

Article

Informer Short-Term PV Power Prediction Based on Sparrow Search Algorithm Optimised Variational Mode Decomposition

Wu Xu ^{1,2}, Dongyang Li ^{1,2}, Wenjing Dai ^{1,2} and Qingchang Wu ^{3,*}¹ School of Electrical and Information Technology, Yunnan Minzu University, Kunming 650504, China; 040326@ymu.edu.cn (W.X.); 15774697105@163.com (D.L.); 15215156275@163.com (W.D.)² Yunnan Key Laboratory of Unmanned Autonomous System, Kunming 650504, China³ Lancang–Mekong International Vocational Institute, Yunnan Minzu University, Kunming 650504, China

* Correspondence: wuqingchang@ymu.edu.cn

Abstract: The output power of PV systems is influenced by various factors, resulting in strong volatility and randomness, which makes it difficult to forecast. Therefore, this paper proposes an Informer prediction model based on optimised VMD for predicting short-term PV power. Firstly, the temporal coding of the Informer model is improved and, secondly, the original sequence is decomposed into multiple modal components using VMD, and then optimisation of the results of VMD in conjunction with the optimisation strategy of SSA improves the characteristics of the time series data. Finally, the refined data are fed into the Informer framework for modelling and prediction, utilising the self-attention mechanism and multiscale feature fusion of Informer to precisely forecast PV power. The power of PV prediction data from the SSA-VMD-Informer model and four other commonly used models is compared. Experimental results indicate that the SSA-VMD-Informer model performs exceptionally well in short-term PV power prediction, achieving higher accuracy than traditional methods. As an example, the results of predicting the PV power on 24 April in a region of Xinjiang are 1.3882 for RMSE, 0.8310 for MSE, 1.14 for SDE, and 0.9944 for R².

Keywords: PV power; time series; Informer; VMD; SSA; short-term forecasting



Citation: Xu, W.; Li, D.; Dai, W.; Wu, Q. Informer Short-Term PV Power Prediction Based on Sparrow Search Algorithm Optimised Variational Mode Decomposition. *Energies* **2024**, *17*, 2984. <https://doi.org/10.3390/en17122984>

Academic Editor: Oscar Barambones

Received: 30 May 2024

Revised: 11 June 2024

Accepted: 15 June 2024

Published: 17 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Energy is the cornerstone of human life. Humankind relies on traditional sources of energy such as fossil fuels, hydropower, and nuclear energy to meet its energy needs, but these methods pose significant environmental challenges. With China's government pursuing a “dual-carbon” goal and a strategy to establish a power system based on new energy sources, accelerating the transformation of the energy structure has become particularly urgent. In this context, the development of new energy sources, especially photovoltaic (PV) power generation [1], has become an important task at present. Although photovoltaic power plants are competitive with traditional fossil fuel power plants, the volatility of solar radiation and the fluctuation of power generation due to various climatic and geographic factors still pose a challenge to the performance of solar power plants and power grids [2]. Therefore, accurate prediction of PV power becomes the key to solving this problem, which can effectively avoid all the adverse effects caused by the discrepancy between the actual PV power and the expected one.

Predicting PV power involves many complex variables, making it crucial to comprehensively consider these factors when building the prediction model and to employ suitable modelling techniques to achieve the most accurate results possible [3]. In the early days, machine learning models were used to predict PV power. They are able to learn laws and patterns from data, enabling automated decision making and prediction. Compared to traditional rule coding methods, machine learning is adaptive and can be adapted and optimised to new data, discovering hidden patterns and correlations in the data, which improves the accuracy of the prediction. Random Forest [4] is an integrated learning

method that improves accuracy by building multiple decision trees and combining their predictions. Differential Integration Moving Average Autoregressive Model (ARIMA) [5] is an autoregressive model that combines difference, integration, and moving average for the modelling and forecasting of time series data. XGBoost [6] is a gradient boosting algorithm that improves the accuracy and generalisation of a model by integrating multiple weak classifiers (usually decision trees). Support Vector Regression (SVR) [7] is a regression method based on support vector machines that fit the data by maximising the boundaries and is suitable for dealing with small samples and nonlinear and high-dimensional data. Machine learning algorithms are more efficient in dealing with some simple prediction problems, but historical PV power generation data are usually a multidimensional, nonlinear dataset containing a large number of time steps, whereas neural networks are more expressive and generalisable in dealing with complex, nonlinear problems. Long Short-Term Memory (LSTM) [8] is a Recurrent Neural Network (RNN) variant suitable for processing and predicting time series data with memory units that capture long-term dependencies. Gated Recurrent Unit (GRU) [9] is an LSTM-like variant of RNN with update gates and reset gates for capturing long-term dependencies in sequential data more efficiently. Convolutional Neural Network (CNN) [10] is a neural network used to process input data with a grid structure to extract and learn features through convolutional and pooling layers. BP Neural Networks (Backpropagation Neural Networks) [11] is a traditional neural network structure that optimises the network parameters by means of backpropagation algorithms for problems such as classification and regression. Deep Belief Networks [12] are neural networks consisting of multiple layers of constrained Boltzmann machines that learn the distribution and feature representation of data through unsupervised pre-training and supervised fine-tuning.

