

---

# Supplementary materials

---

Anonymous Author(s)

Affiliation

Address

email

## A Datasets and Implementation

### A.1 Datasets

We conducted experiments on 8 real-world datasets to evaluate the effectiveness of the our proposed PMLF loss functions across various domains. The detailed dateset information are depicted in Figure 1

- **ETT**(Electricity Transformer Temperature): The ETT dataset contains 7 variables of electricity transformer temperature from July 2016 to july 2018. There are 4 sub datasets: ETTh1, ETTh2, ETTm1, ETTm2, where ETTh recorded hourly and ETTm recorded every 15 minutes.
- **Weather**: Weather contains 21 meteorological variables collected every 10 minutes from the Weather Station of the Max Planck Biogeochemistry Institute in 2020.
- **ECL** : ECL records the hourly electricity consumption data of 321 clients from 2012 to 2014.
- **Traffic**: Traffic collects hourly road occupancy rates measured by 862 sensors of San Francisco Bay area freeways from January 2015 to December 2016.
- **Solar-Energy**: Solar records the solar power production of 137 PV plants in 2006, which are sampled every 10 minutes.

We follow the same data processing and train-validation-test set split protocol used in TimesNet, where the train, validation, and test datasets are strictly divided according to chronological order to make sure there are no data leakage issues. As for the forecasting settings, we fix the length of the lookback series as 96 and the prediction length varies in {96, 192, 336, 720}.

Table 1: Detailed Dataset Descriptions. Dim denotes the variable number of each dataset. Prediction Length denotes the future time steps to be predicted and four prediction setting are included in each dataset. Dataset Size denotes the total number of time steps in (Train, Validation, Test) split respectively. Frequency denotes the sampling interval of time steps.

Dataset	Dim	Prediction Length	Dataset Size	Frequency	Domain
ETTh1	7	{96, 192, 336, 720}	(8545, 2881, 2881)	1 hour	Electricity
ETTh2	7	{96, 192, 336, 720}	(8545, 2881, 2881)	1 hour	Electricity
ETTm1	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15 min	Electricity
ETTm2	7	{96, 192, 336, 720}	(34465, 11521, 11521)	15 min	Electricity
Weather	21	{96, 192, 336, 720}	(36792, 5271, 10540)	10 min	Weather
ECL	321	{96, 192, 336, 720}	(18317, 2633, 5261)	1 hour	Electricity
Traffic	862	{96, 192, 336, 720}	(12185, 1757, 3509)	1 hour	Transportation
Solar-Energy	137	{96, 192, 336, 720}	(36601, 5161, 10417)	10 min	Energy

## 20 A.2 IMPLEMENTATION DETAILS

21 All the experiments are implemented in PyTorch and conducted on four NVIDIA 4090 24GB GPU.  
22 For fair comparison, we set the input size for all models to be uniform, with the batchsize for ETT  
23 and Weather datasets set to 64 and the ECL, Traffic, and Solar datasets set to 16. Except for changing  
24 the learning rate to fully learn new structures, do not change other parameters related to the model.

25 Before calculating the loss, the time series is decomposed into seasonal and trend components using  
26 the moving average method. The kernel size refers to the setting in Autoformer, which is 25. But  
27 in order to avoid fixed kernel sizes affecting time series with different sampling frequencies, we  
28 designed a hybrid expert decomposition mechanism that uses a set of average pooling layers to extract  
29 trends and combines them with learnable weights. The kernel sizes  $\{7, 13, 15, 25, 49\}$  represent the  
30 corresponding periods at different frequencies.

## 31 B Baselines

32 To evaluate the general applicability of PMLF, we compare it with a varied collection of leading time-  
33 series forecasters that span the principal architectural families: state-space (S-Mamba), Transformer  
34 (iTransformer, TimeXer, PatchTST), multilayer perceptron (Amplifier, TimeMixer), and convolutional  
35 neural network (TimesNet). The core ideas of these baselines are outlined below.

