MDMLP-EIA: Multi-domain Dynamic MLPs with Energy Invariant Attention for Time Series Forecasting

Hu Zhang [1,2], **Zhien Dai** [3*], **Zhaohui Tang** [3], **Yongfang Xie** [3]

[1]College of Computer Science and Engineering, Changsha University, China
[2]School of Mathematics and Statistics, Central South University, China
[3]School of Automation, Central South University, China
zhanghu@csu.edu.cn, zhiendai@csu.edu.cn, zhtang@csu.edu.cn, yfxie@csu.edu.cn

Abstract

Time series forecasting is essential across diverse domains. While MLP-based methods have gained attention for achieving Transformer-comparable performance with fewer parameters and better robustness, they face critical limitations including loss of weak seasonal signals, capacity constraints in weight-sharing MLPs, and insufficient channel fusion in channel-independent strategies. To address these challenges, we propose MDMLP-EIA (Multi-domain Dynamic MLPs with Energy Invariant Attention) with three key innovations. First, we develop an adaptive fused dual-domain seasonal MLP that categorizes seasonal signals into strong and weak components. It employs an adaptive zero-initialized channel fusion strategy to minimize noise interference while effectively integrating predictions. Second, we introduce an energy invariant attention mechanism that adaptively focuses on different feature channels within trend and seasonal predictions across time steps. This mechanism maintains constant total signal energy to align with the decomposition-prediction-reconstruction framework and enhance robustness against disturbances. Third, we propose a dynamic capacity adjustment mechanism for channel-independent MLPs. This mechanism scales neuron count with the square root of channel count, ensuring sufficient capacity as channels increase. Extensive experiments across nine benchmark datasets demonstrate that MDMLP-EIA achieves state-of-the-art performance in both prediction accuracy and computational efficiency.

Code — https://github.com/zh1985csuccsu/MDMLP-EIA
Appendix — https://arxiv.org/abs/2511.09924

Introduction

Time series forecasting (TSF) has extensive applications in air quality monitoring (Zheng et al. 2015), traffic flow prediction (Yin et al. 2021), energy management (Amasyali and El-Gohary 2018), economic analysis (Zhang, Aggarwal, and Qi 2017), and industrial indicator forecasting (Zhang et al. 2025). In recent years, deep learning methods for time series prediction have received widespread research attention (Shao et al. 2024). These methods include RNN-based approaches (e.g., LSTNet(Lai et al. 2018), Witran (Jia et al. 2023), DAN (Li, Xu, and Anastasiu 2024)), CNN-based methods (e.g., SciNet (Liu et al. 2022), TimesNet (Wu et al. 2022), Moderntcn (Luo and Wang 2024)), GNN-based methods (e.g., StemGNN (Cao et al. 2020), FourierGNN (Yi et al. 2023a), Msgnet (Cai et al. 2024)), and Transformer-based methods (e.g., Informer (Zhou et al. 2021), PatchTST (Nie et al. 2022), Crossformer (Zhang and Yan 2023), Ada-MSHyper (Shang et al. 2024)).

These deep learning models achieve good prediction results in certain scenarios, but typically suffer from complex structures, slow training/inference speeds, and reduced robustness due to numerous parameters, particularly with limited training data (Zhou et al. 2021). DLinear (Zeng et al. 2023) addresses these issues by introducing a simple single-layer linear model that achieves Transformer-comparable prediction accuracy with fewer parameters. This success has sparked extensive research into MLP-based time series prediction methods, including FreTS (Yi et al. 2023b), CycleNet (Lin et al. 2024a), FilterNet (Yi et al. 2024), and Amplifier (Fei et al. 2025). These methods enhance temporal feature representation and improve MLP performance through trend-seasonal decomposition and frequency domain signal processing. However, several limitations still remain: ***1) Loss of weak seasonal signal***: Softshrink is commonly employed to reduce noise in the frequency domain (Yi et al. 2023b). However, this approach may inadvertently remove weak seasonal signals that have low energy but contain valuable cyclical information for time series prediction (see Figure 1(a)). Although Amplifier (Fei et al. 2025) attempts to address the loss of weak seasonal signals by transferring high-energy components from low-frequency to high-frequency regions through spectrum flipping, it simultaneously amplifies unwanted noise along with useful weak seasonal signals. ***2) Capacity limitation of weight-sharing MLPs and insufficient channel fusion for channel-independent strategy***: The channel-independent strategy employs weight sharing to merge values from different time steps across each channel dimension, and it is considered as a robust prediction strategy with fewer model parameters (Han, Ye, and Zhan 2024). As the channel-independent MLPs need to handle temporal fusion of features across multiple different channels, their capacity is closely related to the fusion performance. However, current channel-independent MLPs usually use fixed capacity, which limits their performance when processing tasks with varying numbers of channels, as shown in Figure 1(b).