In the past few years, the Transformer model [13–15] has injected new vigour into short-term PV power prediction, mainly in terms of its powerful sequence processing ability and efficient parallel computing capability. Through the self-attention mechanism, Transformer is able to capture remote dependencies and complex nonlinear patterns, thus improving prediction accuracy. In addition, its flexible architecture can easily integrate multiple input features with high adaptability, which improves the model's performance in handling multivariate prediction. These advantages enable Transformer to demonstrate significant improvements and innovations in the field of PV power prediction. The literature [16] proposes a Dual Encoder Transformer Model (DualET) that extracts information from image and sequence data by means of wavelet transform and sequence decomposition modules and introduces an attention module to learn the association between temporal features and cloud information. The experimental results of this model on real datasets show that it outperforms other models in short-term PV power prediction.

Due to the volatility and randomness of PV power generation, the accuracy of directly predicting its power is not high. For this reason, the power of PV power generation can be decomposed and each component can be predicted separately to improve the accuracy of the overall prediction. The literature [17] combines sequence decomposition methods and deep learning models to propose a hybrid model (VMD-GA-Conv-A-LSTM) that can significantly improve the accuracy of PV power prediction. The model first decomposes the PV power sequences using VMD with optimised parameters. It then inputs the sub-sequences with preprocessed historical meteorological data into the LSTM model combining 1D convolution and an attention mechanism for prediction, and obtains the overall prediction value by accumulating the prediction results of each sub-sequence, which shows the best prediction performance on several benchmark models.

A single neural network is prone to overfitting, and intelligent optimisation algorithms can optimise the parameters of the neural network, improve its generalisation ability, reduce its sensitivity to noise, and better handle complex and nonlinear data, thus improving the accuracy and stability of PV power prediction. The literature [18] applies the Artificial Fish Schooling Algorithm (AFSA) to optimise the BP neural network, leveraging AFSA's global optimisation and intrinsic parallel computing capabilities to effectively optimise the

network's weights and thresholds, so as to train an efficient PV output power prediction model. The literature [19] proposes a prediction of a short-term model for PV power using Pearson's correlation coefficient, EEMD, sample entropy, SSA, and LSTM, and the experimental results demonstrate that the model achieves a low prediction error for PV power under diverse weather conditions. The literature [20] introduces a system for predicting PV power using a deep convolutional neural network (CNN) coupled with a signal decomposition algorithm. This system extracts deep features using AlexNet for transfer learning, decomposes historical power signals into sub-components, and converts all input parameters into 2D feature maps for CNN input. The literature [21] introduces ADAMS, a new Autoformer model designed for short-term PV power forecasting. ADAMS utilises a multiscale framework and de-stationary attention mechanism. ADAMS outperforms baselines in predicting PV output.

Based on the above, the previous literature has typically used only one of the methods of decomposing raw PV data or intelligent optimisation algorithms when optimising short-term PV power prediction models, which have limited improvement in prediction accuracy. Therefore, this paper proposes a new model, the SSA-VMD-Informer model, which utilises both approaches and aims to improve the prediction accuracy of short-term PV power. The process is outlined as follows:

- Firstly, Informer, a model that uses a self-attentive mechanism with a multilevel temporal prediction mechanism and focuses on processing long series data, is adopted and the temporal coding of the original Informer model is optimised.
- Secondly, Variational Mode Decomposition (VMD) is optimised with the intelligent optimisation algorithm Sparrow Search Algorithm (SSA), which automatically adjusts the key parameters of VMD and decomposes the original PV power generation data with the optimal parameters obtained.
- Finally, the modal components obtained from the SSA-optimised VMD of the original PV power and other environmental factors affecting PV power generation are inputted into the Informer model, and the reliability and superiority of the model are verified by the PV power generation test.

2. Materials and Methods

2.1. Informer Network

The Informer model is a deep learning approach tailored for time series prediction, particularly suited for handling long series data. This model was proposed by researchers at Huazhong University of Science and Technology in 2020 [22]. The Informer model is characterised by its utilisation of a multilevel self-attention mechanism and a Transformer-based structure, enabling it to effectively capture long-term dependencies in time series data.

The overall architecture of Informer is shown in Figure 1. The left side of the architecture represents the encoding process of Informer; the right side represents the decoding process; the encoder extracts features from the input sequence through the multilayer probabilistic sparse self-attention module; and the decoder receives the input of the sequence (with the target part of the prediction set to 0), interacts with the encoded features through the multi-attention, and outputs the prediction results directly. The encoder's task is to convert the input sequence into a high-dimensional feature vector, while the decoder is responsible for transforming this vector into the target sequence. Similarly, to the Transformer, the Informer model uses residual connections and normalisation techniques to enhance convergence speed and improve performance. Additionally, the Informer model is structured with the property of adaptive length, which enables it to effectively process time series data of different lengths.