- 36 • Amplifier: An energy-amplification block heightens weak spectral bands, the spectrum is re-  
37 normalised, and parallel seasonal and trend heads with a lightweight channel-interaction module  
38 enhance performance on low-signal datasets.
- 39 • TimeXer: Targets mixed endogenous and exogenous forecasting. Patch-wise self-attention models  
40 the target series, variate-wise cross-attention injects exogenous cues, and global endogenous  
41 tokens integrate the two streams, improving robustness to abrupt external shocks.
- 42 • S-Mamba: Represents each time step with a single per-channel token and applies a bidirectional  
43 Mamba state-space layer along the channel axis to capture inter-variable dependencies. This  
44 near-linear-time design surpasses Transformer baselines while greatly reducing computational  
45 cost.
- 46 • iTransformer: Reassigns Transformer roles by applying attention across variables to learn cross-  
47 channel links, while the feed-forward block operates along time to model nonlinear temporal  
48 dynamics. This rearrangement lowers memory usage for long horizons and yields interpretable  
49 variable-level attention.
- 50 • TimeMixer: A pure-MLP predictor. Depth-wise convolutions extract multi-scale bands, linear  
51 layers mix these components, and a parallel head simultaneously outputs all future steps. The  
52 absence of attention provides GPU-friendly speed without sacrificing accuracy.
- 53 • PatchTST: Divides long sequences into fixed-length temporal patches that serve as Transformer  
54 tokens and shares encoder parameters among channels. This strategy reduces attention complexity  
55 from  $\mathcal{O}(L^2)$  to  $\mathcal{O}((L/P)^2)$  while retaining local semantics.
- 56 • TimesNet: Converts a one-dimensional series into a two-dimensional time-period grid and  
57 employs heterogeneous CNN kernels to capture intra-period seasonality as well as inter-period  
58 trends. This representation enables direct transfer from vision backbones and delivers strong  
59 results in forecasting, anomaly detection, and classification.

60 These heterogeneous baselines ensure that any improvements attributed to PMLF are not limited to a  
61 single modelling philosophy but instead reflect a broad enhancement of time-series learning.

## 62 C More Experimental Results

### 63 C.1 Robustness Assessment

64 To examine the robustness of our framework, the Amplifier baseline was trained five times with  
65 independent random seeds. Table 2 reports the mean performance together with the corresponding  
66 standard deviations. The consistently small variances confirm that the Amplifier yields repeatable  
67 results, underscoring the robustness of the proposed approach.

Table 2: Robustness of PMLF performance. The results are obtained from five random seeds using the Amplifier as backbone.

Dataset	ETTh1		ETTh2		Weather	
Horizon	MSE	MAE	MSE	MAE	MSE	MAE
96	0.375 $\pm$ 0.002	0.394 $\pm$ 0.001	0.283 $\pm$ 0.000	0.332 $\pm$ 0.000	0.156 $\pm$ 0.001	0.193 $\pm$ 0.001
192	0.423 $\pm$ 0.003	0.425 $\pm$ 0.002	0.353 $\pm$ 0.002	0.379 $\pm$ 0.001	0.209 $\pm$ 0.002	0.243 $\pm$ 0.004
336	0.490 $\pm$ 0.003	0.456 $\pm$ 0.002	0.385 $\pm$ 0.000	0.405 $\pm$ 0.001	0.262 $\pm$ 0.002	0.283 $\pm$ 0.004
720	0.484 $\pm$ 0.003	0.467 $\pm$ 0.002	0.403 $\pm$ 0.002	0.423 $\pm$ 0.001	0.339 $\pm$ 0.002	0.334 $\pm$ 0.004

