EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Short-term photovoltaic power prediction based on dual decomposition with TCN-Informer-xLSTM

Guancheng Jin[1], He Jiang[1,*], Mofan Wei[1], Rui Guo[1]

[1] School of Renewable Energy, Shenyang Institute of Engineering, Shenyang 110136, China

Abstract

As renewable energy generation is increasingly integrated into power grids worldwide, the random nature of renewable energy output poses significant challenges to the stability of power systems. Therefore, it is essential to accurately predict the output of renewable energy sources. In this paper, a dual decomposition algorithm based on variational mode decomposition (VMD) and improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) is proposed to decompose the original photovoltaic power sequence and combine the entropy values of the subsequences to obtain the predicted sequences for the high frequency and low frequency components. Then, different prediction models are used for the high-frequency and low-frequency sequences to predict the photovoltaic outputs, where the Temporal Convolutional Networks (TCN)-Informer model is used for the high-frequency component and the xLSTM model is used for the low-frequency component, and finally, the RIME algorithm is applied to find the optimization of the hyperparameters. The results of simulation analysis show that the quadratic decomposition method proposed in this paper significantly improves the prediction accuracy of photovoltaic sequences and reduces the computational complexity.

Keywords: Photovoltaic Prediction, Dual Decomposition, Deep Learning.

Received on 18 March 2025, accepted on 20 July 2025, published on 29 September 2025

doi: 10.4108/ew.10415

1. Introduction

In order to promote the green and low-carbon energy transition and the construction of new energy power generation in various countries, new energy power generation, as a key link in this process, has important strategic significance for its future development and technological innovation. In the current stage of development, photovoltaic power generation has gradually become a key step on the road to green transformation in various countries[1-3]. Since the output power of photovoltaic power generation is affected by factors such as environmental and climatic conditions, which leads to strong randomness and volatility[4-5], it is therefore of great significance to accurately predict photovoltaic power generation.Accurate photovoltaic power forecasting can contribute to the grid connection of new energy sources, the consumption of new energy sources, the improvement of power market construction and power grid operation, and so on[6].

At present, photovoltaic prediction models are mainly divided into three types: physical, statistical methods, and combined prediction methods[7]. The main method of physical methods is to use numerical weather prediction (NWP)[8] to obtain meteorological characteristics, and at the same time establish a solar radiation model and a photovoltaic panel power generation conversion equation to make predictions. The advantages of physical methods are strong interpretability and no need for historical data. However, their limitations are that they are greatly affected by the accuracy of the collection equipment, have high requirements for the temporal and spatial alignment between each device, and have a lag in short-term predictions[9]. Statistical methods are prediction methods based on historical power generation data of photovoltaic power plants and mathematical models established through time series characteristics[10], such as the autoregressive integrated moving average model (ARIMA)[11] that analyzes the data autocorrelation and differentiation.

*Corresponding author. Email: jianghescholar@163.com

There are also analytical methods based on establishing a mapping relationship between output and characteristics, for example, when establishing a complex relationship between photovoltaic power and environmental factors, the kernel function is used to map nonlinear problems to high-dimensional space support vector machines (SVM)[12]. Statistical methods also include the establishment of probability intervals[13] to quantify the probability of uncertainty in photovoltaic power generation[14]. This method has a relatively simple model and is computationally efficient, making it suitable for online prediction. However, this method has poor prediction results in complex weather regions[15] and cannot handle complex nonlinear feature relationships.

There are many types of combined prediction methods. The idea of the combined approach is essentially to combine the strong and weak points of what each model has to offer and to improve the prediction accuracy through this complementary approach. For example, using the clustering algorithm K-means in machine learning, Markov chains, and ensemble learning XGboost, etc.[16-18]. In the literature[19], the authors use the iterative self-organizing data analysis algorithm (ISODATA) to classify historical data into similar days, and then perform variational mode decomposition to predict each component and super impose them to obtain a prediction structure. This method has a single modeling approach and does not consider the computational efficiency of the model. Reference[20] proposes a TCN-Informer prediction model based on the VMD decomposition algorithm, but the article does not consider the effectiveness of the single decomposition algorithm, the problem of modal overlap, and the model's adaptability to the decomposition sequence. Literature[21] improves the weather feature extraction ability by improving the encoder in the Transformer model to TCN network, which improves the prediction accuracy of the model for strong fluctuating weather, but does not take into account the expression ability of the model in the face of smoother weather conditions.

Based on the above analysis, this study innovatively constructs a combined prediction model (VMD-ICEEMDAN-TCN-Informer) based on the VMD-ICEEMDAN dual decomposition framework and the fusion of TCN-Informer and xLSTM. The model architecture not only improves the performance of data decomposition, but also considers the improvement of the computational efficiency and prediction performance of the network model during the stacking process. Continuity of the time series and generalizability of the model were considered during modeling. Finally, the weather category was not classified, which would have destroyed the time series characteristics. In order to deal with the strong volatility of the photovoltaic power signal, VMD was first considered to decompose it and obtain each mode. Then, the entropy calculation was performed simultaneously to extract the complex components in the signal and apply ICEEMDAN mode to them.This decomposition method not only improves the decomposition quality and reduce mode aliasing, and effectively suppress noise interference by introducing adaptive white noise. The high and low frequency signals are obtained by re-superimposing the decomposed signals, and the modeling of each signal is carried out separately using the high and low frequency signals, which effectively improves the computational efficiency of a single complex model. The high frequency signal prediction model is selected as the TCN-Informer model. Since the time span of the selected dataset in this study is long, the input sequence selected for prediction is 24 steps, and it is a single step prediction task. This model extracts the global spatial characteristics of complex signals while the introduction of Informer reduces the spatial complexity between features. The low-frequency component uses xLSTM to address the limitations of traditional LSTM in processing long sequences, and the hyperparameters are optimized using the frost-ice optimization algorithm. Finally, the results of the high and low frequency components are superimposed to obtain the final prediction result.

2. Dual decomposition algorithm

In this section, we first perform a correlation analysis on the various features that influence photovoltaic output to screen the features and reduce the dimension of the input data. We then input the processed data into the VMD model and perform entropy analysis on the output sub-sequences. Finally, we input the sequence with the highest entropy value into ICEEMDAN for secondary decomposition.

2.1 Correlation of characteristics

Since the output power of photovoltaic is affected by multiple external factors such as solar irradiance, weather temperature, air humidity, global radiation level, direct light and diffuse reflection, etc., taking all these characteristics into account will sharply increase the computational complexity of the model. Therefore, it is necessary to screen the factors affecting photovoltaic power one by one to improve the prediction accuracy of the model. In this paper, the Pearson correlation coefficient method, a statistical method, is used to analyze the linear correlation between the influencing factors of photovoltaic power. The formula for calculating the Pearson correlation coefficient is shown in (1):

$$r(x_i, y_i) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

where x is the observed value of the variable characteristic;y represents the observed value of the dependent variable characteristic; $\bar{x}$, $\bar{y}$ indicate the mean value of the variable characteristics, The Pearson correlation coefficient $r(x_i, y_i)$ ranges from -1 to 1 in its calculated results. Its actual physical meaning is that the larger the correlation coefficient, the stronger the correlation between the characteristic variables. Conversely, the smaller the correlation coefficient, the weaker the correlation between the characteristic variables. In this study, the characteristic data obtained by the collector are the seven characteristics of total irradiance level, direct irradiance level, global irradiance level, air temperature,

atmospheric environment, humidity, and power generation.

