tag{14}$$

This operation is mathematically equivalent to the spectral formulation:

$$\mathbf{y} = \left(\mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)\right)\tilde{\mathbf{x}}, \tag{15}$$

where $\odot$ denotes the Hadamard product, $\tilde{\mathbf{x}} = \mathbf{\Phi}^{-1}\mathbf{x}$ is the spectral input, and $\mathbf{\Lambda} \in \mathbb{C}^{L\times(K+1)}$ is a Vandermonde matrix with entries $\mathbf{\Lambda}_{j,k} = \lambda_j^k$ (where λ_j is the j-th eigenvalue in $\mathbf{D}$).

Proof. We begin by expanding the definition of the output signal $\mathbf{y}$ in the vertex domain. By substituting the spectral decomposition $\mathbf{S}^k = \mathbf{\Phi}\mathbf{D}^k\mathbf{\Phi}^{-1}$ into the filter definition, we obtain:

$$\mathbf{y} = \sum_{k=0}^{K} \text{diag}(\mathbf{w}_k)\mathbf{\Phi}\mathbf{D}^k \underbrace{\mathbf{\Phi}^{-1}\mathbf{x}}_{\tilde{\mathbf{x}}} = \sum_{k=0}^{K} \text{diag}(\mathbf{w}_k)\mathbf{\Phi}\mathbf{D}^k\tilde{\mathbf{x}}. \tag{16}$$

Since matrix operations involving diagonal matrices and Hadamard products are best analyzed element-wise, we proceed by examining the t-th entry of the output vector $\mathbf{y}$, denoted as y_t. Let $\phi_{t,j}$ represent the element of the basis matrix $\mathbf{\Phi}$ at row t and column j. The term $\mathbf{D}^k\tilde{\mathbf{x}}$ scales the j-th frequency component $\tilde{x}_j$ by the eigenvalue power λ_j^k. Thus, the matrix-vector multiplication for the t-th element can be written as:

$$y_t = \sum_{k=0}^{K} w_{t,k} \left(\sum_{j=0}^{L-1} \phi_{t,j}\lambda_j^k\tilde{x}_j \right), \tag{17}$$

where $w_{t,k}$ corresponds to the filter weight at time step t for the k-th shift order (i.e., the entry $\mathbf{W}_{t,k}$).

Assuming finite dimensions, we interchange the order of summation in Eq. (17) to group terms related to the frequency index j:

$$y_t = \sum_{j=0}^{L-1} \phi_{t,j}\tilde{x}_j \left(\sum_{k=0}^{K} w_{t,k}\lambda_j^k \right). \tag{18}$$

We now focus on the inner summation term $\sum_{k=0}^{K} w_{t,k}\lambda_j^k$. Recall the definition of the Vandermonde matrix $\mathbf{\Lambda}$, where $\mathbf{\Lambda}_{j,k} = \lambda_j^k$. The summation can be interpreted as the dot product between the t-th row of the filter matrix $\mathbf{W}$ and the j-th column of the transposed Vandermonde matrix $\mathbf{\Lambda}^\top$:

$$\sum_{k=0}^{K} \mathbf{W}_{t,k}(\mathbf{\Lambda}^\top)_{k,j} = (\mathbf{W}\mathbf{\Lambda}^\top)_{t,j}. \tag{19}$$

Substituting this result back into the expression for y_t, we have:

$$y_t = \sum_{j=0}^{L-1} \underbrace{\phi_{t,j}}_{\mathbf{\Phi}_{t,j}} \underbrace{(\mathbf{W}\mathbf{\Lambda}^\top)_{t,j}}_{\text{Modulation}} \tilde{x}_j. \tag{20}$$

The term $\phi_{t,j}(\mathbf{W}\mathbf{\Lambda}^\top)_{t,j}$ is precisely the definition of the element at (t, j) of the Hadamard product $\mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)$. Let $\mathbf{H}_{TV} = \mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)$ be the Time-Varying spectral kernel. The equation for y_t simplifies to:

$$y_t = \sum_{j=0}^{L-1} (\mathbf{H}_{TV})_{t,j}\tilde{x}_j. \tag{21}$$

In matrix notation, this summation corresponds to the matrix-vector product:

$$\mathbf{y} = \mathbf{H}_{TV}\tilde{\mathbf{x}} = \left(\mathbf{\Phi} \odot (\mathbf{W}\mathbf{\Lambda}^\top)\right)\tilde{\mathbf{x}}. \tag{22}$$

This completes the proof, demonstrating the equivalence between the vertex-domain Time-Varying recursion and the proposed spectral formulation. □

Expressive Generalization over Static Filtering. Beyond the spectral equivalence above, we further characterize the expressive relation between TVCF and conventional shift-invariant filtering.

Proposition 1 (TVCF Generalizes Static (Fixed) Filtering). *Let* $\mathbf{S}$ *be the directed cyclic shift operator. Define the static and time-varying filter families as*

$$\mathcal{F}_{\text{static}} = \left\{ \sum_{k=0}^{K} h_k \mathbf{S}^k \right\}, \qquad \mathcal{F}_{\text{TVCF}} = \left\{ \sum_{k=0}^{K} \text{diag}(\mathbf{w}_k)\mathbf{S}^k \right\}. \tag{23}$$

Then $\mathcal{F}_{\text{static}} \subsetneq \mathcal{F}_{\text{TVCF}}$.

Proof. For any static filter, setting $\mathbf{w}_k = h_k\mathbf{1}$ yields $\text{diag}(\mathbf{w}_k) = h_k\mathbf{I}$, recovering $\sum_{k=0}^{K} h_k\mathbf{S}^k$. Hence, $\mathcal{F}_{\text{static}} \subseteq \mathcal{F}_{\text{TVCF}}$. To establish strict inclusion, consider a nonconstant $\mathbf{w}_0$ and $\mathbf{w}_k = \mathbf{0}$ for $k > 0$. The resulting operator $\text{diag}(\mathbf{w}_0)$ generally does not commute with $\mathbf{S}$, whereas every static polynomial filter does. Therefore, it cannot be represented by any static shift-invariant filter. □

This establishes that TVCF enlarges the filtering function class by relaxing temporal shift invariance while retaining conventional static filtering as a special case.

B Spectral Foundation of the Directed Cyclic Graph

In this section, we provide the mathematical rationale for modeling MTSF on a DCG combined with a zero-padding mechanism. We demonstrate that this topological structure satisfies two critical requirements for efficient Time-Varying Convolutional Filter: (1) **Spectral Diagonalizability**, which enables the use of the Fast Fourier Transform, and (2) **Mitigation of Circular Boundary Artifacts**, which preserves the temporal causality of the sequence.

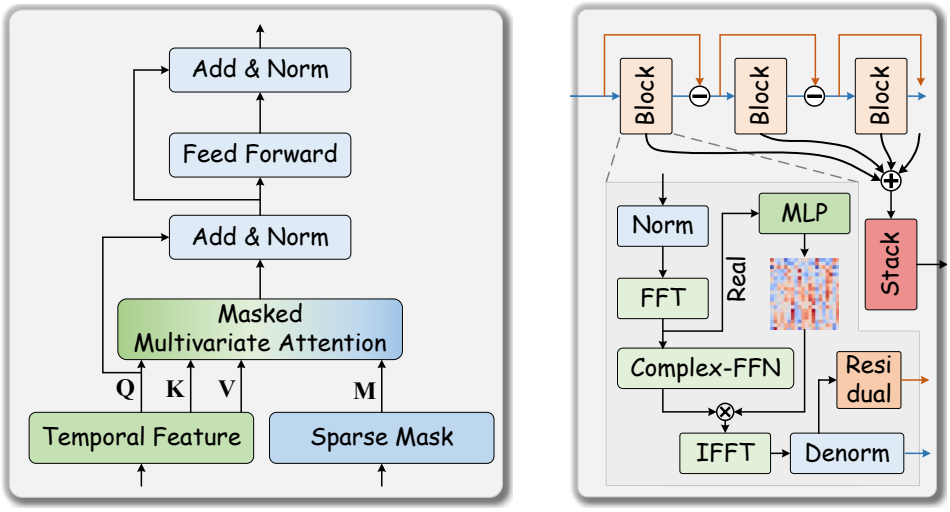
1. Spectral Diagonalizability. Standard time series modeling often assumes a line graph structure. However, the adjacency matrix of a line graph is not diagonalizable by a unified, computationally efficient basis. In contrast, consider a DGC $\mathcal{G}_{cyc}$ with L_{pad} nodes. Its graph shift operator (adjacency matrix), denoted as $\mathbf{S}_{cyc} \in \{0,1\}^{L_{pad}\times L_{pad}}$, is a circulant matrix:

$$\mathbf{S}_{cyc} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}. \tag{24}$$

A fundamental property of circulant matrices is that they are diagonalized by the Discrete Fourier Transform (DFT) matrix $\mathbf{U}$. Specifically, $\mathbf{S}_{cyc} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U}_{k,n} = \frac{1}{\sqrt{L_{pad}}} e^{-j\frac{2\pi kn}{L_{pad}}}$. This property is pivotal for our TVCF (Eq. 4), as it allows the graph convolution to be performed via element-wise multiplication in the frequency domain with complexity $\mathcal{O}(L_{pad}\log L_{pad})$, avoiding the prohibitive $\mathcal{O}(L_{pad}^3)$ cost of general graph eigendecomposition.

2. Mitigation of Circular Boundary Artifacts. While the cyclic topology enables spectral efficiency, it introduces a "wrap-around" artifact where the last time step $x_{L_{pad}}$ influences the first time step x_1 (due to the entry $\mathbf{S}_{cyc}[1, L_{pad}] = 1$). This violates the temporal properties of non-periodic sequences. We resolve this by introducing the zero-padding strategy Shin et al. (2025); Stanković et al. (2025). Let the original signal be $\mathbf{x} = [x_1, x_2, \dots, x_L]^\top$. We construct a padded signal $\mathbf{x}_{pad} \in \mathbb{R}^{L_{pad}}$ where $L_{pad} \geq L + K$ (K being the filter order):

$$\mathbf{x}_{pad} = [x_1, x_2, \dots, x_L, \underbrace{0, \dots, 0}_{P \text{ zeros}}]^\top. \tag{25}$$



(a) Topology-Guided Fusion Module (b) Frequency Domain Residual Prediction

Figure 7: **Detailed architectural schematics.** (a) The Topology-Guided Fusion Module aggregates spatial features using the learned anisotropic mask. (b) The Frequency Domain Residual Prediction (FDRP) Module employs a spectral decomposition strategy to refine forecasting accuracy.

Consider the k-th order graph shift operation $\mathbf{y} = \mathbf{S}_{cyc}^k \mathbf{x}_{pad}$. The operator $\mathbf{S}_{cyc}$ shifts elements "downwards," wrapping the last element to the first position. Since the last P elements of $\mathbf{x}_{pad}$ are zeros, the values wrapped around to the beginning of the sequence are guaranteed to be zero for any shift order $k \leq P$.

Mathematically, this effectively imposes a zero-boundary condition (zero-initialization) for the temporal sequence:

$$(\mathbf{S}_{cyc}^k \mathbf{x}_{pad})_t = \begin{cases} 0 & \text{if } t \leq k \\ x_{t-k} & \text{if } t > k \end{cases} \tag{26}$$

for the valid time steps. By truncating the output $\mathbf{Z}$ to the original length L (Eq. 4), our TVTS framework effectively implements a rigorous linear convolution using the spectral machinery of the cyclic graph. This ensures that the learned spectral representations capture intrinsic temporal dynamics without contamination from artificial circular correlations. □

C Theoretical Analysis of Directed Geometric Clustering

We further characterize the expressive capacity of Directed Geometric Clustering (DGC) relative to conventional symmetric similarity learning. The key distinction arises from the decoupled source and target embeddings, which remove the reciprocity constraint imposed by symmetric similarity.

Proposition 2 (DGC Generalizes Symmetric Similarity). *Let $\mathbf{E}_s, \mathbf{E}_t \in \mathbb{R}^{N \times d}$ denote the row-normalized source and target embeddings, and define the DGC affinity as*

$$\mathbf{A} = \tanh\left(\alpha \mathbf{E}_s \mathbf{E}_t^\top\right). \tag{27}$$

Conventional symmetric cosine similarity is a strict special case of the DGC affinity family.

Proof. When $\mathbf{E}_s = \mathbf{E}_t = \mathbf{E}$, the pre-activation matrix becomes $\mathbf{E}\mathbf{E}^\top$, which is symmetric. Since $\tanh(\cdot)$ is applied element-wise, the resulting affinity matrix remains symmetric. Thus, conventional symmetric similarity is contained in the DGC affinity family.

Conversely, decoupled source and target embeddings generally satisfy

$$\mathbf{E}_s \mathbf{E}_t^\top \neq \mathbf{E}_t \mathbf{E}_s^\top,$$

allowing $A_{ij} \neq A_{ji}$. Such asymmetric affinities cannot be represented by a symmetric Gram-based similarity matrix, establishing strict inclusion. □

Therefore, DGC strictly extends symmetric similarity by relaxing the reciprocity constraint while retaining symmetric interactions as a special case. This provides the representational capacity required to model non-reciprocal predictive dependencies.

D Overall Forward Procedure of TVTS

To complement the mathematical formulations in Sections 4.2–4.4, we summarize the overall forward procedure of TVTS in Algorithm 1. The algorithm consists of four stages. First, the *TVCF* extracts dynamic temporal representations by performing position-specific spectral filtering on a padded directed cyclic graph. Second, *DGC* learns an asymmetric and sparse topology that captures non-reciprocal inter-variable dependencies. Third, the *TGF* injects the learned topology into attention to selectively aggregate cross-variable information. Finally, *FDRP* iteratively refines the forecast by decomposing the prediction into reconstruction and future residual components in the frequency.

E Additional Architectural Details

In this section, we provide detailed schematics and functional descriptions of two critical auxiliary components within the TVTS framework, as illustrated in Figure 7.

Topology-Guided Fusion Module. Figure 7 (a) elucidates the internal mechanism of the topology-guided fusion module. This component incorporates a masked multi-head attention mechanism Qiu et al. (2025), serving as a vital bridge between the learned temporal dynamics (Section 4.2) and spatial structure (Section 4.3). By leveraging the sparse mask attention matrix $\mathbf{M}$ generated by the DGC module (Eq. 7), the fusion layer performs selective feature aggregation. It dynamically weights contributions from neighboring nodes based on the learned directional correlations, effectively enhancing the representation of relevant spatial dependencies while suppressing noise from irrelevant sensors. This ensures that the subsequent filtering process operates on spatially contextualized signals.

Frequency Domain Residual Prediction (FDRP) Implementation Details. As summarized in the main text, the FDRP module employs a stack of I blocks to iteratively refine forecasts by capturing global temporal dependencies in the spectral domain. Detailed in Figure 7b, the workflow of the i-th block consists of three mathematically distinct phases:

1. Normalization and Spectral Transformation. To mitigate distribution shifts and stabilize spectral learning, we first apply Instance Normalization to the input residual $\mathbf{R}^{(i-1)} \in \mathbb{R}^{N\times L}$. The time-domain signal is then transformed into the frequency domain using the FFT:

$$\bar{\mathbf{R}} = \frac{\mathbf{R}^{(i-1)} - \mu}{\sqrt{\sigma^2 + \epsilon}}, \quad \mathbf{S}_{in} = \text{FFT}(\bar{\mathbf{R}}) \in \mathbb{C}^{N\times F_{in}}, \tag{28}$$

where $\mu, \sigma \in \mathbb{R}^{N\times 1}$ represent the instance-wise mean and standard deviation, respectively. Since the input is real-valued, we only retain the positive frequency components ($F_{in} = \lfloor L/2 \rfloor + 1$).

2. Complex Upsampling and Dynamic Modulation. We address the dimension gap between the look-back window L and the prediction horizon H via a complex-valued network operating.

Complex Upsampler $\phi(\cdot)$: We design a complex-valued MLP to map the spectrum from F_{in} to $F_{out} = \lfloor (L+H)/2 \rfloor + 1$. The network comprises two complex linear layers interspersed with ComplexReLU activation and ComplexDropout:

$$\mathbf{S}_{up} = \phi(\mathbf{S}_{in}) = \mathbf{W}_{c2}\,(\text{CDropout}(\text{CReLU}(\mathbf{W}_{c1}\mathbf{S}_{in}))) \in \mathbb{C}^{N\times F_{out}}, \tag{29}$$

where $\mathbf{W}_{c1}, \mathbf{W}_{c2}$ are learnable complex weights. The *ComplexReLU* (CReLU) is defined as:

$$\text{CReLU}(z) = \text{ReLU}(\text{Real}(z)) + j \cdot \text{ReLU}(\text{Imag}(z)). \tag{30}$$

Spectral Modulator $\mathbf{W}_{mask}(\cdot)$: To generate a content-aware frequency mask, a real-valued MLP $\mathcal{K}$ processes the magnitude of the input spectrum:

$$\mathbf{W}_{mask} = \text{Sigmoid}\,(\mathcal{K}(|\mathbf{S}_{in}|)) \in \mathbb{R}^{N\times F_{out}}, \quad \tilde{\mathbf{S}} = \mathbf{S}_{up} \odot \mathbf{W}_{mask}. \tag{31}$$

Algorithm 1 Overall Forward Procedure of TVTS

Input: Historical sequence $\mathbf{X} \in \mathbb{R}^{L \times N}$, number of TVCF layers N_{T}, filter order K, basis size M, neighbor budget k, scaling factor α, and number of FDRP blocks I.
Parameters: spectral bases $\{\mathbf{B}^{(\ell)}\}_{\ell=1}^{N_{\mathrm{T}}}$, position-specific coefficients $\{\mathbf{C}^{(\ell)}\}_{\ell=1}^{N_{\mathrm{T}}}$, source/target embeddings $\mathbf{E}_1, \mathbf{E}_2 \in \mathbb{R}^{N \times d}$, projection matrices $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$, spectral modulator $\mathcal{W}$.
Output: Forecast $\hat{\mathbf{Y}} \in \mathbb{R}^{H \times N}$.
$L_{\mathrm{pad}} = L + K - 1, \quad \mathbf{P} = \mathbf{0}_{(K-1)\times N}$ ▷ Define the zero-padding buffer.
$\mathbf{S}_{\mathrm{cyc}} = \mathrm{DirectedCyclicShift}(L_{\mathrm{pad}})$ ▷ Construct the directed cyclic graph.
$(\mathbf{U}, \{\lambda_r\}_{r=1}^{L_{\mathrm{pad}}}) = \mathrm{Eigendecompose}(\mathbf{S}_{\mathrm{cyc}})$ ▷ Obtain the GFT basis.
$\mathbf{\Lambda}_{r,q} = \lambda_r^q, \quad q = 0, \ldots, K$ ▷ Construct the Vandermonde matrix.
Time-Varying Convolutional Filtering (TVCF) (Section 4.2):
for $\ell = 1$ to N_{T} **do**
  $\mathbf{X}_{\mathrm{pad}}^{(\ell)} = \mathrm{Concat}(\mathbf{X}^{(\ell-1)}, \mathbf{P})$ ▷ Extend the sequence.
  $\widetilde{\mathbf{X}}_{\mathrm{pad}}^{(\ell)} = \mathbf{U}^\top \mathbf{X}_{\mathrm{pad}}^{(\ell)}$ ▷ Transform into the graph spectral domain.
  $\overline{\mathbf{B}}_{m,:}^{(\ell)} = \dfrac{\mathbf{B}_{m,:}^{(\ell)}}{\|\mathbf{B}_{m,:}^{(\ell)}\|_2 + \varepsilon}, \quad m = 1, \ldots, M$ ▷ Normalize spectral bases.
  $\mathbf{F}^{(\ell)} = \mathbf{U} \odot \left([\mathbf{C}^{(\ell)} \overline{\mathbf{B}}^{(\ell)}] \mathbf{\Lambda}^\top \right)$ ▷ Form the time-varying spectral kernel.
  $\mathbf{Z}_{\mathrm{pad}}^{(\ell)} = \mathrm{Re}\left(\mathbf{F}^{(\ell)} \widetilde{\mathbf{X}}_{\mathrm{pad}}^{(\ell)} \right)$ ▷ Reconstruct filtered temporal features.
  $\mathbf{Z}^{(\ell)} = \mathcal{T}_L \left(\mathbf{Z}_{\mathrm{pad}}^{(\ell)} \right)$ ▷ Remove the padding buffer.
  $\mathbf{X}^{(\ell)} = \mathrm{LayerNorm}\big(\mathbf{X}^{(\ell-1)} + \mathrm{Dropout}\big(\mathrm{ReLU}(\mathbf{Z}^{(\ell)}) \big) \big)$ ▷ Residual temporal refinement.
end for
$\mathbf{Z} = \mathbf{X}^{(N_{\mathrm{T}})}$
Directed Geometric Clustering (DGC) (Section 4.3):
$\overline{\mathbf{E}}_1 = \mathrm{RowNormalize}(\mathbf{E}_1), \quad \overline{\mathbf{E}}_2 = \mathrm{RowNormalize}(\mathbf{E}_2)$ ▷ Source and target embeddings.
$\mathbf{A} = \tanh\left(\alpha \, \overline{\mathbf{E}}_1 \overline{\mathbf{E}}_2^\top \right)$ ▷ Compute asymmetric affinities.
for $i = 1$ to N **do**
  $\mathcal{I}_i = \text{Top-K}(\mathbf{A}_{i,:}, k)$ ▷ Select outgoing neighbors.
  $\widetilde{\mathbf{A}}_{ij} = \begin{cases} \mathbf{A}_{ij}, & j \in \mathcal{I}_i, \\ 0, & j \notin \mathcal{I}_i, \end{cases}$ ▷ Enforce directed sparsity.
  $\mathbf{M}_{i,:} = \mathrm{Softmax}(\widetilde{\mathbf{A}}_{i,:})$ ▷ Normalize retained dependencies.
end for
Topology-Guided Fusion (TGF) (Section 4.4):
$\mathbf{Z}' = \mathrm{Linear}(\mathbf{Z}^\top)$ ▷ Project temporal features into variable tokens.
$\mathbf{Q} = \mathbf{Z}'\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{Z}'\mathbf{W}_K, \quad \mathbf{V} = \mathbf{Z}'\mathbf{W}_V$ ▷ Construct queries, keys, and values.
$\mathbf{Z}_{\mathrm{TGF}} = \mathrm{Softmax}\left(\dfrac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{D}} + \psi(\mathbf{M}) \right) \mathbf{V}$ ▷ Inject topology as structural bias.
$\hat{\mathbf{Z}} = \mathrm{LayerNorm}(\mathbf{Z}' + \mathrm{FFN}(\mathbf{Z}_{\mathrm{TGF}}))$ ▷ Refine the topology-aware representation.
Frequency-Domain Residual Prediction (FDRP) (Section 4.4):
$\mathbf{R}^{(0)} = \hat{\mathbf{Z}}, \quad \hat{\mathbf{Y}} = \mathbf{0}$ ▷ Initialize residual and forecast.
for $i = 1$ to I **do**
  $\tilde{\mathbf{O}}^{(i)} = \mathcal{F}^{-1}\big(\phi\big(\mathcal{F}(\mathbf{R}^{(i-1)}) \big) \odot \mathcal{W}\big(\mathcal{F}(\mathbf{R}^{(i-1)}) \big) \big)$ ▷ Residual forecasting.
  $[\hat{\mathbf{X}}^{(i)}, \hat{\mathbf{Y}}^{(i)}] = \mathrm{Split}(\tilde{\mathbf{O}}^{(i)})$ ▷ Separate reconstruction and prediction.
  $\mathbf{R}^{(i)} = \mathbf{R}^{(i-1)} - \hat{\mathbf{X}}^{(i)}$ ▷ Update the residual.
  $\hat{\mathbf{Y}} = \hat{\mathbf{Y}} + \hat{\mathbf{Y}}^{(i)}$ ▷ Accumulate the forecast.
end for
return $\hat{\mathbf{Y}}$