Dataset	ETTh1		ETTh2		ECL	
Horizon	MSE	MAE	MSE	MAE	MSE	MAE
96	0.311 $\pm$ 0.002	0.342 $\pm$ 0.002	0.172 $\pm$ 0.002	0.250 $\pm$ 0.001	0.148 $\pm$ 0.000	0.242 $\pm$ 0.000
192	0.371 $\pm$ 0.004	0.372 $\pm$ 0.003	0.237 $\pm$ 0.001	0.293 $\pm$ 0.002	0.161 $\pm$ 0.001	0.243 $\pm$ 0.000
336	0.399 $\pm$ 0.003	0.395 $\pm$ 0.003	0.299 $\pm$ 0.002	0.332 $\pm$ 0.002	0.170 $\pm$ 0.000	0.263 $\pm$ 0.001
720	0.481 $\pm$ 0.005	0.436 $\pm$ 0.004	0.391 $\pm$ 0.003	0.387 $\pm$ 0.004	0.199 $\pm$ 0.001	0.288 $\pm$ 0.001

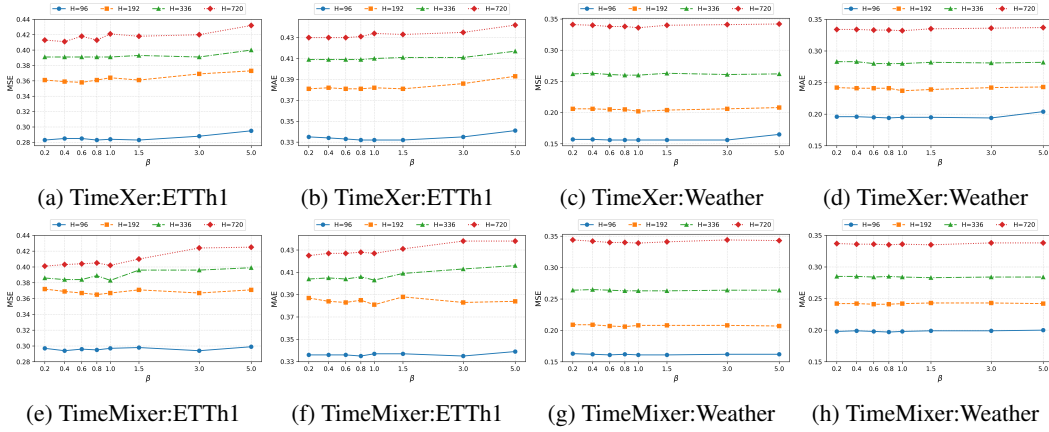


Figure 1: Sensitivity of the dynamic-weighting coefficient  $\beta$ . The first row shows TimeXer performance on ETTh2 (a, b) and Weather (c, d) as  $\beta$  varies, and the second row gives the corresponding results for TimeMixer.

## 68 C.2 Parameter Sensitivity

69 To examine how the dynamic-weighting coefficient  $\beta$  influences forecasting accuracy, we conducted  
70 a grid search over  $\beta \in \{0.2, 0.4, 0.6, 0.8, 1, 1.5, 3, 5\}$  for two representative networks: TimeXer and  
71 TimeMixer. Figure 1 reports the resulting MSE and MAE on the ETTh2 and Weather datasets. For  
72 TimeXer (top row) and TimeMixer (bottom row), the error curves remain nearly flat across the entire  
73 range, and the optimal  $\beta$  values cluster around 1.0 on both datasets. The maximum deviation from  
74 the best MSE and MAE is below 2.5%, indicating that the proposed dynamic weighting scheme is  
75 insensitive to the precise choice of  $\beta$ . These results confirm the robustness of our framework with  
76 respect to this hyper-parameter.