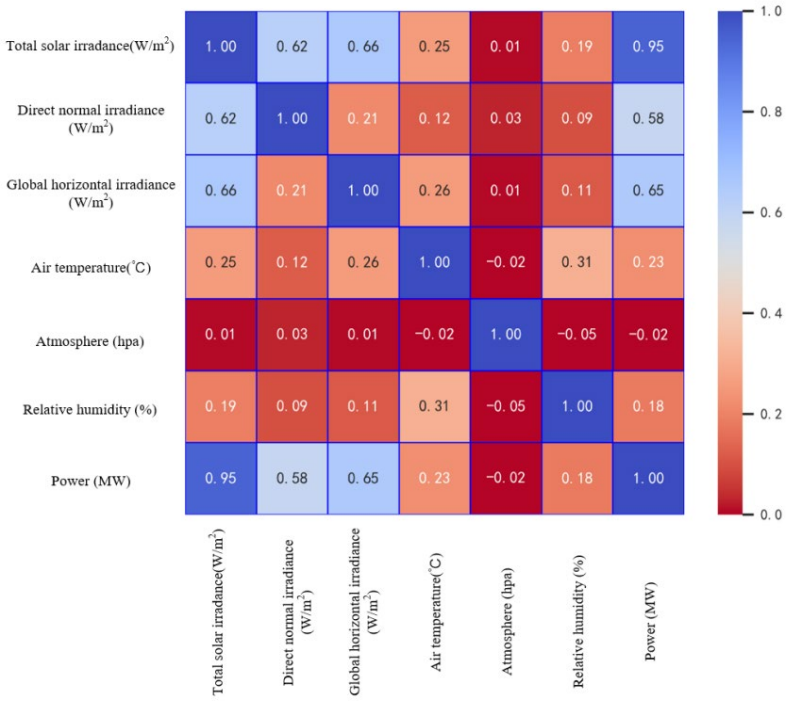

Figure. **1**. Pearson's correlation coefficient heat map

The correlation between each feature is shown in Fig 1. It can be seen that the correlation coefficients between the photovoltaic output power and the total radiation level, direct radiation level, and global radiation level are the largest, indicating a strong correlation. The correlation coefficients between the photovoltaic output power and the air temperature and relative humidity are smaller, indicating a weak correlation. Therefore, these features are selected as the input features of the model in this study. The smallest correlation coefficient with the atmospheric environment indicates that the correlation between the photovoltaic power and the atmospheric environment cannot be determined.

2.2 Variational Mode Decomposition

In photovoltaic forecasting tasks, photovoltaic power generation sequences exhibit extreme volatility due to factors such as weather conditions and geographical location. Based on this characteristic, VMD can decompose complex input signals into a set of IMFs, each with different frequency characteristics. VMD minimizes the bandwidth constraints of the mode solutions, thereby effectively extracting the intrinsic features of the input signal and improving the model's ability to represent photovoltaic sequences, ultimately enhancing prediction accuracy.

(1) First, the core of VMD is to decompose the raw photovoltaic output sequence $s(t)$ into several modal function $m(t)$ by establishing a constrained optimization problem. The objective function of the optimization problem is:

$$\min_{\{m_{k(t)}\},\{\omega_k(t)\}} \left\{ \sum_K \left\| \frac{\partial}{\partial t} \left[\left(\delta(t) + \frac{j}{\pi t} \right) m_k(t) \right] \exp\left(-j\omega_k\right) t \right\|_2^2 \right\} \quad (2)$$

where $s(t)$ is the original PV Power Sequence; ω_k is the center frequency of the kth modal function; m_k is the kth modal function; $\delta(t)$ is the Dirac function; K is the number of modes.

(2) Second, to solve the decomposition problem, VMD uses frequency modulation to represent each modes.

$$m_k(t) = a_k(t) e^{j\omega_k t} \quad (3)$$

where $a_k(t)$ is the complex envelope function to represent the spectrum of the localized signal.

(3) Third, by means of Lagrange multipliers λ and quadratic penalty terms β and the constraint $\sum_K m_k(t) = s(t)$, VMD transforms the constrained variational problem into an unconstrained variational problem. By minimizing the bandwidth constraint and the regularization term, the optimization problem becomes:

$$L(\{m_k\},\{\omega_k\},\lambda) = \sum_k \left\| \frac{\partial}{\partial t}\left[\left(\delta(t)+\frac{j}{\pi t}\right) m_k(t)\right] \exp(-j\omega_k t) \right\|_2^2 + \beta \left\| s(t) - \sum_{k=1}^{K} m_k(t) \right\|_2^2 + \left\langle \lambda(t), s(t) - \sum_{k=1}^{K} m_k(t) \right\rangle \tag{4}$$

(4) Fourth，The process variable is mainly used to solve this optimization problem with constraints, and to reduce the complexity of the computation, the time-domain equational constraints are transformed into frequency-domain equational constraints $\sum_{k=1}^{K} \hat{m}_k(t) = \hat{s}(t)$ and the alternating direction multiplier method is used to update these process variables for , and , respectively, with the following update equations:

$$\hat{m}_k^{n+1}(\omega) = \frac{2\left(\hat{m}(\omega) - \sum_{i \neq k} \hat{m}_i(\omega)\right) + \hat{\lambda}(\omega)}{2 + 4\beta(\omega - \omega_k)^2} \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \eta\left(\hat{m}(\omega) - \sum_{k=1}^{K} \hat{m}_k^{n+1}(\omega)\right) \tag{6}$$

$$\omega_k^{n+1} = \frac{\int_{-\infty}^{\infty} \omega \left|\hat{m}_k(\omega)\right|^2 d\omega}{\int_{-\infty}^{\infty} \left|\hat{m}_k(\omega)\right|^2 d\omega} \tag{7}$$

where $\hat{m}(\cdot)$ and $\hat{\lambda}(\cdot)$ denote the Fourier transform; η is the learning rate or step size to control the update rate of the Lagrange multipliers

2.3 Entropy

The entropy value of each modal row obtained after VMD decomposition is used to evaluate the complexity of each modal , and by calculating the entropy value, we can identify which modal rows contain more ordered information or noise. Modes with low complexity may be useful signals, while modes with high complexity may be noise. In this study, the PV power signal is finally decomposed into four modes by VMD, $s(t) = \sum_{k=1}^{4} m_k(t)$ and the entropy value is calculated for each mode by the formula:

$$H(m_k) = -\sum_{i=1}^{N} p_i \log(p_i) \tag{8}$$

where p_i is the probability of each discrete event in the probability distribution of each mode $m_k(t)$; N is the number of discretization intervals for the signal values, which is taken to be 100 in this study.

2.4 Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Different from the traditional EEMD method, ICEEMDAN adopts an adaptive strategy in the process of noise generation and addition, adjusting the noise size according to the feedback results of each iteration in order to better maintain the original characteristics of the signal. For modes with high entropy values, ICEEMDAN's adaptive noise mechanism and multiple iterations decomposition can effectively reduce the overlapping phenomenon between modes, further optimize the frequency components of each mode and improve the accuracy of subsequent modeling data. The specific steps are as follows:

(1) The new signal $m_{new}(t)$ is obtained by adding Gaussian white noise conforming to normal distribution to the VMD high entropy value modal sequence as shown in equation (9):

$$m_{new}(t) = m_{high}(t) + \varepsilon_0 \xi_j(t), \quad j = 1,2,3,\cdots,\mathbf{N} \tag{9}$$

where $m_{high}(t)$ is the sequence of VMD high entropy values; ε_0 is the standard deviation of Gaussian white noise; $\xi_j(t)$ is Gaussian white noise; $\mathbf{N}$ is the number of experiments.