The traditional Transformer model, which relies on the self-attention mechanism, has shown significant predictive capabilities in time series prediction, but still suffers from a number of problems. The Informer model in the Transformer model introduces several innovative improvements:

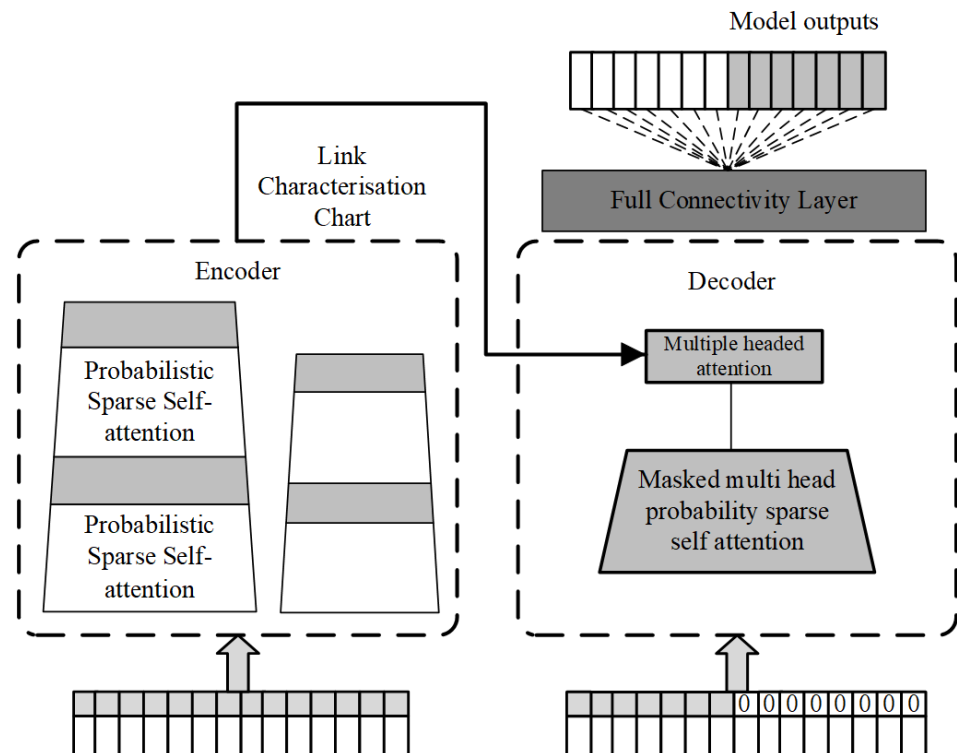


Figure 1. Informer network structure.

2.1.1. ProbSparse Self-Attention

In the traditional Transformer, there are only a few points in the self-attention mechanism that have strong correlations between them. Therefore, if it is possible to remove those irrelevant queries during the attention computation process, the amount of computation can be reduced. Experiments have proved that deleting some of the useless queries does not affect the model accuracy. As shown in Figure 2, in Transformer, each query needs to be computed with each value in the sequence, whereas in Informer, a query only needs to be computed with the values at the positions marked in red in the graph (indicating the positions where its correlation is strong).

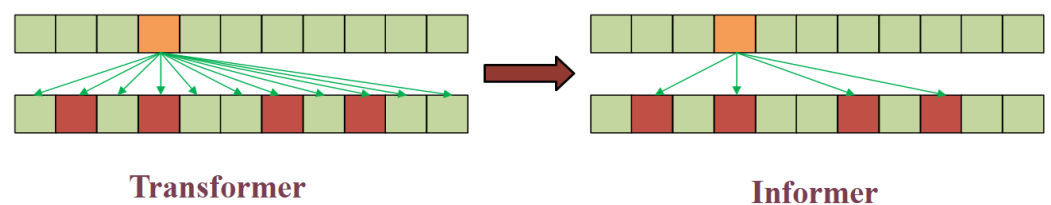


Figure 2. Transformer and Informer attention comparison.

2.1.2. Self-Attention Distilling

The output data processed through the probabilistic sparse self-attention module is fed into the distillation layer. The self-attention distillation mechanism is commonly used to compress or optimise Transformer class models to reduce the model size and computational complexity while maintaining the model performance as much as possible, as shown in Equation (1).

$$X_{j+1}^t = \text{Maxpool}(\text{ELU}(\text{Conv1d}(\left[X_j^t\right]_{AB}))) \quad (1)$$

Conv1d represents a 1D convolution operation applied to a time series.

2.1.3. One Forward Operation

In the Transformer model, the output predictions are made in sequential order; the first prediction is output first based on the time series, then the second prediction is output based on the first prediction, and so on. The speed is slow, as shown in Figure 3.

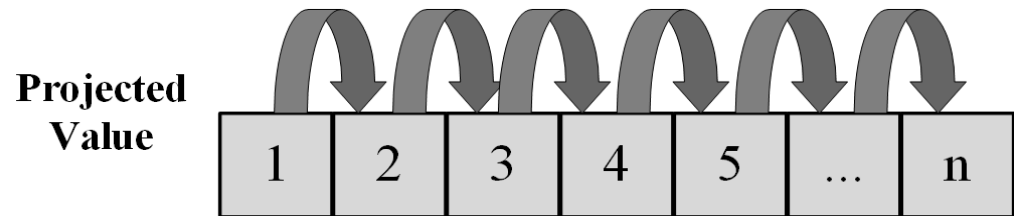


Figure 3. Transformer output predictive value process.

The Informer model enables the decoder to output all the predictions at once after referring to the bootstrap values by introducing the bootstrap values, i.e., giving the results of the previous sequence in the decoder, as shown in Figure 4.

