A EXPERIMENTAL DETAILS

A.1 DATA DESCRIPTIONS

We conduct extensive experiments on nine widely-used time series datasets. (1) *ETT* Zhou et al. (2021) contains data collected from electricity transformers located in two regions of China, between July 2016 and July 2018. The dataset includes two different granularities: ETTh for 1 hour and ETTm for 15 minutes. Each data point consists an 'oil temperature' value and six external power load features. (2) *Electricity*[1] contains electricity consumption data for 321 clients in kilowatts (kW) of 1 hour from 2012 to 2014, which is released from the UCL Machine Learning Repository. (3) *Exchange-Rate* Lai et al. (2018) records the daily exchange rates of eight different countries, spanning from 1990 to 2016. (4) *Traffic*[2] contains hourly road occupancy rates recorded by the real-time individual detectors located on San Francisco Bay area freeways between 2015 to 2016. (5) *Weather*[3] contains 21 meteorological indicators recorded every 10 minutes in Germany during the year 2020, including air temperature, wind velocity, etc. (6) *ILI*[4] consists of the ratio of patients seen with influenza-like illness to the total number of patients, which is collected by the United States Centers for Disease Control and Prevention on a weekly basis from 2002 to 2021. Following the standard protocol outlined in Wu et al. (2021), we partition all datasets into training, validation, and testing sets chronologically, with a ratio of 6:2:2 for the ETT dataset and 7:1:2 for other datasets.

A.2 EVALUATION METRICS

In this paper, we perform zero-mean normalization to the data and select the Mean Absolute Error (MAE) and the Mean Squared Error (MSE) as evaluation metrics:

$$\text{MAE} = \frac{1}{H}\sum_{k=0}^{H} ||\boldsymbol{x}_{t+k} - \hat{\boldsymbol{x}}_{t+k}||, \tag{11}$$

$$\text{MSE} = \frac{1}{H}\sum_{k=0}^{H} ||\boldsymbol{x}_{t+k} - \hat{\boldsymbol{x}}_{t+k}||^2, \tag{12}$$

where H is the output length, $\hat{\boldsymbol{x}}_{t+k}$ is the prediction of the model, and $\boldsymbol{x}_{t+k}$ is the ground truth.

A.3 BASELINE MODELS

We compare MPPN with a variety of state-of-the-art models in the domain of multivariate long-term time series forecasting, including Linear-based models, CNN-based models and Transformer-based models. We summarize them as follows.

- PatchTST Nie et al. (2023): PatchTST is a Transformer-based model with two key innovations: patching and channel-independence. Patching involves segmenting time series data into subseries-level patches, while channel-independence ensures that each input token exclusively contains information from a single channel.
- NLinear Zeng et al. (2023): NLinear performs a simple normalization, where the input is first subtracted by the last value of the sequence, then goes through a single-layer linear model. NLinear demonstrates an effective capability to address distribution shifts within datasets.
- DLinear Zeng et al. (2023): DLinear first decomposes the raw time series into a trend component by moving average and a seasonal component. Then it adopts two one-layer linear layers to each component and sum up the two features to produce the final prediction.
- TimesNet Wu et al. (2022): TimesNet introduces two types of temporal variations, namely intraperiod- and interperiod-variants, and unifies these representations into 2D space. Subsequently, an inception block with multi-scale 2D kernels is employed to effectively model these temporal variations.

[1] https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014
[2] http://pems.dot.ca.gov/
[3] https://www.bgc-jena.mpg.de/wetter/
[4] https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

- MICN Wang et al. (2023): MICN initially decomposes the input series into trend-cyclical and seasonal parts. It then uses down-sampled convolution and isometric convolution to capture both the local and global correlations presented in time series.
- SCINet Liu et al. (2022a): SCINet is a hierarchical convolution network with downsample-convolve-interact operations. The input sequence is downsampled into multiple resolution sub-sequences, which are then iteratively extracted and exchanged to facilitate the learning of an effective representation for the time series.
- FEDformer Zhou et al. (2022): FEDformer incorporates a set of decomposition blocks with different sizes to effectively capture the global profile of the time series. Additionally, it performs the attention mechanism with low-rank approximation in the frequency domain with Fourier transform to enhance the performance for long-term prediction.
- Autoformer Wu et al. (2021): Autoformer is a deep decomposition architecture that incorporates the built-in decomposition block and the series-wise attention based on auto-correlation mechanism. The decomposition block is designed to progressively isolate the long-term trend information from the predicted hidden variables.
- Crossformer Zhang & Yan (2023): Crossformer embeds the input series into a 2D vector with time and dimension features, and two attention layers are proposed to capture the cross-time and cross-dimension dependencies, respectively.
- Informer Zhou et al. (2021): Informer introduces the ProbSparse self-attention mechanism and self-attention distilling operation to address the problems of quadratic time complexity and quadratic memory usage in vanilla Transformer.
- LogTrans Li et al. (2019): LogTrans introduces convolutional self-attention to involve local context and LogSparse Transformer to reduce space complexity in vanilla Transformer models.

A.4 HYPERPARAMETER

By default, MPPN employs multi-resolution patching (non-overlapping patches) and the multiple temporal resolutions (convolutional kernel sizes) are set as [1, 3, 4, 6]. Referring to the results in Table 2, for datasets with obvious periodicity, we select the top-2 period ($k = 2$) computed by the method of FFT as discussed in Section 3 for periodic pattern mining, including Electricity, Weather, Traffic, ILI, ETTh1 and ETTm1. For datasets with less obvious periodicity, we set the period to the default value of 24 in the case of Exchange-Rate, ILI, and ETTh2. For ETTm2 with a granularity of 15 minutes, we set the period to 96, which corresponds exactly to the default daily period. At last, we set the hyperparameter $D = 48$ which is the hidden state of the series for all datasets.