(2) Perform empirical modal decomposition (EMD) on the updated signal $m_{new}(t)$. EMD decomposes the signal into a number of IMFs and residual signals, and the update formula is:

$$m_{new}(t) = \sum_{k=1}^{K} IMF_k(t) + r(t) \tag{10}$$

where K is the total number of eigenmode functions decomposed; $r(t)$ denotes the residual signal.

(3) In order to enhance the stability and accuracy of the decomposition, step 1 to step 2 are repeated several times, each time adding a different Gaussian white noise $\varepsilon_j \xi_j(t)$, and EMD decomposition is performed, and a set of different intrinsic modal functions are obtained through several noise injections and decompositions:

$$\left\{IMF_k^{(1)}(t), IMF_k^{(2)}(t), ..., IMF_k^{(N)}(t)\right\} \tag{11}$$

(4) For each mode $IMF_k(t)$, the results of multiple decompositions are averaged as in the following equation to obtain the final eigenmodes:

$$IMF_k(t) = \frac{1}{M} \sum_{i=1}^{M} IMF_k^{(i)}(t) \tag{12}$$

3. Combined model prediction method based on RIME-TCN-Informer-xLSTM model

Under the combined influence of other factors such as temperature, cloud activity, weather type, etc, the real-time power generation of PV power plants often reflects extreme volatility and randomness, while the data of PV power generation itself is strongly time-dependent, and the traditional time-series analysis methods may have large prediction errors when facing complex data, so this section propose a combination prediction method based on TCN-Informer-xLSTM combined prediction method. The prediction flowchart is shown in Fig. 2.

After the above data processing, the subsequence is rearranged to obtain the high and low frequency components separately, where the high frequency component is input to the TCN-informer model and the low frequency component is input to the xLSTM model. For the high-frequency components, the sequences show strong volatility, and the TCN-Informer model not only combines the time-dependent ability of TCN in capturing long sequences, but also improves the prediction accuracy of the model in dealing with complex sequences and makes the model computation more efficient according to the characteristics of TCN that is easy to parallelize the computation in the structure. Compared with the traditional Transformer model, the Informer model is computationally efficient due to the structure of sparse attention mechanism, structurally improves the generalization of the model when dealing with long sequences due to the structure of attention distillation mechanism, and speeds up the response speed of the model inference due to the generative coding structure in the decoder.

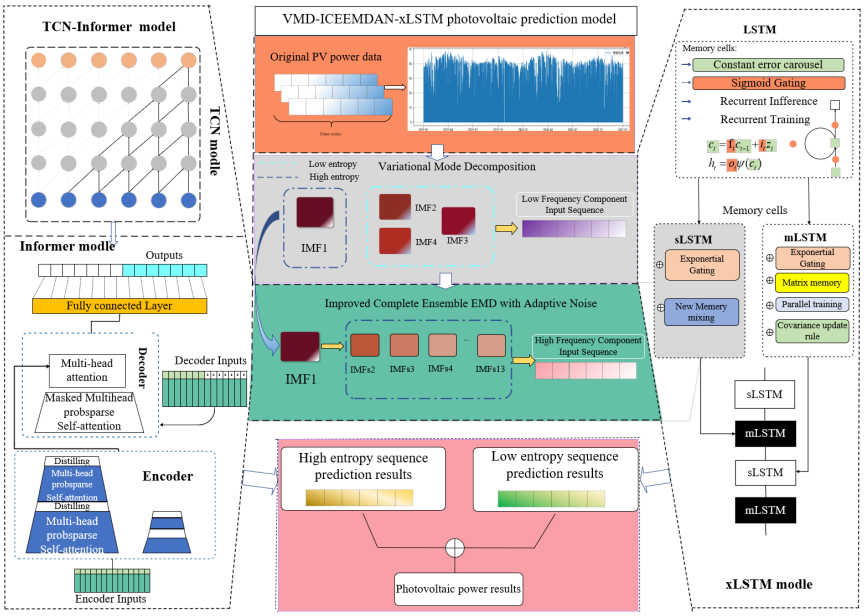

Figure. 2. VMD-ICEEMDAN-TCN-Informer-xLSTM frame

Traditional LSTM models, although capable of handling long-term dependencies to some extent, may suffer from vanishing or exploding gradients when confronted with longer time series, resulting in a degradation of the model's performance. In this case, xLSTM can more precisely identify which information is critical for long-term trend prediction through the improved forgetting gate and input gate mechanisms to better fit the trend changes in the low-frequency components. xLSTM, due to its extended structure using mLSTM and sLSTM, makes it easier for the model to maintain the ability of long-term memory when dealing with sequences with a large time span, and is able to better capture the long-term dependencies in the low-frequency components, and thus effectively learn the trend and periodic features in the sequences.

3.1 Temporal Convolutional Networks

TCN is a variant model based on traditional convolutional frameworks such as CNN, RNN, specialized in processing time-series data, TCN is more than traditional convolutional networks which capture the temporal dependencies of long sequences through causal convolution, inflationary convolution, and residual connections rather than cyclic connections. First of all causal convolution belongs to a special form of one-dimensional convolution, its main feature is that the convolution kernel will only utilize the current and previous time step data when sliding, while the right part of the convolution kernel will be filled with 0, which ensures the before-and-after logic on the time sequence, located in the t time step, the

index number is i, and the size of the convolution kernel is n, then the output value is:

$$o(t) = \sum_{i=0}^{n-1} w(i) \cdot s(t-i) \tag{13}$$

where $w(i)$ is the weight value of the convolution kernel at the ith index position, s denotes the sequence of high-frequency components, and $s(t-i)$ is the data point at the first i time step in the high-frequency component, t is the time step, and n is the size of the convolution kernel. In order to improve the modeling ability of the model when dealing with long time series tasks, the sensory field is increased in causal convolution by introducing an expansion factor d to make the network able to cover longer time span data, where the relation is given by $d = 2^m$.

The overall structure of TCN is shown in Fig. 3, It can be seen that if the number of convolutional layers is set too much, it will lead to the gradient disappearance or gradient explosion when the model is being trained, while the introduction of residual connection can skip the input signal directly to the convolutional layer part and establish a direct connection with the output signal, which not only ensures that the training gradient can be effectively transferred to the shallow layer of the model when the model is being back-propagated, but also ensures that the depth of the network in the model is effectively guaranteed.

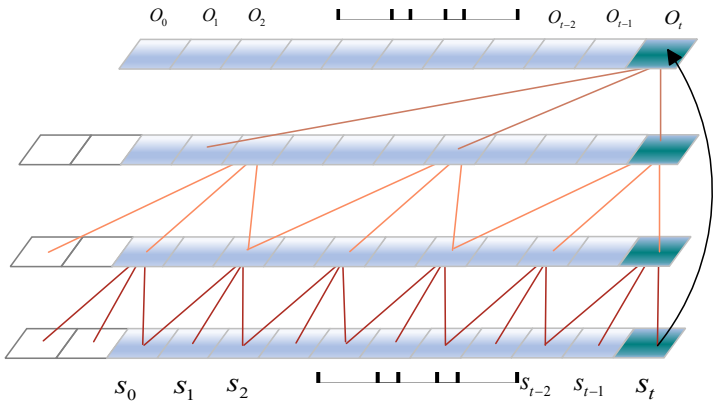

Figure.3. TCN Convolutional Structure

3.2 Informer model

Informer model as a variant model of Transformer, its most important feature is that the traditional attention mechanism in the attention Q、K、V in the calculation meets a longer input sequence, the complexity of the weight matrix with the increase in the length of the sequence N need to calculate the current time step and the other time step of the attention matrix (N × N), and the attention scoring mechanism there are a part of the The query vector q and key vector k show weak correlation, based on the above problems informer model reduces the computational complexity by sparse self-attention mechanism to filter out the key time steps, which reduces the computational complexity from $O(N^2)$ to $O(N \log N)$,which significantly improves the computational efficiency compared to the traditional transformer, and has a better suitability for long sequences of timings.