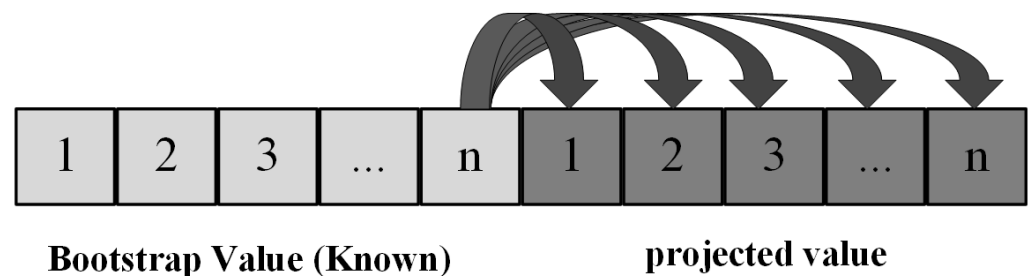


Figure 4. Informer output predictive value process.

The input provided to the decoder by the One Forward Operation is shown in Equation (2).

$$X_{de}^t = \text{Concat}(X_{token}^t, X_0^t) \in R^{(L_{token}+L_y) \times d_{model}} \quad (2)$$

where X_{de}^t denotes the input to the decoder; X_{token}^t is the starting token, which does not use all of the output dimensions; $X_0^t \in R$ represents a placeholder with a value of 0 for the target sequence, which can be used to focus on both the historical information of the sequence and the future state of the sequence when predicting future sequences at the time of multicentre attention; and Concat denotes that X_{token}^t will be spliced together with X_0^t before inputting it to the encoder for prediction.

2.2. Variational Mode Decomposition

Photovoltaic power data are highly volatile and noisy. VMD [23], as an adaptive signal processing method, can break down complex signals into intrinsic modal functions (IMFs), effectively reducing noise and extracting significant features. Additionally, PV power data are typically nonlinear and nonsmooth. VMD adaptively decomposes the signal through a variational approach, which can better handle these characteristics and improve the performance of the prediction model and effectively suppresses modal aliasing. While ensuring that the sum of all modes equals the original signal, VMD continuously updates the modal functions and their centre frequencies to ensure that each modal component has a centre frequency and a limited bandwidth. At the same time, it achieves optimisation by minimising the sum of the estimated bandwidths of each sub-modality. This decomposition process can be regarded as a constrained variational model, as in Equations (3) and (4).

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \partial_t \left\| \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(k) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad (3)$$

$$s.t. \sum_{k=1}^K u_k = x(t) \quad (4)$$

where u_k and w_k are, in order, the k -th modal component and the corresponding centre frequency, K is the number of sub-modal functions, $\delta(t)$ is the Dirac function, $*$ is the sign of the convolution operation, and $x(t)$ is the raw PV power data. To convert a constrained optimisation problem into an unconstrained optimisation problem solution, a quadratic penalty factor α and Lagrange multipliers λ can be introduced to form the original Equations (3) and (4) into augmented Lagrange equations to obtain Equation (5).

$$L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_{k=1}^K \left\| \partial_t \left[(\delta(t) + \frac{j}{\pi t}) * u_k(t) \right] e^{-j\omega t} \right\|_2^2 + \|x(t) - \sum_{k=1}^K u_k(t)\|_2^2 + (\lambda(t), x(t) - \sum_{k=1}^K u_k(t)) \quad (5)$$

An iterative update of the modal components and their centre frequencies is performed using ADMM to find the saddle points of the unconstrained model. The update equations are given in Equations (6) and (7).

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{x}(\omega) - \sum_{i \neq k} \hat{u}_i(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (6)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^n(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^n(\omega)|^2 d\omega} \quad (7)$$

where $\hat{u}_k^{n+1}(\omega)$, $\hat{x}(\omega)$, $\hat{u}(\omega)$, and $\hat{\lambda}(\omega)$ are the Fourier transforms of $u_k^{n+1}(t)$, $x(t)$, $u(t)$, and $\lambda(t)$ in that order, ω denotes the frequency, and n represents the number of iterations.

A single time series prediction model is difficult to fully capture the complex spatio-temporal relationships in the PV power system during the prediction process; it is susceptible to the interference of high-frequency noise, and it may affect the real-time performance and robustness of the system in practical application scenarios. The introduction of Variational Mode Decomposition (VMD) helps to overcome the above problems, and the adaptive decomposition feature of VMD enables it to better deal with nonlinear and nonsmooth signals, and to extract the eigenmode functions at different frequencies. By taking each eigenmode function obtained from VMD as a factor affecting the PV power, the spatial and temporal relationships of the various influencing factors can be considered more comprehensively and the model's prediction accuracy can be enhanced. Simultaneously, the VMD results aid in eliminating high-frequency noise, which makes the model more robust and more adaptable to the complexity of practical applications.

2.3. Sparrow Search Algorithm

SSA [24] is an innovative optimisation algorithm that simulates the behaviour of sparrow groups in nature during foraging and predation. The sparrow, as a flocking animal by nature, has a clear division of labour within the group. A portion of the sparrows is energetic, responsible for finding and collecting food, and guiding the entire population to forage; the other part of the sparrows cooperates with the former as a follower to forage for food. When a sparrow detects a competitor while foraging, it sends out an alarm signal, and the entire population will engage in anti-feeding behaviour. The sparrows responsible for finding food act as guides for the foragers, constantly updating their positions from memory in order to search for a wider range of food. Because sparrows are at risk of being captured by predators, scouts are selected in the population to provide early warning.