For MPPN, we employ different input length on different datasets. Specifically, experiments conducted on the Weather, Traffic, and Electricity dataset employ an input length of 720. For the Exchange-Rate and ETT dataset, we opt an input length of 336. Lastly, for the ILI dataset, we employ the input length as 96. We adopt the default hyperparameters of each baseline model mentioned in their papers to train them. Increasing the length of the input length offers substantial performance benefits for Linear-based models, while Transformer-based models tend to overfit temporal noises rather than extracting temporal information Zeng et al. (2023). We set the input length $L = 96$ for Transformer-based, and $L = 336$ for CNN-based models and Linear-based models. Since the ILI dataset sampled at weekly granularity is smaller in size, we use a different set of parameters ($L = 36$ for Transformer-based models and CNN-based models, and $L = 104$ for Linear-based models).

B MORE RESULTS

B.1 FULL BENCHMARK OF MPPN

In this section, we present a comprehensive benchmark of multivariate long-term time series forecasting results in Table 5, which serves as an expanded version of Table 3. We incorporate two recently released models from arXiv, RLinear and RMLP Li et al. (2023), which currently represent the state-of-the-art in Linear-based approaches, along with the classical Informer Zhou et al. (2021), for the purpose of comparative analysis.

Table 5: **Multivariate** long-term time series forecasting results with different prediction length $O \in \{24, 36, 48, 60\}$ for ILI dataset and $O \in \{96, 192, 336, 720\}$ for others. The SOTA results are **bolded**, while the sub-optimal results are underlined.

Models		MPPN		RLinear		RMLP		Informer	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	**0.144**	**0.196**	0.175	0.225	0.151	0.202	0.375	0.437
	192	**0.189**	**0.240**	0.218	0.260	0.194	0.241	0.483	0.496
	336	**0.240**	**0.281**	0.265	0.294	0.245	0.286	0.588	0.542
	720	**0.310**	**0.333**	0.339	0.428	0.324	0.335	1.061	0.755
Traffic	96	0.387	0.271	0.409	0.281	**0.386**	**0.270**	0.746	0.416
	192	**0.396**	**0.273**	0.414	0.282	0.399	0.276	0.483	0.496
	336	**0.410**	**0.279**	0.428	0.290	0.414	0.285	0.588	0.542
	720	**0.449**	**0.301**	0.463	0.310	0.450	0.304	1.046	0.588
Electricity	96	**0.131**	**0.226**	0.143	0.240	0.132	0.227	0.323	0.409
	192	**0.145**	**0.239**	0.153	0.250	0.146	0.242	0.347	0.431
	336	**0.162**	**0.256**	0.171	0.268	0.165	0.260	0.349	0.432
	720	**0.200**	**0.289**	0.211	0.300	0.204	0.293	0.394	0.457
Exchange	96	**0.089**	**0.204**	**0.089**	0.209	0.096	0.222	0.947	0.771
	192	**0.177**	**0.295**	0.191	0.309	0.219	0.334	1.244	0.882
	336	**0.344**	**0.418**	0.363	0.434	0.369	0.438	1.792	1.070
	720	**0.929**	**0.731**	0.963	**0.731**	1.000	0.735	2.936	1.415
ILI	24	1.796	0.860	1.762	0.879	**1.443**	**0.793**	5.248	1.580
	36	1.748	0.840	1.775	0.857	**1.444**	**0.794**	5.057	1.561
	48	**1.692**	**0.840**	1.731	0.863	1.849	0.932	5.110	1.564
	60	**1.840**	**0.881**	1.863	0.909	1.872	0.947	5.397	1.606
ETTh1	96	0.371	0.393	**0.366**	**0.391**	0.390	0.410	0.934	0.764
	192	0.405	0.413	**0.403**	**0.412**	0.431	0.433	1.006	0.785
	336	0.426	0.425	**0.420**	**0.423**	0.441	0.441	1.036	0.783
	720	**0.436**	**0.452**	0.442	0.456	0.506	0.496	1.174	0.856
ETTh2	96	0.278	**0.335**	**0.274**	0.337	0.310	0.361	2.978	1.360
	192	**0.344**	**0.380**	0.345	0.387	0.375	0.409	6.203	2.078
	336	0.362	0.400	**0.335**	**0.393**	0.360	0.408	5.437	1.961
	720	**0.393**	**0.434**	0.412	0.441	0.437	0.460	4.115	1.692
ETTm1	96	**0.287**	**0.335**	0.301	0.343	0.298	0.346	0.624	0.558
	192	**0.330**	**0.360**	0.341	0.367	0.345	0.376	0.725	0.618
	336	**0.369**	**0.382**	0.374	0.386	0.381	0.399	1.006	0.750
	720	**0.426**	**0.414**	0.430	0.418	0.448	0.444	0.967	0.741
ETTm2	96	**0.162**	**0.250**	0.164	0.253	0.175	0.260	0.382	0.463
	192	**0.217**	**0.288**	0.219	0.290	0.245	0.308	0.849	0.724
	336	**0.273**	**0.325**	**0.273**	0.326	0.297	0.340	1.423	0.915
	720	0.368	**0.383**	**0.366**	0.385	0.368	0.389	3.929	1.469

B.2 Univariate forecasting

We present the univariate time-series forecasting results in Table 6. Following the results from MICN Wang et al. (2023), we include LogTrans Li et al. (2019) in comparison. Our MPPN model continues to demonstrate superior performance in comparison to the baseline models, particularly in the context of the weather dataset with obvious periodicity. MPPN achieves significant reductions in MSE, with decreases of **62.07%**(0.0029→0.0011), **33.33%**(0.0021→0.0014), **30.43%**(0.0023→0.0016), and **32.26%**(0.0031→0.0021) observed for forecast horizons of 96, 192, 336, and 720, respectively. The experimental results provide evidence for the effectiveness of MPPN in extracting essential characteristics of time series and performing univariate forecasting.

