


# Wavelet-TimesNet: Improving Long-Term Solar Power Forecasting via Adaptive Wavelet Transform and Multi-Scale Residual Networks


**Guohui Liu**

(Beijing Union University, BeiJing, China)

 <https://orcid.org/0009-0004-4472-3361>, [20241081210204@buu.edu.cn](mailto:20241081210204@buu.edu.cn))


**Huan Zhang**

(Beijing Union University, BeiJing, China)

 <https://orcid.org/0000-0003-1265-2447>, [huan@buu.edu.cn](mailto:huan@buu.edu.cn) )


**Jianghong Li**

(Beijing Union University, BeiJing, China)

 <https://orcid.org/0009-0001-2721-3968>, [20231081210209@buu.edu.cn](mailto:20231081210209@buu.edu.cn))


**Yanling Zhao**

(Beijing Union University, BeiJing, China)

 <https://orcid.org/0009-0001-1166-4860>, [20231081210202@buu.edu.cn](mailto:20231081210202@buu.edu.cn))

**Xin Liu**

(Beijing Union University, BeiJing, China)

 <https://orcid.org/0009-0004-1529-4897>, [20242085410215@buu.edu.cn](mailto:20242085410215@buu.edu.cn))

**Abstract:** Long-term photovoltaic power prediction is crucial for the optimal dispatch of energy systems and the stability of power grids. However, existing methods are limited in accuracy when dealing with non-stationary signals such as intermittent fluctuations in light due to issues like spectral leakage and the rigidity of fixed convolutional kernel feature extraction. To address these challenges, this paper proposes a novel model, Wavelet-TimesNet, which integrates adaptive wavelet transform and multi-scale residual networks, aiming to enhance the robustness of long-term predictions. This model dynamically adjusts the parameters of the wavelet basis function to achieve multi-resolution analysis of local periodic features, effectively suppressing noise interference. It constructs a multi-scale residual network to capture local details such as hourly irradiance mutations and global trends such as seasonal power variations using parallel convolutional kernels of different sizes. An adaptive wavelet attention mechanism is introduced to dynamically weight and fuse frequency-domain and time-domain features, enhancing the focus on key information. Experiments were conducted based on photovoltaic datasets from Xinjiang's temperate continental climate and a subtropical monsoon climate region in China. The results show that in 96-hour predictions, Wavelet-TimesNet reduces the mean absolute error (MAE) by 4.97% and 3.29% in Xinjiang and China, respectively, and the mean squared error (MSE) by 7.80% and 5.75%. In 192-hour predictions, the mean squared percentage error (MSPE) of the Xinjiang dataset is reduced by 36.42%. Compared with advanced models such as Transformer and Informer, this model demonstrates significant advantages in handling non-stationary signals and capturing long-term trends, especially in extreme weather scenarios like sudden sandstorms and continuous rainy days, where prediction accuracy is notably improved. The research results provide an efficient solution for the precise dispatch of photovoltaic power stations, which is of

great significance for reducing peak shaving costs in power grids, promoting the consumption of renewable energy, and facilitating the transformation of energy structures.

**Keywords:** Photovoltaic power generation; Long-term prediction; Adaptive wavelet transform; Multi-scale residual network; TimesNet

**Categories:** H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

**DOI:** 10.3897/jucs.155783

## 1 Introduction

With the continuous growth of global energy demand, the consumption of traditional fossil fuels has exacerbated environmental pollution and resource depletion. According to the International Energy Agency (IEA), global energy demand is projected to increase by approximately 30% by 2040. Meanwhile, global warming has intensified the impacts of climate change, further promoting the development of renewable energy, especially photovoltaic (PV) power generation, which, as a green and low-carbon energy form, is playing an increasingly important role in the energy structure transformation. According to the IEA's 2024 Renewable Energy Report, global renewable energy installed capacity is expected to triple by 2030 compared to the past six years, with photovoltaic power generation accounting for as much as 80%, becoming the core driving force for clean energy growth. Its emission reduction benefits are equivalent to reducing 2.5 billion tons of carbon dioxide emissions annually.

However, the intermittency and randomness of photovoltaic power generation remain a challenge for grid integration. The intermittent changes in photovoltaic power generation depend on weather conditions such as solar radiation, temperature, humidity, cloud cover, and wind speed, all of which are unpredictable and uncontrollable [Loganathan, 21]. Notably, changes in temperature and solar radiation have a substantial impact on power generation [Alay, 24]. The nature of these variables can lead to instability in photovoltaic power generation, causing sudden surpluses or shortages in power output, which may result in an imbalance between power generation and load demand, triggering grid control and operation issues. If the power generation can be accurately predicted, operational optimization strategies such as peak shaving can be effectively utilized to reduce the uncertainty of the power generation system. For photovoltaic systems, accurate power generation prediction is crucial, as it can enhance the quality of power supply to the grid and reduce costs associated with power generation fluctuations. Studies have shown that for every 1% reduction in photovoltaic power generation prediction error, the peak shaving cost of the grid can be reduced by approximately 7%, significantly improving the economic efficiency and stability of the energy system [Kardakos, 13].

The main contributions of this study are as follows:

- The adaptive wavelet transform is adopted to replace the traditional Fast Fourier Transform (FFT) for period detection. The traditional FFT has difficulties in capturing local non-stationary signals and the fixed period assumption deviates from the actual multi-scale characteristics when processing photovoltaic power data. The adaptive wavelet transform can precisely capture the time-varying period features of local non-stationary

signals by dynamically adjusting the scale and translation factors of the wavelet basis function and performing multi-resolution decomposition on the input sequence, effectively suppressing noise interference.

- A multi-scale residual network is constructed, using convolutional kernels of different sizes ( $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ) to process the input data in parallel. The original TimesNet model has limitations in extracting multi-scale features with fixed convolutional kernels. However, by constructing a multi-scale residual network, comprehensive extraction of multi-scale features from photovoltaic power data can be achieved. Small convolutional kernels focus on local detail features, such as hourly irradiance mutations; large convolutional kernels capture macroscopic trend features, such as seasonal power variations.
- An adaptive wavelet attention mechanism is introduced to weight the features from both the frequency domain and the time domain. The original TimesNet model lacks the ability to distinguish the importance of different features during the feature processing, treating all features equally. By introducing the adaptive wavelet attention mechanism, features that are more important for photovoltaic power prediction can be automatically identified and their weights enhanced, while the influence of irrelevant features is suppressed, avoiding equal treatment during feature fusion. This enables the model to focus more on key information, such as high-frequency power fluctuations corresponding to extreme weather, improving the accuracy and stability of the prediction.
- The overall model adopts a three-stage processing paradigm of "dynamic decomposition - multi-granularity modeling - intelligent fusion". The three innovative modules work together to form the MSResNet Block module. The adaptive wavelet transform performs dynamic decomposition, the multi-scale residual network achieves multi-granularity modeling, and the adaptive wavelet attention mechanism completes intelligent fusion. This effectively addresses the rigidity of non-stationary signal processing and feature extraction, providing a solution with higher accuracy and robustness for photovoltaic power prediction.

The remaining organization of the paper is as follows: Section 2 provides a detailed introduction to the research status of photovoltaic power prediction at home and abroad; Section 3 elaborates on the proposed Wavelet-TimesNet model, including its structural design, wavelet period detection method, adaptive wavelet attention mechanism, and the implementation of the multi-scale residual network, and briefly describes the original model; Section 4 summarizes the dataset information and evaluation metrics, and presents the experimental results and analysis; Section 5 summarizes the research findings of this paper.

## 2 Related Work

Research on photovoltaic power generation prediction models can be classified into five major categories: physical methods, empirical methods, statistical methods, machine learning methods, and hybrid methods. Physical methods are based on numerical weather prediction (NWP) and physical equations of photovoltaic systems,