Productors are food seekers and their positions are updated, as in Equation (8).

$$X_i(t+1) = X_i(t) + r_1 \times (X_{best}(t) - X_{mean}(t)) \times e^{-r_2 \times t} \quad (8)$$

where $X_i(t)$ denotes the position of the i -th productor at time t , $X_{best}(t)$ is the current best position, $X_{mean}(t)$ is the average of all current productor positions, and r_1 and r_2 are randomly generated coefficients to control the search step size and direction.

The scrounger's position update formula is shown in Equation (9).

$$X_i(t+1) = X_i(t) + r_3 \times (X_{worst}(t) - X_i(t)) \quad (9)$$

where $X_{worst}(t)$ denotes the current worst position and r_3 is a randomly generated coefficient; the formula helps the scrounger to approach the global optimal solution.

The equation for updating the scout's position is presented in Equation (10).

$$X_i(t+1) = X_{rand}(t) \pm r_4 \times (X_{rand}(t) - X_i(t)) \quad (10)$$

where $X_{rand}(t)$ is a randomly chosen location for a particular sparrow and r_4 is a random coefficient.

3. Model Design

3.1. SSA-VMD-Informer Short-Term PV Power Prediction Model

The traditional Transformer model depends on the self-attention mechanism and has demonstrated strong prediction ability in time series, but there are still some problems: the self-attention mechanism needs to evaluate the similarity between a certain moment and all other moments during the calculation process, which leads to higher time complexity; for long sequence data, in order to deal with these data, it is often necessary to use multiple parallel multihead attention structures, which takes up a lot of memory resources; in addition, when using the Transformer model, the prediction speed is reduced when dealing with long sequences because the decoder needs to rely on the output of the previous encoder when making predictions. In this regard, this paper proposes an SSA-VMD-Informer short-term PV power prediction model. This model captures information in time series data more comprehensively, thereby enhancing prediction accuracy. In this paper, the prediction model will be constructed from the following aspects:

- Improvements are made to the temporal coding of the Informer model to improve the model's ability to understand and process time series data.
- Decomposition of raw PV power data using variational modal decomposition techniques, which provides support for more in-depth data analysis and processing by increasing the dimensionality of the data.
- SSA is introduced to optimise VMD to ensure that VMD can achieve the best decomposition, thus improving the efficiency and effectiveness of data processing.
- The components obtained from the SSA-VMD decomposition, as well as the set of relevant factor features of the raw PV power data, are inputs into the improved time-coded Informer model for prediction.

The model structure is shown in Figure 5.

3.2. Improvement of Informer Time Code

In long sequence timing modelling problems, it is necessary to consider not only local timing information, but also hierarchical timing information, such as day of the week, month, and year, as well as timestamped information for unexpected events such as holidays.

As shown in Figure 6, the original Informer model location embedding is divided into three types: feature vector, local timestamp (location encoding in Transformer), and global timestamp. The global timestamp part contains week, month, and holiday parts.

In photovoltaic power generation, the amount of power generated is greatly influenced by panel temperature and radiation levels. Panel temperatures follow a descending order from winter to summer and radiation levels follow a descending order from winter to autumn. Hence, the global timestamps for PV power forecasts should include the seasons of each PV's historical power, along with weeks, months, and holidays.

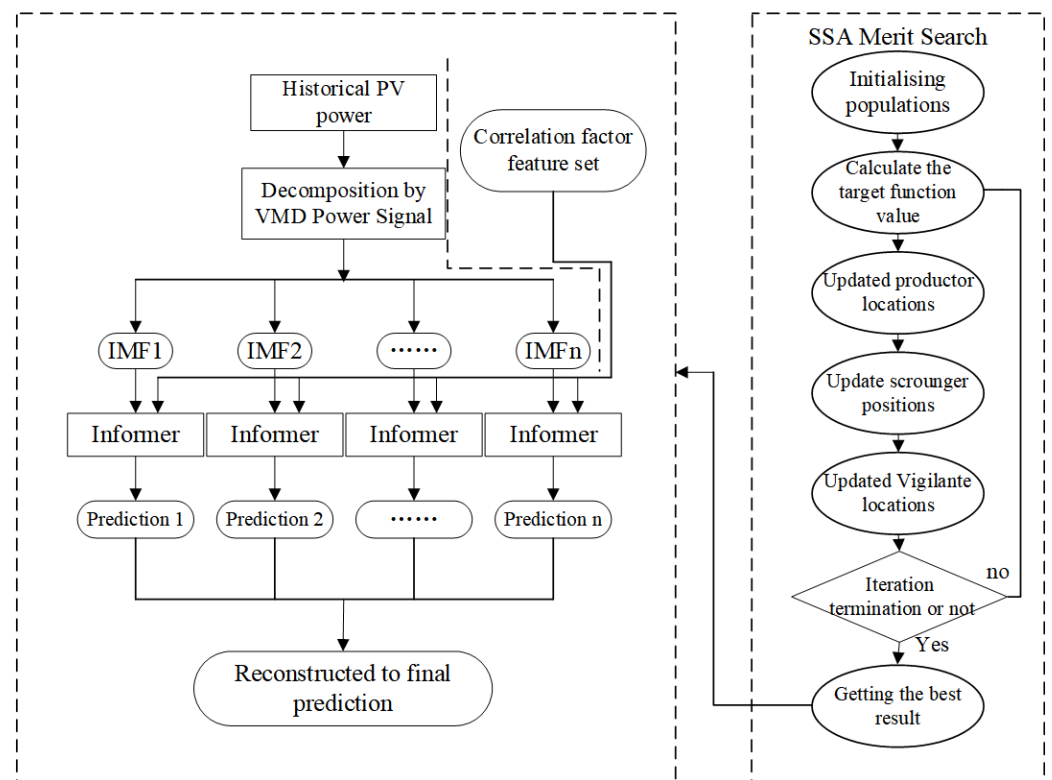


Figure 5. SSA-VMD-Informer model structure.