*Corresponding author

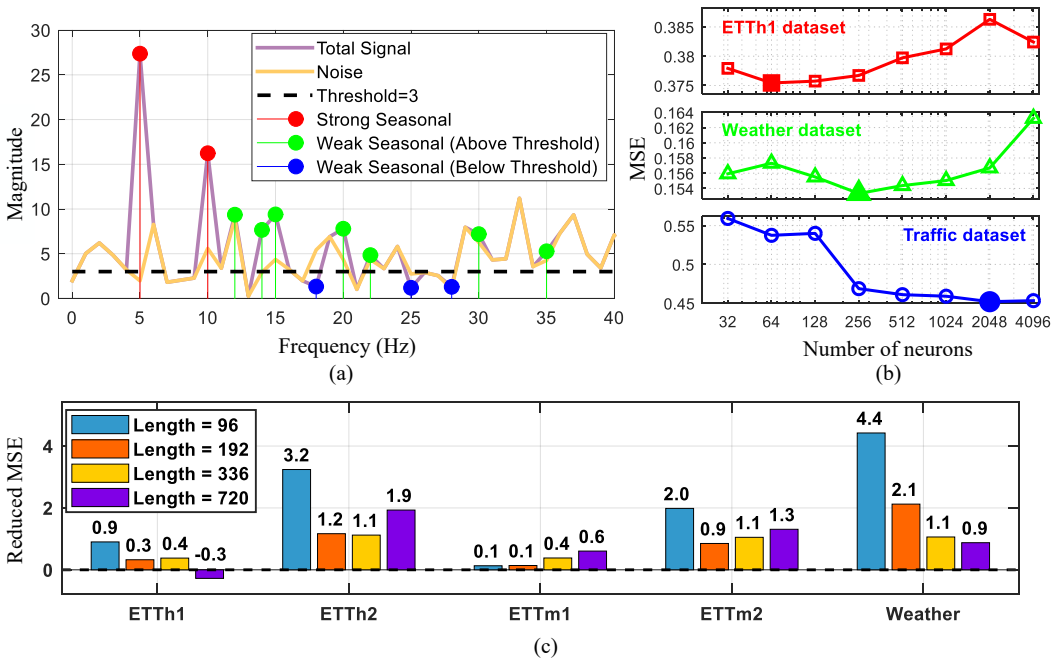


Figure 1: Limitations of current MLP-based methods. (a) Loss of weak seasonal signals: Weak seasonal signals closely resemble noise signals and are difficult to distinguish from them. When applying frequency domain amplitude restrictions to reduce noise, some weak seasonal signals are inevitably lost. (b) Different capacity requirements of channel-independent MLPs for prediction tasks with varying numbers of channels: For prediction tasks with larger numbers of channels (such as Traffic), MLPs require greater capacity to meet the predictive demands of each channel. Conversely, for prediction tasks with fewer channels (such as ETTh1), MLPs require smaller capacity to prevent model overfitting. (c) Reduced Mean Squared Errors (MSEs) achieved by the proposed energy invariant attention at varied prediction lengths in some datasets.

Furthermore, the channel-independent strategy may result in insufficient channel fusion for TSF, particularly for popular forecasting methods via seasonal-trend decomposition (Fei et al. 2025)(Wang et al. 2024)(Stitsyuk and Choi 2025)(Wu et al. 2021a)(Zhou et al. 2022). These methods directly sum trend and seasonal predictions to generate final outputs, potentially limiting the model's ability to represent different channels across time steps, as shown in Figure 1(c).

To address these issues, we propose MDMLP-EIA with the following main contributions:

1) To effectively capture weak seasonal signals while minimizing noise interference, we design an ***adaptive fused dual-domain seasonal MLP***. We partition seasonal signals into strong and weak components, processing them through frequency-domain MLP and standard MLP respectively. Especially, since both noise and weak seasonal signals exhibit similar low energy characteristics in the frequency domain, we develop an ***Adaptive Zero-initialized Channel Fusion (AZCF) strategy*** that selectively integrates strong and weak seasonal predictions while suppressing noise contamination. The idea is different with the spectrum flipping in Amplifier (Fei et al. 2025), which transfers high-energy components from low-frequency to high-frequency regions and unavoidably amplifies unwanted noise along with useful weak seasonal signals.

2) We propose a ***dynamic capacity adjustment (DCA) mechanism*** for channel-independent MLPs. This mechanism adjusts neuron count based on the square root of channel numbers, ensuring adequate MLP capacity scaling. For small channel counts, minimal inter-channel redundancy requires rapid capacity growth; for large channel counts, increased redundancy permits slower growth, preventing parameter explosion and overfitting.

3) To address insufficient channel fusion, we introduce an ***energy invariant attention mechanism (EIA)*** that adaptively weights trend and seasonal predictions while maintaining theoretical consistency with decomposition frameworks. Unlike conventional weighted fusion that may suffer from energy inconsistency and signal distortion, EIA preserves total signal energy, ensuring stable fusion and enhanced robustness.

4) Nine benchmark experiments achieve state-of-the-art accuracy and efficiency.

Related Works

MLP-based Forecasting Methods N-BEATS (Oreshkin et al. 2019) stacks multiple fully connected layers with forward and backward residuals for univariate TSF. LightTS (Zhang et al. 2022) applies an MLP-based structure on top of two delicate down-sampling strategies for fast TSF. DEPTS (Fan et al. 2022) employs local and seasonal MLP modules to capture local trends and global periodicity in time series, respectively. NHITS (Challu et al. 2023) introduces

multi-rate data sampling and hierarchical interpolation to focus on time series components of varying frequencies. TimeMixer (Chen et al. 2023) uses feature-dimension and time-dimension weight-sharing MLPs for efficient time series prediction with minimal parameters. DLinear (Zeng et al. 2023) designs a simple single-layer linear model to establish dynamic relationships between input and output time series. SOFTS (Han et al. 2024) employs a centralized structure to extract global key representations distributed to individual channels, effectively capturing channel correlations.

Forecasting in the Frequency Domain FITS (Xu, Zeng, and Xu 2023) introduces a low-pass filter and implements MLP interpolation in the frequency domain for efficient time series prediction. FilterNet (Yi et al. 2024) designs plain and contextual shaping filters to capture key temporal patterns in time series. CycleNet (Lin et al. 2024a) proposes a residual cycle forecasting technique that iteratively extracts periodic patterns from time series. Peri-midFormer (Wu et al. 2024) decomposes time series by cycle and utilizes a pyramid structure to merge and predict signals of different periodicities. SparseTSF (Lin et al. 2024b) simplifies time series prediction to cross-period trend prediction by decomposing series into sub-sequences through fixed-period downsampling. FreTS (Yi et al. 2023b) uses a frequency-domain MLP with dual frequency learners across channel and temporal dimensions for time series representation and prediction.

Forecasting via Seasonal-trend Decomposition Autoformer (Wu et al. 2021a) replaces self-attention with autocorrelation for trend-seasonal decomposition. FEDformer (Zhou et al. 2022) utilizes frequency enhanced blocks and attention to improve time series representation in the frequency domain. TimeMixer (Wang et al. 2024) applies multiscale downsampling to time series signals followed by trend-seasonal decomposition and prediction for each downsampled series. Amplifier (Fei et al. 2025) enhances low-energy components through spectrum flipping during frequency-domain processing. xPatch (Stitsyuk and Choi 2025) utilizes patching and channel-independence techniques with Exponential Moving Average (EMA) for seasonal-trend decomposition in time series prediction.

MDMLP-EIA

Figure 2 illustrates our proposed MDMLP-EIA framework (*Appendix A shows the overall pseudocode*). Let *L* denote input historical timesteps, *Q* future timesteps, and *C* channels per timestep. The input series $x \in \mathbb{R}^{L\times C}$ undergoes RevIN normalization, and EMA decomposes *x* into trend component x_1 and seasonal component x_2. Then, a trend MLP generates trend prediction y_1 from x_1 (*See Appendix B for details*). Meanwhile, our proposed ***1) adaptive fused dual-domain seasonal MLP*** produces seasonal prediction y_2 from x_2. After that, an ***2) energy-invariant attention*** module is designed to integrate y_1 and y_2. Finally, the output series $y \in \mathbb{R}^{Q\times C}$ is obtained by inverse RevIN normalization. All the above multi-domain MLPs are channel-independent and adopt our proposed ***3) DCA mechanism***.