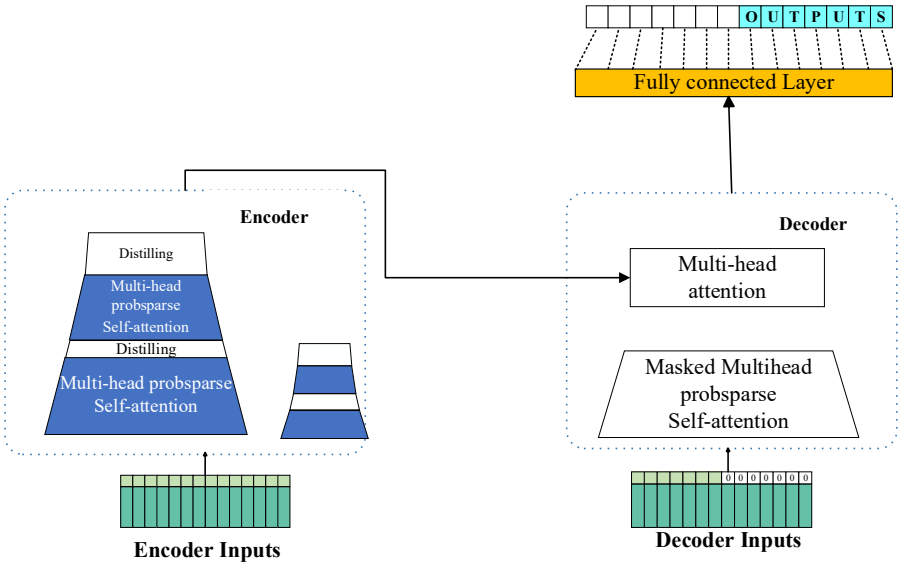

Figure. 4 Informer encoder-decoder structure

Query Sparsity Measure: In order to examine the traditional attention mechanism in the calculation process which Q, K have strong correlation between, through the introduction of probabilistic form of kernel smoother on the original base formula to define the attention of the ith query as shown in equation (14):

$$A(q_i,K,V)=\sum_J k(q_i,k_j)\frac{v_j}{\sum_l k(q_i,k_l)}=E_{p(k_j|q_i)}\left[v_j\right] \quad (14)$$

where $k(q_i,k_j)$ is the similarity function between the query point q_i and the k_j of the remaining points. The value of the output is the product of the expectation $E_{p(k_j|q_i)}$ and the probability weighted value v_j of the query point q_i for all k_j. $p(k_j|q_i)$ is the attentional probability of q_i over k_j .If the attention probability obeys a uniform distribution $q(k_j|q_i)=\frac{1}{L_K}$, the self-attention computation will be significantly reduced. To measure the degree of dispersion of the distribution p and the probability q is measured by introducing the Kullback-Leibler scattering, neglecting the constant term then the sparsity measure equation (15) is:

$$M(q_i,K)=\ln(\sum_{j=1}^{L_K}e^{\frac{q_ik_j^T}{\sqrt{d}}})-\frac{1}{L_K}\sum_{j=1}^{L_K}\frac{q_ik_j^T}{\sqrt{d}} \quad (15)$$

Probsparse self-attention mechanism: Based on the above metric Informer attention mechanism is calculated as in equation (16):

$$A(Q,K,V)=\text{Softmax}(\frac{QK^T}{\sqrt{d}})\text{V} \quad (16)$$

Each K in the equation is associated only with Q that accounts for strong correlations, and the Q matrix is a mapping of the first h strong correlations q after the sparse metric, so the size of the sparse matrix is equal to the size of q. The sparsity h is controlled by a constant sampling factor c according to the number of queries as L_Q . The control formula is $u=C\ln L_Q$.

In order to further improve the computational efficiency, it is necessary to set the sparsity limit value in it, which finally results in the sparsity formula using the empirical formula as:

$$M(q_i,K)=\max_j\left\{\frac{q_ik_j^T}{\sqrt{d}}\right\}-\frac{1}{L_K}\sum_{j=1}^{L_K}\frac{q_ik_j^T}{\sqrt{d}} \quad (17)$$

In practice, the input lengths of the query and key are usually equal in the self-attention computation, as $L_Q=L_K=L$ .For the remaining query vectors after filtering out the strongly correlated q in the above process, the average input value is used for all the remaining query vectors.

Encoder structure: The encoder is shown in the left part of Fig.4, the encoder accepts a large number of long-time sequences at the same time after the above Probsparse attention

mechanism, in which the dark green part of the original input vector, the light green part of the Local timestamp and global timestamp by the composition.

In order to further reduce the amount of computation to increase the computational efficiency of Informer through the introduction of distillation mechanism to further filter the important information, distillation layer contains a one-dimensional convolutional layer (Conv1D) and the maximum pooling layer , where the activation function for the ELU, the role of maximum pooling layer for the downsampling operation, so the length of the sequence will be reduced by half after each distillation layer, distillation The operation process is given in the following equation:

$$X_{k+1}^{t} = \text{MaxPool}(\text{ELU}(\text{Conv1D}(\left[X_{k}^{t}\right]_{AB}))) \tag{18}$$

where X_k denotes the input of the kth layer network, $[\square]_{AB}$ denotes the Probsparse attention module, and in order to keep the output of each encoder stack of the same dimension, the encoder stack method is used for cascade.

Decoder structure: The decoder is shown by the right half of Fig. 4 . The input to the decoder can be represented as $X_{de} = \{X_{token}, X_0\}$, where the two parts are the data from the previous time step of the prediction sequence $X_{\text{token}}^{\text{t}} \in \square^{L_{token} \times d_{\text{model}}}$ and the prediction sequence placeholder $X_0 \in \square^{L_y \times d_{\text{model}}}$, thus the overall dimension of the input to the decoder is $X_{de} \in \square^{(L_{token}+L_y) \times d_{\text{model}}}$, which provides the basis for the subsequent multi-head attention computation. Ultimately, the model performs a dimensional transformation of the output through a fully connected layer to generate the final prediction.

3.3 xLSTM model

The xLSTM is an improved model based on the traditional Long Short-Term Network (LSTM). xLstm still has difficulty in capturing dependencies over long time spans when dealing with long time sequences, and therefore the LSTM model is unable to capture long term trends among sequences for low-frequency signals. Therefore, xLstm has been developed to improve the accuracy of the model by introducing a gating mechanism and a memory structure. xLSTM is an extension of two variants of LSTM (mLSTM and sLSTM).

The computation and structure of sLSTM and mLSTM are shown in Fig. 5. sLSTM introduces exponential gating and normalized states n_t in the data transfer between the input gate i_t and the forgetting gate f_t , and stabilizes the structure so as to help the LSTM to learn more complex time-dependent and nonlinear patterns, and to reduce data fluctuations during training to accelerate convergence speed.