Position Code

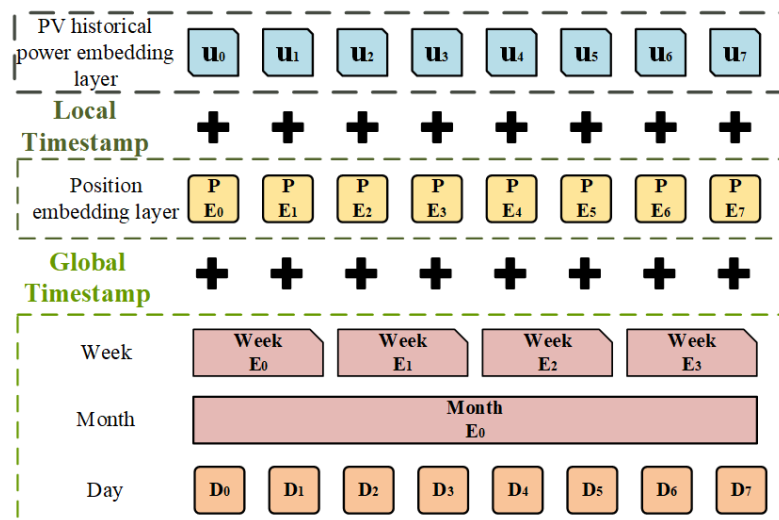


Figure 6. Original Informer location coding.

The PV power mainly depends on the light intensity, so according to the sunrise and sunset time of each month in the area where the PV power plant is located, the PV historical power generation is divided into night and day. The PV power generation data used in this study are from the power generation of the PV power plant in Xinjiang for one year. For example, the sunrise time in January in Xinjiang is about 9:30 and the sunset time is about 19:00, so the data from 0:00 to 9:30 and 19:00 to 24:00 are coded as night, while the data from 9:30 to 19:00 are coded as day. The improved time coding is shown in Figure 7.

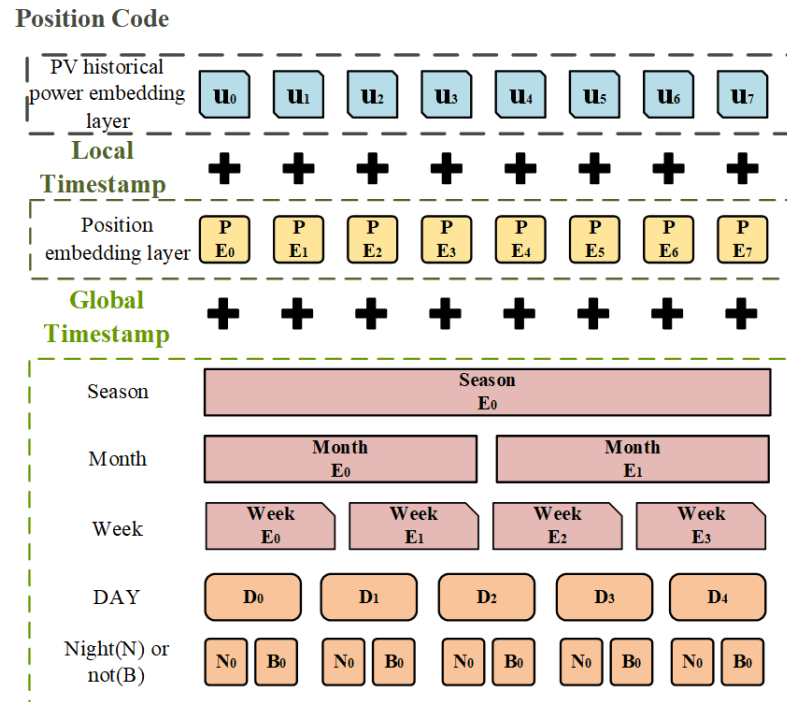


Figure 7. Improved Informer location coding.

The location coding of the Informer model is shown in Equation (11).

$$\begin{cases} PE(P_{\text{pos}}, 2i) = \sin\left(\frac{P_{\text{pos}}}{1000 \cdot d_{\text{model}}^{2i}}\right) \\ PE(P_{\text{pos}}, 2i + 1) = \cos\left(\frac{P_{\text{pos}}}{1000 \cdot d_{\text{model}}^{2i}}\right) \end{cases} \quad (11)$$

where PE represents the position encoding, P_{pos} denotes the absolute position of the vector, d_{model} denotes the total dimension of the vector, i denotes the i -th dimension in the vector, and each dimension in the position encoding has its corresponding sinusoidal signal.

3.3. Intelligent Optimisation Algorithm Selection

The primary objective of an intelligent optimisation algorithm is to locate a solution within the search space that minimises or maximises the value of the fitness function, contingent upon the problem's characteristics. The fitness function $f(x)$ maps a solution x in the problem space to a real value that represents the degree of merit of that solution, the fitness value. The case of lower fitness is usually referred to as a minimisation problem, while higher fitness is usually referred to as a maximisation problem.