Table 6: **Univariate** long-term time series forecasting results with different prediction length $O \in \{24, 36, 48, 60\}$ for ILI dataset and $O \in \{96, 192, 336, 720\}$ for others. The SOTA results are **bolded**, while the sub-optimal results are underlined. Correspondingly, IMP. shows the percentage improvement of MPPN over the state-of-the-art baseline models.

Models		**MPPN**		DLinear		MICN*		FEDformer*		Autoformer*		Informer*		LogTrans*		IMP.
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
Weather	96	**0.0011**	**0.025**	0.0057	0.063	0.0029	0.039	0.0062	0.062	0.011	0.081	0.0038	0.044	0.0046	0.052	62.07%
	192	**0.0014**	**0.027**	0.0062	0.066	0.0021	0.034	0.0060	0.062	0.0075	0.067	0.0023	0.040	0.0056	0.060	33.33%
	336	**0.0016**	**0.030**	0.0064	0.068	0.0023	0.034	0.0041	0.050	0.0063	0.062	0.0041	0.049	0.0060	0.054	30.43%
	720	**0.0021**	**0.034**	0.0068	0.070	0.0048	0.054	0.0055	0.059	0.0085	0.070	0.0031	0.042	0.0071	0.063	32.26%
Traffic	96	**0.115**	**0.191**	0.126	0.202	0.158	0.241	0.207	0.312	0.246	0.346	0.257	0.353	0.226	0.317	8.73%
	192	**0.118**	**0.204**	0.129	0.208	0.154	0.236	0.205	0.312	0.266	0.370	0.299	0.376	0.314	0.408	8.53%
	336	**0.118**	**0.200**	0.130	0.213	0.165	0.243	0.219	0.323	0.263	0.371	0.312	0.387	0.387	0.453	9.23%
	720	0.146	0.237	**0.142**	**0.226**	0.182	0.264	0.244	0.344	0.269	0.372	0.366	0.436	0.491	0.437	\
Electricity	96	**0.195**	**0.304**	0.203	0.315	0.310	0.398	0.253	0.370	0.341	0.438	0.484	0.538	0.288	0.393	3.94%
	192	**0.230**	**0.329**	0.233	0.337	0.300	0.394	0.282	0.386	0.345	0.428	0.557	0.558	0.432	0.483	1.29%
	336	**0.266**	**0.360**	0.268	0.364	0.323	0.413	0.346	0.431	0.406	0.470	0.636	0.613	0.430	0.483	0.75%
	720	0.316	0.413	**0.307**	**0.408**	0.364	0.449	0.422	0.484	0.565	0.581	0.819	0.682	0.491	0.531	\
Exchange	96	**0.095**	**0.232**	0.155	0.288	0.099	0.240	0.154	0.304	0.241	0.387	0.591	0.615	0.237	0.377	4.04%
	192	0.204	**0.339**	**0.194**	0.351	0.198	0.354	0.286	0.420	0.300	0.369	1.183	0.912	0.738	0.619	\
	336	0.431	0.492	0.420	0.508	**0.302**	**0.447**	0.511	0.555	0.509	0.524	1.367	0.984	2.018	1.070	\
	720	1.274	0.858	0.814	0.727	**0.738**	**0.662**	1.301	0.879	1.260	0.867	1.872	1.072	2.405	1.175	\
ILI	24	**0.630**	**0.627**	0.726	0.681	0.674	0.671	0.708	**0.627**	0.948	0.732	5.282	2.050	3.607	1.662	6.53%
	36	**0.517**	**0.573**	0.793	0.745	0.712	0.733	0.584	0.617	0.634	0.650	4.554	1.916	2.407	1.363	11.47%
	48	**0.621**	**0.640**	0.886	0.815	0.823	0.803	0.717	0.697	0.791	0.752	4.273	1.846	3.106	1.575	13.39%
	60	**0.695**	**0.688**	0.960	0.860	0.992	0.892	0.855	0.774	0.874	0.797	5.214	2.057	3.698	1.733	18.71%

Results* are from MICNWang et al. (2023); Other results are implemented by us.

C ERROR BARS EVALUATION

To evaluate the robustness of MPPN across different settings, we conduct experiments based on three independent runs on ETTh1, Weather and Electricity. As depicted in Table 7, the standard deviation (Std.) is basically no more than 3% of the mean values (Mean). These results highlight the robustness of MPPN in the face of diverse initialization settings.

Table 7: The error bars of MPPN with 3 runs, input length $I = 720$ and output length $O \in \{96, 192, 336, 720\}$ on ETTh1, Weather and Electricity.