achieving predictions by simulating atmospheric radiation, cloud movement, and other processes [Zhi, 23], or by introducing physical modeling intermediate variables in data preprocessing to consider the physical characteristics of photovoltaic data [Tao, 24]. These methods rely on high-precision meteorological inputs and have high computational complexity, making them suitable for medium and long-term predictions. However, complex atmospheric conditions make physical models structurally complex, limiting their wide application in practice. In contrast, empirical methods rely on historical patterns or expert experience, using algorithm-optimized parameters based on empirical equations to estimate average global solar radiation for predicting photovoltaic power generation [Tolabi, 14]. They are simple and easy to implement but have limited modeling capabilities for nonlinear relationships. Statistical methods mainly include time series analysis (such as ARIMA) [Xie, 22] and regression models (such as Bayesian averaging) [Tahir,24], which are highly interpretable but struggle to handle the non-stationarity and multi-scale periodic characteristics of photovoltaic power. In recent years, machine learning methods, including support vector machine (SVM) models [Doubleday,20] and random forests [Rafati, 21], have become a research hotspot due to their excellent nonlinear modeling capabilities. Compared to physical, empirical, and statistical models, they offer higher accuracy and broader applicability [Wu, 21], and have had a significant impact in the field of prediction. Loganathan et al. integrated satellite images and meteorological data and used machine learning models such as neural networks to estimate solar radiation. The results showed that the model integrating data performed better in predicting solar radiation than models using only meteorological data or image data, significantly improving prediction accuracy [Ahn, 21]. As an evolutionary extension of machine learning, deep learning can effectively solve complex problems and achieve good prediction accuracy by increasing the complexity and depth of the model. RNN has certain advantages in short sequence prediction tasks [Park, 21], but when applied to long sequence data, it is prone to problems of vanishing and exploding gradients. Unlike RNN, LSTM does not repeatedly cover the input information. By introducing gating mechanisms and cell states [Harrou, 20], LSTM effectively solves the problems of vanishing and exploding gradients in RNN when processing long sequence data, enhancing its ability to capture long-term dependencies. However, a common problem with traditional RNN and LSTM is their limited parallelism, which causes efficiency-related issues. To address the shortcomings of single models, hybrid models have emerged, combining the advantages of multiple methods and significantly improving model performance and application. Common hybrid combinations include deep learning models combined with deep learning models, statistical methods combined with deep learning models, and data decomposition methods combined with deep learning models. Lim et al. proposed the CNN-LSTM model, which can simultaneously capture the spatial and temporal features of data by integrating spatial feature extraction and dynamic modeling, thereby improving prediction accuracy [Lim, 22]. Wu et al. proposed the CNN-Informer model, which uses CNN to extract feature information and the Informer model to establish temporal feature relationships for photovoltaic power prediction. This model demonstrated high accuracy and stability in photovoltaic power prediction [Zhuang, 24]. Liu et al. proposed the FCM-LSTM model, which uses the fuzzy C-means clustering algorithm to classify data and then builds long short-term memory neural network prediction models for different categories. This model can effectively handle data under different weather conditions and reduce prediction errors

[Liu, 24]. AlKandari et al. proposed a hybrid model that combines deep learning methods (such as LSTM and GRU) with the Theta statistical method for solar power prediction, which demonstrated excellent predictive performance under different climatic conditions [AlKandari, 24]. Zhang et al. combined similar day clustering, CNN, and LSTM for ultra-short-term photovoltaic power prediction, and this model exhibited outstanding overall performance in time series modeling and information capture, enhancing the understanding of the seasonal, periodic, and irregular patterns of photovoltaic power generation [Zhang, 24]. Cui et al. combined weather type clustering, AHA-VMD-MPE decomposition and reconstruction, and an improved Informer model for short-term photovoltaic power prediction, and this model demonstrated high accuracy and stability in short-term photovoltaic power prediction [Cui, 24]. Sarah et al. proposed a method of multi-level data fusion and deep neural networks, integrating spatial and temporal information for regional-level solar power prediction. This method outperformed existing deep learning methods in terms of prediction accuracy and provided new ideas for regional-level solar power prediction [Almaghrabi, 24]. Branco et al. proposed a hybrid Wavelet-LSTM model that combines wavelet transform with long short-term memory networks for fault forecasting in electrical power grids. By enhancing the modeling of nonlinear time series through noise reduction, their approach outperformed standard LSTM models in both accuracy and robustness, offering a reliable solution for fault prediction in power systems [Branco, 22]. Although hybrid approaches often surpass single models in point accuracy, their performance typically hinges on external signal decomposition and manually tuned hyperparameters, which prevents end-to-end, data-driven adaptation. The resulting multi-stage pipeline also introduces error propagation and nontrivial computational overhead. Moreover, in our PV-forecasting setting, these hyperparameters are sensitive to noise levels, missing-data patterns, and regional climatic heterogeneity, leading to degraded transferability across regions and seasons.

With the continuous development of deep learning technology, single models for time series have made significant progress in handling long sequence data. The emergence of the Transformer model in 2017 marked the beginning of a new era [Vaswani, 17]. By introducing the self-attention mechanism, Transformer overcame the sequential dependency of traditional RNNs and could effectively model long-range dependencies in long sequences. This innovation not only significantly improved the performance of the model but also laid the foundation for the development of subsequent models. However, Transformer still faces issues of computational efficiency and memory usage when dealing with complex time series data. In 2021, the Informer model was further optimized based on Transformer by introducing a sparse self-attention mechanism [Zhou, 21]. This improvement not only reduced the computational complexity but also maintained the ability to model long sequences. By sparsifying the attention matrix, Informer significantly enhanced the efficiency and accuracy of the model, making it perform well in solar irradiance prediction. Nevertheless, Informer still has room for improvement in handling dependencies between different time steps. In the same year, the Autoformer model brought a new breakthrough to time series prediction [Wu, 21]. Autoformer introduced decomposition blocks and the Fast Fourier Transform (FFT) to more effectively extract long-term trend information from time series. This innovation not only improved the model's performance in long-term prediction tasks but also provided new ideas for handling multi-scale features. Autoformer outperformed Informer in long-term prediction tasks

but still faced challenges in handling multi-scale features. In 2023, the iTransformer model combined multi-task learning with the Transformer architecture, enabling it to handle multiple resource components simultaneously [Liu, 23]. This innovation not only enhanced the model's performance in processing multi-source data but also provided new methods for modeling complex time series data. By introducing multi-task learning, iTransformer further strengthened the model's robustness and accuracy in handling multi-source data.

Against this backdrop, new single models are beginning to emerge. Unlike hybrid models, these new single models enhance performance through innovations in their architecture and algorithms. They abandon the complex combination methods of hybrid models and focus on optimizing from within a single model, avoiding the problems of complex structure and high computational cost of hybrid models. In terms of cycle detection and feature extraction, the new single models explore more efficient and accurate methods, demonstrating unique advantages and gradually becoming a research hotspot in the field of photovoltaic power generation prediction. Current research shows two major trends: first, multi-modal feature fusion, such as combining meteorological data, decomposing sub-sequences and classifying weather types; second, optimizing time series modeling, enhancing the ability to capture long-term dependencies through attention mechanisms and multi-scale convolution. However, most models rely on the Fast Fourier Transform (FFT) for global cycle detection, which is prone to spectral leakage for non-stationary signals of photovoltaic power (such as intermittent fluctuations under cloudy weather), leading to the extraction of pseudo-cycles. Moreover, feature extraction mostly uses fixed convolution kernels or simple residual connections, making it difficult to dynamically adapt to different cycle patterns. As an emerging single time series model, TimesNet maps one-dimensional time series into two-dimensional tensors through multi-period transformation, and combines convolution to mine intra-period and inter-period changes, performing exceptionally well in long-term prediction tasks [Wu,22]. Building on these limitations, we propose Wavelet-TimesNet, a unified end-to-end architecture for long-horizon PV power forecasting. The model replaces external, fixed decompositions with learnable wavelet parameters inside the network and couples a multi-scale residual backbone with wavelet-aware attention fusion. This design enhances adaptability to non-stationary signals and multi-scale dependencies, while mitigating the error accumulation and inference-latency overhead characteristic of multi-stage hybrid pipelines.

### 3 Methodology

This paper proposes an improved Wavelet-TimesNet model for time series prediction tasks. The model has undergone multiple enhancements on the basis of the original TimesNet architecture, including the introduction of a new periodicity detection method, a multi-scale residual network, a two-dimensional feature extraction module with an adaptive wavelet transform attention mechanism, and the use of residual connections to optimize the network structure. The following sections provide a detailed description of each module of the original model, each module of the improved model, and the overall network framework, as follows.

### 3.1 TimesNet

In the real world, time series often exhibit multiple periodicities and are influenced by the interactions among these periods, leading to highly complex variations. This complexity poses challenges for modeling time series data in short-term load forecasting. Traditional methods in time series analysis mainly focus on the continuity and recursiveness of time series, while ignoring the underlying multi-periodicity. In contrast, TimesNet regards individual time points in a time series as relatively independent. On one hand, it considers overlapping time periods, capturing the interactions between different periods and emphasizing the changes between periods. On the other hand, it examines the relationships among time points within each period, with a focus on the changes between adjacent moments within the same period, referred to as intra-period changes. However, traditional one-dimensional space time series modeling methods cannot aggregate time points from different periods into the same stage but rather disperse them throughout the time series, making it difficult to achieve mutual attention. Moreover, these methods cannot identify the influence relationships between different periods, making it hard to simultaneously capture both types of changes in one-dimensional space modeling. To address this issue, TimesNet extends the analysis by transforming time variations into a two-dimensional space, as shown in Figure 1. The 1D time series is reshaped into a 2D tensor, where each column contains the time points within a period, and each row contains the time points corresponding to the same phase but from different periods.