Adaptive Fused Dual-Domain Seasonal MLP

The adaptive fused dual-domain seasonal MLP uses two parallel networks: one processes strong seasonal signals, the other extracts weak seasonal patterns.

Frequency MLP and Strong Seasonal MLP Unlike xPatch (Stitsyuk and Choi 2025) which uses CNN, we employ a frequency-temporal learner (Yi et al. 2023b) with embedding, real FFT, frequency MLP, and inverse real FFT to reconstruct features with strong seasonal patterns $f_{s1} \in \mathbb{R}^{C\times L\times E}$. A strong seasonal MLP then learns predictions $y_{21} \in \mathbb{R}^{Q\times C}$ from f_{s1}. *See Appendix C for details*.

Weak Seasonal MLP The frequency-temporal learner (Yi et al. 2023b) uses softshrink processing to reduce noise but inevitably loses weak seasonal signals with low amplitude. Therefore, we construct a weak seasonal MLP (*detailed in Appendix D*) to extract weak seasonal predictions $y_{22} \in \mathbb{R}^{Q\times C}$ from seasonal signal $x_2 \in \mathbb{R}^{C\times L}$.

AZCF Mechanism Existing time-series signal fusion methods typically use attention mechanisms to weight signals across channel and temporal dimensions. Given strong seasonal prediction y_{21} and weak seasonal prediction y_{22}, the complete seasonal prediction y_2 is calculated as

$$y_2 = \alpha_1 \odot y_{21} + \alpha_2 \odot y_{22}, \tag{1}$$

where $\alpha_1, \alpha_2 \in \mathbb{R}^{Q\times C}$ are obtained indirectly through Query, Key, and Value transformations. However, this indirect approach introduces unnecessary complexity and neglects the significant signal-to-noise ratio differences between strong and weak seasonal predictions. Thus, we propose AZCF mechanism with three key characteristics:

Single-parameter fusion. We introduce an adaptive weight coefficient α to modulate the weak seasonal prediction's contribution:

$$y_2 = y_{21} + \alpha \odot y_{22}, \tag{2}$$

where $\alpha \in \mathbb{R}^{Q\times C}$. This single-parameter fusion first ensures strong signal quality, then selectively enhances weak signals while reducing model parameters.

Channel-dimension fusion. Since both y_{21} and y_{22} are obtained through channel-independent methods, the fusion strategy should respect this separation by making independent decisions along the channel dimension, learning which features have more reliable weak seasonal signals. Fusion along the time dimension would disrupt seasonal patterns' temporal coherence and may propagate noise across channels. Therefore, we define:

$$\alpha = \{\alpha_1, \alpha_2, ..., \alpha_C\} \in \mathbb{R}^{1\times C} \tag{3}$$

and expand it to $\alpha \in \mathbb{R}^{Q\times C}$ by broadcasting across the output time step dimension:

$$\alpha[q, c] = \alpha[c], \quad \forall q \in \{1, 2, ..., Q\}. \tag{4}$$

Zero initialization. We initialize α as a zero vector, starting from the most reliable strong seasonal prediction:

$$\alpha_{\text{init}} = \mathbf{0} \in \mathbb{R}^{1\times C}. \tag{5}$$

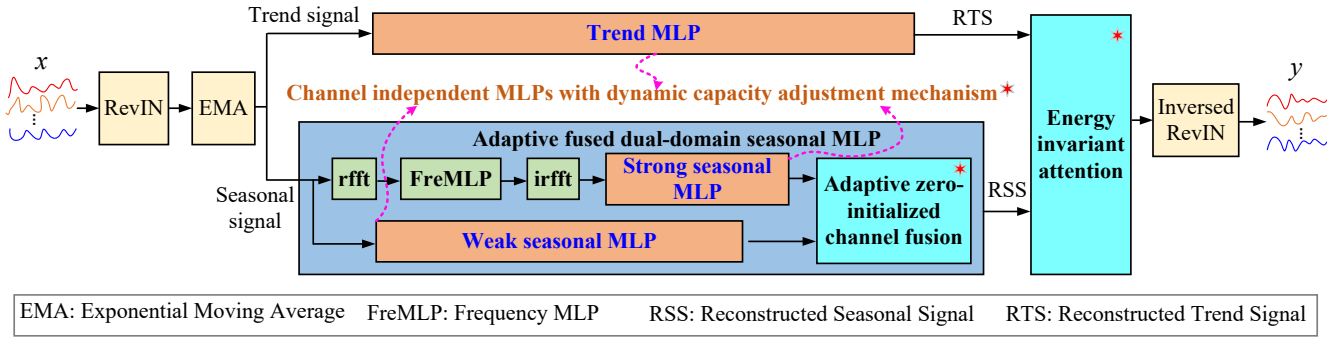


Figure 2: MDMLP-EIA overall architecture. (i) RevIN normalization and EMA decompose input series into trend and seasonal components; (ii) Trend component feeds into trend MLP; seasonal component processes through our **adaptive fused dual-domain seasonal MLP** with **adaptive zero-initialized channel fusion**); (iii) Our **energy invariant attention module** merges trend and seasonal predictions; (iv) Our **Dynamic capacity adjustment mechanism** optimizes multi-domain MLPs.

This gives us the initial prediction:

$$y_2^{\text{init}} = y_{21} + \alpha_{\text{init}} \odot y_{22} = y_{21}. \tag{6}$$

Equations (2) and (6) together essentially constitute a ***progressive learning strategy***. This strategy starts from a reliable foundation and allows α to gradually increase only when weak seasonal predictions genuinely improve results, preventing over-reliance on potentially noisy signals during early training. Compared to conventional attention mechanisms, our method requires only C parameters versus significantly more (typically $O(C^2)$ for Query, Key, Value transformations). With C ranging from tens to hundreds, this represents substantial reduction, making our model more lightweight and reducing overfitting risk. *See Appendix E for theoretical proof.*

EIA Mechanism

Conventional decomposition-based methods typically generate the final prediction y_3 by directly summing the trend prediction y_1 and seasonal prediction y_2 ($y_3 = y_1 + y_2$). However, they may limit the model's adaptive capacity to handle varying temporal dynamics and feature importance across channels and time steps.