Dataset		Weather					Electricity					ETTh1				
Metric		Seed1	Seed2	Seed3	Mean	Std.	Seed1	Seed2	Seed3	Mean	Std.	Seed1	Seed2	Seed3	Mean	Std.
O=96	MSE	0.1441	0.1459	0.1479	0.1460	0.0016	0.1311	0.1310	0.1312	0.1311	0.0001	0.3715	0.3773	0.3751	0.3746	0.0024
	MAE	0.1969	0.1986	0.2030	0.1995	0.0026	0.2263	0.2262	0.2263	0.2263	0.0001	0.3931	0.3993	0.3974	0.3966	0.0026
O=192	MSE	0.1893	0.1899	0.1898	0.1897	0.0002	0.1460	0.1459	0.1460	0.1460	0.0001	0.4055	0.4117	0.4118	0.4097	0.0029
	MAE	0.2404	0.2409	0.2418	0.2410	0.0006	0.2398	0.2399	0.2399	0.2399	0.0000	0.4135	0.4201	0.4208	0.4181	0.0033
O=336	MSE	0.2401	0.2406	0.2396	0.2401	0.0004	0.1621	0.1621	0.1618	0.1620	0.0001	0.4269	0.4505	0.4321	0.4365	0.0101
	MAE	0.2814	0.2817	0.2808	0.2813	0.0004	0.2565	0.2567	0.2566	0.2566	0.0001	0.4257	0.4492	0.4327	0.4359	0.0099
O=720	MSE	0.3104	0.3105	0.3106	0.3105	0.0001	0.2005	0.2006	0.2007	0.2006	0.0001	0.4366	0.4518	0.4455	0.4446	0.0062
	MAE	0.3332	0.3330	0.3346	0.3336	0.0007	0.2898	0.2898	0.2902	0.2900	0.0002	0.4528	0.4619	0.4571	0.4573	0.0037

D HYPERPARAMETER SENSITIVITY

D.1 INPUT LENGTH SELECTION

In this section, we evaluate the impact of input length on model performance, as presented in Table 8. The experimental results indicate that the impact of input length on model performance varies across different datasets. Due to the differences in granularity and periodicity among datasets, fine-tuning the input length for each specific dataset is preferable over adopting a one-size-fits-all approach in real-world applications. For example, in datasets with simple patterns like ETTm1, an input length of 336 is sufficient to capture substantial information. But for datasets characterized by intricate time series patterns, such as Electricity and Weather, a longer input length is required to extract more comprehensive information that can effectively contribute to the prediction task. Besides, the performances of MPPN exhibit a notable growth as the size of lookback window increases, which highlights the strong temporal relation extraction capability of our method Zeng et al. (2023).

Table 8: Multivariate long-term series results on Weather, Electricity and ETTm1 dataset with prediction length 96, and input lengths I in $\{96, 192, 336, 720\}$. Four differnet variants of MPPN are evaluated, with the best results highlighted in **bold**.

Dataset	Weather		Electricity		ETTm1	
Metric	MSE	MAE	MSE	MAE	MSE	MAE
I=96	0.174	0.220	0.189	0.268	0.339	0.368
I=192	0.158	0.206	0.144	0.237	0.301	0.343
I=336	0.150	0.201	0.136	0.231	**0.287**	**0.335**
I=720	**0.144**	**0.196**	**0.131**	**0.226**	0.304	0.351

D.2 Multiple resolution selection

To evaluate the impact of multiple resolutions on prediction performance, we select three sets of distinct kernel sizes. The selection of different resolutions better aligns with the typical diurnal rhythm of daily activities. For Electricity at an hourly granularity, we have chosen a convolutional filter with a size of 6 to partition a day into "morning, afternoon, evening, and late night". For Weather at a 10-minute granularity, convolutional filters with sizes of 3 and 6 can be employed to uncover patterns and variations in the temporal data at intervals of every half-hour or every hour. The corresponding results are presented in Table 9. It can be seen from Table 9 that the first set of kernel size enjoys the best performance across different scenarios for both Weather and Electricity. As a consequence, we adopt $[1, 3, 4, 6]$ as the default kernel sizes in our main experimental results for the nine public datasets. In practice, the selection of the optimal resolution depends on the specific application context, data granularity, and periodicity, which requires a case-by-case analysis.

Table 9: Multivariate long-term series prediction results on Weather and Electricity with input length 720 and prediction length in $\{96, 192, 336, 720\}$. Three different combinations of convolutional kernel sizes are evaluated, with the best results highlighted in bold.

Multiple resolutions		[1,3,4,6]		[1,3,5,7]		[2,4,6,8]	
	Metric	MSE	MAE	MSE	MAE	MSE	MAE
Weather	96	**0.144**	**0.196**	0.147	0.203	0.153	0.211
	192	**0.189**	**0.240**	0.190	0.242	0.198	0.250
	336	**0.240**	**0.281**	0.241	0.283	0.244	0.287
	720	**0.310**	**0.333**	0.312	0.335	0.311	0.335
Electricity	96	**0.131**	**0.226**	0.132	0.227	0.140	0.241
	192	**0.145**	**0.239**	0.146	0.241	0.155	0.255
	336	**0.162**	**0.256**	0.162	0.257	0.169	0.267
	720	**0.200**	**0.289**	**0.200**	0.290	0.210	0.304

E Ablation of patching

This ablation study investigates the effect of not using patches, using overlapped and non-overlapped patches to the forecasting performance. In the case of overlapped patches, we set $stride = 1$ in Conv1d of multi-resolution patching. From Table 10, the prediction performance of the Traffic dataset experiences a significant decline when patches are not employed. Conversely, for the Electricity dataset, the impact of using or not using patches is minimal. This observation underscores the distinctive sensitivity of datasets towards patches. As a result, the design of patching emerges as an important factor in improving forecasting performance, as well as optimizing running time and memory utilization. Notably, prediction results on the Electricity dataset exhibits better than baseline models even without the utilization of patches, highlighting the effectiveness of the multi-periodic pattern mining and channel adaptive module in MPPN. Besides, the MSE and MAE scores show minimal variations between the models utilizing overlapped and non-overlapped patches. This

Table 10: Ablation studies: multivariate long-term series results on Electricity and Traffic with input length 720 and prediction lengths O in $\{96, 192, 336, 720\}$. Three variants of patching are evaluated, with the best results highlighted in **bold**.