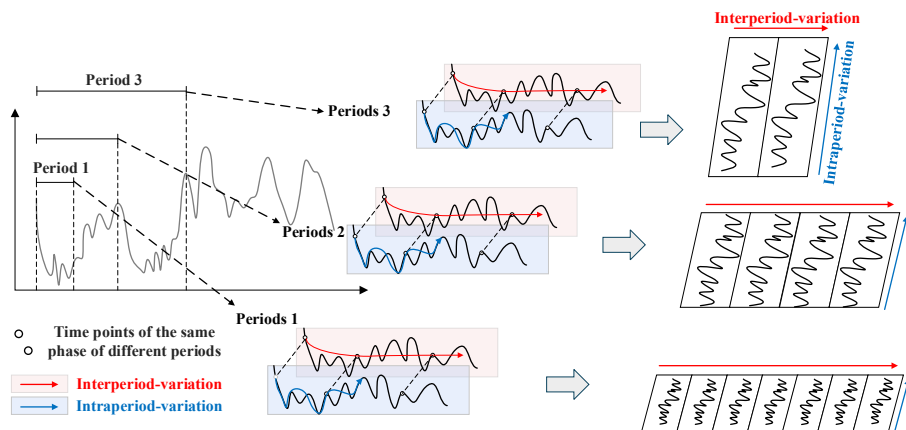


Figure 1: The multi-periodicity and 2D temporal variation of time series

The TimesBlock aims to integrate the above processes and effectively model both types of changes using a parameter-efficient initial block. The configuration of TimesNet will be described in detail below.

### 3.1.1 Theory of Transformation from 1D to 2D

To consistently capture and represent the temporal variations within and between periods, it is necessary to study the periodicity of time series. This can be achieved by applying the Fast Fourier Transform (FFT) along the time dimension of the 1D time series, denoted as  $X_{1D} \in \mathbb{R}^{T \times d_{\text{model}}}$ , where  $T$  represents the time length and  $d_{\text{model}}$  indicates the input feature dimension. The FFT calculation along the time dimension enables us to identify the periodic patterns presented in the time series as follows:

$$\begin{cases} A = \text{Avg}(\text{Amp}(\text{FFT}(X_{1D}))) \\ \{f_1, \dots, f_k\} = \text{arg}_{f_s \in \{1, \dots, \lfloor \frac{T}{2} \rfloor\}} \text{Top}k(A) \\ p_i = \lceil T/f_i \rceil, i \in \{1, 2, \dots, k\} \end{cases}$$

Where  $A \in \mathbb{R}^T$  represents the intensity of each frequency component in  $X_{1D}$ . The  $k$  largest frequencies  $\{f_1, \dots, f_k\}$  correspond to the  $k$  most significant period lengths  $\{p_1, \dots, p_k\}$  in the time series. This process can be concisely described as follows:

$$A, \{f_1, \dots, f_k\}, \{p_1, \dots, p_k\} = \text{Period}(X_{1D})$$

As shown in Figure 2, the original 1D time series  $X_{1D}$  is folded according to the selected period. This process can be formalized as follows:

$$X_{2D}^i = \text{Reshape}_{p_i, f_i}(\text{Padding}(X_{1D})) \quad i \in \{1, \dots, k\}$$

where  $\text{Padding}(\cdot)$  refers to the process of appending zeros to the end of the sequence to ensure that the sequence length is divisible by  $p_i$ . By performing the above operation, we obtain a set of 2D tensors  $\{X_{2D}^1, X_{2D}^2, \dots, X_{2D}^k\}$ , each  $X_{2D}^i$  corresponding to the 2D temporal variations dominated by the period  $p_i$ .

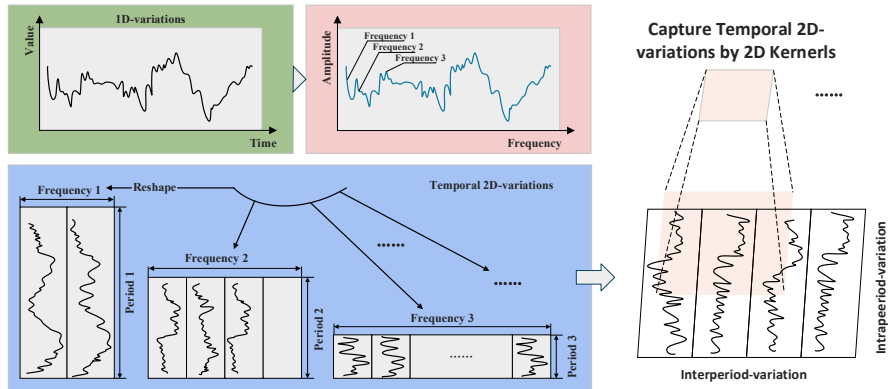


Figure 2: 2D structural time series and the variation of 2D nuclear capture time

### 3.1.2 TimesBlock

The architecture of TimesNet is shown in Figure 3. For the  $l$ -th layer of TimesBlock, the input is  $X_{1D}^{l-1} \in \mathbb{R}^{T \times d_{\text{model}}}$ , and 2D temporal variations can be extracted by using 2D convolution processing. The process can be formalized as:

$$X_{1D}^l = \text{TimesBlock}(X_{1D}^{l-1}) + X_{1D}^{l-1}$$

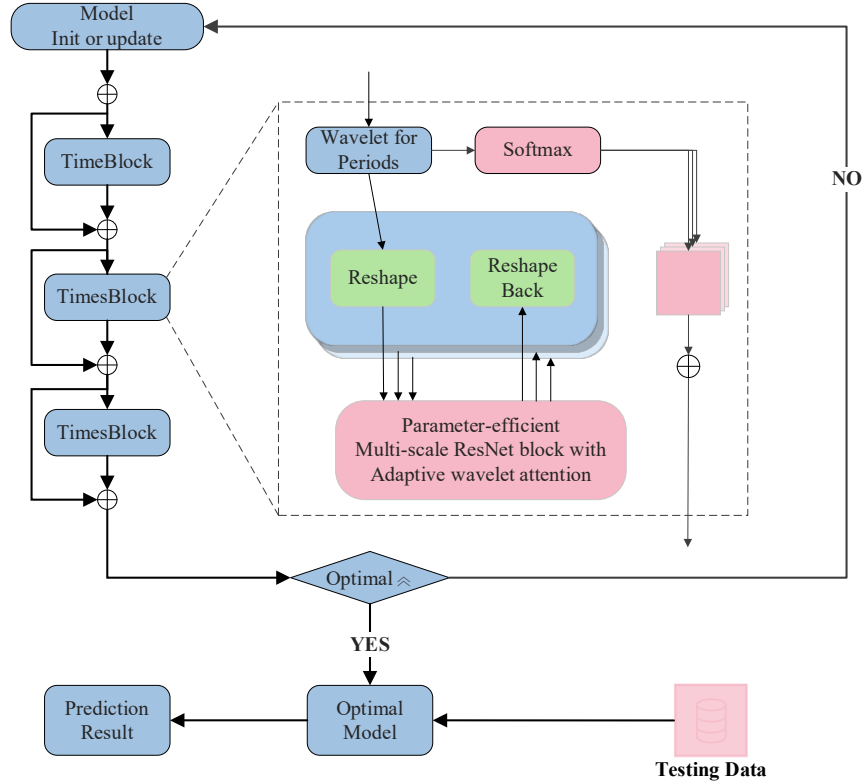


Figure 3: The overall architecture of TimesNet

The whole process of TimesBlock for capturing temporal 2D variations of time series will be elaborated in detail, which consists of the following four sub-processes:

(1) Transforming 1D time series into 2D tensors.

Firstly, the 1D temporal features of the input are extracted for periods and transformed them into 2D tensors to represent the 2D temporal changes. According to the process described in the previous discussion, the formula can be rewritten as:

$$A^{l-1}, \{f_1, \dots, f_k\}, \{p_1, \dots, p_k\} = \text{Period}(X_{1D}^{l-1})$$

$$X_{2D}^{l,i} = \text{Reshape}_{p_i, f_i}(\text{Padding}(X_{1D}^{l-1})) \quad i \in \{1, \dots, k\}$$

(2) Capturing 2D temporal variations representation.

For the 2D tensors, which possess 2D locality, we can utilize 2D convolutions to extract information. We adopt the classical Inception model.

$$\hat{X}_{2D}^{l,i} = \text{Inception}(X_{2D}^{l,i})$$

(3) Transforming 2D tensor to 1D Space.

After capturing the temporal features, we convert them back to 1D space for information aggregation.

$$\hat{X}_{1D}^{l,i} = \text{Trunc}\left(\text{Reshape}_{1,(p_i \times f_i)}(\hat{X}_{2D}^{l,i})\right) \quad i \in \{1, \dots, k\}$$

Where  $\hat{X}_{1D}^{l,i} \in R^{T \times d_{model}}$ ,  $\text{Trunc}(\cdot)$  represents the removal of the zeros added during the  $\text{Padding}(\cdot)$  operation in step (1).

(4) Adaptive aggregation.

Similar to the design in Autoformer, the adaptive fusion is performed by taking the weighted sum of the obtained 1D representations based on their corresponding frequency values, resulting in the final output.

$$\hat{A}_{f_1}^{l-1}, \dots, \hat{A}_{f_k}^{l-1} = \text{Softmax}(A_{f_1}^{l-1}, \dots, A_{f_k}^{l-1})$$

$$X_{1D}^l = \sum_{i=1}^k \hat{A}_{f_i}^{l-1} \times \hat{X}_{1D}^{l,i}$$

TimesBlock is capable of simultaneously capturing fine-grained and multi-scale temporal 2D patterns. As a result, compared to directly extracting features from 1D time series, TimesNet achieves more efficient representation learning.