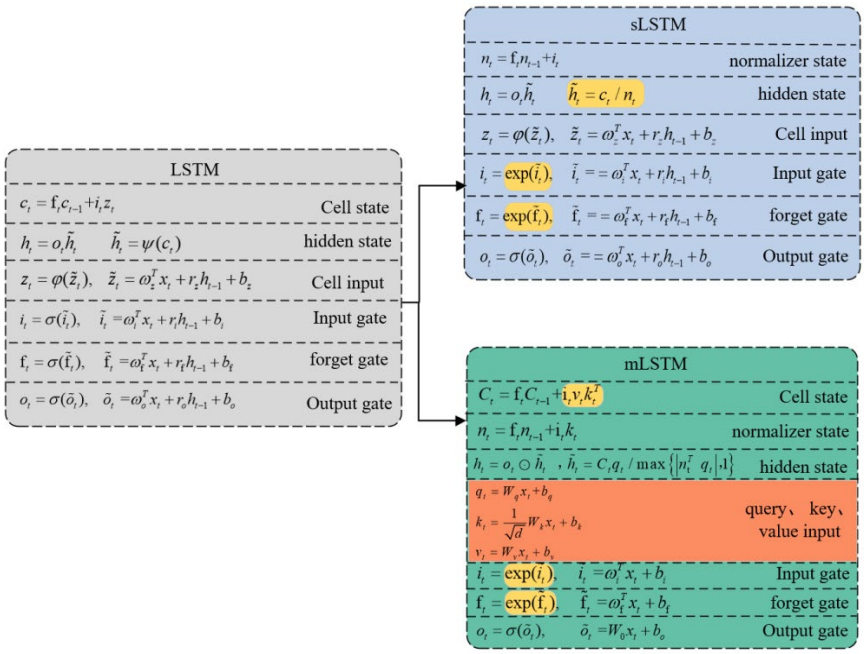

Figure. 5 Diagram of xLSTM memory cell and gating structure

The design of mLSTM is to add an attention mechanism similar to the Transformer model to the traditional LSTM, by introducing *k*, *q*, and *v* computations to update the memory cell as a way to capture the complex relationship between features, and the model structure is similar to the sLSTM network described above, but the amount of state of each cell is changed from a vector to a matrix, which makes it greatly enhance the memory cell's storage capacity.

3.4 RIME optimization algorithm

RIME optimization algorithm is to simulate the formation process of frost and ice in nature to introduce the concept of soft frost and hard frost, and the introduction of attachment coefficients to simulate the probability of freezing of atmospheric particles and the establishment of the search space and the threshold value of the proposed optimization mechanism, the core principle is divided into three steps: the soft frost search strategy, the hard frost perforation mechanism, and the positive greedy mechanism.

The soft frosting search strategy:

$$R_{ij}^{new} = R_{best,j} + \alpha \cos\theta r_1 \left[h(b_{\max_{ij}} - b_{\min_{ij}}) \right], r_2 < E \tag{19}$$

where R_{ij}^{new} is the updated particle position; *i* and *j* denote the *j*th particle in the *i*th particle population, then $R_{best,j}$ denotes the *i*th particle in the optimal population; The parameter r_1 is a (-1~1) random number to indicate the update direction of the particle, $\cos\theta$ is updated as the number of iterations increases, and the update equation is shown in equation (20), α is the environmental factor that simulates the external fluctuation situation during each iteration, and the iterative update equation is shown in equation (21),.*h* takes values in the range (0,1) representing the distance between the centers of the two particles, $b_{\max_{ij}}$ and $b_{\min_{ij}}$ are the upper and lower thresholds of the escape space used to limit the particle update domains. *E* is the attachment coefficient, representing the update probability of the particle with respect to the number of iterations; as shown in equation (22), r_2 is a (0,1) random number that controls the position update of the particle along with E.

$$\theta = \pi \frac{t}{10 \cdot T} \tag{20}$$

where *t* is the number of immediate iterations of the algorithm; *T* is the maximum number of iterations of the algorithm.

$$\alpha = 1 - \left[\frac{w \cdot t}{T} \right] / w \tag{21}$$

where α is the step function; $[\cdot]$ represents the rounding function; *w* is the number of step function segments.

$$E = \sqrt{(t/T)} \tag{22}$$

Hard frost perforation mechanism: The purpose of this mechanism is to avoid the algorithm from falling into local optimal solution by using the replacement of particles to improve the convergence of the algorithm, the replacement formul a is shown below:

$$R_{ij}^{new} = R_{best,j}, r_3 < F^{normal}(S_i) \tag{23}$$

where S_i is the selection probability of the particle cluster and F^{normal} denotes the normalization of it; r_3 is a (-1,1) random number determined in concert with F^{normal} to determine the particle update probability.

Positive greed mechanism: The fitness values of individuals before and after iteration are firstly compared and analyzed, based on which the decision of individual retention or replacement is dynamically adjusted and the solution vectors of related individuals are synchronously updated. This mechanism effectively selects the retention of representative particles to promote the optimization performance of the overall solution set, and at the same time, it also guides the population to shift to a more optimal solution space during each iteration through the dual updating mechanism, thus realizing the double improvement of the convergence efficiency and the solution set.

4. Example analysis

4.1 Introduction to the experimental dataset and assessment metrics

The dataset selected for this experiment adopts the 2019-2020 power generation dataset of a 40MW PV power plant in the north of China, the time scale of the dataset is 15 min, and a total of 70,096 time points of data are recorded, which is a real and reliable source, with a long time span, and provides high-frequency and fine-grained time-series data covering diurnal and seasonal cycle variations, which reflects the PV power generation's long-term trends and fluctuation characteristics. Second, the diversified weather conditions (e.g., sunny, cloudy, rainy, snowy, etc.) in the northern region make the data rich in environmental characteristics, which provides a data basis for studying the laws of PV power generation affected by the environment. Once again, the long time series data support the development of short-term and long-term power prediction models, and at the same time help to improve the robustness and generalization ability of the models. The prediction model proposed in this experiment is written

in PyCharm integrated development environment using Python programming language and implemented based on PyTorch deep learning framework.

Four evaluation metrics were used in this experiment to assess the predictive effectiveness of the model. The evaluation formula is as follows:

$$E_{\mathrm{MAE}} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{24}$$

$$E_{\mathrm{MAPE}} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{25}$$

$$E_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{26}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - \hat{y}_i\right)^2} \tag{27}$$

where MAE is the mean absolute error; MAPE is the mean absolute percentage error; RMSE is the root mean square error.

4.2 Pre-processing of data

Due to the large time span of the dataset used in this experiment, as shown in Fig. 6, the fluctuation of PV output in the summer of each year may be larger compared to other seasons due to the effect of cloud cover, and the PV output in the winter is also strongly fluctuating due to the effect of snow and ice cover and presents a large amount of noise in the dataset.6 The dataset contains a large amount of noise as shown in Fig. 6, and it is necessary to carry out the operation of separating the data and reducing the noise. Therefore, it is necessary to separate the data and reduce the noise.

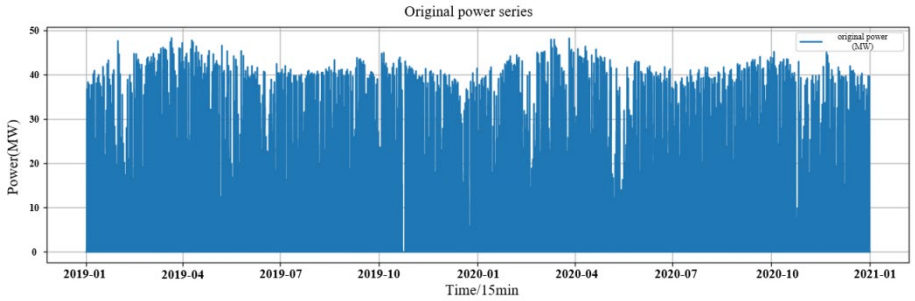

Figure. 6. Original Input Signal

In order to evaluate the complexity of each IMF after VMD decomposition of the original signal, the method of sequence entropy calculation is introduced for this analysis. In practice, the original signal is decomposed into four modes by VMD and the entropy value of each mode is calculated as shown in Fig. 7.