Patching		Overlap		Non-overlap		w/o Patch	
Metric		MSE	MAE	MSE	MAE	MSE	MAE
Electricity	96	**0.130**	**0.225**	0.131	0.226	0.131	0.225
	192	0.145	0.239	0.145	0.239	0.146	0.239
	336	0.162	0.256	0.162	0.256	0.162	0.256
	720	0.200	0.289	0.200	0.289	0.200	0.289
Traffic	96	**0.386**	**0.270**	0.387	0.271	0.414	0.305
	192	0.397	**0.272**	**0.396**	0.273	0.457	0.341
	336	0.410	0.280	0.410	**0.279**	0.457	0.336
	720	0.459	0.315	**0.449**	**0.301**	0.509	0.363

observation highlights the insensitivity of our model to changes in patch configuration, thus attesting to its robustness. It is advisable to flexibly select different patching methods based on the characteristics of the dataset, which enables a more tailored and effective modeling of the specific dataset.

F Spatial and temporal embedding

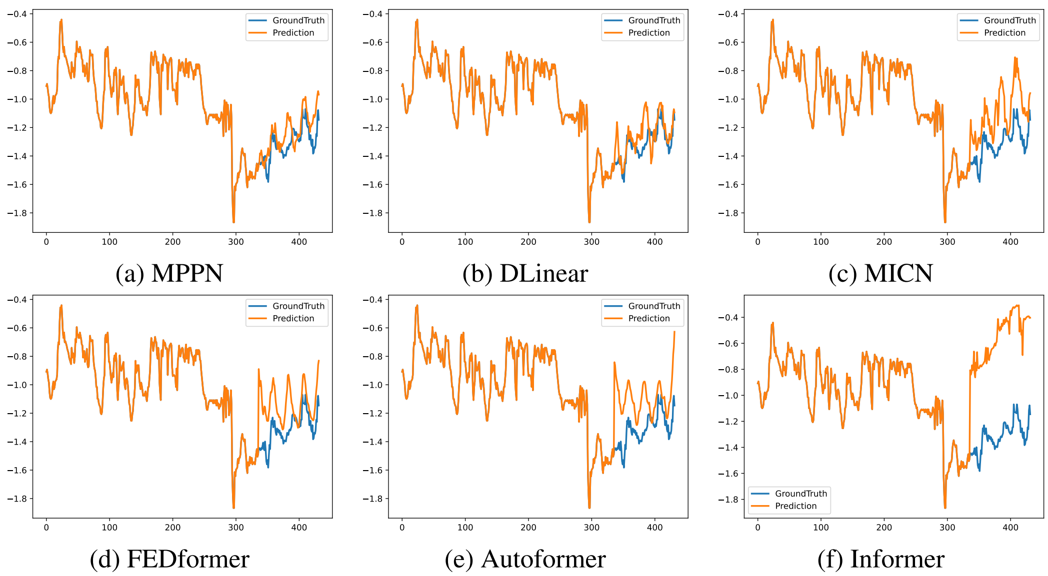
Table 11: Ablation studies: multivariate long-term series prediction results on Traffic, Weather and Electricity with input length 720 and prediction length in $\{96, 192, 336, 720\}$. The original MPPN together with four embedding approaches is compared, with the best results highlighted in bold.

Methods		MPPN		+ L embedding		+H embedding		+ L & H embedding		+ S & H embedding	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Traffic	96	**0.387**	**0.271**	0.426	0.314	0.423	0.312	0.397	0.288	0.410	0.300
	192	**0.396**	**0.273**	0.433	0.312	0.414	0.295	0.406	0.287	0.424	0.308
	336	**0.410**	**0.279**	0.438	0.315	0.416	0.287	0.419	0.294	0.421	0.298
	720	**0.449**	**0.301**	0.467	0.326	0.478	0.336	0.457	0.316	0.454	0.312
Weather	96	**0.144**	**0.196**	0.147	0.203	0.177	0.258	0.146	0.201	0.187	0.270
	192	**0.189**	**0.240**	0.193	0.253	0.213	0.284	0.192	0.245	0.240	0.315
	336	**0.240**	**0.281**	0.253	0.303	0.255	0.311	0.255	0.305	0.326	0.357
	720	**0.310**	**0.333**	0.330	0.366	0.329	0.363	0.326	0.357	0.328	0.360
Etth1	96	**0.371**	**0.393**	0.395	0.417	0.391	0.415	0.387	0.410	0.397	0.421
	192	**0.405**	**0.413**	0.429	0.435	0.438	0.440	0.421	0.431	0.436	0.439
	336	**0.426**	**0.425**	0.453	0.453	0.460	0.456	0.469	0.463	0.470	0.462
	720	**0.436**	**0.452**	0.484	0.487	0.495	0.498	0.511	0.508	0.491	0.496

In this section, we perform ablation studies on Traffic, Weather and Electricity dataset to investigate the effect of different spatial and temporal embedding mechanisms: 1) **L embedding**: we add time stamp encoding of the historical L-step input data (Month, Day, Weekday, Hour, etc.); 2) **H embedding**: we add time stamp encoding of the H-step prediction output into MPPN; 3) **L & H embedding**: we map the data at each time-step into a low dimensional vector both in the input and the output. 4) **S & H embedding**: we consider spatial encoding for different channels of time series after pattern extraction together with the **H embedding**. It can be observed from Table 11 that none of the introduced spatio-temporal encoding mechanisms yields performance improvement for the original MPPN. This indicates that time encoding may be not necessary for time series modeling, as the multi-resolution and multi-period patterns extracted by MPPN could well capture the temporal features of time series, thus enabling efficient and accurate long-term time series prediction.