In the case of a minimisation problem, the fitness function represents the objective or cost of the problem and the task of the optimisation algorithm is to find the solution that minimises the fitness function. Therefore, when the fitness is lower, it means that the solution is closer to the optimal solution of the problem or better. Optimising VMD with an intelligent optimisation algorithm is a minimisation problem.

We selected the first 10 days of data of raw PV power of the Xinjiang PV power plant in 2019 as an example for VMD and decomposed VMD optimised by three different commonly used optimisation algorithms with minimum sample entropy as the fitness function. The parameters for the intelligent optimisation algorithm are as follows: a maximum of 20 iterations, a population size of 15, and a dimensionality of 2 (α and K). Here, α represents the regularisation parameter, which controls the smoothing degree of the signals and the stability of the decomposition results, while K indicates the number of modal functions to be decomposed.

The VMD of different intelligent optimisation algorithms are compared with the optimised VMD of SSA and the trend of the fitness value curves of different optimisation algorithms is obtained, as shown in Figure 8. The figure shows that SSA achieves its lowest fitness value of 0.049073 in the initial iteration, while WOA and GWO reach the lowest in the third iteration and PSO reaches the lowest in the seventh iteration.

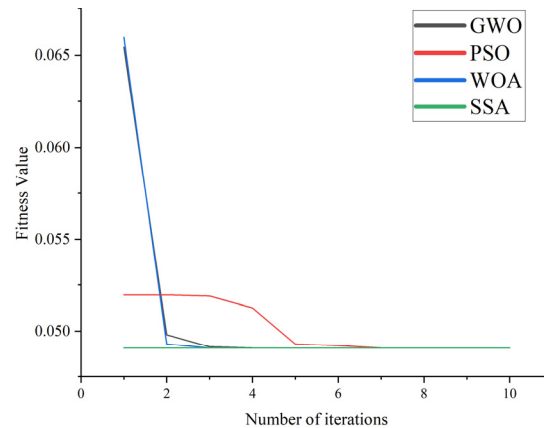


Figure 8. Variation trend of fitness curves of different optimisation algorithms.

After SSA optimisation, the parameters of VMD are $K = 9$ and $\alpha = 1896$. According to Figure 9, the nine modal components generated by SSA-VMD decomposition are clearly distinguished in frequency. Each component not only retains the characteristics of the original PV power sequence, but also effectively inhibits mode aliasing. Thus, optimising VMD with SSA enhances the accuracy of PV power prediction.

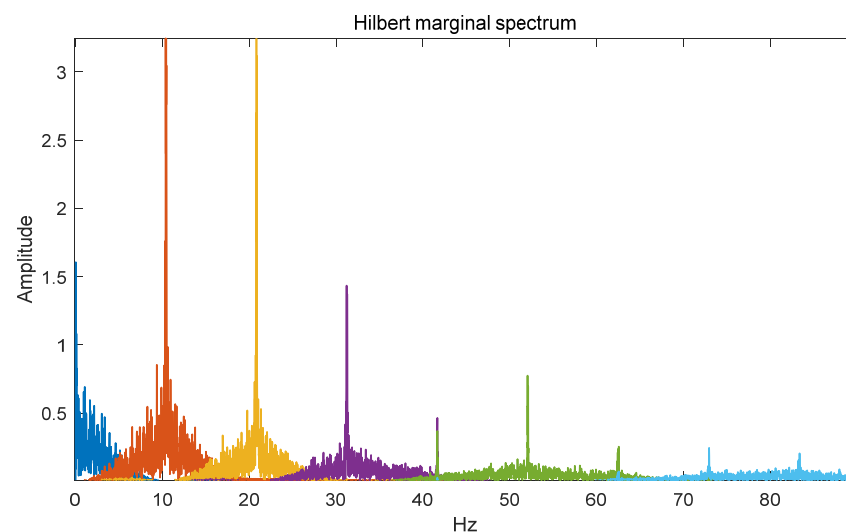


Figure 9. SSA-VMD decomposes the IMF marginal spectrum.

Therefore, this research component will optimise the parameters of VMD by means of the sparrow optimisation algorithm, with a view to obtaining a combination of the VMD parameters that is more suitable for PV power data, thus improving the performance of the prediction model. This approach is expected to be advantageous in the application of VMD to provide more accurate and reliable decomposition results for PV power prediction.

3.4. SSA Optimisation VMD

In order to optimise VMD using the SSA, we minimise the Sample Entropy (SampEn) as a fitness function. In this case, the key parameters of VMD are the modal number K and the penalty parameter α . These parameters need to be optimised by SSA to minimise the sample entropy:

Firstly, the formula for defining the sample entropy is as in Equation (12). Assuming that the intrinsic modal functions (IMFs) obtained after VMD are $\{IMF_1, IMF_2, \dots, IMF_k\}$, we weight and sum the sample entropy of each IMF as the total sample entropy.

$$SampEn(IMF) = -\log\left(\frac{A(IMF)}{B(IMF)}\right) \quad (12)$$

where $A(IMF)$ is the number of pairs of sequences that satisfy the specified similarity metric and $B(IMF)$ is the number of pairs of sequences that satisfy the slightly relaxed condition.