Note that the output of $\mathcal{K}$ is projected to dimension F_{out} to match the upsampled spectrum.

Table 5: Theoretical complexity breakdown per layer for Transformer backbones. In this context, the look-back window size is symbolized by L. The term N corresponds to the quantity of variables, and P indicates the patch length for patch-wise models.

Models	PatchTST (2023)	Crossformer (2023)	iTransformer (2024c)	CARD (2024b)	Fredformer (2024)	Leddam (2024)	DUET (2025)	TVTS (Ours)
Complexity	$\mathcal{O}((\frac{L}{P})^2)$	$\mathcal{O}(N^2 + (\frac{L}{P})^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}((\frac{L}{P})^2 N)$	$\mathcal{O}(\frac{L}{P} N^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)$	$\mathcal{O}(N^2)$

3. Reconstruction and Recursive Update. The refined spectrum $\tilde{\mathbf{S}}$ is converted back to the time domain via IFFT. To compensate for energy dispersion caused by extending the sequence length in the frequency domain (implicit zero-padding), we introduce a scaling factor $\alpha = (L + H)/L$. The signal is reconstructed and denormalized as:

$$\tilde{\mathbf{O}}^{(i)} = \Big(\text{IFFT}(\tilde{\mathbf{S}}) \cdot \alpha\Big) \cdot \sqrt{\sigma^2 + \epsilon} + \mu. \tag{32}$$

Finally, $\tilde{\mathbf{O}}^{(i)} \in \mathbb{R}^{N \times (L+H)}$ is sliced into reconstruction $\hat{\mathbf{X}}^{(i)}$ (first L steps) and prediction $\hat{\mathbf{Y}}^{(i)}$ (last H steps). As described in the main text, $\hat{\mathbf{X}}^{(i)}$ updates the residual for the next block, while $\hat{\mathbf{Y}}^{(i)}$ is accumulated into the final forecast.

F Computational Complexity Analysis

Table 5 summarizes the theoretical computational complexity per layer. Channel-independent methods, such as PatchTST Nie et al. (2023), achieve high efficiency by processing variables in isolation, typically $\mathcal{O}((\frac{L}{P})^2)$; however, this approach inherently neglects critical inter-variable correlations. In contrast, multivariate models like iTransformer Liu et al. (2024c) and DUET Qiu et al. (2025) incur a complexity of $\mathcal{O}(N^2)$ (where N is the number of variables), allowing them to explicitly capture spatial dependencies and better model complex system dynamics.

Our proposed TVTS aligns with the multivariate category with a complexity of $\mathcal{O}(N^2)$, which is primarily driven by the spatial attention mechanism required for modeling variable interactions. This design choice is computationally feasible because our spectral-based TVCF scales linearly or log-linearly with the sequence length L, ensuring a low temporal computational cost. Consequently, TVTS maintains an efficiency tier comparable to state-of-the-art models like DUET, successfully striking a balance between robust time series modeling and computational feasibility.

G Experimental Setups.

G.1 Datasets

We elaborate on the datasets employed in TVTS with the following details:

- **ETT** (Electricity Transformer Temperature) Zhou et al. (2021): This dataset captures critical power system metrics, specifically recording the oil temperature and six distinct power load features from two electricity transformers (Unit 1 and Unit 2). The data spans a two-year period from July 2016 to July 2018. Following standard protocols, we utilize four subsets based on sampling granularity: **ETTh1**, **ETTh2**, **ETTm1** and **ETTm2**.
- **Electricity** (ECL) Liu et al. (2024c): This dataset monitors the hourly electricity consumption patterns of 321 distinct clients. The sequence covers a temporal range from 2012 to 2014, serving as a representative benchmark for MTSF.
- **Weather** Wu et al. (2021): Sourced from the Max Planck Institute for Biogeochemistry, this dataset comprises climatological data recorded throughout 2020. It contains 21 meteorological indicators (e.g., humidity, pressure) collected at 10-minute intervals, offering a comprehensive view of regional atmospheric dynamics.
- **Traffic** Wu et al. (2021): This dataset consists of hourly road occupancy rates captured by 862 sensors deployed across the San Francisco Bay Area freeway system. The data, collected between 2015 and 2016, effectively reflects complex spatial-temporal traffic patterns.

- **Exchange** Lai et al. (2018): Focusing on financial stability, this dataset tracks the daily exchange rates of eight major national currencies against the U.S. dollar. The data spans a long historical period from 1990 to 2016.
- **PEMS** (Caltrans Performance Measurement System) Liu et al. (2022a): This large-scale traffic dataset aggregates flow data sampled every 5 minutes from distinct districts in California. We utilize four specific subsets: **PEMS03**, **PEMS04**, **PEMS07**, **PEMS08**, each representing a different topological structure of the sensor network.
- **DowJones** Yue et al. (2025a): This financial dataset tracks the daily stock prices of 27 component companies of the Dow Jones Industrial Average (DJIA). The collection period extends from January 1999 to March 2025, providing a volatile environment for testing long-range dependency modeling.
- **NASDAQ** [2]: Similar to DowJones but broader in scope, this dataset includes the daily NASDAQ index alongside key macroeconomic indicators. Spanning from 2010 to 2024, it serves as a testbed for multivariate financial forecasting.
- **Wiki** [3]: This dataset logs the daily web traffic volume for Wikipedia articles over a two-year duration (2018–2019). We utilize the top 99 prominent channels to evaluate the model's ability to handle erratic user behavior patterns.
- **METR-LA** [4]: A benchmark for spatial-temporal modeling, this dataset records traffic speed statistics from 207 sensors in the Los Angeles highway network.
- **COVID-19** Chen et al. (2022): Provided by Johns Hopkins University, this epidemiological dataset contains daily hospitalization records related to the COVID-19 pandemic in California during 2020.

G.2 Detailed Evaluation Metrics

To provide a comprehensive assessment of prediction performance, we utilize five distinct metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), Coefficient of Determination (R^2), Mean Absolute Scaled Error (MASE), and Pearson Correlation Coefficient (r). Let H denote the prediction horizon. Given the ground-truth $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_H\}$ and the predicted $\hat{\mathbf{Y}} = \{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_H\}$, where $\mathbf{y}_t, \hat{\mathbf{y}}_t \in \mathbb{R}^N$ represent true and predicted vectors at step t, the formulas are as follows:

- **Mean Squared Error (MSE):** MSE measures the average squared difference between the estimated values and the actual value. It is sensitive to outliers and penalizes larger errors more heavily.

$$\mathrm{MSE} = \frac{1}{H}\sum_{t=1}^{H} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2 \tag{33}$$

- **Mean Absolute Error (MAE):** MAE calculates the average magnitude of errors in a set of predictions, without considering their direction. It provides a robust measure of accuracy that is less sensitive to outliers compared to MSE.

$$\mathrm{MAE} = \frac{1}{H}\sum_{t=1}^{H} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_1 \tag{34}$$

- **Coefficient of Determination (R^2):** R^2 indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. A value closer to 1 implies a better fit.

$$R^2 = 1 - \frac{\sum_{t=1}^{H} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_2^2}{\sum_{t=1}^{H} \|\mathbf{y}_t - \bar{\mathbf{y}}\|_2^2} \tag{35}$$

- **Mean Absolute Scaled Error (MASE):** MASE is a scale-independent metric that compares the model's forecast error against the mean absolute error of a naive one-step random walk method.

$$\mathrm{MASE} = \frac{\frac{1}{H}\sum_{t=1}^{H} \|\mathbf{y}_t - \hat{\mathbf{y}}_t\|_1}{\frac{1}{H-1}\sum_{t=2}^{H} \|\mathbf{y}_t - \mathbf{y}_{t-1}\|_1} \tag{36}$$

[2] https://www.kaggle.com/datasets/sai14karthik/nasdq-dataset

[3] https://www.kaggle.com/datasets/sandeshbhat/wikipedia-web-traffic-201819

[4] https://github.com/liyaguang/DCRNN

Table 6: Dataset specifications. "Prediction Length" indicates the future time steps to be predicted. "Split" indicates the (Train, Validation, Test) ratio. "Frequency" specifies the sampling interval of timestamps.

Dataset	Dim	Prediction Length	Total length	Split	Frequency	Information
ETTm1	7	[96, 192, 336, 720]	69680	6:2:2	15 mins	Electricity
ETTm2	7	[96, 192, 336, 720]	69680	6:2:2	15 mins	Electricity
ETTh1	7	[96, 192, 336, 720]	17420	7:1:2	Hourly	Electricity
ETTh2	7	[96, 192, 336, 720]	17420	7:1:2	Hourly	Electricity
Electricity (ECL)	321	[96, 192, 336, 720]	26304	7:1:2	Hourly	Electricity
Traffic	862	[96, 192, 336, 720]	17544	7:1:2	Hourly	Transportation
Weather	21	[96, 192, 336, 720]	52696	7:1:2	10 mins	Weather
Exchange	8	[96, 192, 336, 720]	7588	7:1:2	Daily	Economy
PEMS03	358	[12, 24, 48, 96]	26209	6:2:2	5 mins	Transportation
PEMS04	307	[12, 24, 48, 96]	16992	6:2:2	5 mins	Transportation
PEMS07	883	[12, 24, 48, 96]	28224	6:2:2	5 mins	Transportation
PEMS08	170	[12, 24, 48, 96]	17856	6:2:2	5 mins	Transportation
DowJones	27	[3, 6, 9, 12] [24, 36, 48, 60]	6577	7:1:2	Daily	Finance
Wiki	99	[3, 6, 9, 12] [24, 36, 48, 60]	730	7:1:2	Daily	Web
METR-LA	207	[3, 6, 9, 12] [24, 36, 48, 60]	34272	7:1:2	5 mins	Transportation
COVID-19	55	[3, 6, 9, 12] [24, 36, 48, 60]	335	7:1:2 6:2:2	Daily	Health
NASDAQ	12	[3, 6, 9, 12] [24, 36, 48, 60]	3914	7:1:2	Daily	Finance

- **Pearson Correlation Coefficient (r):** The Pearson correlation coefficient measures the linear correlation between the predicted and observed series, reflecting the model's ability to capture temporal trends regardless of scale.

$$r = \frac{\sum_{t=1}^{H}(\mathbf{y}_t - \bar{\mathbf{y}})^T(\hat{\mathbf{y}}_t - \bar{\hat{\mathbf{y}}})}{\sqrt{\sum_{t=1}^{H}\|\mathbf{y}_t - \bar{\mathbf{y}}\|_2^2}\sqrt{\sum_{t=1}^{H}\|\hat{\mathbf{y}}_t - \bar{\hat{\mathbf{y}}}\|_2^2}} \tag{37}$$

In the above equations, $\|\cdot\|_2^2$ and $\|\cdot\|_1$ denote the L_2 and L_1 norms, respectively. $\bar{\mathbf{y}} = \frac{1}{H}\sum_{t=1}^{H}\mathbf{y}_t$ and $\bar{\hat{\mathbf{y}}} = \frac{1}{H}\sum_{t=1}^{H}\hat{\mathbf{y}}_t$ represent the temporal averages of the ground truth and predicted sequences.

G.3 Implementation Details

Computational Protocol. All experiments are implemented in PyTorch and conducted on a single NVIDIA A100 GPU (80GB). Consistent with standard evaluation protocols, we integrate TVTS into the widely adopted *Time-Series-Library* benchmark repository [5]. Optimization is performed using the Adam optimizer with an early stopping mechanism to mitigate overfitting. For comparative fairness, baseline results are either reproduced using their official open-source implementations under identical settings or cited directly from the original literature.

[5] https://github.com/thuml/Time-Series-Library

Table 7: Hyperparameter settings for different datasets. See the Appendix G.3 section for detailed information on the parameters.

Parameter	ETTm1	ETTm2	ETTh1	ETTh2	ECL	Traffic	Weather	Exchange	PEMS03	PEMS04	PEMS07	PEMS08
lr	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.0005	0.0005	0.0005	0.001
bs	16	16	16	16	16	16	32	16	16	16	16	16
D	512	512	512	512	512	512	512	512	512	512	512	512
N_{l1}	3	4	4	2	4	3	4	2	3	3	3	3
N_{l2}	1	1	2	1	3	3	3	2	2	2	2	2
I	3	2	2	2	2	3	3	2	2	3	4	2
M	64	64	64	32	128	128	128	64	64	64	128	64
β	0.1	0.1	0.3	0.3	0.7	0.7	0.7	0.1	0.7	0.7	0.3	0.3
Top-k	2	2	2	2	10	25	2	2	15	10	10	15

Hyperparameter Configuration. To ensure optimal performance, hyperparameters are selected via a rigorous grid search. The general search space includes learning rates (lr) in $\{$1e-4, 5e-4, 1e-3$\}$ and batch sizes (bs) in $\{4, 8, 16, 32\}$. Structurally, we tune the hidden dimension $D \in \{128, 256, 512\}$ and the layer depths for TVCF (N_{l1}), DGC (N_{l2}), and FDRP (I) modules within $\{1, 2, 3, 4\}$. Key model-specific parameters are configured as follows: the spectral basis size M is searched from $\{16, 32, 64, 128, 256\}$, the filter order is fixed at $K = L/2$, and the loss balancing coefficient is tuned within $\beta \in [0.1 \sim 0.9]$. We set the regularization weight $\lambda = 1 \times 10^{-4}$ and saturation factor $\alpha = 3$ by default. Notably, the sparsity threshold Top-k is adapted to dataset dimensionality: we search Top-$k \in [1 \sim 10]$ for lower-dimensional datasets (e.g., ETTm1/m2/h1/h2, Exchange) and Top-$k \in [5 \sim 30]$ for high-dimensional ones (e.g., ECL, Traffic, Weather, PEMS03/04/07/08). Specific optimal configurations for each dataset are detailed in Table 7.

H Detailed Results and Other Comparative Analysis

H.1 Main Experiments.

H.1.1 Long-term Forecasting Experiments

Table 8 comprehensively enumerates the detailed performance metrics for long-term forecasting tasks. These experiments were conducted with a fixed look-back window of $L = 96$, extending across four distinct prediction horizons $H \in \{96, 192, 336, 720\}$. As evidenced by the results, TVTS consistently achieves superior performance across the vast majority of evaluated scenarios. Specifically, out of the 40 total experimental settings, our framework secures the SOTA position in **33** for MSE and **38** for MAE. This dominance validates the holistic efficacy of TVTS as proposed in our introduction: effective forecasting requires addressing both temporal non-stationarity and spatial anisotropy. By combining TVCF to represent position-dependent temporal patterns with DGC to capture directional cross-variable dependencies, TVTS mitigates the error accumulation typically seen in baselines. This dual capability ensures high-fidelity predictions, even as the forecasting horizon expands.

H.1.2 Short-term Forecasting Experiments

Table 10 reports the comprehensive short-term forecasting results on the PEMS03/04/07/08 traffic flow datasets. The experiments utilize a fixed look-back window of $L = 96$ across varying prediction horizons $H \in \{12, 24, 48, 96\}$. To ensure a rigorous evaluation, we have augmented the comparison with several formidable baselines, including TimeMixer++ Wang et al. (2025b), CARD Wang et al. (2024b), and Fredformer Piao et al. (2024). Remarkably, TVTS consistently achieves state-of-the-art performance, securing the lowest MSE and MAE across **all 20 evaluated scenarios**. This sweeping dominance underscores the model's capability to adeptly leverage historical contexts through the TVCF, which serves as a dynamic spectral learner for temporal evolution. Furthermore, by integrating the DGC method, TVTS explicitly captures intricate inter-sensor correlations, effectively modeling the underlying spatial impacts inherent in complex traffic networks.

Table 8: Comparison of long-term MTSF performance. The look-back length L is 96, and prediction horizons $H \in \{96, 192, 336, 720\}$. The best and second-best results are highlighted in **bold** and underlined, respectively.