Normalized Attention Fusion A possible solution is to introduce normalized attention for adaptive component fusion using weighted combination:

$$y_3 = \beta \odot y_1 + (1 - \beta) \odot y_2, \tag{7}$$

where $\beta \in \mathbb{R}^{Q \times C}$ represents the learnable attention weight for y_1, and $\odot$ denotes element-wise multiplication. Since $\beta + (1 - \beta) = 1$ for any $\beta \in [0, 1]$, $\beta \odot y_1 + (1 - \beta) \odot y_2$ represents a convex combination providing enhanced expressiveness by adaptively balancing trend and seasonal contributions. However, this creates **energy inconsistency challenges**. While time series decomposition theory reconstructs signals through $x = x_1 + x_2$ (preserving total magnitude), learned weights make signal magnitude variable and β-dependent. This will cause: 1) ***amplification of reconstruction errors*** with extreme β values, 2) ***inconsistent signal scaling*** across time steps and channels, and ***3) noise amplification*** from uncontrolled magnitude variations that violate decomposition framework guarantees.

EIA To address these limitations while preserving normalized attention fusion capabilities, we propose *EIA* that generates the fused prediction $y_3 \in \mathbb{R}^{Q \times C}$ as:

$$y_3 = 2 \times (\beta \odot y_1 + (1 - \beta) \odot y_2), \tag{8}$$

where $2\times$ ensures energy invariance by compensating for implicit normalization in the convex combination, maintaining consistent signal magnitude regardless of β values. The proposed EIA mechanism provides: 1) ***Perfect reconstruction preservation***: When $\beta = 0.5$, it reduces to $y_3 = y_1 + y_2$, matching traditional decomposition methods; 2) ***Stable adaptive fusion***: When $\beta \neq 0.5$, the model adaptively emphasizes components while maintaining constant energy, avoiding signal amplification/attenuation issues in standard weighted fusion. This ***energy conservation property*** inherently provides noise suppression and acts as a natural regularizer: while allowing adaptive emphasis on different components, it prevents excessive amplification that could propagate noise. Meanwhile, the complementary weights ensure that information from both components is always preserved, creating robust fusion that is less sensitive to individual component errors. *See Appendix F for detailed β calculation and theoretical proof.*

DCA Mechanism

As shown in the middle of Figure 2, the trend MLP, strong seasonal MLP, and weak seasonal MLP all employ a DCA mechanism that links the number of neurons *n* in the MLP to the number of channels *C* through a dynamic adjustment coefficient *cof*:

$$cof = \lfloor \sqrt{C}/\tau \rfloor, \tag{9}$$

where $\lfloor \cdot \rfloor$ represents the ceiling operation, and τ is an adjustable coefficient (set to 5 in this paper). *See Appendix G for detailed DCA mechanism.*

Model Loss and Learning Rate Adjustment Scheme

We implement the arctangent loss function and sigmoid learning rate adjustment for time series prediction. Please refer details in xPatch (Stitsyuk and Choi 2025).

Experiments

Experimental Setup

Datasets We conduct extensive experiments on nine real-world multivariate time series datasets: the ETT benchmark (specifically ETTh1, ETTh2, ETTm1, and ETTm2) (Zhou et al. 2021), Weather (Wu et al. 2021b), Traffic (Sen, Yu, and Dhillon 2019), Electricity (Wu et al. 2021b), Exchange-Rate (Lai et al. 2018), and Solar-Energy (Lai et al. 2018). All datasets are preprocessed and normalized as described in (Nie et al. 2022) and (Liu et al. 2023), and split into training, validation, and test sets with a 7:2:1 ratio.

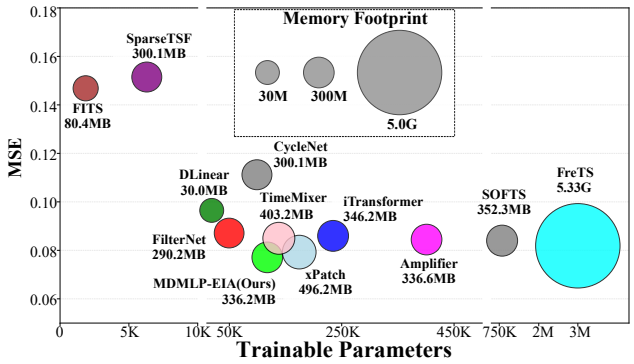
Baselines Our experimental baselines include recent state-of-the-art TSF models: xPatch (2025) (Stitsyuk and Choi 2025), Amplifier (2025) (Fei et al. 2025), TimeMixer (2024) (Wang et al. 2024), CycleNet (2024) (Lin et al. 2024a), SOFTS (2024) (Han et al. 2024), iTransformer (2024) (Liu et al. 2023), FilterNet (2024) (Yi et al. 2024), FITS (2024) (Xu, Zeng, and Xu 2023), SparseTSF (2024) (Lin et al. 2024b), DLinear (2023) (Zeng et al. 2023), and FreTS (2023) (Yi et al. 2023b).

Implementation Details All the experiments are implemented using PyTorch (Paszke et al. 2019), and conducted on 4 A100 GPUs. We evaluate performance using MSE and Mean Absolute Error (MAE). For fair comparisons, several proven effective strategies (Stitsyuk and Choi 2025), including RevIN, arctangent loss, and sigmoid learning rate adjustment, are applied to all baseline models.

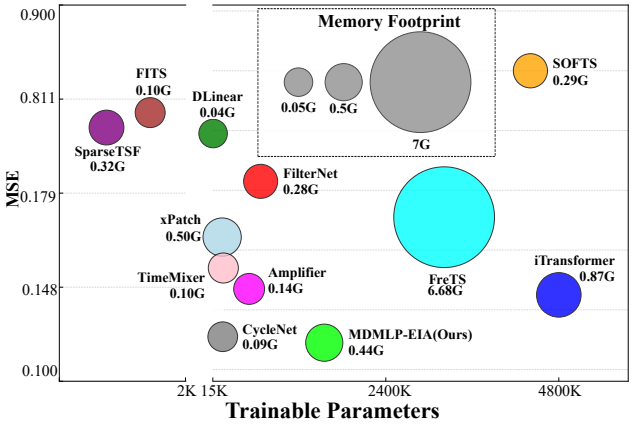
Main Results

Unified Experimental Settings We use unified settings based on the forecasting protocol proposed by TimesNet (Wu et al. 2023): a lookback window of $L = 96$ and prediction horizons of $T = \{96, 192, 336, 720\}$ for all datasets. Table 1 details the long-term forecasting results with the unified settings. MDMLP-EIA (Ours) achieves the best MSE on 5 out of 9 datasets and the best MAE on 7 out of 9 datasets. In particular, compared to the two most recent methods, Amplifier (2025) (Fei et al. 2025) and xPatch (2025) (Stitsyuk and Choi 2025), MDMLP-EIA (ours) surpasses them by approximately 4.14% and 2.91% in MSE, and 2.29% and 1.37% in MAE, respectively. Furthermore, compared to CycleNet (2024) (Lin et al. 2024a), the method with the most comparable performance, MDMLP-EIA (ours) shows an advantage of approximately 2.68% in MSE and 4.66% in MAE. *Full results under unified experimental settings are provided in Appendix H.*