### 3.2 Wavelet-TimesNet

The core design philosophy of the Wavelet-TimesNet model is to achieve high-precision long-term forecasting of photovoltaic power by leveraging dynamic multi-scale feature analysis and adaptive weighted fusion. The overall architecture of the model is illustrated in Figure 4, which outlines the complete pipeline from data preprocessing and periodicity analysis to model training based on a Multi-Scale Residual Network (MSResNet Block) and an adaptive wavelet attention mechanism. The model adopts a three-stage processing paradigm of “dynamic decomposition – multi-granularity modeling – intelligent fusion”, enabling robust representation learning across varying temporal resolutions.

Firstly, by replacing the traditional FFT with an adaptive wavelet period detector [Ding, 24], the input sequence is decomposed into multiple resolutions based on the dynamically adjusted wavelet basis function parameters (scale and translation factor), accurately capturing the time-varying periodic characteristics of local non-stationary signals (such as intermittent fluctuations under cloudy weather) and suppressing noise interference. Secondly, a parallel convolutional branch is constructed based on a multi-scale residual network, using convolutional kernels of different sizes ( $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ ) to simultaneously extract local details (such as hourly irradiance mutations) and global trends (such as seasonal power variations) of the time series, breaking through the limitations of fixed convolutional kernels in extracting multi-scale features. Finally, an adaptive wavelet attention mechanism is employed to achieve dynamic weighted fusion of frequency domain features (wavelet transform output) and time domain features (original signal), automatically focusing on key information (such as high-frequency power fluctuations corresponding to extreme weather) through a dual frequency-domain and time-domain attention mechanism, avoiding equalized processing during feature fusion.

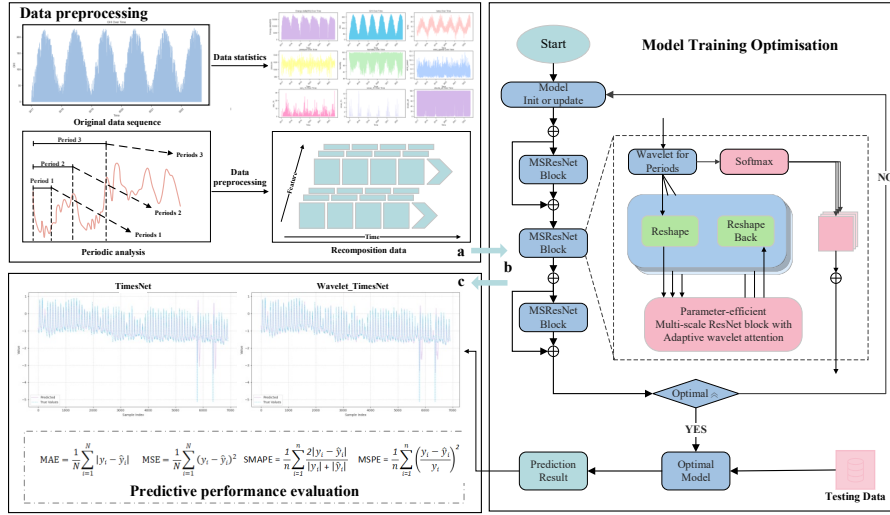


Figure 4: The overall processing flow of Wavelet-TimesNet

### 3.2.1 Period Detection Module

The traditional TimesNet model employs the Fast Fourier Transform (FFT) for period detection. However, the approach encounters two primary issues when dealing with photovoltaic power data: Firstly, the global frequency analysis of FFT struggles to capture the local non-stationary signals caused by sudden weather changes. Secondly, the fixed period assumption deviates from the actual multi-scale characteristics of photovoltaic power generation. This paper proposes an Adaptive Wavelet Transform, whose structural formula is as follows.

Improved wavelet transform formula:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^* \left( \frac{t-b}{a} \right) dt$$

Where  $a$  is the scale parameter,  $b$  is the translation parameter, and  $\psi(t)$  is the adaptive wavelet basis function.

Adaptive wavelet basis function:

$$\psi(t) = \sum_{i=1}^n w_i \cdot \psi_i(t)$$

where  $w_i$  is the weight of the  $i$ -th wavelet basis function, and  $\psi_i(t)$  is the predefined wavelet basis function.

Dynamic scale adjustment:

$$a = f(input_{data})$$

Where  $f$  is a dynamic adjustment function based on input data, obtained through neural network learning.

Improved wavelet transform output:

$$F_{wavelet}(t) = \sum_{j=1}^m W(a_j, b_j)$$

Where  $a_j$  and  $b_j$  represent the  $j$ -th scale and translation parameters.

### 3.2.2 Multi-scale Residual Network Module

The Inception module in the original TimesNet model mainly relies on fixed-scale convolution operations for feature extraction. This fixed-scale convolution approach can only capture feature information at specific scales, and its ability to capture the complex and variable multi-scale features in photovoltaic power data, especially the trend and fluctuation features at different time granularities, is limited. For example, it is difficult for fixed-scale convolution to simultaneously handle short-term sudden changes in light intensity and long-term seasonal power variations, leading to poor performance of the model in processing photovoltaic data with a wide range of scales. The introduction of the Multi-scale Residual Network (MSResNet) module effectively addresses the above issues. This module processes the input data in parallel using convolution kernels of different scales, enabling it to capture feature information at different time scales simultaneously. Convolution kernels of different sizes can cover different time windows, with small kernels focusing on local detail features and large kernels capturing more macroscopic trend features, thereby achieving comprehensive extraction of multi-scale features in photovoltaic power data. Its structure is shown in the following Figure 5.

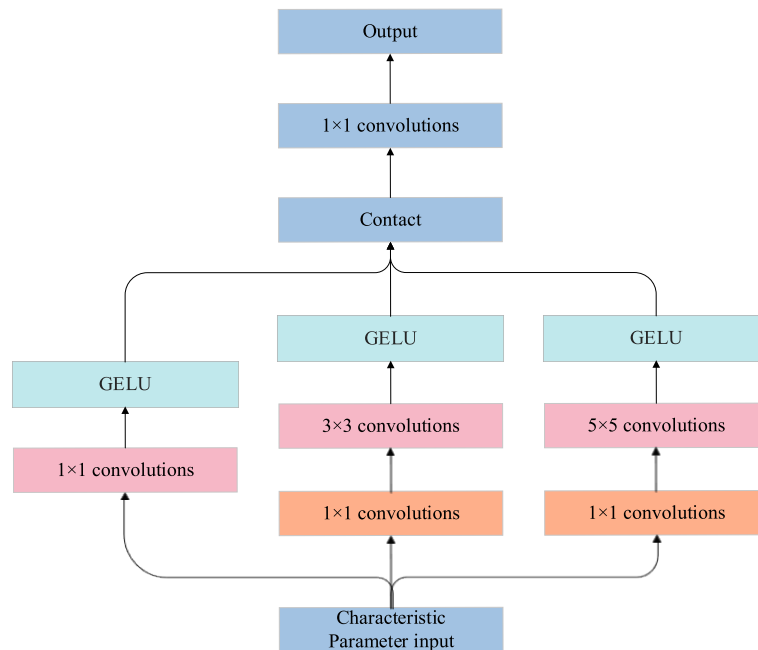


Figure 5: Multi-scale Residual Network

The processing flow of MSResNet is as follows when the input  $X \in \mathbb{R}^{B \times T \times N}$  is given:

Multi-scale convolutional branch:

$$\begin{cases} Y_1 = GELU(Conv1D_{k=1}(X)) \\ Y_3 = GELU(Conv1D_{k=3}(X)) \\ Y_5 = GELU(Conv1D_{k=5}(X)) \end{cases}$$

Feature fusion:

$$Y_{contact} = W_{fuse} \cdot Concat(Y_1, Y_3, Y_5) + b_{fuse}$$

Where  $W_{fuse} \in \mathbb{R}^{3d_{model} \times d_{model}}$  are learnable parameters and  $b_{fuse}$  is the bias term.

Residual connection:

$$Y_{out} = LayerNorm(Y_{fused} + X)$$

Stabilize the training process through Layer Normalization .

### 3.2.3 Adaptive Wavelet Attention Mechanism

The original TimesNet model lacks the ability to distinguish the importance of different features during the feature processing stage. When fusing features, all features are treated equally without considering the differences in their contributions to photovoltaic power prediction in different frequency and time domains. In some periods, high-frequency light intensity fluctuation features may be more important for prediction, while in other periods, low-frequency seasonal trend features may be more critical. The original model cannot adaptively adjust the focus on different features based on the real-time data situation. The introduction of the Adaptive Wavelet Attention mechanism addresses this deficiency. This mechanism can weight features separately in the frequency and time domains, automatically identify the features that are more important for photovoltaic power prediction, enhance their weights, and suppress the influence of irrelevant features. In this way, the model can focus more on key information, improving the accuracy and stability of the prediction. Its structure is shown in Figure 6.