Datasets	Model	TVTS (Ours)		FACT (2026b)		DistDF (2026a)		GTR (2026)		TimePro (2025)		TFPS (2025)		FreDF (2025a)		FilterTS (2025c)		FilterNet (2024)		iTrans. (2024c)		MSGNet (2024)		CrossGNN (2023)		PatchTST (2023)		DLinear (2023)	
	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	MSE	MAE	MSE	MAE
ETTm1	96	**0.313**	**0.350**	0.327	0.367	0.316	0.357	0.316	0.357	0.326	0.364	0.327	0.367	0.324	0.362	0.321	0.360	0.321	0.361	0.334	0.368	0.319	0.366	0.335	0.373	0.329	0.367	0.345	0.372
	192	**0.356**	**0.369**	0.365	0.372	0.358	0.380	0.358	0.380	0.367	0.383	0.374	0.395	0.373	0.385	0.363	0.382	0.367	0.387	0.377	0.391	0.376	0.397	0.372	0.390	0.367	0.385	0.380	0.389
	336	**0.384**	**0.397**	0.397	0.409	0.392	0.404	0.392	0.404	0.402	0.409	0.401	0.408	0.402	0.404	0.395	0.403	0.401	0.409	0.426	0.420	0.417	0.422	0.403	0.411	0.399	0.410	0.413	0.413
	720	**0.445**	**0.429**	0.451	0.433	0.448	0.437	0.448	0.437	0.469	0.446	0.479	0.456	0.469	0.444	0.462	0.438	0.477	0.448	0.491	0.459	0.481	0.458	0.461	0.442	0.454	0.439	0.474	0.453
	Avg	**0.375**	**0.386**	0.385	0.395	0.378	0.394	0.378	0.394	0.391	0.400	0.395	0.406	0.392	0.399	0.385	0.396	0.392	0.401	0.407	0.410	0.398	0.411	0.393	0.404	0.387	0.400	0.403	0.407
ETTm2	96	**0.170**	**0.250**	0.175	0.255	0.174	0.256	0.174	0.256	0.178	0.260	**0.170**	0.255	0.173	0.252	0.172	0.255	0.175	0.258	0.180	0.264	0.177	0.262	0.176	0.266	0.175	0.259	0.193	0.292
	192	**0.233**	**0.291**	0.239	0.299	0.239	0.298	0.239	0.298	0.242	0.303	0.235	0.296	0.241	0.298	0.237	0.299	0.240	0.301	0.250	0.309	0.247	0.307	0.240	0.307	0.241	0.302	0.284	0.362
	336	**0.289**	**0.330**	0.311	0.344	0.300	0.338	0.300	0.338	0.303	0.342	0.297	0.335	0.298	0.334	0.299	0.398	0.311	0.347	0.311	0.348	0.312	0.346	0.304	0.345	0.305	0.343	0.369	0.427
	720	**0.387**	**0.387**	0.398	0.395	0.397	0.394	0.397	0.394	0.400	0.399	0.401	0.397	0.398	0.393	0.397	0.394	0.414	0.405	0.412	0.407	0.414	0.403	0.406	0.400	0.402	0.400	0.554	0.522
	Avg	**0.270**	**0.315**	0.281	0.323	0.277	0.321	0.277	0.321	0.281	0.326	0.276	0.321	0.278	0.319	0.276	0.321	0.285	0.328	0.288	0.332	0.288	0.330	0.282	0.330	0.281	0.326	0.350	0.401
ETTh1	96	**0.367**	**0.389**	0.370	0.392	0.373	0.393	0.373	0.393	0.375	0.398	0.398	0.413	0.382	0.400	0.374	0.391	0.382	0.402	0.386	0.405	0.390	0.411	0.382	0.398	0.414	0.419	0.386	0.400
	192	**0.421**	**0.419**	0.428	0.430	0.428	0.425	0.428	0.425	0.427	0.429	0.423	0.423	0.430	0.427	0.424	0.421	0.430	0.429	0.441	0.436	0.442	0.442	0.427	0.425	0.460	0.445	0.437	0.432
	336	**0.463**	**0.438**	0.476	0.449	0.466	0.445	0.466	0.445	0.472	0.450	0.484	0.461	0.474	0.451	0.464	0.441	0.472	0.451	0.487	0.458	0.480	0.468	0.465	0.445	0.501	0.466	0.481	0.459
	720	0.461	0.457	0.477	0.471	**0.453**	**0.453**	**0.453**	**0.453**	0.476	0.474	0.488	0.476	0.463	0.462	0.470	0.466	0.481	0.473	0.503	0.491	0.494	0.488	0.472	0.468	0.500	0.488	0.519	0.516
	Avg	**0.428**	**0.426**	0.438	0.436	0.430	0.429	0.430	0.429	0.438	0.438	0.448	0.443	0.437	0.435	0.433	0.430	0.441	0.439	0.454	0.447	0.452	0.452	0.437	0.434	0.469	0.454	0.456	0.452
ETTh2	96	**0.281**	**0.332**	0.303	0.350	0.287	0.336	0.290	0.342	0.293	0.345	0.313	0.355	0.289	0.337	0.290	0.338	0.293	0.343	0.297	0.349	0.309	0.359	0.302	0.348	0.292	0.342	0.333	0.387
	192	**0.355**	**0.379**	0.361	0.391	0.358	0.381	0.362	0.389	0.367	0.394	0.405	0.410	0.363	0.385	0.374	0.390	0.374	0.396	0.380	0.400	0.402	0.414	0.390	0.406	0.387	0.400	0.477	0.476
	336	0.405	0.416	0.429	0.433	0.408	0.421	0.414	0.429	0.419	0.431	**0.392**	**0.415**	0.419	0.426	0.406	0.420	0.417	0.430	0.428	0.432	0.435	0.443	0.426	0.444	0.426	0.433	0.594	0.541
	720	**0.410**	**0.431**	0.433	0.448	0.416	0.435	0.423	0.442	0.427	0.445	**0.410**	0.433	0.415	0.437	0.418	0.437	0.449	0.460	0.427	0.445	0.417	0.441	0.445	0.464	0.431	0.446	0.831	0.657
	Avg	**0.363**	**0.390**	0.382	0.406	0.367	0.393	0.372	0.400	0.377	0.403	0.380	0.403	0.371	0.396	0.372	0.396	0.383	0.407	0.383	0.407	0.396	0.417	0.393	0.418	0.384	0.405	0.559	0.515
ECL	96	**0.132**	**0.223**	0.135	0.231	0.137	0.235	0.134	0.229	0.139	0.234	0.149	0.236	0.144	0.233	0.151	0.245	0.147	0.245	0.148	0.240	0.165	0.274	0.173	0.275	0.161	0.250	0.197	0.282
	192	**0.148**	**0.240**	0.154	0.248	0.159	0.257	0.152	0.245	0.156	0.249	0.162	0.253	0.159	0.247	0.163	0.256	0.160	0.250	0.162	0.253	0.184	0.292	0.195	0.288	0.199	0.289	0.196	0.285
	336	**0.162**	**0.254**	0.165	0.261	0.178	0.272	0.171	0.264	0.172	0.267	0.200	0.310	0.172	0.263	0.180	0.274	0.173	0.267	0.178	0.269	0.195	0.302	0.206	0.300	0.215	0.305	0.209	0.301
	720	**0.187**	**0.279**	0.194	0.290	0.208	0.300	0.212	0.302	0.209	0.299	0.220	0.320	0.204	0.294	0.224	0.311	0.210	0.309	0.225	0.317	0.231	0.332	0.231	0.335	0.256	0.337	0.245	0.333
	Avg	**0.157**	**0.249**	0.162	0.257	0.172	0.267	0.166	0.260	0.169	0.262	0.183	0.280	0.170	0.259	0.180	0.271	0.173	0.268	0.178	0.270	0.194	0.300	0.201	0.300	0.208	0.295	0.212	0.300
Exchange	96	**0.080**	**0.199**	0.090	0.207	–	–	–	–	0.085	0.204	0.083	0.205	0.091	0.212	0.081	**0.199**	0.091	0.211	0.086	0.206	0.102	0.230	0.084	0.203	0.088	0.205	0.088	0.218
	192	**0.169**	**0.293**	0.188	0.303	–	–	–	–	0.178	0.299	0.174	0.297	0.193	0.314	0.171	0.294	0.186	0.305	0.177	0.299	0.195	0.317	0.171	0.294	0.176	0.299	0.176	0.315
	336	**0.298**	**0.397**	0.307	0.409	–	–	–	–	0.328	0.414	0.310	0.398	0.372	0.443	0.321	0.409	0.380	0.449	0.331	0.417	0.359	0.436	0.319	0.407	0.301	**0.397**	0.313	0.427
	720	**0.790**	**0.669**	0.806	0.696	–	–	–	–	0.817	0.679	–	0.756	0.893	0.715	0.837	0.688	0.896	0.712	0.847	0.691	0.940	0.738	0.805	0.677	0.901	0.714	0.839	0.695
	Avg	**0.334**	**0.390**	0.348	0.404	–	–	–	–	0.352	0.399	0.395	0.414	0.387	0.421	0.352	0.397	0.388	0.419	0.360	0.403	0.399	0.430	0.345	0.395	0.367	0.404	0.354	0.414
Traffic	96	0.395	**0.262**	0.396	0.265	**0.380**	**0.262**	0.440	0.263	–	–	–	–	0.391	0.265	0.448	0.309	0.430	0.294	0.395	0.268	0.605	0.344	0.570	0.310	0.446	0.283	0.650	0.396
	192	0.417	**0.271**	0.418	**0.271**	**0.407**	0.275	0.454	0.274	–	–	–	–	0.410	0.273	0.455	0.307	0.452	0.307	0.417	0.276	0.613	0.359	0.577	0.321	0.540	0.354	0.598	0.370
	336	0.435	**0.278**	0.434	0.280	0.429	0.284	0.472	0.282	–	–	–	–	**0.424**	0.280	0.472	0.313	0.470	0.316	0.433	0.283	0.642	0.376	0.588	0.324	0.551	0.358	0.605	0.373
	720	0.466	**0.297**	0.467	**0.297**	**0.452**	**0.297**	0.514	0.301	–	–	–	–	0.460	0.298	0.508	0.332	0.498	0.323	0.467	0.302	0.702	0.401	0.597	0.337	0.586	0.375	0.645	0.394
	Avg	0.428	**0.277**	0.429	0.278	**0.417**	0.279	0.470	0.280	–	–	–	–	0.421	0.279	0.471	0.315	0.463	0.310	0.428	0.282	0.641	0.370	0.583	0.323	0.531	0.343	0.625	0.383
Weather	96	**0.154**	**0.197**	0.159	0.204	0.164	0.209	0.164	0.209	0.166	0.207	**0.154**	0.202	0.164	0.202	0.162	0.207	0.162	0.207	0.174	0.214	0.163	0.212	0.159	0.218	0.177	0.218	0.196	0.255
	192	**0.203**	**0.244**	0.209	0.253	0.212	0.252	0.212	0.252	0.216	0.254	0.205	0.249	0.220	0.253	0.209	0.252	0.210	0.250	0.221	0.254	0.212	0.254	0.211	0.266	0.225	0.259	0.237	0.296
	336	**0.262**	**0.287**	0.264	0.291	0.270	0.295	0.270	0.295	0.273	0.296	**0.262**	0.289	0.275	0.294	0.263	0.292	0.265	0.290	0.278	0.296	0.272	0.299	0.267	0.310	0.278	0.297	0.283	0.335
	720	**0.342**	**0.338**	0.360	0.351	0.348	0.345	0.348	0.345	0.351	0.346	0.344	0.342	0.356	0.347	0.344	0.344	**0.342**	0.340	0.358	0.349	0.350	0.348	0.352	0.362	0.354	0.348	0.345	0.381
	Avg	**0.240**	**0.267**	0.248	0.275	0.248	0.275	0.248	0.275	0.251	0.276	0.241	0.271	0.254	0.274	0.244	0.274	0.245	0.272	0.258	0.279	0.249	0.278	0.247	0.289	0.259	0.281	0.265	0.317
1st count		**33**	**38**	0	2	5	3	1	1	0	0	5	1	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0

H.1.3 Robustness and Generalization Analysis.

To rigorously assess the generalization capability of TVTS beyond standard benchmarks, we extended our evaluation to five dynamic datasets with varying dimensionality ($D \in [12, 207]$): *DowJones*, *Wiki*, *METR-LA*, *COVID-19*, and *NASDAQ*. As detailed in Table 9, we adopted a dual-protocol evaluation strategy to probe performance across varying temporal scales:

- **Avg$_s$ (Short-term):** Fixed look-back window $L = 12$ with horizons $H \in \{3, 6, 9, 12\}$.
- **Avg$_l$ (Extended Short-term):** Fixed look-back window $L = 36$ with horizons $H \in \{24, 36, 48, 60\}$.

Benchmarked against cutting-edge multivariate baselines such as OLinear Yue et al. (2025a) and Leddam Yu et al. (2024), TVTS demonstrates exceptional versatility, securing state-of-the-art average prediction performance in **17 of 20** metrics.

Specifically, the model's resilience to high volatility is evidenced by a **3.2%** average MSE reduction on the erratic *COVID-19* dataset. This improvement highlights the effectiveness of TVCF in representing

position-dependent spectral variations under highly non-stationary sequences inherent in healthcare and web traffic data. Furthermore, consistent superiority across diverse dimensions, ranging from *NASDAQ* ($D = 12$) to *METR-LA* ($D = 207$), confirms that the DGC module effectively constructs robust, latent anisotropic topologies independent of feature density. This cross-domain consistency serves as strong evidence of TVTS's ability to learn intrinsic invariant spectral representations, ensuring sustained high predictive fidelity regardless of input dimensionality or sampling frequency.

H.1.4 Univariate Forecasting Evaluation on M4 Benchmark

Dataset Setup. To rigorously assess the generalization capability of TVTS beyond multivariate scenarios, we conducted extensive evaluations on the M4 benchmark dataset. This large dataset contains time series from different domains (e.g., finance, demographics, and macroeconomics). The data is organized into subsets based on sampling frequencies: *Yearly*, *Quarterly*, and *Monthly*. Following established protocols Wu et al. (2023), we aggregate the remaining high-frequency subsets (*Weekly*, *Daily*, and *Hourly*) into the *Others* category. For experimental consistency, we strictly adhere to the standard M4 training/testing evaluation splits, setting the horizon H between 6 and 48 depending on the specific frequency subset.

Evaluation Metrics. We employ three standard metrics to quantify forecasting performance: Symmetric Mean Absolute Percentage Error (**SMAPE**), Mean Absolute Scaled Error (**MASE**), and the Overall Weighted Average (**OWA**). The OWA metric is particularly critical as it provides a holistic ranking by averaging the relative SMAPE and MASE compared to a Naïve2 baseline. Given the ground truth y_t and the model prediction $\hat{y}_t$ at time step t, the metrics are formulated as follows:

$$\begin{aligned} \text{SMAPE} &= \frac{200}{H} \sum_{i=1}^{H} \frac{|y_{T+i} - \hat{y}_{T+i}|}{|y_{T+i}| + |\hat{y}_{T+i}|}, \\ \text{MASE} &= \frac{1}{H} \sum_{i=1}^{H} \frac{|y_{T+i} - \hat{y}_{T+i}|}{\frac{1}{T+H-m} \sum_{j=m+1}^{T+H} |y_j - y_{j-m}|}, \\ \text{OWA} &= \frac{1}{2} \left[\frac{\text{SMAPE}}{\text{SMAPE}_{\text{Naïve2}}} + \frac{\text{MASE}}{\text{MASE}_{\text{Naïve2}}} \right], \end{aligned} \tag{38}$$

where m denotes the periodicity of the seasonality (e.g., 12 for monthly data).

Performance Analysis. As detailed in Table 11, TVTS delivers exceptional forecasting performance, ranking first in **12 out of the 15** evaluated scenarios across SMAPE, MASE, and OWA metrics. It consistently outperforms both competitive univariate baselines, such as N-BEATS Oreshkin et al. (2020) and N-HiTS Challu et al. (2023), and recent foundation models like TimesNet Wu et al. (2023) and Koopa Liu et al. (2023).

Notably, TVTS excels on the *Yearly* and *Quarterly* subsets, which are prone to evolving trends and non-stationary shifts (e.g., achieving a record OWA of **0.779** on Yearly data). This robustness stems from the TVCF, which dynamically captures spectral variations over time, allowing for superior adaptation to trend shifts compared to static spectral methods. Moreover, in the *Others* category, which is dominated by high-frequency and volatile data, TVTS attains a remarkable SMAPE of **4.658**, surpassing the runner-up by a significant margin. These results underscore the versatility of our linear spectral operator, demonstrating its ability to efficiently reconstruct underlying dynamical systems under varying sampling rates and data characteristics.

H.1.5 Comprehensive Metric Evaluation.

To ensure a holistic scrutiny of predictive fidelity, we extend our evaluation to include three complementary indicators: the Coefficient of Determination (R^2), Pearson Correlation Coefficient (r), and Mean Absolute Scaled Error (MASE), with calculation details provided in Appendix G.3. These metrics offer distinct perspectives, rigorously assessing the model's goodness-of-fit, temporal correlation alignment, and scale-independent robustness, respectively.

The comparative results on the challenging ECL and Traffic datasets are detailed in Table 12. TVTS exhibits strict dominance, achieving state-of-the-art performance across all evaluated horizons and metrics. Notably, regarding the goodness-of-fit, TVTS significantly outperforms the strongest

Table 9: Performance evaluation on datasets characterized by high volatility and varying variable dimensions ($D \in [12 \sim 207]$). We adopt two settings: Avg_s ($L = 12, H \in \{3, 6, 9, 12\}$) for short-term forecasting, and Avg_l ($L = 36, H \in \{24, 36, 48, 60\}$) for extended short-term horizons.