Hyperparameter Search We assess how lookback length affects model performance to identify optimal values for each baseline. Models benefiting from longer sequences undergo additional hyperparameter optimization following xPatch (Stitsyuk and Choi 2025). We use official implementations



(a) Exchange (8 Variables, with training batch size is 256)



(b) Electricity (321 Variables, with training batch size 8)

Figure 3: Model effectiveness and efficiency comparison.

with original configurations as baselines, then tune lookback length and other hyperparameters accordingly. The results are shown in Table 2. Our model achieves the most first-place results, with best MSE on 6/9 datasets and best MAE on 6/9 datasets. Notably, MDMLP-EIA (ours) surpasses the two most recent methods, Amplifier (2025) (Fei et al. 2025) and xPatch (2025) (Stitsyuk and Choi 2025), by approximately 3.48% and 2.20% in MSE, and 1.73% and 1.15% in MAE, respectively. Furthermore, compared to CycleNet (2024) (Lin et al. 2024a) and TimeMixer (2024) (Wang et al. 2024), the methods with the most comparable performance, MDMLP-EIA (ours) shows advantages of approximately 4.82% and 3.27% in MSE, and 3.37% and 2.17% in MAE, respectively. *Full results with hyperparameter search are provided in Appendix H.*

More visualization results are provided in Appendix I.

Model Analysis

Validation of Adaptive Weak Signal Fusion To validate the effectiveness of introducing the weak seasonal MLP in the MDMLP-EIA model and the proposed AZCF mechanism for fusing strong and weak seasonal predictions, we compare AZCF against the following six ablation cases:

1. **w/o WS**: Without the weak seasonal MLP;

Method	MDMLP-EIA (**Ours**)		xPatch (2025)		Amplifier (2025)		TimeMixer (2024)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)		FilterNet (2024)		FITS (2024)		SparseTSF (2024)		FreTS (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
ETTh1	0.436	**0.424**	0.442	0.429	0.446	0.426	0.449	0.428	**0.432**	0.427	0.435	0.428	0.452	0.441	0.450	0.430	0.483	0.463	0.450	0.426	0.471	0.455	0.447	0.438
ETTh2	**0.361**	**0.386**	0.373	0.394	0.370	0.391	0.369	0.391	0.383	0.404	0.371	0.395	0.380	0.401	0.369	0.391	0.385	0.405	0.385	0.398	0.462	0.462	0.426	0.429
ETTm1	**0.378**	**0.376**	0.385	0.402	0.385	0.380	0.386	0.380	0.386	0.394	0.386	0.383	0.402	0.398	0.386	0.382	0.406	0.391	0.408	0.393	0.405	0.401	0.395	0.388
ETTm2	**0.271**	**0.314**	0.279	0.319	0.275	0.317	0.272	**0.314**	**0.271**	**0.314**	0.278	0.317	0.286	0.324	0.273	0.315	0.287	0.331	0.281	0.320	0.290	0.341	0.287	0.322
electricity	**0.167**	**0.257**	0.185	0.267	0.173	0.260	0.183	0.265	0.169	0.260	0.870	0.763	0.178	0.264	0.205	0.281	0.226	0.304	0.216	0.281	0.197	0.277	0.213	0.288
Exchange	0.361	0.403	0.371	0.407	0.382	0.414	0.363	0.402	0.398	0.430	**0.358**	0.402	0.377	0.412	0.410	0.437	0.451	0.470	0.433	0.467	0.455	0.424	0.362	**0.394**
Solar	0.238	**0.237**	0.244	0.240	0.281	0.262	**0.225**	0.240	0.235	0.269	0.264	0.250	0.236	0.239	0.280	0.266	0.427	0.360	0.303	0.282	0.277	0.271	0.335	0.318
Traffic	0.481	0.287	0.500	0.283	0.504	0.300	0.509	0.280	0.485	0.312	0.428	0.252	**0.414**	**0.248**	0.459	0.288	0.691	0.427	0.589	0.322	0.579	0.329	0.654	0.353
Weather	**0.239**	**0.262**	0.247	0.266	0.246	0.267	0.243	0.263	0.254	0.279	0.250	0.267	0.256	0.272	0.244	0.265	0.258	0.277	0.261	0.278	0.247	0.276	0.271	0.292
count of 1^{st}	5	7	0	0	0	0	1	1	2	1	1	0	1	1	0	0	0	0	0	0	0	0	0	1

Table 1: Average long-term forecasting results with lookback $L = 96$ and prediction lengths $T \in \{96, 192, 336, 720\}$. **Bold**: best, underline: second-best. *Full results: Table 1, Appendix H.*