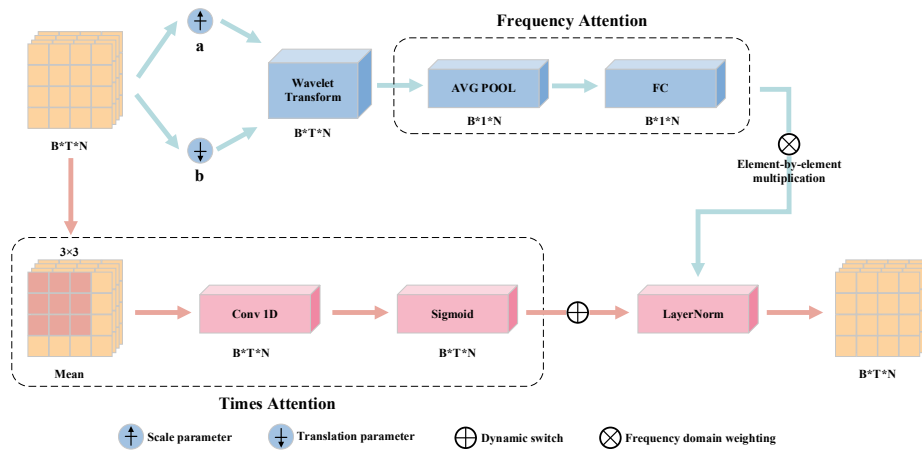


Figure 6: Structural diagram of adaptive wavelet attention mechanism

Adaptive wavelet transform:

$X \in \mathbb{R}^{B \times T \times N}$  adjusts the wavelet odd function through the dynamically generated scale  $a$  and translation  $b$ :

$$\begin{cases} a = \text{Sigmoid}(W_{scale} \cdot X + b_{scale}) \\ b = \text{Sigmoid}(W_{shift} \cdot X + b_{shift}) \end{cases}$$

Where  $W_{scale}$  and  $W_{shift}$  are learnable parameters.

The output of the wavelet transform:

$$W(a, b) = \sum_{t=1}^T X(t) \cdot \psi\left(\frac{t-b}{a}\right)$$

Frequency attention:

$$\begin{aligned} \alpha_{time} &= \text{Softmax}(W_{time} \cdot X) \\ Y_{time} &= \alpha_{time} \odot X \end{aligned}$$

Times attention:

$$Y_{out} = \text{LayerNorm}(Y_{freq} + Y_{time})$$

## 4 Experiments Results and Analysis

To evaluate the performance of the proposed Wavelet-TimesNet model, by integrating photovoltaic power generation and related meteorological data, and using multiple evaluation indicators, ablation experiments, multi-scale prediction experiments and multi-model comparison experiments were conducted to comparatively analyze the model's performance under different datasets.

### 4.1 Dataset analysis

To comprehensively verify the performance of the Wavelet-TimesNet model, this paper combines photovoltaic power generation-related data with other relevant meteorological parameters such as temperature, relative humidity, and wind speed. The datasets used in this study are from photovoltaic power stations in Xinjiang and another specific region in China. These two datasets have different sources and each presents unique climatic and power generation characteristics.

The dataset from Xinjiang was collected from multiple representative photovoltaic power stations in Xinjiang. Xinjiang is vast with diverse climate types, mainly characterized by a temperate continental climate, and some areas have an arid desert climate. It has abundant solar resources but large temperature differences between day and night. The meteorological conditions in different regions have significant impacts on photovoltaic power generation. These data were collected by the local energy monitoring department, and the original data were recorded in minutes. CSV format raw data were obtained from the relevant database and aggregated and integrated by hour using Python.

The dataset from a specific region in China was collected from a photovoltaic power station in that region. This area has a subtropical monsoon climate, with hot and rainy summers and mild and dry winters. The light conditions vary significantly across different seasons, and its meteorological environment is quite different from that of Xinjiang. This dataset was also collected by the professional monitoring agency in that

region, initially recorded in minutes, and then processed into hourly data by me using Python.

For the dataset from Xinjiang, it was divided into training, validation, and test sets in a 7:1:2 ratio. For the dataset from the specific region in China, it was also divided in a 7:1:2 ratio, with part of the data from 2018 to 2019 used as the training set, part of the data from 2019 for validation, and the remaining part for testing. This division method not only makes full use of the datasets but also helps to comprehensively evaluate the model's generalization ability under different climatic and power generation conditions.

## 4.2 Model Evaluation Criteria

When evaluating the predictive performance of a model, choosing the appropriate evaluation metrics is crucial. To comprehensively assess the model's performance, this paper employs multiple evaluation metrics. The following introduces five commonly used metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination ( $R^2$ ), Symmetric Mean Absolute Percentage Error (SMAPE), and Mean Squared Percentage Error (MSPE). These metrics are used to objectively measure the difference between the predicted values and the actual values of the model. The calculation results of these metrics represent the average value of all time steps within the prediction range. These metrics respectively measure the model's prediction accuracy, robustness, and computational efficiency from different perspectives.

**Mean Absolute Error (MAE):** Mean Absolute Error refers to the average of the absolute errors between the predicted values generated by the model and the actual values of the samples, that is, the average distance. It can intuitively reflect the average degree to which the predicted values deviate from the actual values. When calculating MAE, all errors are treated equally and are not affected by the direction ... of the error.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Where  $N$  is the total number of samples,  $\hat{y}_i$  is the  $i$ -th predicted value, and  $y_i$  is the  $i$ -th actual value.

The smaller the MAE value, the smaller the average deviation between the predicted values and the actual values, and the better the model's predictive performance.

**Mean Squared Error (MSE):** The mean squared error is the average of the squared differences between the predicted values and the actual values. During the calculation process, squaring the errors amplifies the influence of data points with larger errors on the overall result, making the model more sensitive to larger errors.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

The smaller the MSE value is, the smaller the sum of squared errors between the predicted values and the actual values of the model is, and the better the fitting effect of the model is. It is of great significance in measuring the overall error degree of the model.

**Coefficient of determination ( $R^2$ ):** An important indicator for measuring the goodness of fit of a model, it is used to evaluate the overall consistency between predicted values and actual values. It represents the proportion of the total variation in

the dependent variable that the model can explain, reflecting the model's ability to capture data fluctuations from the perspective of relative error.

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

The closer the  $R^2$  value is to 1, the stronger the model's explanatory power for the data, and the more variable trends it can capture; the smaller the value, the poorer the model's fitting effect, and it may miss key features or have significant deviations.

Symmetric Mean Absolute Percentage Error (SMAPE): Symmetric Mean Absolute Percentage Error is a relative indicator for measuring the accuracy of prediction, which overcomes the problem that the traditional Mean Absolute Percentage Error (MAPE) may have an infinitely large error when the actual value is close to 0. It calculates the average percentage of prediction error by standardizing the absolute value of the sum of the predicted value and the actual value.

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|} \times 100\%$$

The SMAPE value ranges from 0% to 200%. A value of 0% indicates a perfectly accurate prediction, and the closer the value is to 0%, the higher the prediction accuracy of the model. The larger the value, the greater the prediction error. This metric performs well in evaluating the accuracy of the relative changes between predicted and actual values and is often used in scenarios such as sales forecasting and time series forecasting.

Mean Squared Percentage Error (MSPE): MSPE is the average of the squared percentage errors between predicted and actual values, measuring the prediction performance of the model from the perspective of relative error. It takes into account the relative relationship between predicted and actual values and provides better comparability for data of different magnitudes.

$$MSPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2 \times 100\%$$

The smaller the MSPE value, the smaller the sum of the squared relative errors between the predicted values and the actual values of the model, indicating that the model has a stronger ability to fit and predict data. It is particularly suitable for comparing the prediction accuracy of different data scales or different variables.

Relative Error Reduction (RER): To quantify, in an intuitive way, the improvement of our method over a strong baseline, we report the RER metric. Let  $Err$  denote the error measure. Let  $b$  be the baseline model (fixed to Transformer throughout) and  $o$  other model.

$$RER = \frac{Err_b - Err_o}{Err_b} \times 100\%$$

RER is a dimensionless percentage that enables fair comparison across datasets and forecasting horizons: larger is better; 0% indicates parity with the baseline; positive values indicate an error reduction relative to the baseline, whereas negative values indicate inferior performance to the baseline.

### 4.3 Experimental results

To comprehensively evaluate the performance of the Wavelet-TimesNet model, ablation experiments, multi-scale prediction experiments, and comparative experiments with other advanced models were conducted. The experimental results were based on two datasets from Xinjiang and a certain region in China.