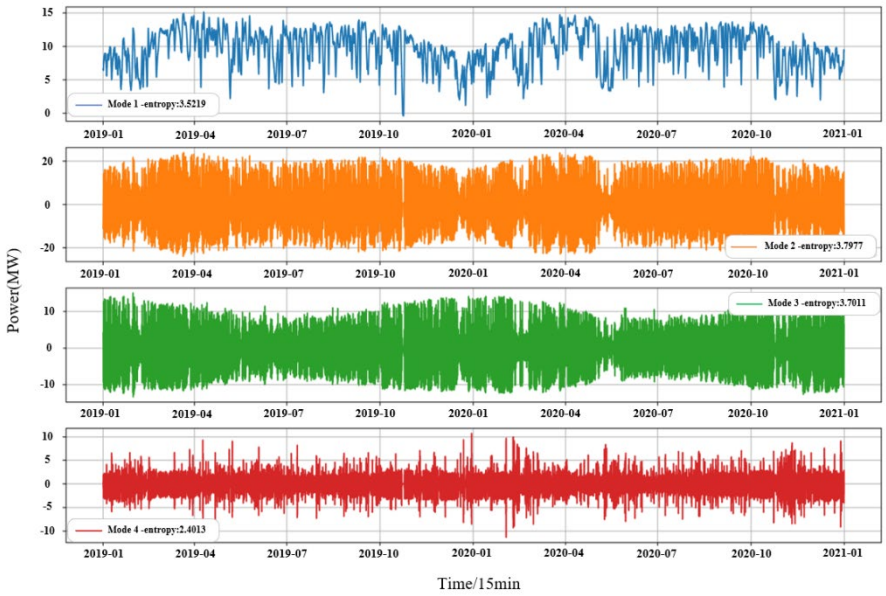

Figure. 7 VMD decomposition sequence and entropy value

The entropy value of mode 2 is the highest and therefore the signal contains more noise, in order to improve the prediction effect of the model, it is finally chosen to carry out the decomposition of the sequence once again.

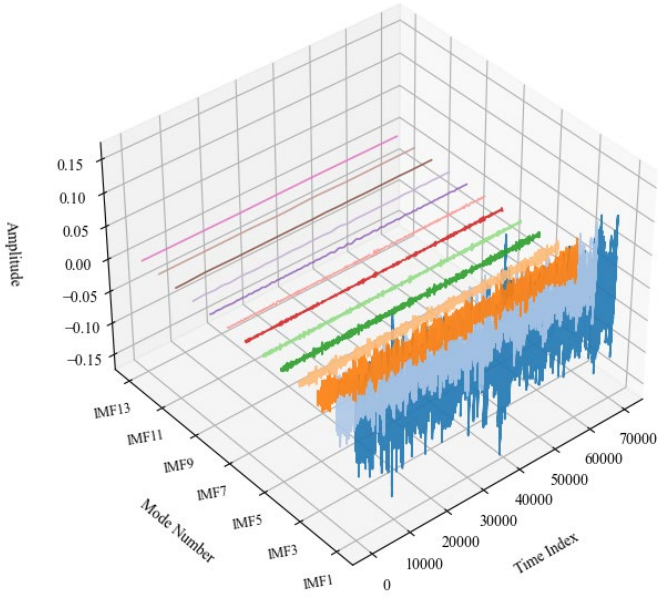

Figure. 8. Sequence diagram of ICEEMDAN decomposition

The high-frequency component is decomposed by ICEEMDAN to obtain 13 intrinsic mode functions, and the decomposition results are shown in Fig. 8. Due to the separation of more IMFs, in order to reduce the amount of input signal computation of the model, the above components and the original dataset of weather features are superimposed and reconstructed to form a new low-frequency and high-frequency components, and the two components are modeled

separately. The superimposed reconstructed high and low-frequency signals are shown in Fig. 9, and it is clearly found that the high-frequency signal fluctuation amplitude is larger, the noise has a strong randomness, and the prediction complexity is higher; and the low-frequency component is a smoother curve, which can more realistically reflect the long-term trend of the photovoltaic power change with the season and time.

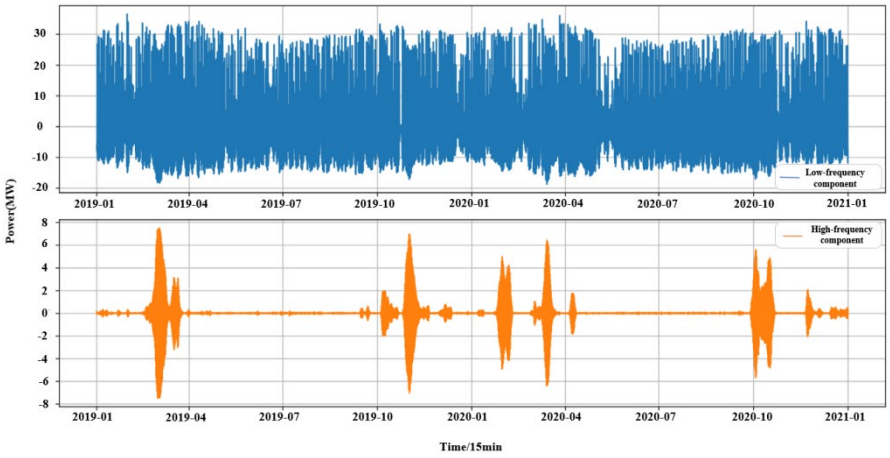

Figure. 9 Reconstruction of high and low frequency signals

4.3 Experimental model parameter settings

In this experiment, the RIME algorithm is used to optimize some parameters of the model, the algorithm enhances the search performance by simulating the frost and ice growth mechanism in nature, using the soft frost search mechanism and the puncture mechanism of the hard frost, the parameters are set as shown in Table 1.

Table 1. Parameters setting

TCN-Informer		xLSTM	
Kernel size	2	Number of head attention	8
No blocks	2	Hidden size	64
Dilations	[1,2]	layers	['s','m']
Number of head attention	8	Project factor slstm	4/3
Probsparse attention	5	Project factor mlstm	2
Encoder layers	2	Out-size	(512,32)
Decoder layers	1	Learning rate	0.05

4.4 Simulation results and analysis

In order to verify the superiority of the combined model PV short-term power prediction of TCN-Informer and xLSTM under the dual decomposition of VMD and ICEEMDAN, the comparative models proposed in this section are the TCN-Informer model with and without dual decomposition and the traditional Transformer and its improved model. Under the principle of ensuring control variables, the number of multi-attention heads in Transformer is 4, and the length of the input sequence is 24, and the number of hidden layer layers in the TCN model is 4.

The simulation results are shown in Figs. 10 and 11, which show that the VMD-ICEEMDAN data decomposition method provides smooth and noiseless input data, while the The short-term feature extraction ability of TCN makes up for the shortcomings of the traditional Transformer and Informer models when dealing with fluctuating signals with large amplitude, the TCN network is responsible for extracting the short-term signals, and the Informer and xLSTM models have a better ability to extract the long-term dependence when dealing with longer sequences than the traditional Transformer and Lstm models, and the combination of these models makes the VMD-ICEEMDAN data decomposition method provide smooth and noisy input data. The combination of these models makes the VMD-ICEEMDAN-TCN-Informer after the optimization of hyperparameters by the RIME optimization algorithm the best performing model in the PV prediction experiments.