Model		TVTS (Ours)		OLinear (2025a)		TimeMix.++ (2025b)		TimeMix. (2024a)		FilterNet (2024)		FITS (2024)		Fredfm. (2024)		Leddam (2024)		CARD (2024b)		iTrans. (2024c)		PatchTST (2023)		TimesNet (2023)		DLinear (2023)	
Metrics		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
DowJones	3	**1.532**	**0.271**	1.550	0.276	1.555	0.278	1.568	0.276	1.587	0.278	1.611	0.283	1.563	0.277	1.538	0.277	1.563	0.276	1.551	0.276	1.556	0.276	2.269	0.325	1.574	0.280
	6	**2.512**	**0.359**	2.569	0.361	2.591	0.364	2.593	0.362	2.634	0.364	2.661	0.370	2.619	0.364	2.588	0.364	2.579	0.362	2.589	0.363	2.588	0.362	3.934	0.430	2.613	0.367
	9	**3.558**	**0.430**	3.565	**0.430**	3.572	0.433	3.564	0.431	3.651	0.434	3.704	0.443	3.575	0.432	3.644	0.440	3.569	**0.430**	3.599	0.433	3.585	0.432	4.715	0.492	3.621	0.436
	12	4.501	**0.485**	4.517	0.490	4.630	0.496	**4.472**	0.490	4.634	0.496	4.666	0.503	4.605	0.493	4.622	0.498	4.528	0.490	4.579	0.493	4.533	0.492	5.999	0.561	4.586	0.497
	Avg_s	**3.026**	**0.386**	3.050	0.389	3.087	0.393	3.049	0.390	3.126	0.393	3.160	0.400	3.090	0.391	3.098	0.395	3.060	0.389	3.079	0.391	3.065	0.390	4.229	0.452	3.099	0.395
	24	**7.401**	**0.662**	7.432	0.664	8.283	0.689	8.327	0.683	8.000	0.683	7.974	0.690	7.758	0.672	8.029	0.679	7.416	0.665	7.925	0.677	7.641	0.670	11.535	0.834	7.590	0.670
	36	**10.727**	0.802	10.848	0.799	14.754	0.856	11.192	0.813	12.011	0.823	11.907	0.837	11.456	0.808	11.962	0.828	10.799	**0.798**	12.087	0.827	11.210	0.807	16.922	0.982	10.986	0.803
	48	13.964	**0.910**	14.045	0.914	16.893	0.970	15.278	0.945	14.814	0.933	15.821	0.969	14.696	0.921	15.266	0.937	**13.881**	0.912	14.787	0.930	14.866	0.935	19.501	1.093	14.157	0.922
	60	**16.840**	**1.002**	16.959	1.017	18.881	1.059	20.997	1.067	18.932	1.054	19.320	1.077	18.058	1.032	18.407	1.045	17.257	1.021	18.298	1.041	17.947	1.036	22.804	1.177	18.018	1.035
	Avg_l	**12.233**	**0.844**	12.321	0.848	14.703	0.893	13.948	0.877	13.439	0.873	13.755	0.893	12.992	0.858	13.416	0.872	12.338	0.849	13.274	0.869	12.916	0.862	17.690	1.021	12.688	0.857
Wiki	3	6.120	0.372	6.161	**0.368**	6.149	0.389	6.209	0.392	6.234	0.402	7.470	0.496	6.190	0.387	6.148	0.383	6.183	0.378	6.237	0.393	**6.112**	0.380	7.597	0.510	6.254	0.438
	6	**6.417**	**0.385**	6.453	**0.385**	6.436	0.401	6.475	0.402	6.460	0.401	8.326	0.544	6.696	0.404	6.455	0.397	6.465	0.393	6.484	0.400	6.425	0.395	7.962	0.515	6.579	0.467
	9	**6.655**	**0.395**	6.666	0.398	6.714	0.420	6.702	0.418	6.697	0.416	8.869	0.564	6.768	0.411	6.687	0.412	6.714	0.415	6.689	0.411	6.743	0.426	8.150	0.524	6.776	0.508
	12	6.831	**0.404**	6.834	0.406	6.852	0.421	6.902	0.426	6.899	0.426	9.394	0.608	7.168	0.424	6.899	0.424	6.852	0.415	6.868	0.419	**6.814**	0.414	8.117	0.533	6.927	0.513
	Avg_s	**6.506**	**0.389**	6.528	**0.389**	6.538	0.408	6.572	0.409	6.572	0.411	8.515	0.553	6.705	0.406	6.547	0.404	6.553	0.400	6.569	0.405	6.523	0.404	7.956	0.520	6.634	0.481
	24	6.861	**0.422**	6.894	0.423	6.902	0.460	6.900	0.446	6.907	0.443	7.887	0.613	**6.531**	0.432	6.919	0.450	6.925	0.440	6.886	0.437	6.858	0.430	8.023	0.612	6.883	0.520
	36	6.412	0.440	6.446	**0.439**	6.539	0.473	6.520	0.467	6.514	0.467	7.774	0.660	**5.935**	0.453	6.456	0.457	6.463	0.451	6.431	0.452	6.400	0.445	7.229	0.595	6.393	0.538
	48	5.944	**0.445**	6.004	0.446	6.115	0.487	6.108	0.484	6.135	0.478	7.737	0.706	**5.871**	0.464	6.031	0.468	6.031	0.460	6.101	0.483	5.959	0.449	7.184	0.641	5.940	0.547
	60	5.658	**0.450**	5.705	0.454	5.736	0.497	5.732	0.476	5.811	0.482	7.855	0.746	**5.389**	0.463	5.740	0.478	5.723	0.463	5.681	0.462	5.633	0.452	6.805	0.645	5.605	0.552
	Avg_l	6.219	**0.439**	6.262	0.440	6.323	0.479	6.315	0.468	6.342	0.467	7.813	0.681	**5.931**	0.453	6.286	0.463	6.285	0.453	6.275	0.458	6.212	0.444	7.310	0.623	6.205	0.539
METR-LA	3	**0.201**	**0.168**	0.207	0.171	0.204	0.189	0.205	0.192	0.204	0.189	0.211	0.198	0.205	0.188	0.204	0.191	0.210	0.180	0.205	0.188	0.204	0.190	0.221	0.204	0.218	0.231
	6	**0.293**	**0.207**	0.301	**0.207**	0.298	0.230	0.297	0.230	0.296	0.228	0.306	0.237	0.298	0.227	**0.293**	0.227	0.311	0.219	0.300	0.229	0.298	0.227	0.308	0.238	0.307	0.278
	9	**0.366**	**0.231**	0.382	0.238	0.384	0.266	0.381	0.264	0.377	0.261	0.392	0.271	0.385	0.263	0.369	0.264	0.401	0.253	0.386	0.265	0.382	0.263	0.387	0.273	0.386	0.316
	12	0.445	**0.261**	0.452	0.263	0.456	0.294	0.455	0.295	0.449	0.290	0.467	0.300	0.457	0.292	**0.442**	0.292	0.474	0.281	0.460	0.295	0.456	0.292	0.462	0.298	0.452	0.353
	Avg_s	**0.326**	**0.217**	0.335	0.220	0.335	0.245	0.334	0.245	0.331	0.242	0.344	0.251	0.336	0.242	0.327	0.243	0.349	0.233	0.338	0.244	0.335	0.243	0.344	0.253	0.341	0.294
	24	0.659	0.347	0.650	**0.337**	**0.617**	0.394	0.671	0.413	0.670	0.402	0.698	0.416	0.676	0.408	0.680	0.405	0.700	0.378	0.700	0.413	0.679	0.410	0.698	0.415	0.645	0.458
	36	0.811	0.401	0.800	**0.388**	**0.781**	0.457	0.841	0.480	0.824	0.471	0.874	0.490	0.852	0.477	0.841	0.471	0.874	0.448	0.867	0.480	0.845	0.484	0.856	0.475	0.785	0.533
	48	0.904	0.463	0.905	**0.427**	**0.842**	0.520	0.964	0.531	0.955	0.521	1.013	0.546	0.982	0.526	0.963	0.528	1.017	0.498	1.017	0.539	0.972	0.536	0.972	0.518	0.885	0.585
	60	1.005	0.493	0.999	**0.457**	**0.958**	0.551	1.047	0.573	1.050	0.563	1.122	0.589	1.084	0.569	1.029	0.556	1.126	0.541	1.079	0.572	1.077	0.578	1.033	0.543	0.959	0.623
	Avg_l	0.845	0.426	0.838	**0.402**	**0.799**	0.480	0.881	0.499	0.875	0.489	0.927	0.510	0.898	0.495	0.878	0.490	0.929	0.466	0.916	0.501	0.893	0.502	0.890	0.488	0.819	0.550
COVID-19	3	1.109	0.502	**1.100**	**0.487**	1.298	0.584	1.237	0.547	1.195	0.555	2.039	0.790	1.165	0.548	1.216	0.570	1.103	0.521	1.193	0.561	1.220	0.573	2.021	0.704	2.386	0.909
	6	1.683	**0.615**	1.750	0.619	1.833	0.682	2.003	0.739	1.839	0.711	2.683	0.919	**1.465**	0.685	1.782	0.689	1.919	0.735	1.933	0.755	1.982	0.762	2.405	0.808	3.220	1.053
	9	**2.122**	**0.727**	2.239	0.734	2.472	0.822	2.594	0.860	2.537	0.897	3.147	1.050	2.145	0.845	2.407	0.866	2.358	0.841	2.441	0.879	2.633	0.916	2.858	0.969	3.803	1.160
	12	**2.470**	**0.822**	2.538	0.831	3.273	1.084	3.103	0.981	2.782	0.956	3.630	1.156	2.833	0.984	2.851	0.991	2.857	0.971	2.819	0.984	3.050	1.030	2.993	0.964	4.524	1.288
	Avg_s	**1.846**	**0.667**	1.907	0.668	2.219	0.793	2.234	0.782	2.088	0.780	2.875	0.979	1.902	0.765	2.064	0.779	2.059	0.767	2.096	0.795	2.221	0.820	2.569	0.861	3.483	1.102
	24	4.498	1.195	**4.474**	**1.180**	6.539	1.618	6.335	1.554	5.926	1.517	8.248	1.829	4.799	1.347	4.860	1.342	5.133	1.394	4.715	1.321	5.528	1.450	5.634	1.442	9.780	1.851
	36	**7.114**	**1.663**	7.241	1.670	7.986	1.770	8.222	1.787	7.696	1.733	10.345	2.050	7.536	1.727	7.378	1.708	7.377	1.725	7.299	1.681	8.351	1.830	9.114	1.848	12.804	2.083
	48	9.991	1.977	10.076	1.985	11.655	2.156	11.669	2.157	11.572	2.141	11.999	2.197	**9.833**	**1.951**	10.051	1.999	11.013	2.103	10.141	2.012	11.259	2.114	10.940	2.033	14.244	2.189
	60	11.761	2.140	12.079	2.182	12.734	2.235	12.188	2.173	**11.311**	**2.066**	13.185	2.282	12.455	2.209	11.467	2.119	12.528	2.227	11.871	2.156	12.666	2.225	12.888	2.186	15.472	2.275
	Avg_l	**8.341**	**1.744**	8.467	1.754	9.728	1.945	9.604	1.918	9.126	1.864	10.944	2.089	8.656	1.808	8.439	1.792	9.013	1.862	8.506	1.792	9.451	1.905	9.644	1.877	13.075	2.099
NASDAQ	3	**0.035**	**0.091**	0.036	0.092	0.036	0.098	**0.035**	0.093	0.038	0.100	0.042	0.111	0.039	0.102	0.040	0.103	0.037	0.096	0.04	0.105	0.038	0.099	0.049	0.123	0.044	0.123
	6	**0.049**	0.120	**0.049**	**0.117**	0.050	0.121	**0.049**	0.118	0.052	0.126	0.056	0.134	0.053	0.126	0.054	0.128	0.052	0.123	0.054	0.129	0.053	0.124	0.061	0.142	0.062	0.155
	9	**0.061**	**0.137**	0.062	**0.137**	0.063	0.144	0.062	0.139	0.065	0.145	0.069	0.154	0.067	0.147	0.066	0.147	0.063	0.141	0.068	0.150	0.065	0.145	0.073	0.161	0.082	0.189
	12	**0.071**	**0.153**	0.073	0.154	0.075	0.161	0.073	0.156	0.076	0.162	0.081	0.172	0.079	0.165	0.078	0.164	0.075	0.158	0.078	0.165	0.077	0.161	0.088	0.179	0.1	0.215
	Avg_s	**0.054**	**0.125**	0.055	**0.125**	0.056	0.131	0.055	0.126	0.058	0.133	0.062	0.142	0.059	0.135	0.059	0.135	0.057	0.130	0.060	0.137	0.058	0.132	0.068	0.151	0.072	0.17
	24	**0.118**	**0.216**	0.121	**0.216**	0.132	0.233	0.122	0.221	0.130	0.230	0.140	0.244	0.128	0.226	0.125	0.222	0.124	0.220	0.137	0.237	0.127	0.224	0.198	0.299	0.155	0.274
	36	**0.160**	**0.258**	0.163	0.261	0.177	0.278	0.183	0.279	0.175	0.273	0.184	0.284	0.17	0.268	0.174	0.271	0.167	0.266	0.184	0.280	0.174	0.269	0.229	0.326	0.196	0.306
	48	0.203	**0.295**	0.205	0.296	0.216	0.311	**0.200**	0.298	0.224	0.314	0.234	0.324	0.218	0.306	0.222	0.312	0.218	0.307	0.229	0.318	0.225	0.314	0.267	0.352	0.244	0.344
	60	**0.235**	**0.328**	0.259	0.336	0.249	0.337	0.238	**0.328**	0.259	0.340	0.282	0.357	0.262	0.339	0.264	0.341	0.264	0.341	0.279	0.352	0.265	0.339	0.327	0.394	0.318	0.401
	Avg_l	**0.179**	**0.274**	0.187	0.277	0.193	0.29	0.186	0.281	0.197	0.289	0.21	0.302	0.194	0.285	0.196	0.286	0.193	0.284	0.207	0.297	0.198	0.286	0.255	0.343	0.228	0.331
1st count		**29**	**37**	3	17	5	0	4	1	1	1	0	0	7	1	2	0	1	2	0	0	2	0	0	0	0	0

baselines (e.g., iTransformer and CARD), boosting the average R^2 on the ECL dataset to 0.648 compared to the runner-up's 0.614. Furthermore, the consistently superior Pearson correlation (r) scores demonstrate that TVTS effectively captures dynamic temporal trends and synchronizes well with ground-truth fluctuations, avoiding the phase shift issues common in long-term forecasting. This

Table 10: Comparison of Short-term MTSF performance. The look-back length L is 96, and prediction horizons $H \in \{12, 24, 48, 96\}$. The best and second-best results are highlighted **bold** and underlined, respectively.

Datasets	Model	TVTS (Ours)		TimeMix.++ (2025b)		FilterNet (2024)		FITS (2024)		CARD (2024b)		Fredfm. (2024)		TimeMix. (2024a)		PatchTST (2023)		TimesNet (2023)		Crossfm. (2023)		Dlinear (2023)	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
PEMS03	12	**0.061**	**0.163**	0.097	0.208	0.071	0.177	0.117	0.226	0.072	0.177	0.068	0.174	0.076	0.188	0.099	0.216	0.085	0.192	0.090	0.203	0.122	0.243
	24	**0.082**	**0.189**	0.120	0.230	0.102	0.213	0.235	0.324	0.107	0.217	0.093	0.202	0.113	0.226	0.142	0.259	0.118	0.223	0.240	0.201	0.317	0.085
	48	**0.129**	**0.239**	0.170	0.272	0.162	0.272	0.541	0.521	0.194	0.302	0.146	0.258	0.191	0.292	0.211	0.319	0.155	0.260	0.202	0.317	0.333	0.425
	96	**0.209**	**0.311**	0.274	0.342	0.244	0.340	1.062	0.790	0.323	0.402	0.228	0.330	0.288	0.363	0.269	0.370	0.228	0.317	0.262	0.367	0.457	0.515
	Avg	**0.120**	**0.226**	0.165	0.263	0.145	0.251	0.489	0.465	0.174	0.275	0.135	0.243	0.167	0.267	0.180	0.291	0.147	0.248	0.169	0.282	0.278	0.375
PEMS04	12	**0.071**	**0.169**	0.099	0.214	0.082	0.190	0.129	0.239	0.089	0.194	0.085	0.189	0.092	0.204	0.105	0.224	0.087	0.195	0.098	0.218	0.148	0.272
	24	**0.082**	**0.177**	0.115	0.231	0.110	0.224	0.246	0.337	0.128	0.234	0.117	0.224	0.128	0.243	0.153	0.275	0.103	0.215	0.131	0.256	0.224	0.340
	48	**0.120**	**0.231**	0.144	0.261	0.160	0.276	0.568	0.539	0.224	0.321	0.174	0.276	0.213	0.315	0.229	0.339	0.136	0.250	0.205	0.326	0.355	0.437
	96	**0.173**	**0.283**	0.185	0.297	0.234	0.343	1.181	0.843	0.382	0.445	0.273	0.354	0.307	0.384	0.291	0.389	0.190	0.303	0.402	0.457	0.452	0.504
	Avg	**0.112**	**0.215**	0.136	0.251	0.146	0.258	0.531	0.489	0.206	0.299	0.162	0.261	0.185	0.287	0.195	0.307	0.129	0.241	0.209	0.314	0.295	0.388
PEMS07	12	**0.057**	**0.150**	0.090	0.197	0.064	0.163	0.109	0.222	0.068	0.166	0.063	0.158	0.073	0.184	0.095	0.207	0.082	0.181	0.094	0.200	0.115	0.242
	24	**0.079**	**0.177**	0.119	0.219	0.093	0.200	0.230	0.327	0.103	0.206	0.089	0.192	0.111	0.219	0.150	0.262	0.101	0.204	0.139	0.247	0.210	0.329
	48	**0.111**	**0.216**	0.149	0.256	0.137	0.248	0.551	0.531	0.165	0.268	0.136	0.241	0.237	0.328	0.253	0.340	0.134	0.238	0.311	0.369	0.398	0.458
	96	**0.166**	**0.268**	0.258	0.359	0.198	0.306	1.112	0.809	0.258	0.346	0.197	0.298	0.303	0.354	0.346	0.404	0.181	0.279	0.396	0.442	0.594	0.553
	Avg	**0.103**	**0.203**	0.152	0.258	0.123	0.229	0.500	0.472	0.149	0.247	0.121	0.222	0.181	0.271	0.211	0.303	0.124	0.225	0.235	0.315	0.329	0.396
PEMS08	12	**0.070**	**0.169**	0.119	0.222	0.080	0.182	0.122	0.233	0.080	0.181	0.081	0.185	0.091	0.201	0.168	0.232	0.112	0.212	0.165	0.214	0.154	0.276
	24	**0.095**	**0.194**	0.149	0.249	0.114	0.219	0.236	0.330	0.118	0.220	0.112	0.214	0.137	0.246	0.224	0.281	0.141	0.238	0.215	0.260	0.248	0.353
	48	**0.154**	**0.249**	0.206	0.292	0.184	0.284	0.562	0.540	0.199	0.289	0.174	0.267	0.265	0.343	0.321	0.354	0.198	0.283	0.315	0.355	0.440	0.470
	96	**0.259**	**0.305**	0.329	0.355	0.309	0.356	1.216	0.846	0.405	0.431	0.277	0.335	0.410	0.407	0.408	0.417	0.320	0.351	0.377	0.397	0.674	0.565
	Avg	**0.145**	**0.229**	0.200	0.279	0.172	0.260	0.534	0.487	0.201	0.280	0.161	0.250	0.226	0.299	0.280	0.321	0.193	0.271	0.268	0.307	0.379	0.416

Table 11: Univariate forecasting results on the M4 dataset. The M4 dataset comprises six subsets, three of which are classified into the "Others" category. We strictly adhere to the data processing and forecasting length settings (len $\in [6, 48]$) established in the TimesNet benchmark Wu et al. (2023). **Bold** indicates the best results, and underlined denotes the second best, .

Freq.	Metrics	TVTS	FilterNet	Koopa	KNF	SCINet	TimesNet	N-HiTS	N-BEATS	PatchTST	FiLM	LightTS	DLinear	FED.	Stationary	Auto.	Pyra.	In.
Yearly	SMAPE	**13.267**	13.627	13.352	13.986	13.717	13.394	13.418	13.436	16.463	17.431	14.247	16.965	13.728	13.717	13.974	15.530	14.727
	MASE	**2.968**	3.040	2.997	3.029	3.076	3.004	3.045	3.043	3.967	4.043	3.109	4.283	3.048	3.078	3.134	3.711	3.418
	OWA	**0.779**	0.799	0.786	0.804	0.807	0.787	0.793	0.794	1.003	1.042	0.827	1.058	0.803	0.807	0.822	0.942	0.881
Quarterly	SMAPE	10.116	11.004	10.159	10.343	10.845	**10.101**	10.202	10.124	10.644	12.925	11.364	12.145	10.792	10.958	11.338	15.449	11.360
	MASE	1.181	1.328	1.189	1.202	1.295	1.183	1.194	**1.169**	1.278	1.664	1.328	1.520	1.283	1.325	1.365	2.350	1.401
	OWA	0.890	0.984	0.895	0.965	0.965	0.890	0.899	**0.886**	0.949	1.193	1.000	1.106	0.958	0.981	1.012	1.558	1.027
Monthly	SMAPE	**12.696**	14.092	12.730	12.894	13.208	12.866	12.791	12.801	13.399	15.407	14.014	13.514	14.260	13.917	13.958	17.642	14.062
	MASE	**0.951**	1.115	0.953	1.023	0.999	0.964	0.969	0.955	1.031	1.298	1.053	1.037	1.102	1.097	1.103	1.913	1.141
	OWA	**0.883**	1.013	0.901	0.985	0.928	0.894	0.899	0.893	0.949	1.144	0.981	0.956	1.012	0.998	1.002	1.511	1.024
Others	SMAPE	**4.658**	5.829	4.861	4.753	5.432	4.982	5.061	5.008	6.558	7.134	15.880	6.709	4.954	6.302	5.485	24.786	24.460
	MASE	**3.072**	3.803	3.124	3.138	3.583	3.323	3.216	3.443	4.511	5.090	11.434	4.953	3.264	4.064	3.865	18.581	20.960
	OWA	**1.002**	1.213	1.004	1.019	1.136	1.048	1.040	1.070	1.401	1.553	3.474	1.487	1.036	1.304	1.187	5.538	5.879
Avg	SMAPE	**11.727**	12.831	11.863	12.126	12.369	11.930	11.927	11.910	13.152	14.863	13.525	13.639	12.840	12.780	12.909	16.987	14.086
	MASE	**1.592**	1.743	1.595	1.641	1.677	1.597	1.613	1.613	1.945	2.207	2.111	2.095	1.701	1.756	1.771	3.265	2.718
	OWA	**0.858**	0.929	**0.858**	0.874	0.894	0.867	0.861	0.862	0.998	1.125	1.051	1.051	0.918	0.930	0.939	1.480	1.230
1st count		**12**	0	1	0	0	1	0	2	0	0	0	0	0	0	0	0	0

multidimensional superiority confirms that our method not only minimizes point-wise errors but also preserves the intrinsic structural integrity and scale consistency of high-dimensional time series.

Table 12: Performance comparison of the $R^2(\uparrow)$, $r(\uparrow)$, and MASE $(\downarrow)$ metrics on complex multivariate datasets (ECL and Traffic). The symbols $\downarrow$ and $\uparrow$ indicate that lower and higher values are better, respectively. The look-back length L is uniformly set to 96. **Bold** indicates the best results, and underlined denotes the second best, respectively.