Dataset	MDMLP-EIA **Ours**		xPatch (2025)		Amplifier (2025)		TimeMixer (2024)		CycleNet (2024)		SOFTS (2024)		iTransformer (2024)		FilterNet (2024)		FITS (2024)		SparseTSF (2024)		FreTS (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	MSE	MAE
ETTh1	**0.410**	**0.420**	0.421	0.426	0.421	0.425	0.423	0.425	0.452	0.443	0.418	0.430	0.447	0.453	0.428	0.430	0.884	0.642	0.438	0.434	0.480	0.471	0.425	0.433
ETTh2	**0.335**	**0.378**	0.348	0.387	0.344	0.385	0.345	0.387	0.358	0.395	0.373	0.403	0.364	0.402	0.365	0.399	0.339	0.386	0.369	0.401	0.912	0.649	0.488	0.470
ETTm1	0.351	**0.367**	0.354	0.372	0.359	0.371	0.358	0.374	0.356	0.373	**0.349**	0.374	0.366	0.389	0.370	0.381	0.358	0.372	0.359	0.375	0.373	0.384	0.355	0.372
ETTm2	**0.253**	**0.305**	0.255	0.308	0.260	0.312	**0.253**	0.306	0.255	0.307	0.260	0.314	0.267	0.321	0.259	0.309	0.303	0.351	0.256	0.310	0.274	0.333	0.260	0.316
Electricity	0.157	0.249	0.160	0.251	0.162	0.254	0.160	0.251	**0.156**	**0.246**	0.703	0.650	0.160	0.250	0.172	0.267	0.434	0.491	0.171	0.259	0.167	0.261	0.162	0.256
Exchange	0.362	0.404	0.371	0.407	0.383	0.414	0.363	0.403	0.398	0.431	**0.358**	0.403	0.377	0.412	0.410	0.437	0.452	0.471	0.431	0.459	0.456	0.424	0.362	**0.395**
Solar	**0.190**	**0.213**	0.201	0.222	0.207	**0.213**	0.228	0.244	0.199	0.223	0.235	0.261	0.197	0.225	0.217	0.234	0.263	0.272	0.223	**0.213**	0.226	0.228	0.239	0.234
Traffic	**0.390**	0.255	0.392	**0.249**	0.406	0.264	0.397	0.250	0.405	0.266	0.428	0.253	0.396	0.251	0.401	0.269	0.446	0.308	0.432	0.261	0.418	0.287	0.421	0.283
Weather	**0.217**	**0.248**	0.222	0.251	0.221	0.251	0.227	0.255	0.223	0.252	0.226	0.256	0.233	0.265	0.231	0.261	0.238	0.273	0.227	0.259	0.229	0.263	0.240	0.276
1st	6	6	0	1	0	1	1	0	1	1	2	0	0	0	0	0	0	0	0	1	0	0	0	1

Table 2: Long-term forecasting results with hyperparameter optimization, averaged over $T \in \{96, 192, 336, 720\}$. **Bold**/underline: best/second-best. *Full results: Table 2, Appendix H.*

2. **MLP-F**: Uses Multi-Layer Perceptron(MLP) to concatenate strong and weak seasonal predictions for fusion;
3. **CWA-F**: Employs channel-wise attention to weight and fuse strong and weak seasonal predictions;
4. **DWL-F** ***(w/o single-parameter fusion)***: Differs from AZCF only in using two weight coefficients for fusion as shown in Equation (1);
5. **RCF** ***(w/o zero initialization)***: Differs from AZCF only in using random initialization instead of zero initialization;
6. **CTF** ***(w/o channel-dimension fusion)***: Identical to AZCF in Equation (2) but without channel-dimension fusion.

Comparing the results of w/o WS and AZCF demonstrates the effectiveness of introducing the weak seasonal MLP. Our AZCF strategy also outperforms MLP-F and CWA-F in most cases, demonstrating its effectiveness in fusing strong and weak seasonal predictions. Specifically, ablating the three key mechanisms in AZCF (DWL-F, RCF, CTF) shows that when single-parameter fusion, zero initialization, or channel-dimension fusion are respectively removed, the model's prediction performance degrades. This validates the effectiveness of each key mechanism in AZCF. *See Appendices H and I for full results and visualizations.*

Validation of EIA Mechanism In our proposed MDMLP-EIA, the EIA has been designed for effective integration of trend and seasonal signals. To validate the effectiveness of EIA fusion, we compare three methods:

1. **ADD (w/o EIA)**: Directly adds trend and seasonal signals without EIA;
2. **MLP**: Uses MLP-based attention to weight and fuse trend and seasonal signals;
3. **AGM**: Applies Adaptive Gating Mechanism for fusion (Equation (7)).

This analysis evaluates EIA's applicability across MDMLP-EIA and two leading models (Amplifier (Fei et al. 2025) and xPatch (Stitsyuk and Choi 2025)). Results averaged over six datasets are in Table 4. For MDMLP-EIA, EIA fusion delivers substantial improvements: reducing average MSE/MAE by 5.91%/5.12% vs. MLP and 2.1%/1.53% vs. ADD. While AGM performs competitively, EIA achieves best or joint-best results on 5/6 datasets, reducing average MSE/MAE by 1.83%/0.60% vs. AGM. On Solar, EIA shows the largest advantage, reducing MSE/MAE by 7.36%/2.51%. When integrated into xPatch (Stitsyuk and Choi 2025) and Amplifier (Fei et al. 2025), EIA shows similar improvements, validating its effectiveness and transferability. *Full results and visualizations: Appendices H and I.*

Validation of DCA Mechanism We compared the DCA mechanism with fixed capacity (neuron count) strategies, where the number of neurons varies across 32, 64, 128, 256,

Dataset	w/o WS		MLP-F		DWL-F		CWA-F		RCF		CTF		AZCF	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.442	0.428	0.440	0.428	**0.437**	0.429	0.444	0.427	0.441	0.429	0.438	0.427	**0.437**	**0.424**
ETTh2	0.366	0.389	0.368	0.391	0.369	0.392	0.370	0.392	0.370	0.393	0.374	0.394	**0.362**	**0.387**
ETTm1	**0.377**	0.379	0.382	0.379	0.378	0.378	0.382	0.380	0.379	0.379	0.378	**0.377**	0.378	**0.377**
ETTm2	**0.271**	**0.314**	**0.271**	**0.314**	0.277	0.317	0.272	**0.314**	0.275	0.316	0.276	0.316	**0.271**	**0.314**
Weather	0.241	0.263	0.242	0.265	**0.240**	**0.262**	**0.240**	0.263	0.241	0.263	0.241	**0.262**	**0.240**	**0.262**
Average	0.339	0.355	0.341	0.355	0.340	0.356	0.342	0.355	0.341	0.356	0.341	0.355	**0.338**	**0.353**

Table 3: AZCF validation results averaged across prediction lengths. Best in **bold**. *Full results: Table 3, Appendix H.*