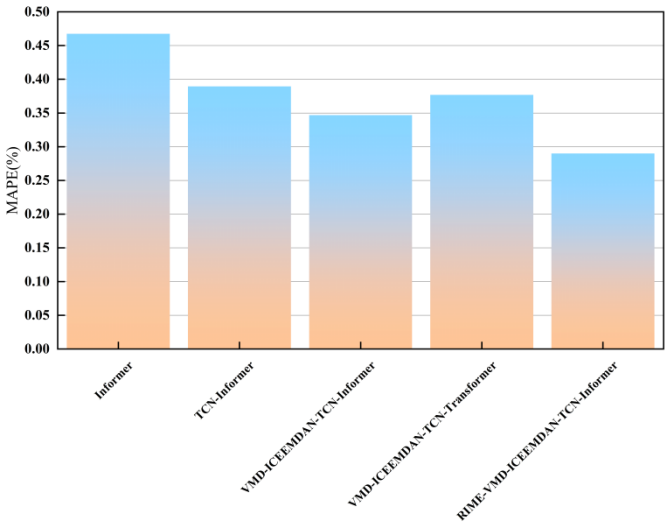

Figure. 10 MAPE indicators for each model

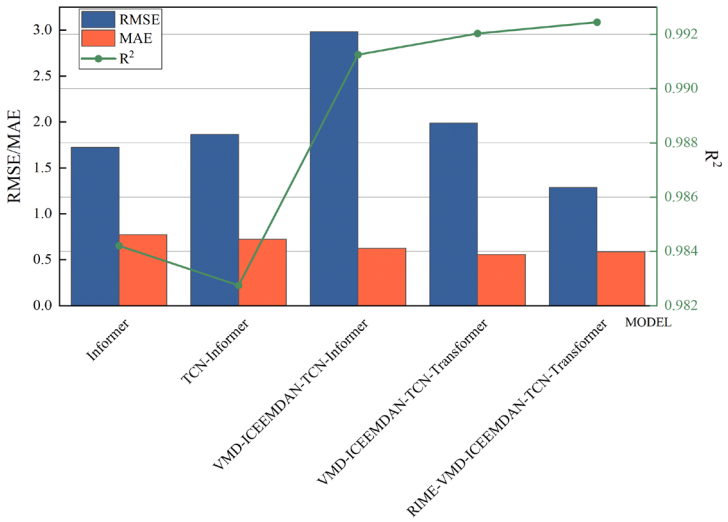

Figure. 11 Comparison of evaluation indicators of each model

The comparison of predicted values and actual values for each model is shown in Figure 12-15. When relying solely on a single model for prediction, the fitting performance is poor, particularly evident during peak periods. If the dataset has not undergone preprocessing, the model may exhibit overfitting, thereby reducing its generalization performance. After decomposition processing, the prediction curves are smoother compared to those of a single model and more closely align with the actual values. especially during peak output periods in the midday hours. Traditional models exhibit lag issues, but after dual decomposition and processing of high-frequency components, the model's predictability for strongly fluctuating signals is significantly improved. During early morning and evening hours, thanks to the mLstm structure of

the xLSTM model, the model achieves good accuracy even under cloudy conditions. Additionally, the ICEEMDAN module in the model's decomposition structure introduces white noise to enhance the model's generalization performance. Due to its parallel computing capability, the TCN-Informer model has shorter computation time and higher computational efficiency compared to the TCN-Transformer model. After parameter optimization of the model in this experiment using the RIME algorithm, the model's prediction accuracy is further improved.

Furthermore, in practice, it was found that the model proposed in this experiment not only performs well in short-term power prediction but also improves its prediction performance as the sequence length increases within a certain range. Therefore, this model can not only be used for short-term power prediction tasks but also adapted to prediction tasks across multiple time scales.