Datasets	Metrics	TVTS (Ours)			CARD (2024b)			iTrans. (2024c)			TimeMix. (2024a)			DLinear (2023)			PatchTST (2023)			TimesNet (2023)		
		R^2	r	MASE	R^2	r	MASE	R^2	r	MASE	R^2	r	MASE	R^2	r	MASE	R^2	r	MASE	R^2	r	MASE
ECL	96	**0.593**	**0.913**	**0.925**	0.533	0.905	0.975	0.550	0.907	0.990	0.456	0.897	1.022	0.120	0.870	1.300	0.426	0.890	1.119	0.405	0.892	1.118
	192	**0.615**	**0.907**	**1.003**	0.547	0.899	1.028	0.590	0.900	1.057	0.575	0.893	1.076	0.215	0.867	1.312	0.509	0.887	1.140	0.503	0.885	1.195
	336	**0.688**	**0.897**	**1.021**	0.678	0.891	1.113	0.678	0.893	1.133	0.640	0.886	1.155	0.247	0.858	1.387	0.604	0.880	1.216	0.592	0.875	1.262
	720	**0.695**	**0.888**	**1.157**	0.653	0.877	1.273	0.638	0.872	1.315	0.634	0.871	1.314	0.148	0.844	1.536	0.605	0.864	1.371	0.603	0.864	1.388
	Avg	**0.648**	**0.901**	**1.027**	0.603	0.893	1.097	0.614	0.893	1.124	0.576	0.887	1.142	0.182	0.860	1.384	0.536	0.880	1.211	0.526	0.879	1.241
Traffic	96	**0.712**	**0.899**	**0.811**	0.690	0.882	0.855	0.690	0.889	0.860	0.640	0.888	0.979	0.441	0.811	1.346	0.668	0.880	0.903	0.625	0.877	1.023
	192	**0.696**	**0.897**	**0.811**	0.683	0.869	0.856	0.689	0.877	0.879	0.660	0.862	0.959	0.551	0.815	1.189	0.677	0.871	0.887	0.635	0.867	1.009
	336	**0.716**	**0.870**	**0.793**	0.702	0.862	0.843	0.713	0.867	0.883	0.685	0.852	0.986	0.582	0.810	1.167	0.700	0.863	0.874	0.654	0.852	1.079
	720	**0.694**	**0.865**	**0.885**	0.679	0.848	0.894	0.693	0.857	0.913	0.672	0.842	0.960	0.564	0.794	1.224	0.681	0.850	0.925	0.661	0.850	1.047
	Avg	**0.705**	**0.883**	**0.825**	0.689	0.865	0.862	0.696	0.873	0.884	0.664	0.861	0.971	0.535	0.808	1.232	0.682	0.866	0.897	0.644	0.862	1.040

H.2 Comparison with LLM-based Models.

Recently, the adaptation of Large Language Models (LLMs) to MTSF has garnered significant attention, leveraging their inherent semantic reasoning capabilities to capture complex temporal dependencies Jin et al. (2024b); Cao et al. (2024); Liu et al. (2024b; 2025a). To situate TVTS within this emerging landscape, we conduct a comprehensive benchmark against SOTA LLM-based methods. We fix the look-back window at $L = 96$ and evaluate performance across prediction horizons $H \in \{96, 192, 336, 720\}$. The empirical results, detailed in Table 13, unequivocally demonstrate the superiority of our proposed framework. Remarkably, out of 30 metric comparisons across the three datasets (15 cases $\times$ 2 metrics), TVTS achieves the SOTA performance in **14** cases for MSE and **15** cases for MAE. Specifically, compared to the strongest baseline TimeCMA, TVTS delivers consistent improvements; for instance, on the Weather dataset, TVTS reduces the average MSE from 0.250 to 0.240, representing a relative improvement of 4.0%. These findings highlight the superiority of our structural paradigm over semantic alignment. Unlike LLMs relying on imprecise textual mapping, TVTS leverages TVCF and DGC to enforce physical inductive biases, specifically evolving dynamics and directed topologies. This constrains the hypothesis space, capturing intrinsic laws from minimal data without the massive dependency of semantic transfer learning.

H.3 Few-Shot Forecasting

Few-shot forecasting has recently emerged as a rigorous benchmark for evaluating the generalization capabilities of time series models Jin et al. (2024a); Yue et al. (2025a). In this domain, LLMs-based models currently occupy a dominant position, leveraging their massive pre-trained knowledge bases and semantic reasoning abilities to capture dependencies from limited samples Jin et al. (2024a). Motivated by this emerging landscape, we investigate the performance of TVTS in such data-scarce regimes to explore whether explicit spectral modeling can offer a data-efficient alternative to computationally expensive LLM-based paradigms. We conduct experiments using only **10%** and **5%** of the training data, with results summarized in Table 14 and Table 15.

Performance Analysis. Empirical results demonstrate that TVTS exhibits exceptional robustness, consistently outperforming leading LLM-based baselines across multiple datasets. In the **10% data scenario** (Table 14), TVTS achieves state-of-the-art performance, securing the best results in **14 out of 15 cases** for MSE and **15 out of 15 cases** for MAE. Specifically, on the ETTh1, it attains an average MSE of **0.537**, outperforming the strongest LLM competitor, TIME-LLM (0.556), by a margin of **3.4%**. In the extreme **5% data scenario** (Table 15), TVTS maintains remarkable competitiveness. While baseline performance degrades due to information scarcity on ETTh1, TVTS

Table 13: Comparison of the performance of LLMs on long-term forecasting tasks. The look-back window is fixed at $L = 96$, with prediction horizons $H \in \{96, 192, 336, 720\}$. **Bold** indicates the best result, and underlined indicates the second best. "1st count" denotes the total number of best results achieved by each model.

Models		TVTS (Ours)		TimeCMA (2025a)		LLM4TS (2025)		TEMPO (2024)		Time-LLM (2024b)		UniTime (2024b)		GPT4TS (2023)	
Metrics		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	**0.170**	**0.250**	0.173	0.258	0.176	0.255	0.185	0.267	0.193	0.280	0.183	0.266	0.190	0.275
	192	**0.233**	**0.291**	0.238	0.301	0.256	0.317	0.243	0.304	0.257	0.318	0.251	0.310	0.253	0.313
	336	**0.289**	**0.330**	0.297	0.338	0.312	0.345	0.309	0.345	0.317	0.353	0.319	0.351	0.321	0.360
	720	0.387	**0.387**	0.393	0.394	0.412	0.405	**0.386**	0.395	0.419	0.411	0.420	0.410	0.411	0.406
	Avg	**0.270**	**0.315**	0.275	0.323	0.289	0.331	0.281	0.328	0.296	0.340	0.293	0.334	0.294	0.339
ETTh2	96	**0.281**	**0.332**	0.286	0.336	0.307	0.345	0.317	0.352	0.295	0.345	0.296	0.345	0.312	0.360
	192	**0.355**	**0.379**	0.363	0.387	0.370	0.386	0.408	0.412	0.386	0.399	0.374	0.394	0.387	0.405
	336	**0.405**	**0.416**	0.406	0.421	0.414	0.432	0.419	0.452	0.419	0.429	0.415	0.427	0.424	0.437
	720	**0.410**	**0.431**	0.417	0.438	0.420	0.438	0.417	**0.431**	0.425	0.442	0.425	0.444	0.433	0.453
	Avg	**0.363**	**0.390**	0.372	0.397	0.378	0.400	0.390	0.412	0.381	0.404	0.378	0.403	0.389	0.414
Weather	96	**0.154**	**0.197**	0.167	0.211	0.204	0.262	0.222	0.265	0.198	0.235	0.171	0.214	0.203	0.244
	192	**0.203**	**0.244**	0.212	0.253	0.223	0.288	0.286	0.314	0.240	0.269	0.217	0.254	0.247	0.277
	336	**0.262**	**0.287**	0.270	0.292	0.381	0.362	0.337	0.349	0.295	0.308	0.274	0.293	0.297	0.311
	720	**0.342**	**0.338**	0.350	0.348	0.427	0.403	0.363	0.321	0.368	0.353	0.351	0.343	0.368	0.356
	Avg	**0.240**	**0.267**	0.250	0.276	0.309	0.329	0.302	0.312	0.275	0.291	0.253	0.276	0.279	0.297
1st count		**14**	**15**	0	0	0	0	1	1	0	0	0	0	0	0

Table 14: **Few-Shot (10% Data)**: Comprehensive performance comparison using only 10% of the training data. Results for baseline methods are reproduced from TIME-LLM Jin et al. (2024a). **Bold** indicates the best result, and underlined indicates the second best.

Datasets	Method	TVTS (Ours)		TIME-LLM (2024a)		GPT4TS (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2022)		Autoformer (2021)		Stationary (2022b)		ETSformer (2023)		LightTS (2022)		Informer (2021)		Reformer (2020)	
	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	MSE	MAE
ETTh1	96	**0.429**	**0.431**	0.448	0.460	0.458	0.456	0.492	0.495	0.516	0.485	0.861	0.628	0.512	0.499	0.613	0.552	0.918	0.639	1.112	0.806	1.298	0.838	1.179	0.792	1.184	0.790
	192	**0.468**	**0.454**	0.484	0.483	0.570	0.516	0.565	0.538	0.598	0.524	0.797	0.593	0.624	0.555	0.722	0.598	0.915	0.629	1.155	0.823	1.322	0.854	1.199	0.806	1.295	0.850
	336	**0.532**	**0.490**	0.589	0.540	0.608	0.535	0.721	0.622	0.657	0.550	0.941	0.648	0.691	0.574	0.750	0.619	0.939	0.644	1.179	0.832	1.347	0.870	1.202	0.811	1.294	0.854
	720	0.717	**0.573**	**0.700**	0.604	0.725	0.591	0.986	0.743	0.762	0.610	0.877	0.641	0.728	0.614	0.721	0.616	0.887	0.645	1.273	0.874	1.534	0.947	1.217	0.825	1.223	0.838
	Avg	**0.537**	**0.487**	0.556	0.522	0.590	0.525	0.691	0.600	0.633	0.542	0.869	0.628	0.639	0.561	0.702	0.596	0.915	0.639	1.180	0.834	1.375	0.877	1.199	0.809	1.249	0.833
ETTm1	96	**0.338**	**0.379**	0.346	0.388	0.390	0.404	0.352	0.392	0.410	0.419	0.583	0.501	0.578	0.518	0.774	0.614	0.761	0.568	0.911	0.688	0.921	0.682	1.162	0.785	1.442	0.847
	192	**0.365**	**0.388**	0.373	0.416	0.429	0.423	0.382	0.412	0.437	0.434	0.630	0.528	0.617	0.546	0.754	0.592	0.781	0.574	0.955	0.703	0.957	0.701	1.172	0.793	1.444	0.862
	336	**0.404**	**0.418**	0.413	0.426	0.469	0.439	0.419	0.434	0.476	0.454	0.725	0.568	0.998	0.775	0.869	0.677	0.803	0.587	0.991	0.719	0.998	0.716	1.227	0.908	1.450	0.866
	720	**0.473**	**0.441**	0.485	0.476	0.569	0.498	0.490	0.477	0.681	0.556	0.769	0.549	0.693	0.579	0.810	0.630	0.844	0.581	1.062	0.747	1.007	0.719	1.207	0.797	1.366	0.850
	Avg	**0.395**	**0.407**	0.404	0.427	0.464	0.441	0.411	0.429	0.501	0.466	0.677	0.537	0.722	0.605	0.802	0.628	0.797	0.578	0.980	0.714	0.971	0.705	1.192	0.821	1.426	0.856
ETTm2	96	**0.175**	**0.261**	0.177	**0.261**	0.188	0.269	0.213	0.303	0.191	0.274	0.212	0.285	0.291	0.399	0.352	0.454	0.229	0.308	0.331	0.430	0.813	0.688	3.203	1.407	4.195	1.628
	192	**0.224**	**0.305**	0.241	0.314	0.251	0.309	0.278	0.345	0.252	0.317	0.270	0.323	0.307	0.379	0.694	0.691	0.291	0.343	0.400	0.464	1.008	0.768	3.112	1.387	4.042	1.601
	336	**0.265**	**0.323**	0.274	0.327	0.307	0.346	0.338	0.385	0.306	0.353	0.323	0.353	0.543	0.559	2.408	1.407	0.348	0.376	0.469	0.498	1.031	0.775	3.255	1.421	3.963	1.585
	720	**0.414**	**0.379**	0.417	0.390	0.426	0.417	0.436	0.440	0.433	0.427	0.474	0.449	0.712	0.614	1.913	1.166	0.461	0.438	0.589	0.557	1.096	0.791	3.909	1.543	3.711	1.532
	Avg	**0.270**	**0.317**	0.277	0.323	0.293	0.335	0.316	0.368	0.296	0.343	0.320	0.353	0.463	0.488	1.342	0.930	0.332	0.366	0.447	0.487	0.987	0.756	3.370	1.440	3.978	1.587

preserves high predictive fidelity (average MSE **0.571**), delivering an **8.9%** relative improvement over TIME-LLM (0.627). These findings suggest that the synergy between TVCF constrained by structured Vandermonde matrices and channel topology learning within the DGC module provides a powerful inductive bias. This spectral prior effectively reduces the hypothesis space, enabling TVTS to capture intrinsic physical laws from minimal observations without relying on the semantic alignment required by LLMs.

Table 15: **Few-Shot (5% Data)**: Comprehensive performance comparison using only 5% of the training data. "-" indicates that the training set was insufficient for the model. **Bold** indicates the best result, and underlined indicates the second best, respectively.

Data	Method	TVTS (Ours)		TIME-LLM (2024a)		GPT4TS (2023)		DLinear (2023)		PatchTST (2023)		TimesNet (2023)		FEDformer (2024)		Autoformer (2021)		Stationary (2022b)		ETSformer (2023)		LightTS (2022)		Informer (2021)		Reformer (2020)	
	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	MSE	MAE
ETTh1	96	**0.456**	**0.444**	0.483	0.464	0.543	0.506	0.547	0.503	0.557	0.519	0.892	0.625	0.593	0.529	0.681	0.570	0.952	0.650	1.169	0.832	1.483	0.910	1.225	0.812	1.198	0.795
	192	**0.535**	**0.491**	0.629	0.540	0.748	0.580	0.720	0.604	0.711	0.570	0.940	0.665	0.652	0.563	0.725	0.602	0.943	0.645	1.221	0.853	1.525	0.930	1.249	0.828	1.273	0.853
	336	**0.722**	0.596	0.768	0.626	0.754	0.595	0.984	0.727	0.816	0.619	0.945	0.653	0.731	**0.594**	0.761	0.624	0.935	0.644	1.179	0.832	1.347	0.870	1.202	0.811	1.254	0.857
	720	**-**	**-**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Avg	**0.571**	**0.510**	0.627	0.543	0.681	0.560	0.750	0.611	0.694	0.569	0.925	0.647	0.658	0.562	0.722	0.598	0.943	0.646	1.189	0.839	1.451	0.903	1.225	0.817	1.241	0.835
ETTm1	96	0.348	0.387	**0.316**	0.377	0.386	0.405	0.332	**0.374**	0.399	0.414	0.606	0.518	0.628	0.544	0.726	0.578	0.823	0.587	1.031	0.747	1.048	0.733	1.130	0.775	1.234	0.798
	192	**0.411**	**0.403**	0.450	0.464	0.440	0.438	0.358	0.390	0.441	0.436	0.681	0.539	0.666	0.566	0.750	0.591	0.844	0.591	1.087	0.766	1.097	0.756	1.150	0.788	1.287	0.839
	336	0.432	**0.411**	0.450	0.424	0.485	0.459	**0.402**	0.416	0.499	0.467	0.786	0.597	0.807	0.628	0.851	0.659	0.870	0.603	1.138	0.787	1.147	0.775	1.198	0.809	1.288	0.842
	720	**0.496**	**0.452**	0.483	0.471	0.577	0.499	0.511	0.489	0.767	0.587	0.796	0.593	0.822	0.633	0.857	0.655	0.893	0.611	1.245	0.831	1.200	0.799	1.175	0.794	1.247	0.828
	Avg	0.422	**0.413**	0.425	0.434	0.472	0.450	**0.400**	0.417	0.526	0.476	0.717	0.561	0.730	0.592	0.796	0.620	0.857	0.598	1.125	0.782	1.123	0.765	1.163	0.791	1.264	0.826
ETTm2	96	0.187	0.271	**0.174**	**0.261**	0.199	0.280	0.236	0.326	0.206	0.288	0.220	0.299	0.229	0.320	0.232	0.322	0.238	0.316	0.404	0.485	1.108	0.772	3.599	1.478	3.883	1.545
	192	0.235	0.308	**0.215**	**0.287**	0.256	0.316	0.306	0.373	0.264	0.324	0.311	0.361	0.394	0.361	0.291	0.357	0.298	0.349	0.479	0.521	1.317	0.850	3.578	1.475	3.553	1.484
	336	0.301	**0.327**	**0.273**	0.330	0.318	0.353	0.380	0.423	0.334	0.367	0.338	0.366	0.378	0.427	0.478	0.517	0.353	0.380	0.552	0.555	1.415	0.879	3.561	1.473	3.446	1.460
	720	**0.430**	**0.399**	0.433	0.412	0.460	0.436	0.674	0.583	0.454	0.432	0.509	0.465	0.523	0.510	0.553	0.538	0.475	0.445	0.701	0.627	1.822	0.984	3.896	1.533	3.445	1.460
	Avg	0.288	0.326	**0.274**	**0.323**	0.308	0.346	0.399	0.426	0.314	0.352	0.344	0.372	0.381	0.404	0.388	0.433	0.341	0.372	0.534	0.547	1.415	0.871	3.658	1.489	3.581	1.487

Table 16: Statistical significance and stability analysis over five independent random seeds. MSE and MAE are averaged over four forecasting horizons $H \in \{96, 192, 336, 720\}$ with $L = 96$. Welch's t-test Welch (1947) compares TVTS with the strongest competing baseline for each metric. * indicates statistical significance at $p < 0.05$.

Dataset	MSE	Std.	MAE	Std.	MSE p-value	MAE p-value
ETTm1	0.375	0.001	0.386	0.001	1.6726×10^{-4} *	7.6944×10^{-6} *
ETTm2	0.270	0.001	0.315	0.001	4.6255×10^{-4} *	2.8319×10^{-4} *
ETTh1	0.428	0.001	0.426	0.001	3.6466×10^{-6} *	1.4067×10^{-2} *
ETTh2	0.363	0.002	0.390	0.001	1.8208×10^{-2} *	1.0684×10^{-3} *
ECL	0.157	0.001	0.249	0.001	2.2533×10^{-6} *	1.6582×10^{-4} *
Exchange	0.334	0.002	0.390	0.001	1.8712×10^{-5} *	1.5132×10^{-3} *
Traffic	0.428	0.002	0.277	0.002	9.8959×10^{-1}	4.0653×10^{-2} *
Weather	0.240	0.004	0.267	0.003	2.5548×10^{-2} *	1.1501×10^{-2} *

H.4 Statistical Significance and Stability.

To assess whether the performance gains of TVTS are robust to random initialization, we repeat all experiments using five independent random seeds and report the mean performance, standard deviation, and statistical significance in Table 16. The standard deviations range from 0.001 to 0.004 for MSE and from 0.001 to 0.003 for MAE, demonstrating consistently stable performance across different runs. We further conduct Welch's t-test Welch (1947) between TVTS and the strongest competing baseline for each metric. TVTS achieves statistically significant improvements at the $p < 0.05$ level on **15 of the 16** evaluated metrics. The only exception is MSE on Traffic ($p = 0.9896$), where TVTS remains competitive but does not yield a statistically significant improvement. Overall, the low run-to-run variance and significance results indicate that the observed gains are systematic and unlikely to arise from random seed variation.

I Ablation Studies

Table 17 details ablation results across seven datasets ($L = 96$). Empirical evidence confirms that both TVCF and DGC are pivotal. Component-wise, the high-dimensional ECL dataset shows greater sensitivity to DGC removal (MSE degrades 0.157→0.184) compared to TVCF removal (→0.172), underscoring DGC's critical role in modeling complex multivariate topological dependencies. Conversely, TVCF dominates on other datasets by capturing intrinsic non-stationary volatility.

Structure & Objectives. The significant error increase with **Random** $\mathbf{M}$ (e.g., +11.8% MSE on ECL) confirms that topology must be semantically learned. Furthermore, our sparse structure outperforms **Full Attention** (0.157 vs. 0.175), successfully avoiding the computational noise and over-smoothing

Table 17: Ablation results with $L = 96$ across the forecast horizons $H \in \{96, 192, 336, 720\}$.