Dataset	MDMLP-EIA								Amplifier								xPatch							
	ADD		MLP		AGM		EIA		ADD		MLP		AGM		EIA		ADD		MLP		AGM		EIA	
	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	0.438	0.425	0.492	0.464	0.439	0.430	**0.437**	**0.424**	0.447	0.426	**0.443**	**0.425**	0.462	0.432	0.454	0.430	0.446	0.431	0.484	0.470	**0.444**	**0.428**	0.448	0.431
ETTh2	0.368	0.391	0.381	0.403	0.370	0.392	**0.362**	**0.387**	**0.369**	**0.391**	0.399	0.406	0.374	0.394	0.371	**0.391**	0.370	**0.393**	0.401	0.419	0.370	**0.393**	**0.369**	**0.393**
ETTm1	0.380	0.378	0.392	0.394	**0.377**	**0.376**	0.378	0.377	0.385	0.380	0.388	0.381	0.386	0.379	**0.383**	**0.378**	0.385	**0.383**	0.417	0.412	**0.384**	0.384	0.385	**0.383**
ETTm2	0.275	0.316	0.280	0.324	0.275	0.315	**0.271**	**0.314**	0.276	0.318	0.285	0.323	**0.273**	**0.315**	**0.273**	0.316	**0.279**	**0.320**	0.295	0.337	0.280	**0.320**	**0.279**	**0.320**
Weather	0.244	0.264	0.247	0.269	**0.240**	**0.262**	**0.240**	**0.262**	0.247	0.267	0.246	0.269	**0.243**	**0.265**	**0.243**	**0.265**	0.247	0.266	0.250	0.272	0.242	0.262	**0.241**	**0.262**
Solar	0.265	0.258	0.256	0.255	0.258	0.243	**0.239**	**0.237**	0.281	0.263	**0.250**	0.255	0.260	**0.243**	0.256	0.247	0.255	0.247	0.257	0.259	0.257	0.247	**0.246**	**0.243**
Average	0.328	0.339	0.341	0.352	0.327	0.336	**0.321**	**0.334**	0.334	0.341	0.335	0.343	0.333	**0.338**	**0.330**	**0.338**	0.330	0.340	0.351	0.362	0.330	**0.339**	**0.328**	**0.339**

Table 4: Validation results for EIA under three frameworks, averaged across prediction lengths. **Bold**: best within each model group. *Full results: Appendix Table 4.*

HS	ETTh1		Solar		Traffic		Weather		Average	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
32	0.448	0.438	0.261	0.262	0.610	0.331	0.240	0.263	0.389	0.323
64	0.439	0.430	0.259	0.258	0.570	0.314	0.240	0.263	0.376	0.316
128	0.438	0.427	0.257	0.254	0.550	0.302	**0.240**	**0.260**	0.372	0.311
256	0.439	0.425	0.249	0.250	0.500	0.294	**0.240**	0.263	0.356	0.308
512	0.440	0.425	0.249	0.249	0.490	0.292	0.240	0.263	0.354	0.307
1024	0.441	0.425	0.248	0.240	0.490	0.289	0.240	0.264	0.354	0.304
2048	0.443	0.424	0.254	0.248	0.490	**0.282**	0.240	0.264	0.357	0.305
DCA	**0.436**	**0.424**	**0.238**	**0.237**	**0.481**	0.287	**0.239**	0.262	**0.349**	**0.302**

Table 5: DCA validation results averaged over $T \in \{96, 192, 336, 720\}$ on datasets with varying channels: ETTh1 ($C = 7$), Weather ($C = 21$), Solar ($C = 137$), Traffic ($C = 862$). Bold: best.

512, 1024, 2048. We selected four datasets with significantly different channel counts (ETTH1 ($C = 7$), Weather ($C = 21$), Solar ($C = 137$), and Traffic ($C = 862$)) for time series prediction tasks with $L = 96$ and prediction lengths of 96, 192, 336, 720. Table 5 presents the average prediction results across all prediction lengths. DCA achieves the lowest MSE and MAE across all datasets, reducing average MSE/MAE by 1.41%/0.07% vs. the second-best (HS=1024). *See Appendix H for detailed results.*

Efficiency Analysis We comprehensively evaluate MDMLP-EIA's efficiency in terms of trainable parameters, memory footprint, and predictive accuracy (MSE), following (Yi et al. 2024). Figure 3 compares our method against representative baselines on datasets with markedly different scales: the Exchange dataset (8 variables, the batch size of training stage is 256) and the Electricity dataset (321 variables, the batch size of training stage is 8) with $L = 96$ and $Q = 96$. MDMLP-EIA demonstrates superior efficiency-accuracy trade-offs across both datasets. On Exchange, it achieves competitive accuracy (MSE ≈ 0.082) with approximately 128K parameters and 336.2MB memory, significantly outperforming parameter-heavy models like iTransformer (224k parameters, 346.2MB memory, MSE $\approx$ 0.086) while maintaining better accuracy than lightweight models like DLinear (18K parameters, 30.0MB memory, MSE ≈ 0.096). On Electricity, MDMLP-EIA maintains strong performance (MSE ≈ 0.140) with 1907K parameters and 0.44G memory, substantially more efficient than FreTS (3236K parameters, 6.68G memory). The results establish MDMLP-EIA's effective balance between model complexity and predictive accuracy across different dataset scales.

Conclusions

We propose MDMLP-EIA, a novel time series forecasting model that addresses critical limitations in existing approaches through three key innovations: a weak seasonal MLP with AZCF fusion mechanism, an EIA mechanism, and a DCA mechanism. These components effectively tackle three prevalent issues: loss of weak seasonal signals, capacity constraints in weight-sharing MLPs, and insufficient channel fusion in channel-independent strategies, enabling efficient and accurate forecasting. Moreover, the proposed AZCF, EIA, and DCA mechanisms demonstrate broad applicability and can enhance other seasonal-trend decomposition-based methods. Future work includes designing unified cross-horizon architectures and enabling domain transfer.

Acknowledgments

We would like to thank all anonymous reviewers for their valuable comments.

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62376040 and 62506047, the Key projects of NSFC under Grant 62233018, the outstanding youth project of Hunan Provincial Department of Education under Grant 24B0797, and the China Postdoctoral Science Foundation under Grant Number 2024T171058 and 2024M753672.

References

Amasyali, K.; and El-Gohary, N. M. 2018. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81: 1192–1205.

Cai, W.; Liang, Y.; Liu, X.; Feng, J.; and Wu, Y. 2024. Msgnet: Learning multi-scale inter-series correlations for multivariate time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 11141–11149.

Cao, D.; Wang, Y.; Duan, J.; Zhang, C.; Zhu, X.; Huang, C.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; et al. 2020. Spectral temporal graph neural network for multivariate time-series forecasting. *Advances in neural information processing systems*, 33: 17766–17778.

Challu, C.; Olivares, K. G.; Oreshkin, B. N.; Ramirez, F. G.; Canseco, M. M.; and Dubrawski, A. 2023. Nhits: Neural hierarchical interpolation for time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 6989–6997.