#### 4.3.1 Xinjiang Dataset

In Table 1, the ablation experiments on the Xinjiang dataset clearly demonstrate the core roles of each module in the Wavelet-TimesNet model. Compared with the base model, after adding the wavelet transform, the MAE decreased by 2.03% to 0.2361, the MSE reduced by 7.27% to 1.0241, and the  $R^2$  increased by 15.93% to 0.3632. This indicates that the wavelet transform, through multi-resolution decomposition, effectively captures the high-frequency fluctuations and low-frequency trends in photovoltaic data, enhancing the model's adaptability to complex climatic conditions. The large diurnal temperature differences and the intense fluctuations in light resources in Xinjiang make the separation of frequency-domain features crucial for prediction accuracy. The wavelet transform module directly optimizes the model's basic fitting ability by filtering out noise and strengthening effective signals. After introducing the multi-scale residual network on the basis of the wavelet transform, the MAE further decreased by 2.46% to 0.2344. Although the MSE slightly increased to 1.0415, the  $R^2$  remained at 0.3524, and the SMAPE dropped to 0.3897. This shows that the residual network, through the cross-layer connection mechanism, enhances the model's ability to integrate features at different time scales, especially suitable for the large-span meteorological variables in the Xinjiang dataset. The final complete model achieved the best performance with MAE = 0.2289 (a 5.02% reduction compared to the base model), MSE = 1.0183 (a 7.80% reduction),  $R^2$  = 0.3668 (a 17.08% increase), and continuous optimization of SMAPE and MSPE. This verifies the complementarity of the wavelet transform and the multi-scale residual network: the former decomposes the frequency components of the data, while the latter integrates multi-scale temporal dependencies. The combination of the two not only retains local details but also captures long-term trends, especially fitting the "high volatility and strong periodicity" power generation characteristics of the Xinjiang region. This indicates that the residual network, through the cross-layer connection mechanism, enhances the model's ability to integrate features at different time scales, and is particularly suitable for the meteorological variables with large spans in the Xinjiang dataset. The final complete model achieved the optimal performance of MAE = 0.2289 (a 5.02% reduction compared to the basic model), MSE = 1.0183 (a 7.80% decrease), and  $R^2$  = 0.3668 (a 17.08% increase), and both SMAPE and MSPE continued to improve. This validates the complementarity of wavelet transform and multi-scale residual network: the former decomposes the frequency components of the data, while the latter integrates multi-scale temporal dependencies. The combination of the two not only retains local details but also captures long-term trends, which is particularly suitable for the "high volatility and strong periodicity" power generation characteristics of the Xinjiang region.

Model	MAE	MSE	R2	SMAPE	MSPE
Basic model	0.2410	1.1044	0.3133	0.3924	1346.9354
+ Wavelet transform	0.2361	1.0241	0.3632	0.3901	1105.5598
+ MSResNet Block	0.2344	1.0415	0.3524	0.3897	956.0420
Complete model	0.2289	1.0183	0.3668	0.3877	865.8474

Table 1: Ablation experiment on the prediction length of 96 hours for Xinjiang Dataset

The results of multi-scale prediction in Table 2 show that Wavelet-TimesNet demonstrates superior robustness across different prediction lengths. The MAE of Wavelet-TimesNet is 4.97% lower than that of TimesNet, with a MSE of 1.0183 and an R<sup>2</sup> of 0.3668, all outperforming the comparison models. At this point, the data fluctuations have not yet fully deviated from the feature space of the training set, and the model's ability to capture high-frequency details becomes crucial. The sudden changes in midday sunlight in Xinjiang leading to fluctuations in peak power generation can be quickly responded to by the model through the precise decomposition of high-frequency components. When the prediction length is extended to 192 hours, the MAE of Wavelet-TimesNet increases by only 1.79%, while that of TimesNet increases by 8.63%, and the MSPE decreases by 36.42%. This indicates that the multi-scale residual network of the model effectively models the long-term trends of solar resources, while TimesNet, lacking a frequency decomposition mechanism, is more susceptible to noise interference when modeling long-term dependencies. Notably, in the 144-hour prediction, the MSE is slightly higher than that of TimesNet, but R<sup>2</sup> and SMAPE still have advantages. This might be due to the "sub-daily cycle" fluctuations in the data at this time scale, and the model's feature weight allocation needs further optimization, but overall, it still maintains a balanced performance in the indicators.

Model	Prediction Length	MAE	MSE	R2	SMAPE	MSPE
TimesNet	96	0.2410	1.1044	0.3133	0.3924	1346.9354
	144	0.2510	1.0685	0.3381	0.4153	982.3337
	192	0.2618	1.0778	0.3348	0.4326	848.4494
Wavelet-TimesNet	96	0.2289	1.0183	0.3668	0.3877	865.8474
	144	0.2558	1.0599	0.3435	0.4312	726.1580
	192	0.2337	1.0611	0.3451	0.3820	539.4543

Table 2: Comparison of Multi-scale Prediction of Xinjiang Dataset

In comparison with other advanced models (Table 3), Wavelet-TimesNet exhibits comprehensive superiority on the Xinjiang dataset. Compared to Transformer (MAE = 0.5444, MSE = 1.2707), Wavelet-TimesNet reduces MAE by 58.04%, MSE by 19.86%, and SMAPE by 63.68%. This can be attributed to the fact that the self-attention mechanism of Transformer incurs high computational costs in long sequences and is prone to the influence of noise. However, the wavelet transform reduces data complexity through prior frequency decomposition, allowing the model to focus on effective features. Although improved models such as Informer and Autoformer

introduce sparse attention, they still find it difficult to outperform the structured feature processing ability of Wavelet-TimesNet under the strong non-stationarity of Xinjiang data (such as the sudden drop in light caused by unexpected sandstorms). When compared with iTransformer, which also has a residual structure, Wavelet-TimesNet reduces MSPE by 58.34% (2077.7969  $\rightarrow$  865.8474), highlighting the optimization effect of wavelet decomposition on residual connections. Through frequency domain preprocessing, the residual network can more efficiently learn residual information at different scales, avoiding the parameter redundancy resulting from iTransformer's direct processing of raw data.

Model	MAE	MSE	R2	SMAPE	MSPE	RER
Transformer	0.5444	1.2707	0.2098	1.06767	1835.3578	0
Informer	0.4590	1.1795	0.2666	0.8295	2392.8996	7.1807
Autoformer	0.4516	1.3560	0.1568	0.7618	3127.3825	-6.7084
iTransformer	0.2318	1.0978	0.3174	0.3806	2077.7969	13.6128
TimesNet	0.2410	1.1044	0.3133	0.3924	1346.9354	13.0911
Wavelet-LSTM	0.3780	1.1170	0.3054	0.6956	1451.5597	12.1000
Wavelet-TimesNet	0.2289	1.0183	0.3668	0.3877	865.8474	19.8601

Table 3: Comparison of Multiple Models on Xinjiang Dataset

The prediction curve in Figure 7 shows that Wavelet-TimesNet has a significantly better fit at peak points (such as the midday power generation peak from sample 3000 to 4000), almost touching the actual values. The radar chart in Figure 8 visually presents its balanced advantages in core indicators such as MAE, MSE, and R<sup>2</sup>, with no obvious weaknesses, indicating that the model has achieved an effective balance in different error dimensions.