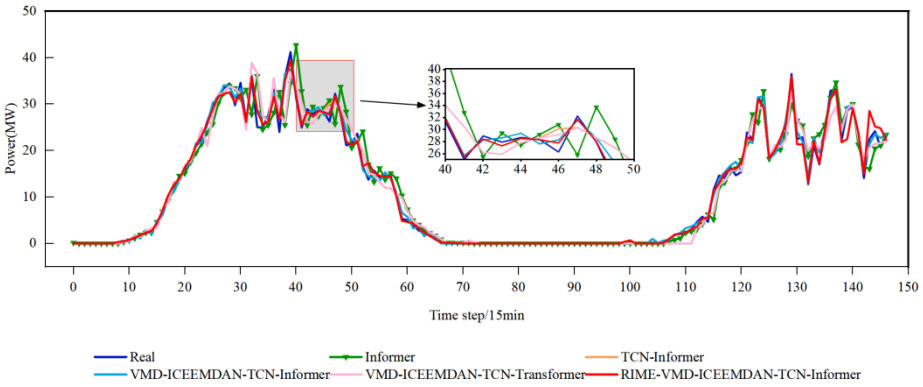

Figure. 12 Spring PV power forecast results comparison

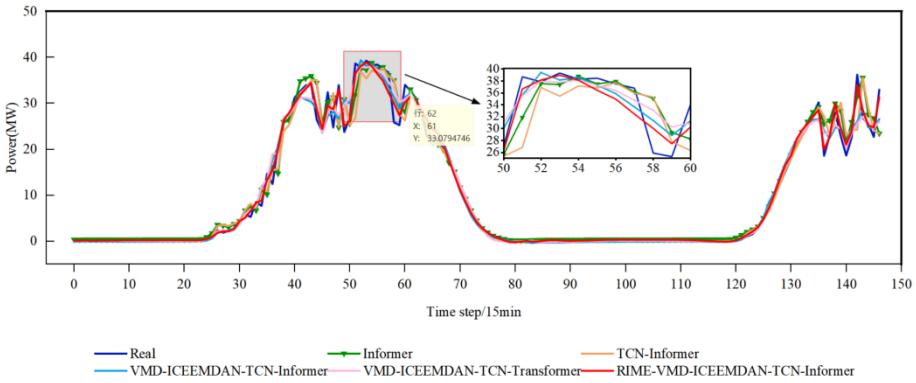

Figure. 13 Summer PV power forecast results comparison

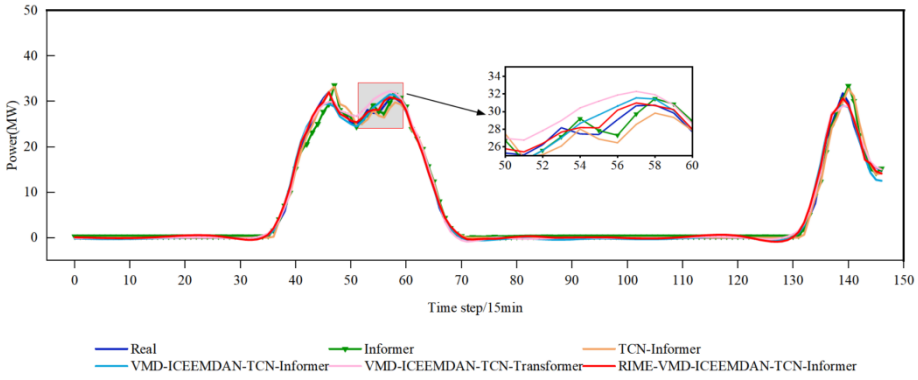

Figure. 14 Autumn PV power forecast results comparison

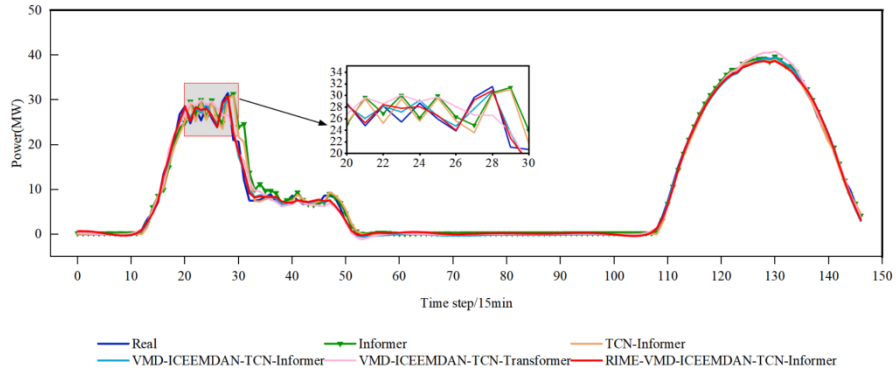

Figure. 15 Winter PV power forecast results comparison

5. Conclusion

To obtain accurate predictions of photovoltaic power output, promote the integration of new energy sources into the grid, and accelerate the development of the electricity market, this study proposes a TCN-Informer-xLSTM prediction model based on dual decomposition. The model was validated using a dataset from multiple climate scenarios in the northern region. The main conclusions are as follows:

(1) The effectiveness of dual decomposition: The VMD-ICEEMDAN dual decomposition framework significantly enhances the feature extraction capability of the original signal. After VMD decomposition, combined with an entropy value screening mechanism, it effectively separates high-frequency noise from low-frequency trend components. Further ICEEMDAN secondary decomposition of high-entropy value modes resolves the mode aliasing and pseudo-mode issues associated with single decomposition, reducing noise interference while enhancing the accuracy of signal reconstruction. Experimental results show that the input data processed by dual decomposition achieves approximately a 10% reduction in the MAPE metric compared to the undecomposed model, validating the critical role of the decomposition strategy in improving prediction accuracy.

(2) The complementary advantages of submodels: high-frequency components utilize the TCN-Informer parallel structure, fully leveraging TCN's local feature capture capabilities and the Informer's global attention mechanism. This ensures computational efficiency while delivering robust prediction results, particularly under complex weather conditions, with the MAPE metric fluctuating within a range of less than 2.5%. This achieves precise fitting of strongly fluctuating signals while maintaining model computational efficiency. Low-frequency components incorporate the xLSTM model, which addresses the gradient vanishing issue in traditional LSTM models for long sequence tasks through an improved gating mechanism and matrix memory structure.

(3) Parameter optimization and computational efficiency: The hyperparameter optimization mechanism of the RIME algorithm effectively balances model complexity and prediction performance. Experimental results show that the optimized model achieves a coefficient of determination (R^2) of 0.9923 on the test set. At the same time, the model achieves computational resource efficiency through parallelization

design (such as causal convolutions in TCN and sparse attention in Informer).

Acknowledgement

This work was supported by National Natural Science Foundation of China (62203311, 62473269), Liaoning Province Key Research and Development Project (2024JH2/102500093), Liaoning Revitalization Talents Program (XLYC2403160), and Basic Scientific Research Project of Liaoning Provincial Department of Education (LJ222411632036).

References

[1] Qazi, A.: A Systematic Review of Renewable Energy Sources. Technologies, and Public Opinions (7), 63837-638512 (2019).
[2] Rasoul, S.: Energy's international history and future. Science 388, 711-711 (2025).
[3] Salman, D., Direkoglu, C., Kusaf, M. et al.: Hybrid deep learning models for time series forecasting of solar power. Neural Comput & Applic 36, 9095–9112 (2024).
[4] Shi, S., Liu, B., Ren, L. et al.: Short time solar power forecasting using P-ELM approach. Sci Rep 14, 30999 (2024).
[5] Singh, P., Singh, N.K. & Singh, A.K.: Wavelet Transform Based Gated-Recurrent Unit Deep Learning Approach for Power Output of Solar Photovoltaic System Forecasting. SN COMPUT. SCI 6, 243 (2025).
[6] Goya, A., Bhattacharya, K.: Design of Multi-Settlement Electricity Markets Considering Demand Response and Battery Energy Storage Systems Participation. Policy and Regulation 2(2), 226-239 (2024).
[7] Bin, H.: A study on the prediction of photovoltaic day-ahead output based on FCM-WS-BP. Control Engineering. 30(12), 2254-2260 (2023).
[8] Bai, M., Wang, R., Lin, C. et al.: Deep multivariable spatial attention CNN-based spatial grid NWP error correction for accurate one-day-ahead photovoltaic power forecast. Earth Sci Inform 18, 427 (2025).
[9] Fu, Y., Chai H., Zhen Z., et al.: Sky image prediction model based on convolutional auto-encoder for minutely solar PV power forecasting. IEEE Transactions on Industry Applications 57(4), 3272-3281 (2021).
[10] R. Asghar, M., Quercio, L., Sabino, A.: A Novel Dual-Stream Attention-Based Hybrid Network for Solar Power Forecasting. IEEE Access 13, 59596-59609 (2025).
[11] Rahimi, N., Park, S., Choi, W. et al.: A Comprehensive Review on Ensemble Solar Power Forecasting Algorithms. Elec. Eng. Technol 18, 719–733 (2023).
[12] Souhe, F.G.Y., Mbey, C.F., Kakeu, V.J.F. et al.: Optimized forecasting of photovoltaic power generation using hybrid deep learning model based on GRU and SVM. Electric Energy 106, 7879–7898 (2024).
[13] Horat, N., Klerings, S. & Lerch, S.: Improving Model Chain Approaches for Probabilistic Solar Energy Forecasting through Post-processing and Machine Learning. Adv. Atmos. Sci 42, 297–312 (2025).
[14] Zhang, X.: Prediction Interval Estimation and Deterministic Forecasting Model Using Ground-Based Sky Image. IEEE Transactions on Industry Applications 59(2), 2210-2224 (2023).
[15] Xu, B., Huang, Y., et al.: Day-ahead probabilistic forecasting of photovoltaic power based on vine copulaquantile regression. Power system technology, 4426-4435 (2022).
[16] C. Yuanjun.: Short-Term Prediction of Photovoltaic Power Based on TCN-LSTM-Attention Model and Kmeans++. 10th International Conference on Power Electronics Systems and Applications 2024, PESA, pp. 1-5, (2024).
[17] Ge, W., Wang, X.: PSO–LSTM–Markov Coupled Photovoltaic Power Prediction Based on Sunny, Cloudy and Rainy Weather. Electric Energy Technology 20, 935–945 (2025).
[18] Zhu, R., Li, T. & Tang, B.: Research on short-term photovoltaic power generation forecasting model based on multi-strategy improved squirrel search algorithm and support vector machine. Sci Rep 14, 14348 (2024).
[19] Zhang, J., Xu, R., et al.: Short-term photovoltaic power prediction based on similar day clustering and PCC-VMD-SSA-KELM model. Solar Energy 45(2), 460-468 (2023).
[20] Qi, L., Xiaoying, R., Zhang, F., Lu, G., Hao, B.: A novel ultra-short-term wind power forecasting method based on TCN and Informer models. Computers and Electrical Engineering 7(A1), 0045-7906 (2024).
[21] Deng, T., Liu, J., Wang, B., et al.: A photovoltaic power prediction method based on multimodal fusion of ground-based cloud maps and meteorological factors. Chinese Journal of Electrical Engineering, 1-14 (2023).