## 77 C.3 Additional Evaluation on Classical Forecasting Networks

78 Additionally, we assesses the architecture-independent effectiveness of PMLF on three widely used  
79 forecasting models: the Transformer-based PatchTST, the selective state-space model S Mamba, and  
80 the convolutional network TimesNet. For each architecture, the original MSE objective was replaced  
81 with PMLF and the models were evaluated on the ETT, Weather, ECL, and Solar datasets under four  
82 prediction horizons  $\{96, 192, 336, 720\}$ . As shown in Figure 2, across all datasets and horizons, the

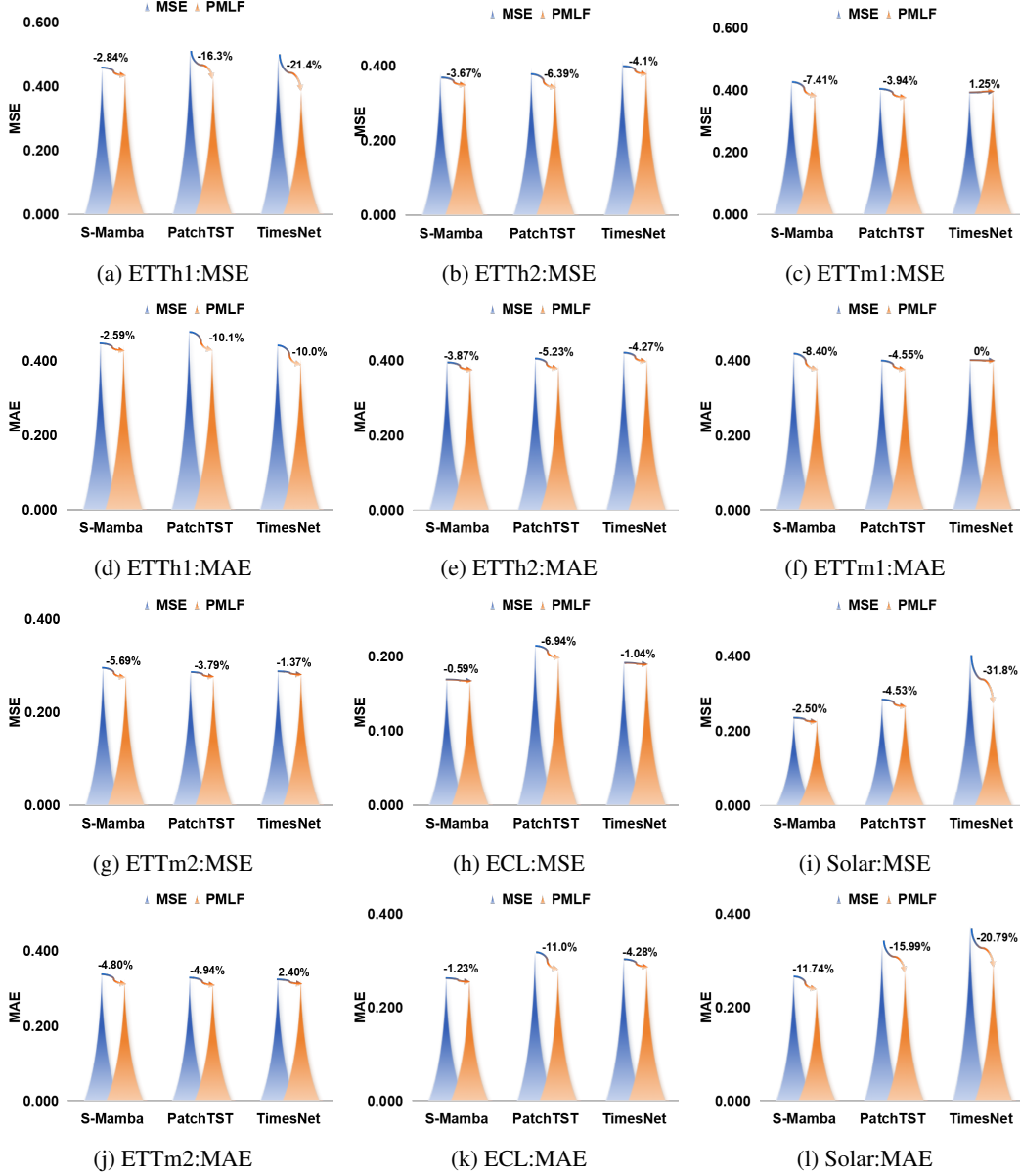


Figure 2: Average forecasting performance of PatchTST, S-Mamba, and TimesNet under four prediction horizons (96, 192, 336, 720) on the ETT (4 subsets), Weather, ECL, and Solar datasets. Models trained with PMLF are contrasted with their original MSE counterparts. Across all horizons and datasets, PMLF consistently reduces MAE, MSE, and RMSE, illustrating its effectiveness on these classical architectures.

substitution of MSE with PMLF consistently reduced the error metrics (MAE, MSE), confirming that PMLF improves forecasting accuracy and robustness even in classical network settings that are not covered by the main set of state-of-the-art architectures.