Chen, S.; Li, C.; Yoder, N.; Arik, S.; and Pfister, T. 2023. Tsmixer: An all-mlp architecture for time series forecasting. arXiv 2023. *arXiv preprint arXiv:2303.06053*.

Fan, W.; Zheng, S.; Yi, X.; Cao, W.; Fu, Y.; Bian, J.; and Liu, T.-Y. 2022. DEPTS: Deep expansion learning for periodic time series forecasting. *arXiv preprint arXiv:2203.07681*.

Fei, J.; Yi, K.; Fan, W.; Zhang, Q.; and Niu, Z. 2025. Amplifier: Bringing Attention to Neglected Low-Energy Components in Time Series Forecasting. *arXiv preprint arXiv:2501.17216*.

Han, L.; Chen, X.-Y.; Ye, H.-J.; and Zhan, D.-C. 2024. Softs: Efficient multivariate time series forecasting with series-core fusion. *arXiv preprint arXiv:2404.14197*.

Han, L.; Ye, H.-J.; and Zhan, D.-C. 2024. The capacity and robustness trade-off: Revisiting the channel independent strategy for multivariate time series forecasting. *IEEE Transactions on Knowledge and Data Engineering*.

Jia, Y.; Lin, Y.; Hao, X.; Lin, Y.; Guo, S.; and Wan, H. 2023. Witran: Water-wave information transmission and recurrent acceleration network for long-range time series forecasting. *Advances in Neural Information Processing Systems*, 36: 12389–12456.

Lai, G.; Chang, W.-C.; Yang, Y.; and Liu, H. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, 95–104.

Li, Y.; Xu, J.; and Anastasiu, D. 2024. Learning from polar representation: An extreme-adaptive model for long-term time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 171–179.

Lin, S.; Lin, W.; Hu, X.; Wu, W.; Mo, R.; and Zhong, H. 2024a. Cyclenet: enhancing time series forecasting through modeling periodic patterns. *Advances in Neural Information Processing Systems*, 37: 106315–106345.

Lin, S.; Lin, W.; Wu, W.; Chen, H.; and Yang, J. 2024b. Sparsetsf: Modeling long-term time series forecasting with 1k parameters. *arXiv preprint arXiv:2405.00946*.

Liu, M.; Zeng, A.; Chen, M.; Xu, Z.; Lai, Q.; Ma, L.; and Xu, Q. 2022. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Advances in Neural Information Processing Systems*, 35: 5816–5828.

Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; and Long, M. 2023. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. *arXiv preprint arXiv:2310.06625*.

Luo, D.; and Wang, X. 2024. Moderntcn: A modern pure convolution structure for general time series analysis. In *The twelfth international conference on learning representations*, 1–43.

Nie, Y.; Nguyen, N. H.; Sinthong, P.; and Kalagnanam, J. 2022. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv preprint arXiv:2211.14730*.

Oreshkin, B. N.; Carpov, D.; Chapados, N.; and Bengio, Y. 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Sen, R.; Yu, H.; and Dhillon, I. S. 2019. Think Globally, Act Locally: A Deep Neural Network Approach to High-Dimensional Time Series Forecasting. In *NeurIPS*, 4838–4847.

Shang, Z.; Chen, L.; Wu, B.; and Cui, D. 2024. Ada-MSHyper: adaptive multi-scale hypergraph transformer for time series forecasting. *Advances in Neural Information Processing Systems*, 37: 33310–33337.

Shao, Z.; Wang, F.; Xu, Y.; Wei, W.; Yu, C.; Zhang, Z.; Yao, D.; Sun, T.; Jin, G.; Cao, X.; et al. 2024. Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis. *IEEE Transactions on Knowledge and Data Engineering*.

Stitsyuk, A.; and Choi, J. 2025. xPatch: Dual-Stream Time Series Forecasting with Exponential Seasonal-Trend Decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 20601–20609.

Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J. Y.; and Zhou, J. 2024. Timemixer: Decomposable multiscale mixing for time series forecasting. *arXiv preprint arXiv:2405.14616*.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.

Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; and Long, M. 2023. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In *International Conference on Learning Representations*.

Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021a. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34: 22419–22430.

Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021b. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *NeurIPS*, 22419–22430.

Wu, Q.; Yao, G.; Feng, Z.; and Shuyuan, Y. 2024. Perimidformer: Periodic pyramid transformer for time series analysis. *Advances in Neural Information Processing Systems*, 37: 13035–13073.

Xu, Z.; Zeng, A.; and Xu, Q. 2023. FITS: Modeling time series with $10k$ parameters. *arXiv preprint arXiv:2307.03756*.

Yi, K.; Fei, J.; Zhang, Q.; He, H.; Hao, S.; Lian, D.; and Fan, W. 2024. Filternet: Harnessing frequency filters for time series forecasting. *Advances in Neural Information Processing Systems*, 37: 55115–55140.

Yi, K.; Zhang, Q.; Fan, W.; He, H.; Hu, L.; Wang, P.; An, N.; Cao, L.; and Niu, Z. 2023a. FourierGNN: Rethinking multivariate time series forecasting from a pure graph perspective. *Advances in neural information processing systems*, 36: 69638–69660.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; and Niu, Z. 2023b. Frequency-domain mlps are more effective learners in time series forecasting. *Advances in Neural Information Processing Systems*, 36: 76656–76679.

Yin, X.; Wu, G.; Wei, J.; Shen, Y.; Qi, H.; and Yin, B. 2021. Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Transactions on Intelligent Transportation Systems*, 23(6): 4927–4943.

Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.

Zhang, H.; Tang, Z.; Xie, Y.; Zheng, Z.; and Gui, W. 2025. Multihorizon KPI Forecasting in Complex Industrial Processes: An Adaptive Encoder–Decoder Framework With Partial Teacher Forcing. *IEEE Transactions on Cybernetics*.

Zhang, L.; Aggarwal, C.; and Qi, G.-J. 2017. Stock price prediction via discovering multi-frequency trading patterns. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2141–2149.

Zhang, T.; Zhang, Y.; Cao, W.; Bian, J.; Yi, X.; Zheng, S.; and Li, J. 2022. Less is more: Fast multivariate time series forecasting with light sampling-oriented mlp structures. arXiv 2022. *arXiv preprint arXiv:2207.01186*.

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

Zheng, Y.; Yi, X.; Li, M.; Li, R.; Shan, Z.; Chang, E.; and Li, T. 2015. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 2267–2276.

Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; and Zhang, W. 2021. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11106–11115.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International conference on machine learning*, 27268–27286. PMLR.