G SUPPLEMENTARY OF MAIN RESULTS

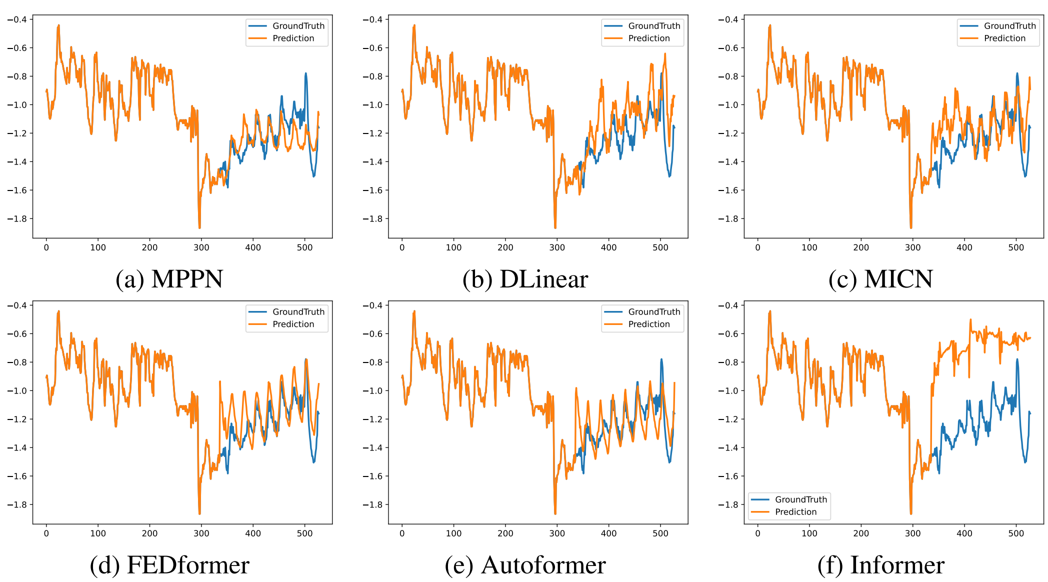
We plot the multivariate forecasting results of our MPPN and several baseline models on the ETTh1 and Weather dataset. All results are from the testing set for qualitative comparison. As depicted in Figure 6-13, our model outperforms the baseline models, exhibiting superior performance in modeling periodicity, short-term fluctuations, and long-term variations of time series. It demonstrates remarkable proficiency in capturing the complex patterns inherent in temporal series data, showcasing its effectiveness in accurately representing and analyzing diverse data characteristics over different time resolutions.



(a) MPPN (b) DLinear (c) MICN

(d) FEDformer (e) Autoformer (f) Informer

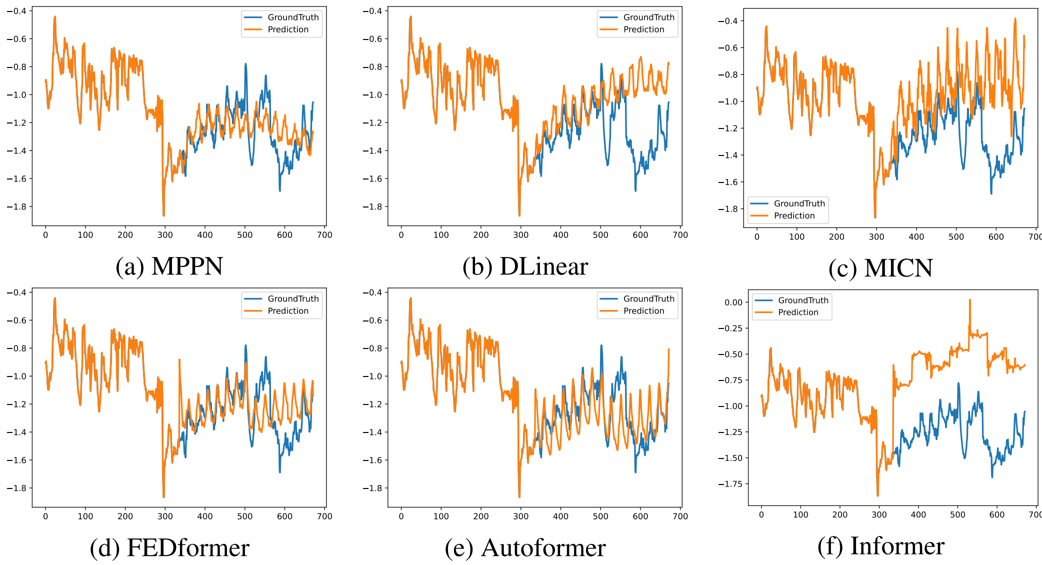
Figure 6: The prediction results on the ETTh1 dataset under the input-336-predict-96 settings.