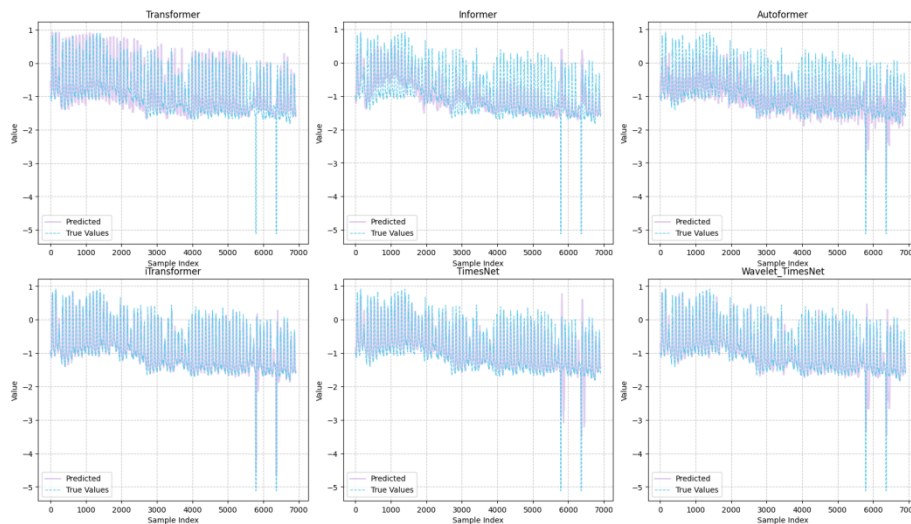


Figure 7: 96-hour prediction comparison on Xinjiang Dataset

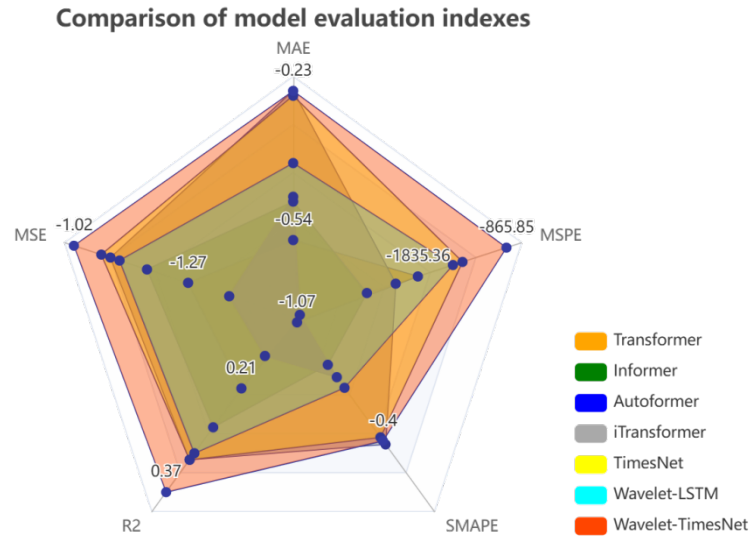


Figure 8: Radar Chart of Multi-Model Comparison for Xinjiang Dataset

#### 4.3.2 Chinese Dataset

In Table 4, the ablation experiments on the Chinese dataset present distinct features from those of the Xinjiang dataset. The base model has an MAE of 0.5551 and an MSE of 0.7393. After adding wavelet transformation, the MAE drops to 0.5475 (down 1.37%), the MSE drops to 0.7098 (down 4.00%), and the  $R^2$  increases to 0.6771. Unlike the high-frequency fluctuations in Xinjiang, the Chinese dataset is influenced by monsoons, showing obvious seasonal cycles. The main role of wavelet transformation here is to separate the seasonal trends from short-term fluctuations, enhancing the model's fitting ability for low-frequency cycles. Notably, the SMAPE slightly increases from 0.7279 to 0.7281, possibly due to the actual power generation being close to zero in some summer periods, causing fluctuations in the relative error calculation, but overall it remains stable. After introducing the residual network, the MAE further drops to 0.5426, but the MSE rises to 0.7149, and the  $R^2$  slightly drops to 0.6748. This reflects the "nonlinear complexity" of the Chinese dataset: under subtropical climate, light intensity is affected by multiple factors such as cloud thickness and precipitation probability. When the residual network is learning cross-scale features, local overfitting may occur due to data noise. However, SMAPE and MSPE continue to improve, indicating that residual connections still have a positive effect on controlling relative errors. The final model's MAE is 3.67% lower than the base model, the MSE is 6.10% lower, the  $R^2$  is 3.10% higher, and the MSPE is 6.98% lower. Although the improvement of individual modules is smaller than that of the Xinjiang dataset, the synergy effect is still significant, indicating that the combination of wavelet transformation and residual networks is universal for different climate data, especially

in handling "non-stationary sequences caused by multi-factor coupling", it can reduce information loss through frequency separation and multi-scale modeling.

Model	MAE	MSE	R2	SMAPE	MSPE
Basic model	0.5551	0.7393	0.6636	0.7279	18647.6289
+ Wavelet transform	0.5475	0.7098	0.6771	0.7281	19574.4941
+ MSResNet Block	0.5426	0.7149	0.6748	0.7316	22050.1621
Complete model	0.5347	0.6942	0.6842	0.7238	17345.6948

Table 4: Ablation experiments on Chinese Dataset

Table 5 shows that the performance of Wavelet-TimesNet varies from that of Xinjiang data under different prediction lengths. When the prediction length is 96 hours, the model's MAE is 0.5347, outperforming TimesNet, with an MSE of 0.6942 and an R<sup>2</sup> of 0.6842. This indicates that in the short-term fluctuations of the monsoon climate, wavelet transform can effectively decompose high-frequency precipitation signals and low-frequency light trends, improving prediction accuracy. For instance, during the sudden clearing periods in the plum rain season, the model captures the sudden changes in light through high-frequency components, avoiding the lagging prediction caused by the smoothing process in TimesNet. When the prediction length is extended to 192 hours, the MAE rises to 0.6042, slightly lower than TimesNet's 0.6053, but the MSPE is 23028.0078, lower than the latter's 24953.2109, indicating that it still has an advantage in long-term trend prediction. However, in the 144-hour prediction, the MSE is 0.7866, higher than TimesNet's 0.7338. The main reason is that this time scale corresponds to the "2-3 day weather system cycle", and when the model deals with such "mesoscale meteorological patterns", the scale division of wavelet decomposition may not match the actual data fluctuation period, leading to feature extraction bias.

Model	Prediction Length	MAE	MSE	R2	SMAPE	MSPE
TimesNet	96	0.5551	0.7393	0.6636	0.7279	18647.6289
	144	0.5579	0.7338	0.6665	0.7356	20247.9121
	192	0.6053	0.8094	0.6304	0.8003	24953.2109
Wavelet-TimesNet	96	0.5347	0.6942	0.6842	0.7238	17345.6948
	144	0.5978	0.7866	0.6425	0.7881	25790.6895
	192	0.6042	0.8071	0.6335	0.7972	23028.0078

Table 5: Comparison of Multi-scale Prediction of Chinese Dataset

Table 6 shows that compared with Transformer (MAE = 0.7998) and Informer (MAE = 0.8362), the model's MAE is reduced by 33.15% - 36.03%, and MSE is decreased by 42.39% - 47.74%, indicating that in nonlinear and multi-noise monsoon climate data, structured feature processing (wavelet frequency division) is more reliable than pure data-driven attention mechanisms. Compared with Autoformer, Wavelet-TimesNet has a lower MSE and a higher R<sup>2</sup>, suggesting that it is superior in balancing

"fitting accuracy" and "interpretability". The SMAPE is 0.7238, slightly higher than Autoformer's 0.7049, which is related to the sensitivity of relative error calculation caused by the low power generation scenario in summer in the Chinese dataset (actual values close to zero). However, the MSPE is 17345.6948, lower than iTransformer's 28531.7285, indicating that the model is better at controlling the mean square of relative errors and avoiding excessive influence from extreme values.

Model	MAE	MSE	R2	SMAPE	MSPE	RER
Transformer	0.7998	1.2055	0.4516	1.0591	19936.9648	0
Informer	0.8362	1.3286	0.3956	1.1050	18807.9023	10.2130
iTransformer	0.7467	1.0312	0.5309	0.9820	28531.7285	37.2707
Autoformer	0.5379	0.7243	0.6705	0.7049	20263.6758	14.4592
TimesNet	0.5517	0.7393	0.6636	0.7279	18647.6289	38.9334
Wavelet-LSTM	0.6951	1.0106	0.5403	0.8480	15880.4638	16.1700
Wavelet-TimesNet	0.5347	0.6942	0.6842	0.7238	17345.6948	42.4137

Table 6: Comparison of Multiple Models on Chinese Dataset

The prediction curve in Figure 9 shows that Wavelet-TimesNet fits more accurately in the recovery stage after continuous precipitation (such as samples 2000-3000), while Transformer-based models often exhibit delayed responses. The radar chart in Figure 10 indicates that the model has significant advantages in the three absolute error metrics of MAE, MSE, and  $R^2$ , and only slightly lags behind Autoformer in SMAPE, reflecting its characteristic of "absolute accuracy priority and controllable relative error", making it suitable for scenarios sensitive to the absolute value of power generation.