## D Visualization

As shown in Figure 3 and 4, we provide a visual comparison between PMLF and MSE on six benchmark datasets: ETTh2, ETTm2, Weather, ECL, Traffic, and Solar. Each row corresponds to one dataset, where the first column presents the overall forecast, and the second and third columns



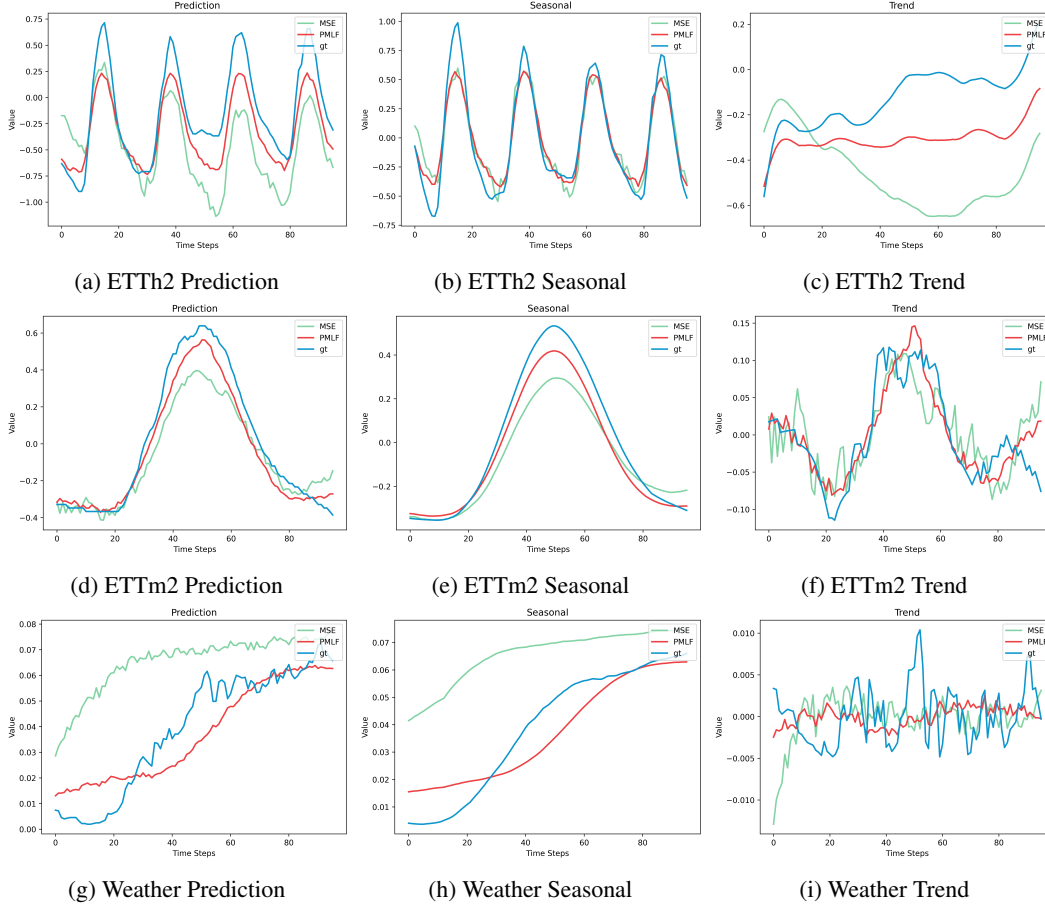


Figure 3: Visual comparison of Amplifier forecasts trained with PMLF and MSE across six benchmark datasets. Each row corresponds to one dataset (ETTh2, ETTm2, Weather), with three columns showing the overall prediction (left), seasonal component (middle), and trend component (right).