(a) MPPN (b) DLinear (c) MICN

(d) FEDformer (e) Autoformer (f) Informer

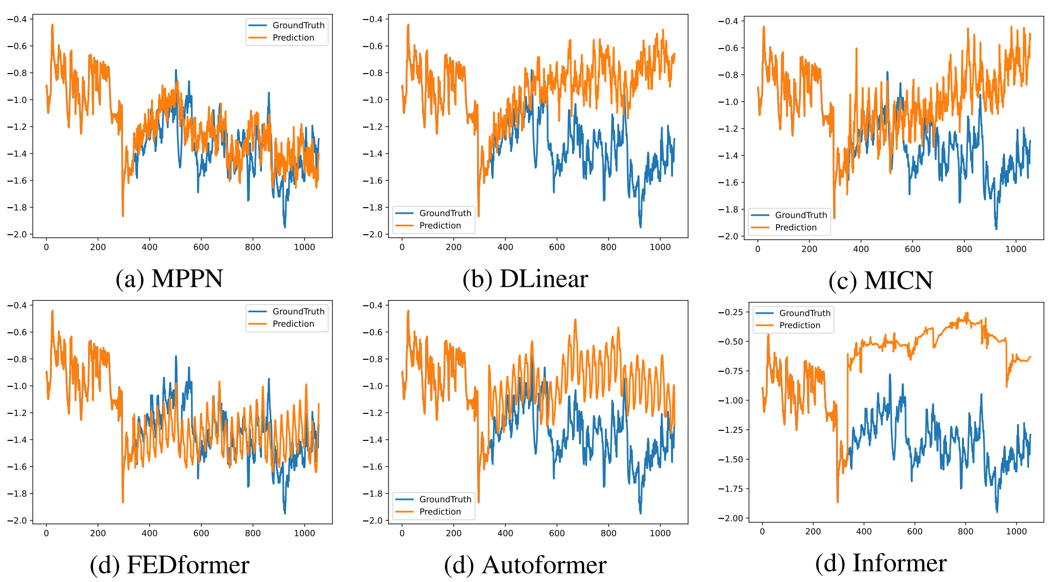
Figure 7: The prediction results on the ETTh1 dataset under the input-336-predict-192 settings.



(a) MPPN (b) DLinear (c) MICN

(d) FEDformer (e) Autoformer (f) Informer

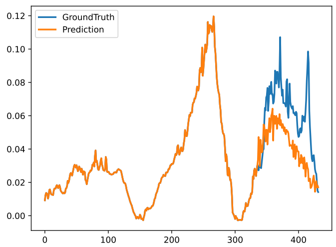
Figure 8: The prediction results on the ETTh1 dataset under the input-336-predict-336 settings.



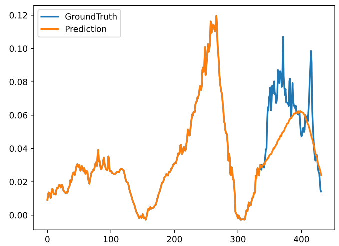
(a) MPPN (b) DLinear (c) MICN

(d) FEDformer (d) Autoformer (d) Informer

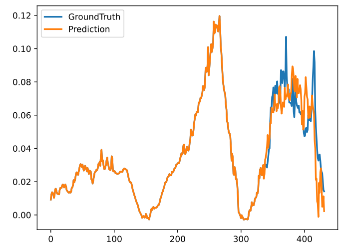
Figure 9: The prediction results on the ETTh1 dataset under the input-336-predict-720 settings.



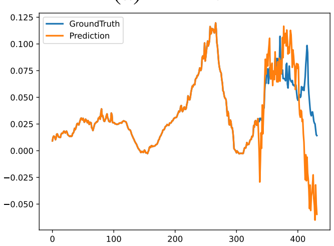
(a) MPPN



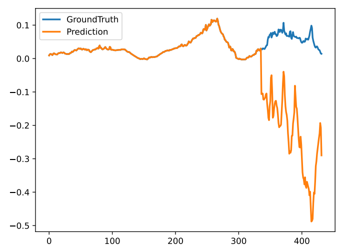
(b) DLinear



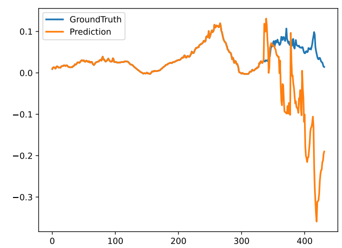
(c) MICN



(d) FEDformer

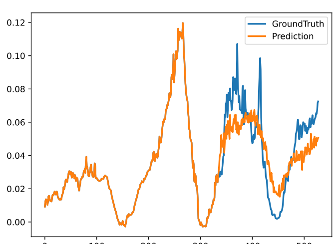


(e) Autoformer

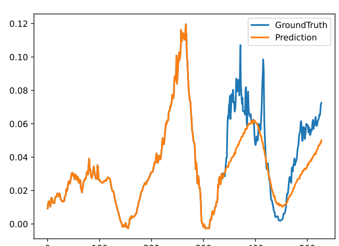


(f) Informer

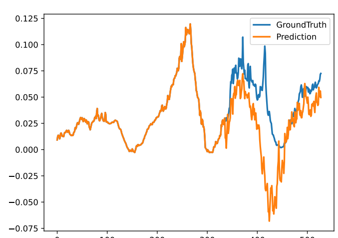
Figure 10: The prediction results on the Weather dataset under the input-336-predict-96 settings.