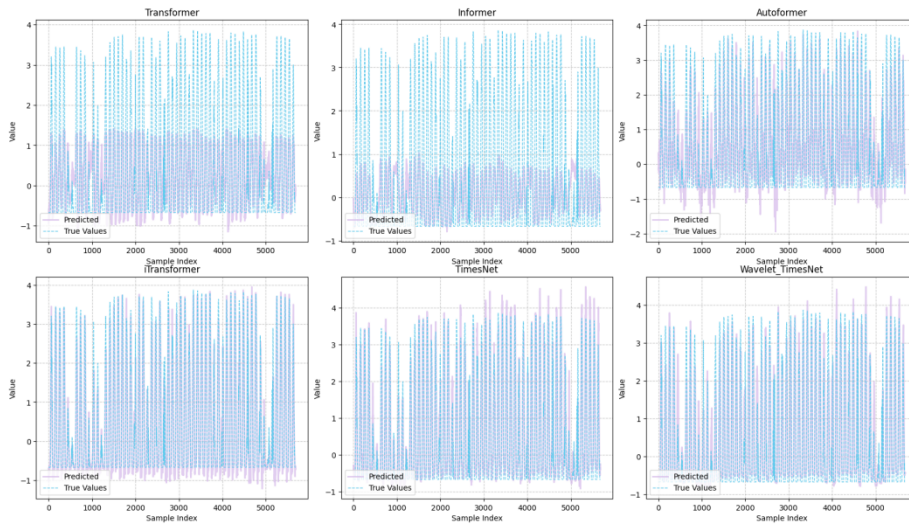


Figure 9: 96-hour prediction comparison on China Dataset

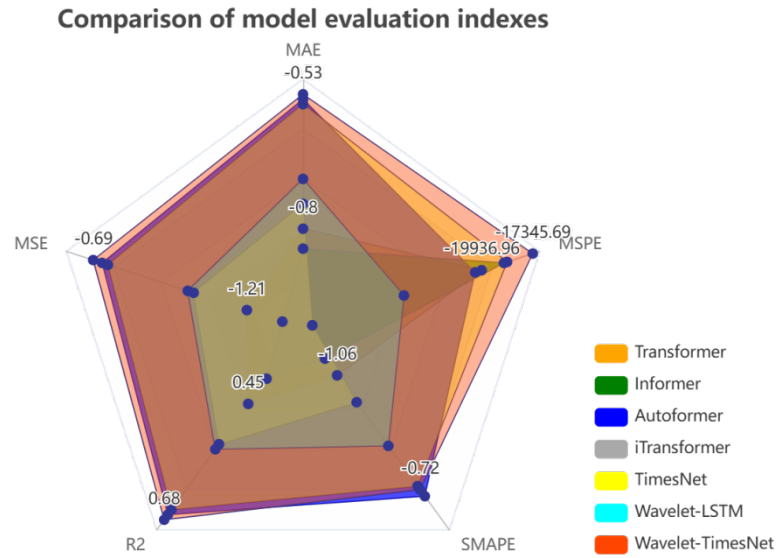


Figure 10: Radar Chart of Multi-Model Comparison on Chinese Dataset

## 5 Conclusion

This paper proposes a hybrid model, Wavelet-TimesNet, based on adaptive wavelet transform and multi-scale residual network, aiming to enhance the robustness of long-term photovoltaic power generation prediction. The model achieves multi-resolution analysis of local periodic features by dynamically adjusting the parameters of the wavelet basis function, thereby suppressing noise interference. It builds a multi-scale residual network to capture local details and global trends at different time granularities using parallel convolution kernels. An adaptive wavelet attention mechanism is introduced to dynamically fuse frequency-domain and time-domain features. Experiments were conducted on two datasets from Xinjiang and a certain region in China. The results show that Wavelet-TimesNet significantly outperforms advanced models such as Transformer and Informer in multiple metrics, especially in handling non-stationary signals and capturing long-term trends. This provides an efficient solution for precise dispatching of photovoltaic power stations and is of great significance for reducing peak shaving costs in the power grid and promoting the transformation of the energy structure.

Although Wavelet-TimesNet has exhibited strong performance in two climate regions, future work will focus on improving its generalizability. This includes evaluating the model on more diverse international datasets, testing across various PV system types and seasonal conditions, and exploring transfer learning for cross-regional adaptability. We also aim to enhance the model's computational efficiency to facilitate deployment in resource-constrained or edge-computing environments.

## References

- [Ahn, 21] Ahn, H. K., & Park, N. (2021). Deep RNN-based photovoltaic power short-term forecast using power IoT sensors. *Energies*, 14(2), 436.
- [Alay, 24] Alay F. D., İlhan N., Güllüoğlu M. T.: A Comparative Study of Data Mining Methods for Solar Radiation and Temperature Forecasting Models, *Journal of Universal Computer Science*, 2024, 30(6): 847.
- [AlKandari, 24] AlKandari, M., & Ahmad, I. (2024). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*, 20(3-4), 231-250.
- [Almaghrabi, 24] Almaghrabi, S., Rana, M., Hamilton, M., & Rahaman, M. S. (2024). Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks. *Information Fusion*, 104, 102180.
- [Branco, 22] Branco, N. W., Cavalca, M. S. M., Stefenon, S. F., & Leithardt, V. R. Q. (2022). Wavelet LSTM for fault forecasting in electrical power grids. *Sensors*, 22(21), 8323.
- [Cui, 24] Cui, S., Lyu, S., Ma, Y., & Wang, K. (2024). Improved informer PV power short-term prediction model based on weather typing and AHA-VMD-MPE. *Energy*, 307, 132766.
- [Ding, 24] Ding, S., Wang, Q., Guo, L., Li, X., Ding, L., & Wu, X. (2024). Wavelet and adaptive coordinate attention guided fine-grained residual network for image denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7), 6156-6166.
- [Doubleday, 20] Doubleday, K., Jascourt, S., Kleiber, W., & Hodge, B. M. (2020). Probabilistic solar power forecasting using bayesian model averaging. *IEEE Transactions on Sustainable Energy*, 12(1), 325-337.
- [Harrou, 20] Harrou, F., Kadri, F., & Sun, Y. (2020). Forecasting of Photovoltaic Solar Power Production Using LSTM. Approach in F. Harrou, & Y. Sun.(Eds.), *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems*, London, 3.
- [Kardakos, 13] Kardakos, E. G., Alexiadis, M. C., Vagropoulos, S. I., Simoglou, C. K., Biskas, P. N., & Bakirtzis, A. G. (2013, September). Application of time series and artificial neural network models in short-term forecasting of PV power generation. In *2013 48th International Universities' Power Engineering Conference (UPEC)* (pp. 1-6). IEEE.
- [Lim, 22] Lim, S. C., Huh, J. H., Hong, S. H., Park, C. Y., & Kim, J. C. (2022). Solar power forecasting using CNN-LSTM hybrid model. *Energies*, 15(21), 8233.
- [Liu, 23] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., & Long, M. (2024, May). itransformer: Inverted transformers are effective for time series forecasting. In *International conference on learning representations* (Vol. 2024, pp. 11116-11140).
- [Liu, 24] LIU, Z., GUO, J., LI, W., JIA, H., & CHEN, Z. (2024). Short-term prediction of concentrating solar power based on FCM-LSTM. *Chinese Journal of Engineering*, 46(1), 178-186.
- [Loganathan, 21] Loganathan, A. K., Stonier, A. A., & Maheswari, Y. U. (2021). Performance enhancement of a photovoltaic module using solar functional coatings. *Journal of Materials Science: Materials in Electronics*, 32(1), 1242-1257.
- [Park, 21] Park, M. K., Lee, J. M., Kang, W. H., Choi, J. M., & Lee, K. H. (2021). Predictive model for PV power generation using RNN (LSTM). *Journal of Mechanical Science and Technology*, 35(2), 795-803.

- [Rafati, 21] Rafati, A., Joorabian, M., Mashhour, E., & Shaker, H. R. (2021). High dimensional very short-term solar power forecasting based on a data-driven heuristic method. *Energy*, 219, 119647.
- [Tahir, 24] Tahir, M. F., Yousaf, M. Z., Tzes, A., El Moursi, M. S., & El-Fouly, T. H. (2024). Enhanced solar photovoltaic power prediction using diverse machine learning algorithms with hyperparameter optimization. *Renewable and Sustainable Energy Reviews*, 200, 114581.
- [Tao, 24] Tao, K., Zhao, J., Tao, Y., Qi, Q., & Tian, Y. (2024). Operational day-ahead photovoltaic power forecasting based on transformer variant. *Applied Energy*, 373, 123825.
- [Tolabi, 14] Tolabi, H. B., Ayob, S. B. M., Moradi, M. H., & Shakarmi, M. (2014). New technique for estimating the monthly average daily global solar radiation using bees algorithm and empirical equations. *Environmental Progress & Sustainable Energy*, 33(3), 1042-1050.
- [Vaswani, 17] Ashish, V., Noam, S., Niki, P., Jakob, U., Llion, J., Aidan, N. G., ... & Illia, P. (2017). Attention is all you need: Advances in neural information processing systems. In *NeurIPS Proceedings*.
- [Wu, 21] Wu, H., Xu, J., Wang, J., & Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Advances in neural information processing systems*, 34, 22419-22430.
- [Wu, 21] Wu, Shuo. (2021). Review of power prediction methods for photovoltaic power generation systems. *Journal of Engineering for Thermal Energy & Power*, 36(8).
- [Wu, 22] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2022). Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186*.
- [Xie, 22] Xie, Zhenxue, Lin, Fan, Wang, Ruogu, Zhang, Yao, Gao, Xin, & Wang, Jianxue. (2022). Ultra-short-term photovoltaic power prediction method based on time-series dynamic regression. *Smart Power*.
- [Zhang, 24] Zhang, M., Han, Y., Wang, C., Yang, P., Wang, C., & Zalhaf, A. S. (2024). Ultra-short-term photovoltaic power prediction based on similar day clustering and temporal convolutional network with bidirectional long short-term memory model: A case study using DKASC data. *Applied Energy*, 375, 124085.
- [Zhi, 23] Zhi Y., Sun T., Yang X.: A physical model with meteorological forecasting for hourly rooftop photovoltaic power prediction, *Journal of Building Engineering*, 2023, 75: 106997.
- [Zhou, 21] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 12, pp. 11106-11115).
- [Zhuang, 24] Zhuang, W., Li, Z., Wang, Y., Xi, Q., & Xia, M. (2024). GCN-Informer: A novel framework for mid-term photovoltaic power forecasting. *Applied Sciences*, 14(5), 2181.