90 show the decomposed seasonal and trend components, respectively. Across all datasets, the forecasts  
 91 generated using PMLF exhibit closer alignment with the ground-truth signals, capturing long-term  
 92 trends and periodic patterns with higher fidelity compared to those produced with MSE.

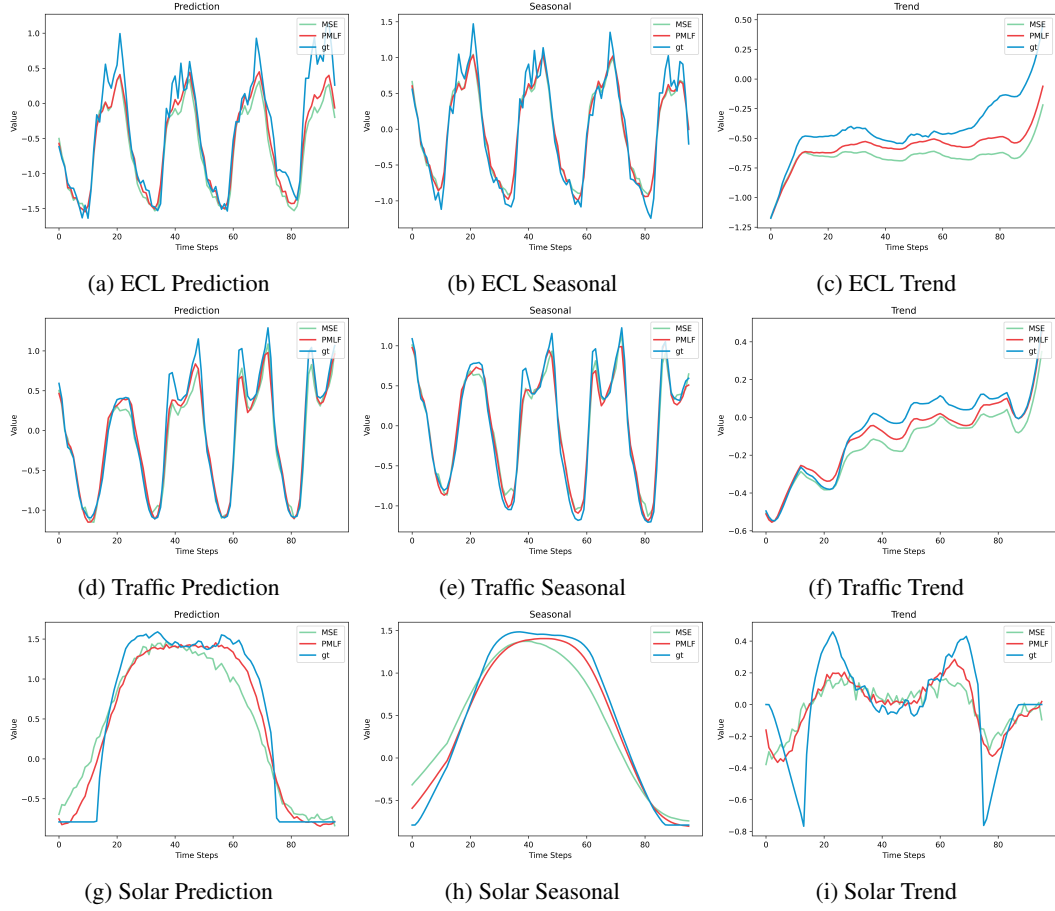


Figure 4: Visual comparison of Amplifier forecasts trained with PMLF and MSE across six benchmark datasets. Each row corresponds to one dataset (ECL, Traffic, and Solar), with three columns showing the overall prediction (left), seasonal component (middle), and trend component (right).