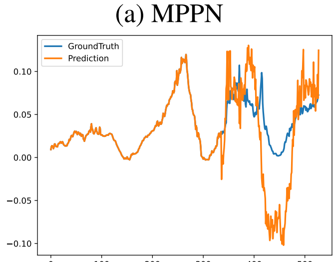
(a) MPPN



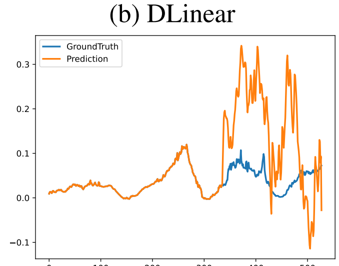
(b) DLinear



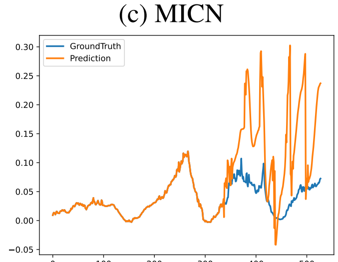
(c) MICN



(d) FEDformer

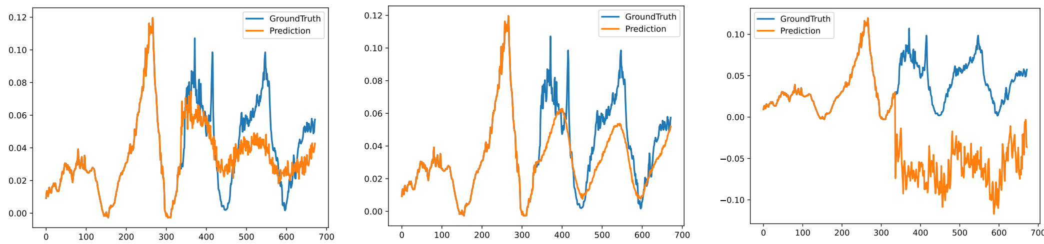


(d) Autoformer

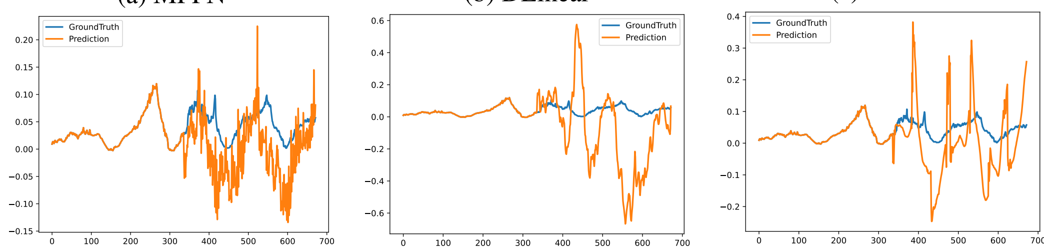


(f) Informer

Figure 11: The prediction results on the Weather dataset under the input-336-predict-192 settings.

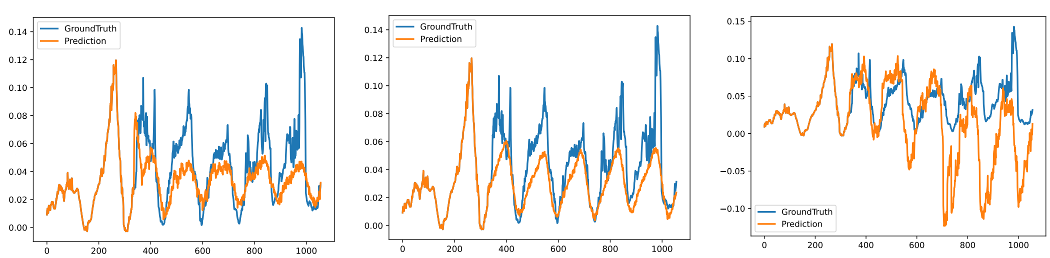


(a) MPPN (b) DLinear (c) MICN

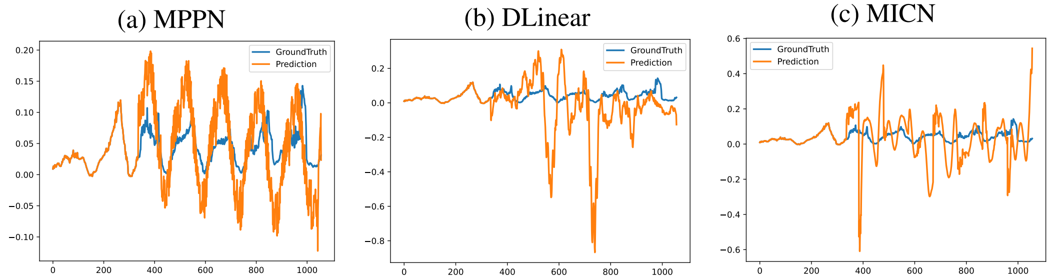


(d) FEDformer (d) Autoformer (d) Informer

Figure 12: The prediction results on the Weather dataset under the input-336-predict-336 settings.



(a) MPPN (b) DLinear (c) MICN



(d) FEDformer (d) Autoformer (d) Informer

Figure 13: The prediction results on the Weather dataset under the input-336-predict-720 settings.

H Limitations and future work

In this paper, we focus on extracting distinctive patterns to represent time series for long-term time series forecasting. We adopt multi-resolution patching and periodic pattern mining mechanism to explicitly extract the predictable patterns from time series data. Although MPPN can effectively capture patterns (e.g., periodicity and trend) in historical time series, most real-world time series can be very complex and influenced by various external factors, which may result in intricate patterns or even previously unseen variations. For example, if a traffic accident occurs by chance on a certain road causing traffic congestion, it will likely affect the speed of the road section; if extreme weather disasters (such as earthquakes and hurricanes) suddenly occur in a certain area, they will greatly affect the weather indicators. Sometimes, accidental events may also cause the peaks and valleys of the time series to arrive earlier or later, such as a customer staying up late to watch a World Cup game, resulting in a delay in electricity usage at home compared to historical baseline. Although MPPN can quickly capture these variations from the data of lookback window, it is difficult to foresee them.

In the future, we will consider how to capture the varying time series patterns and model the external factors. We plan to introduce knowledge graph for incorporating the knowledge of external factors to the prediction model. With the remarkable performance of foundation models in other fields, we will further investigate the general patterns of time series and spatial-temporal series data and construct task-general models to support multiple downstream time series analysis tasks. We believe that this direction is a particularly intriguing and significant.