Data	Model	TVTS		w/o FDRP		w/o DGC		w/o TVCF		Full Attention		Random-$\mathbf{M}$		Based-$\mathbf{H}$		Dual-$\mathbf{A}$		w/o $\mathcal{L}_{\mathcal{F}}$		w/o $\mathcal{L}_{\mathcal{O}}$	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	96	**0.313**	**0.350**	0.320	0.358	0.316	0.353	0.319	0.352	0.321	0.354	0.319	0.355	0.315	0.354	0.314	0.358	0.321	0.357	0.320	0.354
	192	**0.356**	**0.369**	0.359	0.371	0.359	0.377	0.358	0.379	0.359	0.376	0.357	0.376	0.358	0.370	0.358	0.375	0.356	0.376	0.359	0.379
	336	**0.384**	**0.397**	0.387	0.398	0.388	0.402	0.392	0.403	0.389	0.402	0.393	0.402	0.386	0.398	0.387	0.399	0.391	0.404	0.390	0.402
	720	**0.445**	**0.429**	0.447	0.430	0.449	0.438	0.449	0.439	0.448	0.437	0.448	0.437	0.447	0.430	0.446	0.430	0.447	0.442	0.445	0.438
	Avg	**0.375**	**0.386**	0.378	0.389	0.378	0.393	0.380	0.393	0.379	0.392	0.379	0.393	0.377	0.388	0.376	0.391	0.379	0.395	0.379	0.393
ETTm2	96	**0.170**	**0.250**	0.171	0.251	0.171	0.251	0.172	0.252	0.171	0.252	**0.170**	0.251	0.171	0.251	**0.170**	**0.250**	0.175	0.256	**0.170**	**0.250**
	192	**0.233**	**0.291**	0.235	0.295	0.239	0.299	0.240	0.298	0.236	0.294	0.237	0.295	0.235	0.293	**0.233**	0.294	0.236	0.295	0.235	0.294
	336	**0.289**	**0.330**	0.293	0.333	0.294	0.334	0.296	0.333	0.291	0.332	0.291	0.331	0.292	0.332	**0.289**	0.331	0.299	0.337	0.291	0.331
	720	**0.388**	**0.387**	0.389	0.389	0.393	0.390	0.397	0.394	0.393	0.389	0.389	0.388	0.389	0.389	0.389	0.388	0.394	0.391	0.395	0.390
	Avg	**0.270**	**0.315**	0.272	0.317	0.274	0.319	0.276	0.319	0.273	0.317	0.272	0.316	0.272	0.316	**0.270**	0.316	0.276	0.320	0.273	0.316
ETTh1	96	**0.367**	**0.389**	0.369	0.391	0.372	0.392	0.371	0.391	0.369	0.391	0.372	0.392	0.368	**0.389**	0.368	0.390	0.373	0.391	0.369	**0.389**
	192	**0.421**	**0.419**	0.423	0.423	0.426	0.422	0.425	0.421	0.425	0.421	0.428	0.422	0.423	0.420	0.421	0.420	0.428	0.423	0.424	0.419
	336	**0.463**	**0.438**	0.476	0.447	0.475	0.444	0.466	0.443	0.473	0.443	0.472	0.444	0.468	0.440	0.464	0.438	0.474	0.446	0.468	0.444
	720	**0.461**	**0.457**	0.474	0.463	0.484	0.471	0.502	0.486	0.492	0.478	0.573	0.510	0.473	0.460	0.464	0.459	0.503	0.483	0.506	0.485
	Avg	**0.428**	**0.426**	0.436	0.431	0.439	0.432	0.441	0.435	0.440	0.433	0.461	0.442	0.433	0.427	0.429	0.427	0.445	0.436	0.442	0.434
ETTh2	96	**0.281**	**0.332**	0.289	0.334	0.288	0.338	0.291	0.338	0.286	0.338	0.311	0.350	0.285	0.334	0.285	0.333	0.304	0.346	0.304	0.346
	192	**0.355**	**0.379**	0.356	0.380	0.359	0.383	0.360	0.383	0.358	0.383	0.358	0.384	0.356	0.380	0.357	0.380	0.364	0.388	0.359	0.384
	336	**0.405**	**0.416**	0.407	0.420	0.420	0.431	0.417	0.428	0.410	0.421	0.421	0.431	0.409	0.420	0.406	**0.416**	0.417	0.431	0.411	0.424
	720	**0.410**	**0.431**	0.417	0.435	0.423	0.441	0.414	0.435	0.414	0.436	0.419	0.438	0.415	0.433	0.415	0.433	0.419	0.439	0.427	0.444
	Avg	**0.363**	**0.390**	0.367	0.392	0.373	0.398	0.371	0.396	0.367	0.395	0.377	0.401	0.366	0.392	0.366	0.391	0.376	0.401	0.375	0.400
ECL	96	**0.132**	**0.223**	0.133	0.226	0.158	0.246	0.137	0.233	0.135	0.227	0.145	0.239	0.134	0.227	0.135	0.224	0.137	0.232	0.133	0.224
	192	**0.148**	**0.240**	0.150	0.243	0.169	0.256	0.158	0.250	0.159	0.251	0.163	0.257	0.152	0.243	**0.148**	0.242	0.167	0.259	0.150	0.242
	336	**0.162**	**0.254**	0.166	0.263	0.185	0.272	0.175	0.267	0.187	0.274	0.176	0.269	0.166	0.263	0.169	0.265	0.176	0.270	0.174	0.265
	720	**0.187**	**0.279**	0.190	0.283	0.222	0.303	0.219	0.303	0.218	0.301	0.228	0.311	0.192	0.282	0.196	0.281	0.196	0.286	0.190	0.279
	Avg	**0.157**	**0.249**	0.160	0.254	0.184	0.269	0.172	0.263	0.175	0.263	0.178	0.269	0.161	0.254	0.162	0.253	0.169	0.262	0.162	0.253
Exchange	96	**0.080**	0.199	0.084	0.202	0.084	0.202	0.084	0.204	0.085	0.205	0.091	0.213	0.082	0.203	0.083	0.202	0.089	0.210	**0.080**	**0.197**
	192	**0.169**	0.293	0.181	0.302	0.185	0.304	0.175	0.298	0.183	0.304	0.182	0.303	0.182	0.302	0.183	0.301	0.171	0.295	0.170	**0.289**
	336	**0.298**	**0.397**	0.309	0.411	0.342	0.423	0.370	0.444	0.347	0.427	0.342	0.422	0.310	0.411	0.323	0.418	0.344	0.424	0.325	0.411
	720	**0.790**	**0.669**	0.821	0.687	0.823	0.683	0.827	0.684	0.813	0.679	0.836	0.688	0.833	0.689	0.832	0.674	0.824	0.689	0.811	0.676
	Avg	**0.334**	**0.390**	0.349	0.401	0.359	0.403	0.364	0.408	0.357	0.404	0.363	0.407	0.352	0.401	0.355	0.399	0.357	0.405	0.347	0.393
Weather	96	**0.154**	**0.197**	0.157	0.201	0.165	0.204	0.166	0.208	0.165	0.205	0.163	0.207	0.159	0.202	0.155	0.202	0.169	0.211	0.163	0.204
	192	**0.203**	**0.244**	0.206	0.246	0.209	0.246	0.212	0.250	0.212	0.248	0.211	0.250	0.205	0.247	0.205	0.247	0.211	0.250	0.211	0.248
	336	**0.262**	**0.287**	0.265	0.289	0.264	0.287	0.267	0.291	0.267	0.289	0.267	0.291	0.265	0.289	0.263	0.287	0.266	0.291	0.263	0.287
	720	**0.342**	**0.338**	0.346	0.341	0.344	0.342	0.347	0.345	0.350	0.344	0.347	0.345	0.348	0.342	**0.342**	0.339	0.346	0.344	0.345	0.342
	Avg	**0.240**	**0.267**	0.244	0.269	0.246	0.270	0.248	0.274	0.249	0.272	0.247	0.273	0.244	0.270	0.241	0.269	0.248	0.274	0.246	0.270

typical of dense connections. Finally, ablating $\mathcal{L}_{\mathcal{F}}$ or $\mathcal{L}_{\mathcal{O}}$ destabilizes performance, validating their necessity for ensuring spectral consistency and basis diversity, respectively.

J VISUALIZATION

J.1 ANALYZING FILTER BEHAVIOR

To provide an intuitive interpretation of the TVCF formulated in Eq. 4, Figure 8 visualizes the learned spectral weights of both the Fixed Filter and TV Filter on the three datasets ($L = H = 96$).

We observe a stark contrast in the spectral learning patterns. The **Fixed Filter** (Figure 8 (a, c, e)) manifests a static "vertical striping" pattern, indicating that it assigns identical spectral weights across all time steps. This approach inherently assumes temporal invariance, which is often incompatible with real-world dynamics. For instance, electricity load profiles typically exhibit distinct temporal shifts, characterized by surges during peak evening hours and recessions during off-peak dawn intervals; a static filter fails to differentiate these phase-dependent spectral characteristics.

In contrast, our **Time-Varying (TV) Filter** (Figure 8 (b, d, f)) generates a heterogeneous heatmap with dynamic weight distributions. This diversity signifies that the model successfully allocates time-specific spectral learners, dynamically modulating the importance of frequency bands at each time step. Consequently, the TV Filter can promptly amplify critical spectral components relevant to

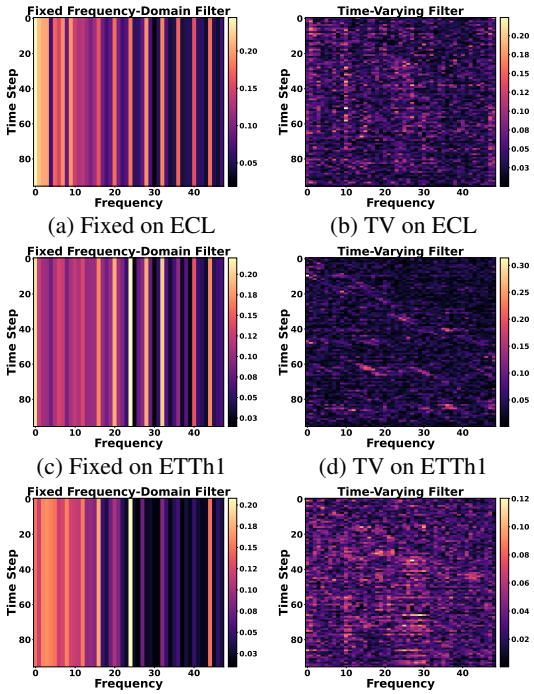
the current state, thereby preserving the fidelity of local temporal trends and enhancing the model's resilience against the non-stationary distribution shifts inherent in complex multivariate systems.

J.2 Analysis of Clustering Performance

J.2.1 Benchmarking Against Channel Clustering Paradigms

To strictly validate the superiority of our clustering strategy, we benchmarked TVTS against state-of-the-art channel learning baselines, specifically DUET Qiu et al. (2025) and BIM^3 Gao et al. (2025) (results detailed in Table 20). These methods predominantly rely on linear pattern extractors to capture temporal features, coupled with Euclidean-based metric learning to perform soft clustering in the frequency under predefined similarity assumptions.

Extensive experiments across six datasets reveal that TVTS maintains a commanding lead, securing the best performance in **29** (MSE) and **28** (MAE) out of 30 cases. This dominance is consistent across varying dimensionalities, from low-variate scenarios to high-dimensional environments like ECL. Quantitatively, on the ETTm1 dataset, TVTS achieves an average MSE of **0.375**, reducing error by **3.1%** compared to the strongest baseline, BIM^3. On the larger ECL dataset, the improvement is even more pronounced with an average MSE of **0.157**, marking an **8.2%** reduction relative to BIM^3.



(a) Fixed on ECL (b) TV on ECL

(c) Fixed on ETTh1 (d) TV on ETTh1

(e) Fixed on ETTh2 (f) TV on ETTh2

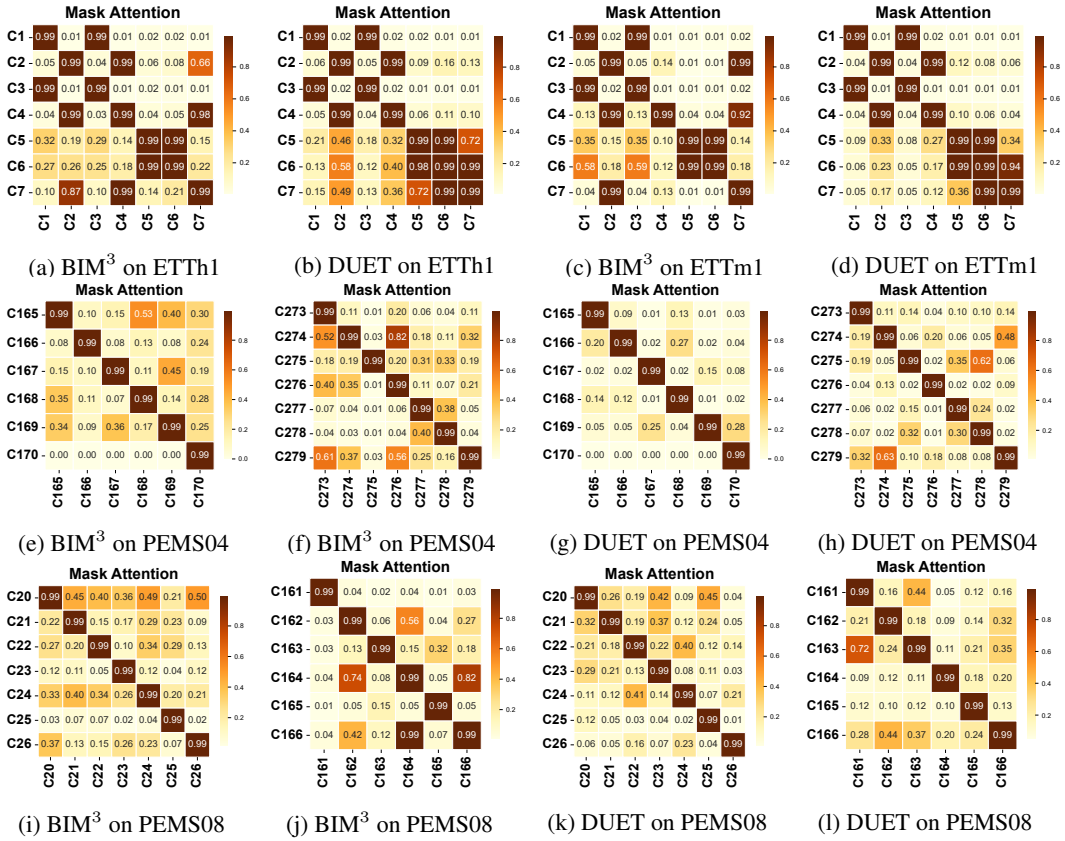
Figure 8: Visualization of learned graph spectral filters on ECL, ETTh1, and ETTh2. The horizontal and vertical axes denote spectral bands and time steps, respectively ($L = 96$, $H = 96$). "Fixed" refers to the static frequency-domain filter Yi et al. (2024), while "TV" denotes our time-varying filter.

These results corroborate our hypothesis regarding the limitations of existing paradigms. The linear time-pattern learning in baselines often yields rigid representations that fail to capture temporal dynamics. Furthermore, their reliance on Euclidean distance tends to capture magnitude proximity while overlooking directional dependencies. In contrast, TVTS leverages the TVCF module to extract dynamic spectral features. Notably, by adopting cosine-based similarity in the DGC module, we prioritize the directional consistency of channel trends rather than mere geometric distance, thereby constructing a more semantic and predictive topology.

J.2.2 Visualization of Clustering Weights.

Visualization of Baseline Channel Masks. Figure 9 visualizes the channel mask attention matrices learned by the state-of-the-art clustering baselines, DUET Qiu et al. (2025) and BIM^3 Gao et al. (2025), across multiple datasets (ETTh1 (N=1-7), ETTm1 (N=1-7), PEMS04 (N=165-170), and PEMS08 (N=161-166)). A ubiquitous characteristic observed across all cases is the strictly symmetric diagonal structure of the feature matrices. This phenomenon stems from the fundamental reliance of these methods on distance-based metric learning (specifically Euclidean or Mahalanobis distance) to measure channel similarity Qiu et al. (2025). Such metrics inherently satisfy the symmetry axiom ($d(x, y) = d(y, x)$), forcing the models to learn bidirectional and undirected representations. Consequently, while these baselines effectively cluster variables with similar magnitudes, they theoretically fail to capture the anisotropic (directional) non-reciprocal dependencies, such as traffic propagation, that are critical for interpreting complex multivariate dynamics.

Extended Visualization of DGC Channel Masks. To further substantiate the robustness of our topological learning, we present extended visualizations of the Geometric Clustering Mask Attention



(a) BIM^3 on ETTh1 (b) DUET on ETTh1 (c) BIM^3 on ETTm1 (d) DUET on ETTm1

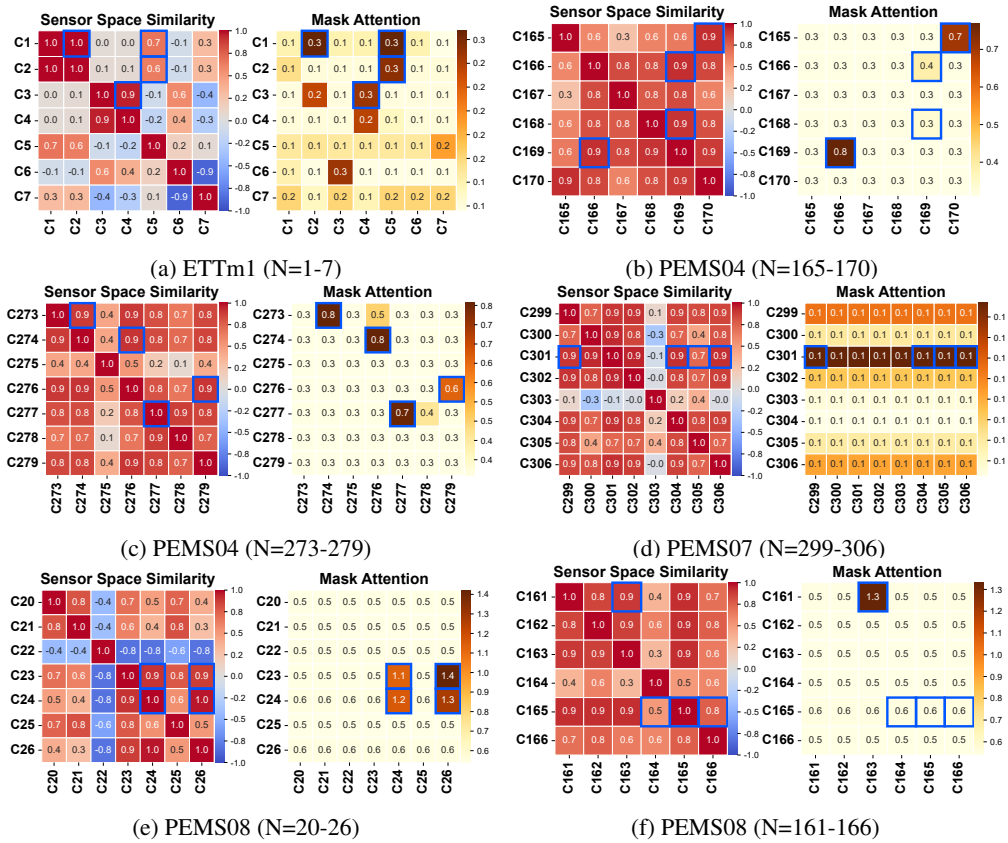
(e) BIM^3 on PEMS04 (f) BIM^3 on PEMS04 (g) DUET on PEMS04 (h) DUET on PEMS04

(i) BIM^3 on PEMS08 (j) BIM^3 on PEMS08 (k) DUET on PEMS08 (l) DUET on PEMS08

Figure 9: Visualization of channel dependency masks learned by BIM^3 and DUET. Note the strictly symmetric diagonal patterns characteristic of distance-based metric learning, which implies these models capture undirected rather than directional dependencies.

across the ETTm1, PEMS04, PEMS07, and PEMS08 datasets in Figure 10. Consistent with the observations in the main text, a distinct contrast remains evident across all scenarios: the static sensor space similarities (left panels) exhibit dense, symmetric patterns, reflecting inherent correlations but high redundancy. In stark contrast, the learned DGC masks (right panels) demonstrate significant sparsity and anisotropy. By breaking the symmetry constraint, the module successfully identifies and amplifies specific directional dependencies (e.g., the asymmetric weights in PEMS07 and PEMS08), thereby effectively pruning noise and reconstructing the latent directed correlation flows required for accurate multivariate modeling.

Topological Visualization of Directed Dependencies. To provide an intuitive view of the latent geometry learned by DGC, we visualize the reconstructed cross-variable topologies as directed graphs in Figure 11. For each dataset, the left panel shows the static cosine-similarity graph, which is inherently symmetric and undirected, whereas the right panel presents the sparse directed topology learned by DGC. Arrow direction denotes dependency orientation, while edge thickness encodes dependency strength. Across datasets, similarity graphs form dense reciprocal webs that obscure directional information, whereas DGC produces markedly sparse and anisotropic structures. By decoupling source and target embeddings, DGC relaxes the reciprocity constraint ($\mathbf{A}_{ij} \neq \mathbf{A}_{ji}$) and recovers directional predictive connectivity beyond pairwise correlation. The learned topology is also dataset-dependent: on traffic benchmarks such as PEMS04 and PEMS08, DGC identifies pronounced unidirectional flows consistent with directional traffic propagation; on ETT/ECL/Exchange benchmarks, many relations remain bidirectional but exhibit unequal strengths in opposite directions. This indicates that DGC does not impose arbitrary orientations, but preserves reciprocal interactions while distinguishing their asymmetric influence. Overall, DGC suppresses redundant symmetric correlations and reconstructs latent directional dependencies that benefit multivariate forecasting.



(a) ETTm1 (N=1-7)

(b) PEMS04 (N=165-170)

(c) PEMS04 (N=273-279)

(d) PEMS07 (N=299-306)

(e) PEMS08 (N=20-26)

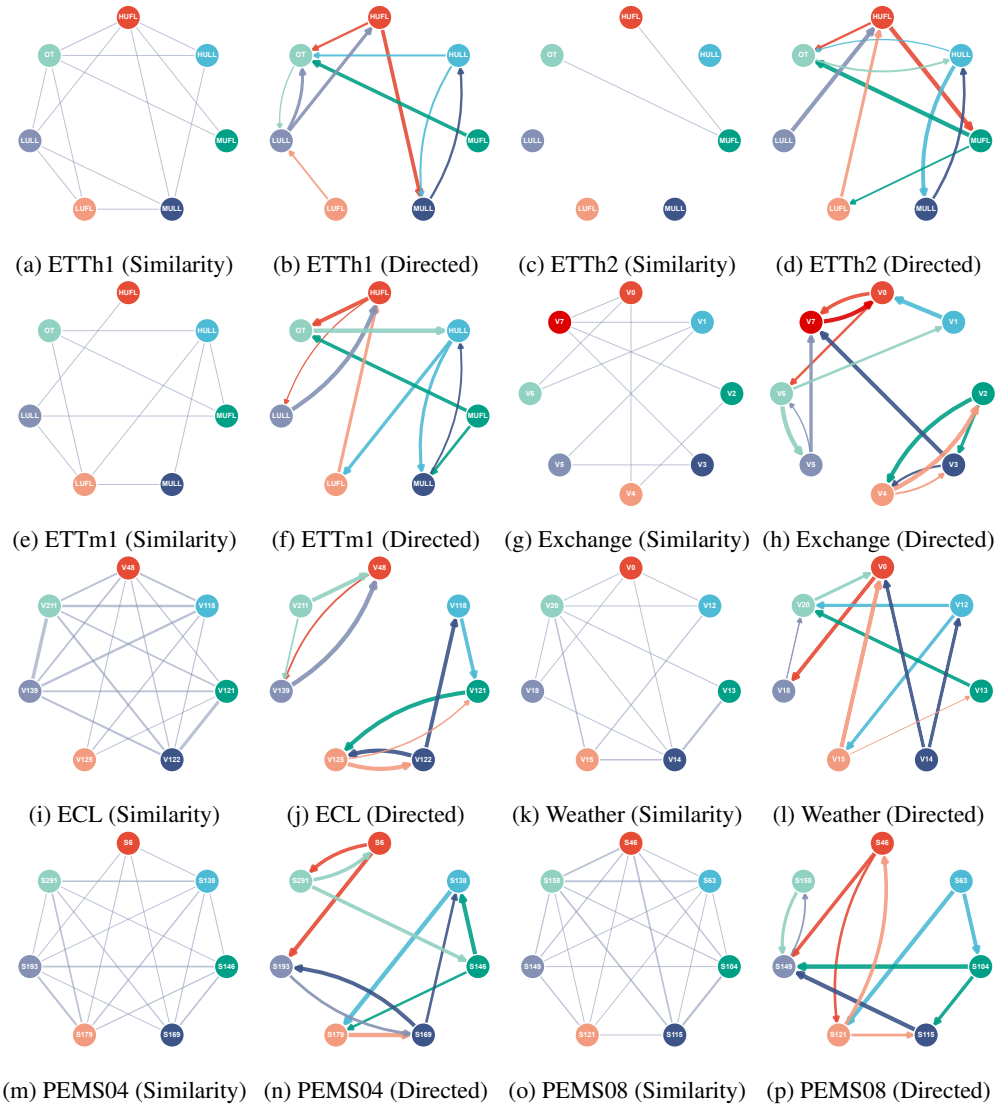
(f) PEMS08 (N=161-166)

Figure 10: Additional topological visualizations across diverse datasets. In each subfigure, the left panel displays the static, symmetric sensor similarity, while the right panel illustrates the learned Geometric Clustering Mask Attention. The results consistently highlight the model's capability to extract sparse and anisotropic dependency flows from dense, static correlations.

Agreement with Granger-Predictive Directions across Temporal Scales. Beyond verifying that DGC learns asymmetric topologies, we further examine whether its preferred directions align with predictive lead–lag relations. We adopt pairwise Granger causality Granger (1969) as an external directional reference, since it tests whether the history of one variable improves the prediction of another. Specifically, we measure the directional agreement between the DGC preference score $\mathbf{D}_{ij} = \mathbf{A}_{ij} - \mathbf{A}_{ji}$ and the corresponding Granger direction score $\mathbf{L}_{ij}$ across different lag orders p, as shown in Figure 12. Rather than consistently fluctuating around the 50% chance-level baseline, the agreement exhibits clear scale-dependent patterns. On ETTh1, it reaches 66.7% at $p = 24$, consistent with the dominant daily dependency in hourly observations. On Exchange, the agreement increases with the lag order and saturates at 71.4%, indicating longer-memory predictive interactions. On PEMS04, it remains between 55% and 65%, while PEMS08 maintains stable agreement of 57–62% across all examined lag orders. These results suggest that DGC does not merely exploit an arbitrary asymmetric parameterization, but learns dataset-specific directional structures aligned with predictive dependencies at informative temporal scales. We therefore interpret this correspondence as evidence of predictive directionality rather than definitive structural causation.

J.3 Visualizing Robustness to Transient Fluctuations

To intuitively validate model robustness, Figure 13 visualizes forecasting trajectories on the challenging ECL dataset, specifically focusing on sharp, non-stationary transient troughs (highlighted in red). Standard Transformer-based baselines (e.g., PatchTST, CARD) exhibit pronounced over-smoothing, failing to capture these rapid drops. This limitation stems from the inherent low-pass filtering nature of canonical self-attention, which tends to average out high-frequency local evolutions. Similarly,



(a) ETTh1 (Similarity) (b) ETTh1 (Directed) (c) ETTh2 (Similarity) (d) ETTh2 (Directed)

(e) ETTm1 (Similarity) (f) ETTm1 (Directed) (g) Exchange (Similarity) (h) Exchange (Directed)

(i) ECL (Similarity) (j) ECL (Directed) (k) Weather (Similarity) (l) Weather (Directed)

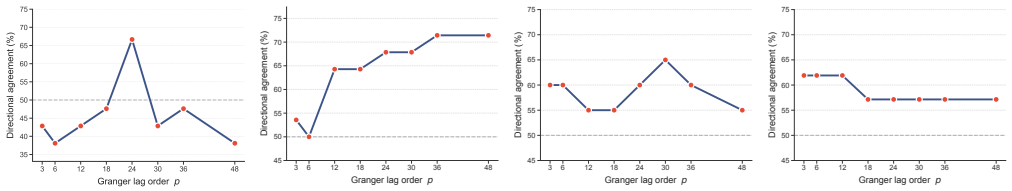
(m) PEMS04 (Similarity) (n) PEMS04 (Directed) (o) PEMS08 (Similarity) (p) PEMS08 (Directed)

Figure 11: Visualization of static similarity graphs and DGC-learned directed topologies across eight datasets. For each dataset, the left panel shows the symmetric cosine-similarity graph, while the right panel presents the directed topology learned by DGC. Arrow direction denotes dependency orientation, and edge thickness indicates dependency strength.

frequency-domain methods like FreDF and FilterNet struggle with phase alignment, as their *fixed* spectral filters apply a shared response across temporal positions. In stark contrast, TVTS reconstructs these irregular shifts with exceptional fidelity. The core advantage lies in the time-varying nature of our TVCF module. Unlike static baselines, TVCF assigns distinct learned spectral responses to different temporal positions, allowing it to instantaneously capture rapid magnitude drops and recovery phases without succumbing to noise. This visualization empirically verifies that our time-varying design effectively reconciles the trade-off between global trend coherence and local detail preservation, resolving the lag and underestimation issues prevalent in static architectures.

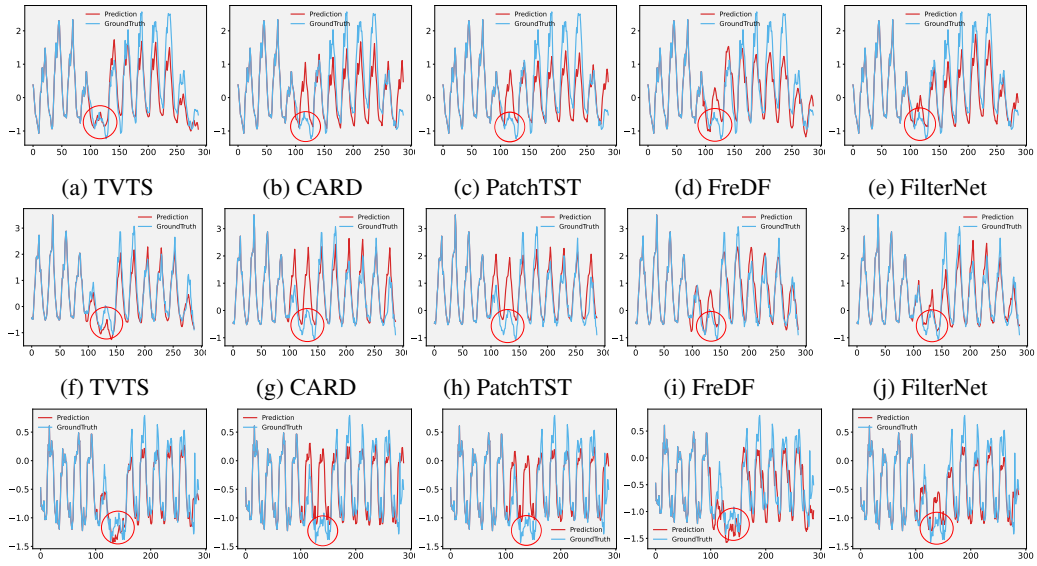
K Empirical Robustness to Variable Permutation

We further examine whether TVTS is sensitive to arbitrary variable ordering. Since DGC uses globally learned variable-specific embeddings, permuting only the input channels does not yield exact permutation equivariance. We therefore treat variable permutation as a structural perturbation and



(a) ETTh1 (b) Exchange (c) PEMS04 (d) PEMS08

Figure 12: Directional agreement between DGC and pairwise Granger-predictive relations Granger (1969) across different lag orders p on the four datasets. Agreement is measured by the directional consistency between $\mathbf{D}_{ij} = \mathbf{A}_{ij} - \mathbf{A}_{ji}$ and the corresponding Granger direction score $\mathbf{L}_{ij}$.



(a) TVTS (b) CARD (c) PatchTST (d) FreDF (e) FilterNet

(f) TVTS (g) CARD (h) PatchTST (i) FreDF (j) FilterNet

(k) TVTS (l) CARD (m) PatchTST (n) FreDF (o) FilterNet

Figure 13: Visualization of forecasting comparisons on ECL with "input-96-prediction-192". The **red circles** highlight regions with sharp non-stationary fluctuations, demonstrating TVTS's superior capability in capturing transient changes compared to baselines.

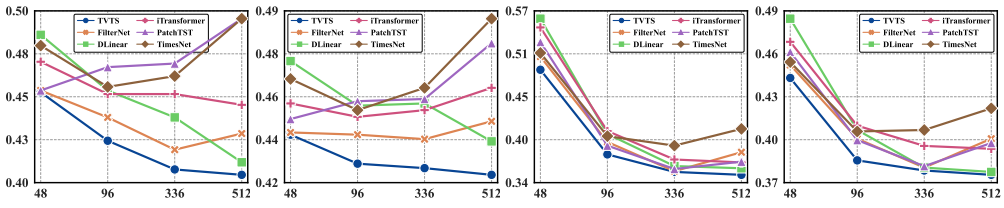
test whether TVTS remains robust under disrupted channel ordering. Formally, given a permutation matrix $\mathbf{P} \in \{0,1\}^{N\times N}$, we construct the perturbed input and target as

$$\mathbf{X}' = \mathbf{P}\mathbf{X}, \qquad \mathbf{Y}' = \mathbf{P}\mathbf{Y}. \tag{39}$$

We conduct this test on ETTh1, ECL, and Weather by randomly permuting 20%- 100% of the variables. As shown in Table 18, the averaged MSE and MAE deviate from the original results by at most $\mathbf{0.003}$. Even under complete permutation, the performance changes only from $0.428/0.426$ to $0.431/0.427$ on ETTh1, from $0.157/0.249$ to $0.158/0.250$ on ECL, and from $0.240/0.267$ to $0.242/0.270$ on Weather. The absence of systematic degradation indicates that TVTS is empirically robust to substantial variable-order perturbations, suggesting that its forecasting performance does not critically depend on a particular channel ordering.

L Impact of Look-back Window Size.

The length of the historical look-back window determines the richness of the temporal context available for forecasting. To assess the capacity to exploit long-term dependencies, we analyze performance trends under varying look-back lengths $L \in \{48, 96, 336, 512\}$ on the ETTh1 and ETTm1 datasets. As illustrated in Figure 14 (numerical details in Table 19), TVTS consistently achieves the lowest error rates. A pivotal observation is the *monotonic improvement* of TVTS: as



(a) MSE on ETTh1 (b) MAE on ETTh1 (c) MSE on ETTm1 (d) MAE on ETTm1

Figure 14: Average forecasting performance with look-back windows $L \in \{48, 96, 336, 512\}$ and horizons $H \in \{96, 192, 336, 720\}$.

Table 18: Robustness of TVTS to variable permutation with $L = 96$. Each percentage denotes the fraction of variables randomly reordered at inference.

Datasets	Model	TVTS		20%		40%		60%		80%		100%	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.367	0.389	0.368	0.389	0.367	0.389	0.368	0.389	0.368	0.389	0.368	0.389
	192	0.421	0.419	0.421	0.419	0.423	0.421	0.422	0.421	0.423	0.421	0.423	0.420
	336	0.463	0.438	0.463	0.441	0.464	0.441	0.468	0.442	0.466	0.440	0.467	0.440
	720	0.461	0.457	0.462	0.458	0.464	0.459	0.465	0.460	0.467	0.459	0.466	0.458
	Avg	0.428	0.426	0.429	0.427	0.430	0.428	0.431	0.428	0.431	0.427	0.431	0.427
ECL	96	0.132	0.223	0.132	0.223	0.132	0.225	0.132	0.225	0.132	0.225	0.132	0.225
	192	0.148	0.240	0.149	0.240	0.150	0.243	0.148	0.240	0.150	0.242	0.149	0.241
	336	0.162	0.254	0.161	0.254	0.167	0.258	0.162	0.255	0.161	0.254	0.164	0.254
	720	0.187	0.279	0.190	0.280	0.188	0.280	0.186	0.280	0.188	0.280	0.187	0.280
	Avg	0.157	0.249	0.158	0.249	0.159	0.252	0.157	0.250	0.158	0.250	0.158	0.250
Weather	96	0.154	0.197	0.156	0.197	0.156	0.201	0.154	0.201	0.154	0.201	0.156	0.202
	192	0.203	0.244	0.203	0.244	0.204	0.245	0.204	0.245	0.205	0.246	0.204	0.250
	336	0.262	0.287	0.263	0.288	0.262	0.290	0.263	0.291	0.263	0.290	0.264	0.288
	720	0.342	0.338	0.342	0.338	0.344	0.341	0.344	0.341	0.346	0.341	0.344	0.340
	Avg	0.240	0.267	0.241	0.267	0.242	0.269	0.241	0.270	0.242	0.270	0.242	0.270

context expands to $L = 512$, our model continuously reduces error, whereas strong baselines (e.g., TimesNet, iTransformer) exhibit saturation or deterioration (a "U-shaped" trend). This divergence validates the effectiveness of our core designs against the limitations highlighted in the Introduction. First, conventional models relying on time-invariant filters or global attention often struggle with the local statistical shifts inherent in longer sequences. In contrast, our TVCF module dynamically modulates filter weights across the sequence, allowing it to adaptively capture evolving temporal patterns rather than applying a rigid, location-agnostic rule. Second, extended windows exacerbate the noise accumulation found in symmetric "all-to-all" structures. TVTS mitigates this via the DGC module, which employs direction-aware metric learning and Top-k sparsification. By constructing a latent anisotropic topology that focuses solely on informative, directed dependencies, TVTS effectively filters out the redundancy that typically degrade performance in long-sequence modeling.

M Sensitivity Analysis

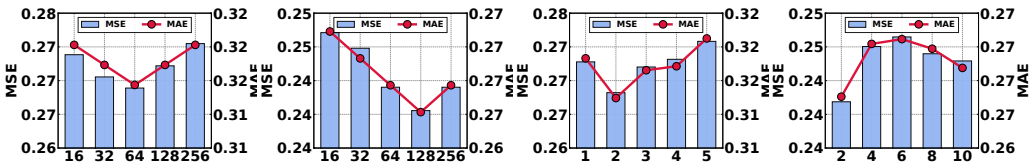
Extended Analysis on M and Top-k. We extend the sensitivity analysis of the spectral basis size M and sparsity threshold k to the ETTm2 and Weather datasets, as shown in Figure 15 (a-d). Corroborating our primary findings, performance metrics follow a distinct convex trajectory. This reaffirms that optimal forecasting demands balancing spectral resolution against noise (via M) and aligning the constructed topology with the dataset's intrinsic connectivity density (via k).

Impact of Balancing Coefficient β. We further investigate the coefficient β (Figure 15 e-h), which modulates the trade-off between time-domain reconstruction and frequency-domain consistency. A divergence in optimal β emerges across datasets:

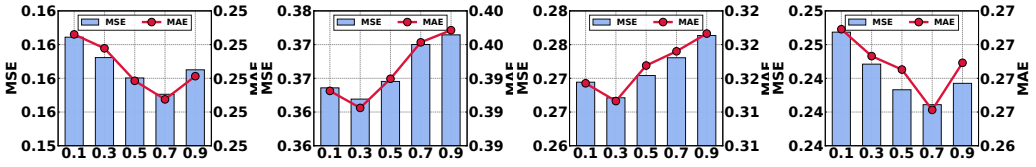
Table 19: Performance comparison across varying look-back windows $L \in \{48, 96, 336, 512\}$ and prediction horizons $H \in \{96, 192, 336, 720\}$. The best results are highlighted in **bold**. TVTS achieves the best MSE in 33 and best MAE in 37 out of 40 cases, with predictive errors progressively decreasing as the look-back window L expands.

L	Dataset	Model	TVTS		FilterNet		DLinear		iTrans.		PatchTST		TimesNet	
		Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
48	ETTh1	96	0.380	**0.393**	**0.379**	0.399	0.419	0.423	0.397	0.405	0.385	0.399	0.424	0.432
		192	0.463	0.441	0.447	0.431	0.469	0.455	0.456	0.440	**0.437**	**0.430**	0.479	0.462
		336	**0.484**	**0.455**	0.503	0.458	0.517	0.485	0.510	0.470	**0.484**	0.459	0.525	0.484
		720	**0.491**	**0.467**	0.496	0.471	0.544	0.526	0.524	0.498	0.517	0.495	0.496	0.477
		Avg	**0.455**	**0.439**	0.456	0.440	0.487	0.472	0.472	0.453	0.456	0.446	0.481	0.464
	ETTm1	96	**0.417**	**0.402**	0.448	0.416	0.503	0.448	0.475	0.428	0.457	0.421	0.481	0.419
		192	0.477	**0.431**	0.479	0.434	0.540	0.469	0.522	0.452	0.502	0.444	**0.467**	0.438
		336	**0.514**	**0.457**	0.532	0.465	0.583	0.497	0.569	0.479	0.566	0.480	0.528	0.468
		720	**0.555**	**0.483**	0.578	0.493	0.618	0.522	0.628	0.511	0.590	0.497	0.579	0.489
		Avg	**0.491**	**0.443**	0.509	0.452	0.561	0.484	0.549	0.468	0.529	0.461	0.514	0.454
336	ETTh1	96	**0.368**	**0.389**	0.379	0.404	0.384	0.405	0.389	0.402	0.401	0.405	0.398	0.418
		192	**0.404**	**0.413**	0.417	0.428	0.430	0.442	0.443	0.439	0.462	0.446	0.447	0.449
		336	**0.431**	**0.432**	0.437	0.443	0.447	0.448	0.479	0.470	0.509	0.474	0.493	0.468
		720	**0.444**	**0.463**	0.458	0.472	0.504	0.515	0.504	0.488	0.513	0.493	0.518	0.504
		Avg	**0.412**	**0.424**	0.423	0.437	0.441	0.453	0.454	0.450	0.471	0.455	0.464	0.460
	ETTm1	96	**0.289**	**0.340**	**0.289**	0.344	0.300	0.345	0.303	0.357	0.294	0.345	0.335	0.380
		192	0.332	**0.366**	**0.331**	0.369	0.336	**0.366**	0.345	0.383	0.334	0.371	0.358	0.388
		336	**0.361**	**0.386**	0.364	0.389	0.372	0.390	0.382	0.405	0.371	0.392	0.406	0.418
		720	0.422	0.422	0.425	0.423	0.427	0.423	0.443	0.439	**0.421**	**0.419**	0.449	0.443
		Avg	**0.351**	**0.379**	0.352	0.381	0.359	0.381	0.368	0.396	0.355	0.382	0.387	0.407

L	Dataset	Model	TVTS		FilterNet		DLinear		iTrans.		PatchTST		TimesNet	
		Metrics	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
96	ETTh1	96	**0.367**	**0.389**	0.382	0.402	0.386	0.400	0.386	0.405	0.414	0.419	0.384	0.402
		192	**0.421**	**0.419**	0.430	0.429	0.437	0.432	0.441	0.436	0.460	0.445	0.436	0.429
		336	**0.463**	**0.438**	0.472	0.451	0.481	0.459	0.487	0.458	0.501	0.466	0.491	0.469
		720	**0.461**	**0.457**	0.481	0.473	0.519	0.516	0.503	0.491	0.500	0.488	0.521	0.500
		Avg	**0.428**	**0.426**	0.441	0.439	0.456	0.452	0.454	0.447	0.469	0.454	0.458	0.450
	ETTm1	96	**0.313**	**0.350**	0.321	0.361	0.345	0.372	0.334	0.368	0.329	0.367	0.338	0.375
		192	**0.356**	**0.369**	0.367	0.387	0.380	0.389	0.377	0.391	0.367	0.385	0.374	0.387
		336	**0.384**	**0.397**	0.401	0.409	0.413	0.413	0.426	0.420	0.399	0.410	0.410	0.411
		720	**0.445**	**0.429**	0.477	0.448	0.474	0.453	0.491	0.459	0.454	0.439	0.478	0.450
		Avg	**0.375**	**0.386**	0.392	0.401	0.403	0.407	0.407	0.410	0.387	0.400	0.400	0.406
512	ETTh1	96	0.370	**0.383**	0.397	0.419	**0.365**	0.394	0.394	0.419	0.386	0.412	0.440	0.455
		192	0.407	0.410	0.426	0.438	**0.400**	0.418	0.429	0.443	0.445	0.457	0.488	0.488
		336	**0.421**	**0.428**	0.436	0.442	**0.428**	0.439	0.440	0.454	0.463	0.463	0.480	0.485
		720	**0.438**	**0.462**	0.470	0.481	0.470	0.492	0.529	0.525	0.691	0.588	0.574	0.537
		Avg	**0.409**	**0.421**	0.432	0.445	0.416	0.436	0.448	0.460	0.496	0.480	0.496	0.491
	ETTm1	96	**0.281**	**0.337**	0.330	0.371	0.303	0.346	0.303	0.359	0.299	0.353	0.338	0.377
		192	**0.328**	**0.365**	0.351	0.387	0.335	0.365	0.343	0.382	0.359	0.397	0.437	0.426
		336	**0.360**	**0.385**	0.380	0.404	0.365	0.385	0.377	0.402	0.370	0.399	0.408	0.426
		720	**0.417**	**0.418**	0.450	0.441	0.419	0.417	0.433	0.434	0.433	0.442	0.458	0.457
		Avg	**0.347**	**0.376**	0.378	0.401	0.356	0.378	0.364	0.394	0.365	0.398	0.410	0.422



(a) M on ETTm2 (b) M on Weather (c) Top-k on ETTm2 (d) Top-k on Weather



(e) β on ECL (f) β on ETTh2 (g) β on ETTm2 (h) β on Weather

Figure 15: Parameter sensitivity analysis regarding basis size M, sparsity Top-k, and balancing coefficient β on the four datasets. Performance is averaged over horizons $H \in \{96, 192, 336, 720\}$ with input length $L = 96$. See Table 21 and Table 22 for full results.

- **Lower-dimensional/Stable Regimes (e.g., ETTh2, ETTm2):** Performance peaks at a lower $\beta \approx 0.3$. This indicates that for series with fewer variates and stable patterns, direct time-domain supervision provides sufficient gradients for convergence, acting as the primary robust objective.
- **High-dimensional/Complex Regimes (e.g., ECL, Weather):** The model favors a higher $\beta \approx 0.7$. These datasets are characterized by intricate periodicities and high-frequency transient fluctuations. Here, stronger spectral supervision becomes critical, guiding the model to capture global trends and rapid variations that are often obscured in purely time-domain objectives.

Therefore, adaptively selecting β based on dataset is crucial for maximizing prediction gains.

N Computational Efficiency Analysis.

Table 23 presents a rigorous computational profiling of TVTS vis-à-vis state-of-the-art baselines, scrutinizing critical metrics including parameter efficiency, FLOPs, and execution costs (training/inference time and memory footprint). Empirical evidence indicates that TVTS achieves a superior equilibrium

Table 20: Performance comparison with clustering-based baselines (DUET Qiu et al. (2025), BIM[3] Gao et al. (2025)). The look-back length L is uniformly set to 96.

Dataset		ETTm1					ETTm2					ETTh1					ETTh2					ECL					Weather					1st
Model	Metric	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg	96	192	336	720	Avg	
DUET (2025)	MSE	0.324	0.369	0.404	0.463	0.390	0.174	0.243	0.304	0.399	0.280	0.377	0.429	0.471	0.496	0.443	0.296	0.368	0.411	0.412	0.372	0.145	0.163	0.175	0.204	0.172	0.163	0.218	0.274	0.349	0.251	0
	MAE	0.354	0.379	0.402	0.437	0.393	0.255	0.302	0.341	0.397	0.324	0.393	0.425	0.446	0.480	0.436	0.345	0.389	0.422	0.434	0.398	0.233	0.248	0.262	0.291	0.259	0.202	0.252	0.294	0.343	0.273	0
BIM[3] (2025)	MSE	0.316	0.371	0.392	0.468	0.387	0.176	0.265	0.293	0.397	0.283	0.369	0.432	0.475	0.476	0.438	0.287	0.390	0.414	0.431	0.381	0.144	0.162	0.174	0.204	0.171	**0.153**	**0.203**	0.264	0.345	0.241	2
	MAE	0.351	0.378	0.391	0.436	0.389	0.259	0.320	0.331	0.392	0.326	0.392	0.425	0.445	0.468	0.433	0.336	0.398	0.427	0.442	0.401	0.231	0.247	0.259	0.292	0.257	**0.195**	**0.242**	0.289	0.341	0.267	2
TVTS (Ours)	MSE	**0.313**	**0.356**	**0.384**	**0.445**	**0.375**	**0.170**	**0.233**	**0.289**	**0.387**	**0.270**	**0.367**	**0.421**	**0.463**	**0.461**	**0.428**	**0.281**	**0.355**	**0.405**	**0.410**	**0.363**	**0.132**	**0.148**	**0.162**	**0.187**	**0.157**	0.154	**0.203**	**0.262**	**0.342**	**0.240**	**29**
	MAE	**0.350**	**0.369**	**0.397**	**0.429**	**0.386**	**0.250**	**0.291**	**0.330**	**0.387**	**0.315**	**0.389**	**0.419**	**0.438**	**0.457**	**0.426**	**0.332**	**0.379**	**0.416**	**0.431**	**0.390**	**0.223**	**0.240**	**0.254**	**0.279**	**0.249**	0.197	0.244	**0.287**	**0.338**	**0.267**	**28**

Table 21: Detailed data presentation of parameter (M, β) sensitivity analysis in the TVTS model.

		Impact of M										Impact of β									
Data	Model	16		32		64		128		256		0.1		0.3		0.5		0.7		0.9	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.172	0.251	**0.170**	**0.250**	**0.170**	**0.250**	0.172	0.252	0.176	0.252	0.171	0.252	**0.170**	**0.250**	**0.170**	0.252	0.171	0.252	0.173	0.253
	192	0.235	0.295	0.235	0.294	**0.233**	**0.291**	0.235	0.293	0.237	0.295	0.235	0.292	**0.233**	**0.291**	0.237	0.295	0.236	0.295	0.237	0.295
	336	0.296	0.333	0.290	0.331	**0.289**	**0.330**	0.291	0.331	0.294	0.332	0.291	0.331	**0.289**	**0.330**	0.293	0.332	0.296	0.334	0.299	0.335
	720	0.388	0.389	**0.387**	**0.387**	0.388	0.388	0.389	0.389	0.390	0.389	0.389	0.388	**0.387**	**0.387**	0.389	0.389	0.394	0.391	0.398	0.394
	Avg	0.273	0.317	0.271	0.316	**0.270**	**0.315**	0.272	0.316	0.274	0.317	0.272	0.316	**0.270**	**0.315**	0.272	0.317	0.274	0.318	0.277	0.319
ETTh2	96	0.284	0.332	**0.281**	**0.332**	0.286	0.336	0.288	0.337	0.290	0.335	0.282	0.333	**0.281**	**0.332**	0.284	0.334	0.288	0.337	0.291	0.339
	192	0.356	0.380	**0.355**	**0.379**	0.357	0.381	0.357	0.381	0.359	0.382	**0.355**	**0.379**	0.356	0.380	0.357	0.381	0.362	0.385	0.363	0.385
	336	0.407	0.418	**0.405**	**0.416**	0.407	0.418	0.409	0.419	0.408	0.420	0.408	0.418	**0.405**	**0.416**	0.409	0.422	0.420	0.429	0.421	0.429
	720	0.419	0.437	**0.410**	**0.431**	0.413	0.433	0.416	0.433	0.420	0.435	0.414	0.436	**0.410**	**0.431**	0.413	0.434	0.416	0.435	0.417	0.438
	Avg	0.367	0.392	**0.363**	**0.390**	0.366	0.392	0.368	0.393	0.369	0.393	0.365	0.392	**0.363**	**0.390**	0.366	0.393	0.372	0.397	0.373	0.398
ECL	96	0.137	0.232	0.135	0.227	0.132	0.224	**0.132**	**0.223**	0.134	0.227	0.135	0.230	0.134	0.227	0.133	0.225	**0.132**	**0.223**	0.134	0.224
	192	0.152	0.244	0.150	0.242	0.149	0.241	**0.148**	**0.240**	0.152	0.248	0.152	0.244	0.152	0.244	0.149	0.241	**0.148**	**0.240**	0.149	0.241
	336	0.169	0.259	0.165	0.256	0.165	0.256	0.164	0.256	**0.162**	**0.254**	**0.162**	**0.254**	0.163	0.255	0.164	0.256	0.164	0.257	0.166	0.258
	720	0.196	0.285	0.194	0.286	0.191	0.282	**0.187**	**0.279**	0.188	0.281	0.196	0.285	0.191	0.284	0.189	0.281	**0.187**	**0.279**	0.188	0.281
	Avg	0.164	0.255	0.161	0.253	0.159	0.251	**0.158**	**0.250**	0.159	0.253	0.161	0.253	0.160	0.253	0.159	0.251	**0.158**	**0.250**	0.159	0.251
Weather	96	0.160	0.202	0.159	0.199	0.157	0.199	**0.154**	**0.197**	0.155	0.198	0.166	0.203	0.161	0.202	0.157	0.199	**0.154**	**0.197**	0.156	0.199
	192	0.214	0.251	0.213	0.249	0.208	0.247	**0.203**	**0.244**	0.209	0.248	0.216	0.253	0.212	0.252	0.205	0.248	**0.203**	**0.244**	0.207	0.249
	336	0.275	0.294	0.273	0.294	0.265	0.288	**0.262**	**0.287**	0.266	0.289	0.269	0.293	0.266	0.291	0.265	0.293	**0.262**	**0.287**	0.264	0.292
	720	0.349	0.345	0.346	0.343	0.343	0.341	**0.342**	**0.338**	**0.342**	**0.339**	0.345	0.342	**0.342**	**0.338**	0.342	0.339	0.343	0.339	0.345	0.341
	Avg	0.250	0.273	0.248	0.271	0.243	0.269	**0.240**	**0.267**	0.243	0.269	0.249	0.273	0.245	0.271	0.242	0.270	**0.241**	**0.267**	0.243	0.270

between predictive performance and computational frugality, maintaining a linear complexity profile comparable to lightweight linear models such as DLinear and TimeMixer, while significantly outperforming complex Transformer-based architectures. This efficiency is intrinsically attributed to the model's formulation as a **Linear Operator**, which circumvents the quadratic bottleneck of canonical attention mechanisms. Crucially, time-varying weights can be learned during training and are static operators for deployment during inference. Specifically, the TVCF leverages the Graph Fourier Transform to expeditiously capture non-stationary temporal dynamics in the spectral domain without incurring heavy recurrent overheads. Furthermore, the DGC module optimizes multivariate modeling by capturing inter-variable dependencies through sparse clustering, effectively mitigating dimensional redundancy. Consequently, TVTS demonstrates exceptional scalability, particularly on high-dimensional datasets like ECL and Traffic, proving its prowess in modeling long-range dependencies with minimal resource consumption.

Table 22: Detailed data presentation of parameter (Top-k) sensitivity analysis in the TVTS model.

Dataset	Model	K=1		K=2		K=3		K=4		K=5		Dataset	Model	K=1		K=2		K=3		K=4		K=5	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE		Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm2	96	0.172	0.251	**0.170**	**0.250**	0.171	0.251	0.171	0.252	0.173	0.251	ETTh2	96	0.283	0.332	**0.281**	**0.332**	0.285	0.335	0.286	0.335	0.290	0.336
	192	0.235	0.295	**0.233**	**0.291**	0.233	0.292	0.235	0.293	0.236	0.295		192	0.356	0.380	**0.355**	**0.379**	0.356	0.382	0.355	0.380	0.359	0.384
	336	0.296	0.333	**0.289**	**0.330**	0.292	0.331	0.292	0.331	0.295	0.331		336	0.408	0.419	**0.405**	**0.416**	0.406	0.417	0.414	0.428	0.407	0.424
	720	0.388	0.389	**0.387**	**0.387**	0.393	0.391	0.394	0.390	0.395	0.396		720	0.419	0.436	**0.410**	**0.431**	0.414	0.434	0.414	0.436	0.420	0.438
	Avg	0.273	0.317	**0.270**	**0.315**	0.272	0.316	0.273	0.317	0.275	0.318		Avg	0.367	0.392	**0.363**	**0.390**	0.365	0.392	0.367	0.395	0.369	0.396

Dataset	Metric	K=5		K=10		K=15		K=20		K=25		Dataset	Metric	K=2		K=4		K=6		K=8		K=10	
ECL	96	0.137	0.231	**0.132**	**0.223**	0.133	0.224	0.137	0.232	0.136	0.229	Weather	96	0.156	0.201	0.163	0.199	0.168	0.205	0.159	0.205	**0.154**	**0.197**
	192	0.151	0.242	0.149	0.240	**0.148**	**0.240**	**0.148**	**0.240**	0.152	0.246		192	**0.203**	**0.244**	0.209	0.247	0.208	0.245	0.209	0.245	0.210	0.249
	336	0.169	0.259	0.164	**0.254**	0.163	**0.254**	**0.162**	**0.254**	0.165	0.257		336	**0.262**	**0.287**	0.269	0.292	0.266	0.288	0.268	0.288	0.267	0.289
	720	0.190	0.281	**0.187**	**0.279**	0.189	0.281	0.194	0.286	0.196	0.286		720	**0.342**	**0.338**	0.345	0.343	0.348	0.344	0.347	0.342	0.347	0.339
	Avg	0.162	0.253	**0.158**	**0.249**	**0.158**	0.250	0.160	0.253	0.162	0.255		Avg	**0.241**	**0.268**	0.247	0.270	0.248	0.271	0.246	0.270	0.245	0.269

Table 23: Computational efficiency evaluation of TVTS against state-of-the-art baselines across five real-world datasets. We report key resource metrics including parameter counts, FLOPs, and execution costs for both training and inference phases (denoted as T.T./I.T. for time and T.M./I.M. for memory). Experiments are conducted with fixed look-back and horizon windows of $L = H = 96$ and a batch size of 16, while other hyperparameters for baseline models follow their official implementations.

	Models	TVTS	DUET	FilterTS	CARD	Crossfm.	iTrans	Fredfm.	TimeMix.	DLinear	TimesNet	PatchTST	FreTS
ETTh1	Params(M)	1.81	4.75	0.68	0.03	36.97	4.83	8.59	0.08	0.02	299.94	3.76	0.42
	FLOPS(M)	20.54	32.75	8.13	1.41	1,507.47	33.99	135.01	10.72	0.13	289,708.00	272.20	2.94
	T.T.(ms/iter)	16.16	25.33	26.60	29.10	90.05	10.11	26.48	27.99	1.42	501.98	8.37	3.63
	T.M.(GB)	0.03	0.08	0.05	0.03	1.11	0.21	0.24	0.05	0.02	5.80	0.14	0.03
	I.T.(ms/iter)	3.28	8.18	9.74	4.05	24.90	1.72	3.73	3.86	0.21	155.97	1.38	0.49
	I.M.(MB)	8.60	22.48	20.38	11.02	201.63	156.00	78.04	18.06	8.48	1,396.24	51.83	14.47
ECL	Params(M)	1.88	8.96	2.75	1.39	9.68	4.83	12.12	0.12	0.02	300.58	3.76	0.42
	FLOPS(GB)	0.67	3.39	0.41	5.07	13.54	1.87	5.55	1.74	0.01	293.98	12.48	0.13
	T.T.(ms/iter)	33.27	226.29	51.96	78.25	587.51	10.66	26.59	57.09	1.37	509.17	65.83	3.78
	T.M.(GB)	0.44	1.75	1.13	5.08	13.15	0.70	1.24	4.24	0.04	5.82	3.02	0.30
	I.T.(ms/iter)	13.26	65.33	25.66	28.21	192.42	3.46	7.08	22.79	0.20	157.13	20.52	0.48
	I.M.(GB)	0.20	1.26	0.47	0.63	0.94	0.23	0.26	0.82	0.02	1.41	0.93	0.23
Exchange	Params(M)	1.81	6.86	0.67	1.39	8.40	4.83	8.59	0.08	0.02	299.94	3.76	0.42
	FLOPS(M)	18.21	53.98	9.29	114.83	380.79	38.88	135.01	12.25	0.15	287,909.00	311.08	3.36
	T.T.(ms/iter)	12.62	30.79	27.83	35.57	87.58	10.14	28.56	27.46	1.47	505.39	9.75	3.51
	T.M.(GB)	0.03	0.12	0.05	0.17	0.47	0.21	0.24	0.06	0.02	5.80	0.15	0.03
	I.T.(ms/iter)	2.95	6.88	6.00	6.02	24.62	1.70	6.86	3.81	0.21	156.33	2.47	0.48
	I.M.(GB)	0.01	0.03	0.02	0.98	0.07	0.16	0.08	0.02	0.01	1.40	0.05	0.02
Traffic	Params(M)	3.49	11.07	7.72	0.98	14.65	4.83	11.09	0.12	0.02	301.69	3.76	0.42
	FLOPS(GB)	4.32	15.60	1.36	10.47	51.52	6.45	14.78	4.66	0.02	288.72	33.52	0.36
	T.T.(s/iter)	0.24	9.93	0.17	0.18	1.81	0.04	0.10	0.17	0.00	5.86	0.17	0.01
	T.M.(GB)	3.47	11.34	3.04	9.50	10.49	2.95	5.76	11.33	0.07	5.86	8.01	0.78
	I.T.(ms/iter)	56.77	1,683.16	80.83	61.21	151.24	13.45	31.60	67.21	0.20	15.64	55.10	3.48
	I.M.(GB)	1.23	8.81	1.21	1.69	0.75	1.27	1.84	2.20	0.04	1.42	2.44	0.59
Weather	Params(M)	1.85	2.13	0.75	0.03	11.21	4.83	0.50	0.10	0.02	299.97	3.76	0.42
	FLOPS(GB)	0.04	0.05	0.02	0.04	1.29	0.10	0.01	0.05	0.00	291.21	0.82	0.01
	T.T.(ms/iter)	19.04	18.01	27.85	28.69	91.59	9.45	37.85	33.30	1.38	505.88	9.89	3.53
	T.M.(GB)	0.04	0.05	0.09	0.05	1.21	0.22	0.05	0.16	0.02	5.80	0.02	0.04
	I.T.(ms/iter)	4.28	5.48	6.00	4.11	29.73	1.66	7.19	5.10	0.20	156.27	2.96	0.48
	I.M.(MB)	10.16	17.05	39.97	14.45	125.18	34.88	18.28	36.03	9.06	1,397.18	90.78	24.37