In order to make VMD optimal with respect to the parameters K and α , we define the optimisation objective function as Equation (13).

$$Fitness(k, \alpha) = \sum_{i=1}^k SampEn(IMF_i) \quad (13)$$

In this formulation, SSA will optimise with the objective of finding the parameters K and α that minimise the total sample entropy.

The steps to optimise VMD using sparrow optimisation algorithm are:

1. Initialisation population: generate an initial population containing multiple sparrow individuals, where each sparrow individual represents a set (K, α) .
2. Fitness calculation: for each sparrow individual, decompose the signal using VMD and calculate the sample entropy of each IMF, where the sum is used as the fitness value.
3. Updating sparrow location: according to the update rule of SSA, update the parameter (K, α) for each sparrow individual so that its fitness value decreases gradually.
4. Iteration process: Repeat the process of fitness calculation and position updating until the maximum number of iterations is reached or the convergence condition is satisfied.
5. Output result: finally output the sparrow individual with the smallest fitness value, i.e., the optimal K and α parameters.

By this method, the global search capability and local search accuracy of SSA can be effectively utilised to optimise the VMD parameter settings, thus improving the decomposition quality and computational efficiency.

4. Experiment and Analysis

4.1. Experimental Data and Evaluation Criteria

In this study, short-term PV power prediction experiments are conducted using the measured PV power data from the Xinjiang PV power plant for the year 2019. The dataset consists of 35,041 sets of data with a sampling interval of 15 min. The data include seven meteorological data items, namely, component temperature, temperature, barometric pressure, humidity, total radiation, direct radiation, scattered radiation at the sampling moment, and the corresponding historical data of PV power generation. There are 34,945 sets of data used for model training and the prediction test is carried out on the remaining 96 sets of data.

To analyse the predictive effectiveness of the model developed in this paper and other models, $RMSE$, MAE , SDE , and R^2 are used to compare and analyse the prediction results, with the formulas expressed as follows.

$$RMSE = \sqrt{\frac{1}{t} \sum_{i=1}^t (\hat{p}_i - p_i)^2} \quad (14)$$

$$MAE = \frac{1}{t} \sum_{i=1}^t |\hat{p}_i - p_i| \quad (15)$$

$$SDE = std(|\hat{p}_i - p_i|) \quad (16)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{p}_i - p_i)^2}{\sum_{i=1}^n (\bar{p} - p_i)^2} \quad (17)$$

where $\hat{p}_i$ and p_i are the predicted and true values of the i -th sampling point, respectively, t represents the number of prediction steps, $\bar{p}$ is the mean of the dependent variable, and n is the sample size. *RMSE* can reflect the overall average error of the model's predicted values. *MAE*—Mean Absolute Error; the smaller the *MAE*, the more stable the model. *SDE* reflects the stability or uncertainty of the sample mean estimation of a set of data, which is a measure of the deviation between the predicted and true values in the PV power generation prediction evaluation indexes; the smaller the *SDE*, the more stable the model is. R^2 can help to assess the accuracy of a PV power prediction model; the larger the R^2 , the stronger the model's predictive ability, and conversely the weaker the model's predictive ability.

4.2. Model Parameter Setting

The main parameters of the Informer model are shown in Table 1.

Table 1. Parameter settings.

Parametric Name	Set Value	Parametric Name	Set Value
<i>seq_len</i>	192	<i>n_heads</i>	8
<i>label_len</i>	96	<i>factor</i>	72
<i>pred_len</i>	96	<i>batch_size</i>	32
<i>enc_in</i>	8	<i>train_epochs</i>	6
<i>dec_in</i>	8	<i>learning_rate</i>	0.001

4.3. Validation of Time Code Validity

To validate the performance of the improved Informer model with time coding, a comparison is made between the predicted values of the original Informer model, the predicted values of the improved time-coded Informer model, and the true values.

In Figure 10, the horizontal coordinates represent the number of forecast steps. Since the time resolution of the dataset is 15 min, each step represents 15 min. The horizontal coordinate ranges from 0 to 96, which means it covers a whole day. (The horizontal coordinates of the subsequent prediction curves are the same as here.) From Figure 10 and Table 2, it can be seen that after improving the time coding for the PV power data, the predicted values obtained by the Informer network are closer to 0 in the black part of the left and right sides, and the original Informer model's prediction of the PV power exists below 0, which is not in line with the real situation, while the Informer model with the improved time coding does not have such a problem, and the predicted values are closer to the real values. The *RMSE* of the improved Informer model decreases by 0.0012, the *MAE* value decreases by 0.0367, and the *SDE* value remains unchanged compared to that before the improved time coding, while the R^2 value increases by 0.0001. It shows that the improved time coding Informer model is more capable of predicting the short-term PV power.

Table 2. Comparison of evaluation indicators of forecasting results before and after Informer improvement.

Model	RMSE	MAE	SDE	R^2
Improved time-coded Informer	3.5188	2.2589	3.30	0.9635
Informer	3.5200	2.2956	3.30	0.9634

Due to the characteristics of the dataset, the PV power generation dataset of a place in Xinjiang used in this study has a generation value at night that is basically 0. Therefore, in this study, most of the predicted values at night are close to 0 or slightly higher than 0. If the generation value at night in the dataset used is not 0, the model is able to make the corresponding predictions as well.

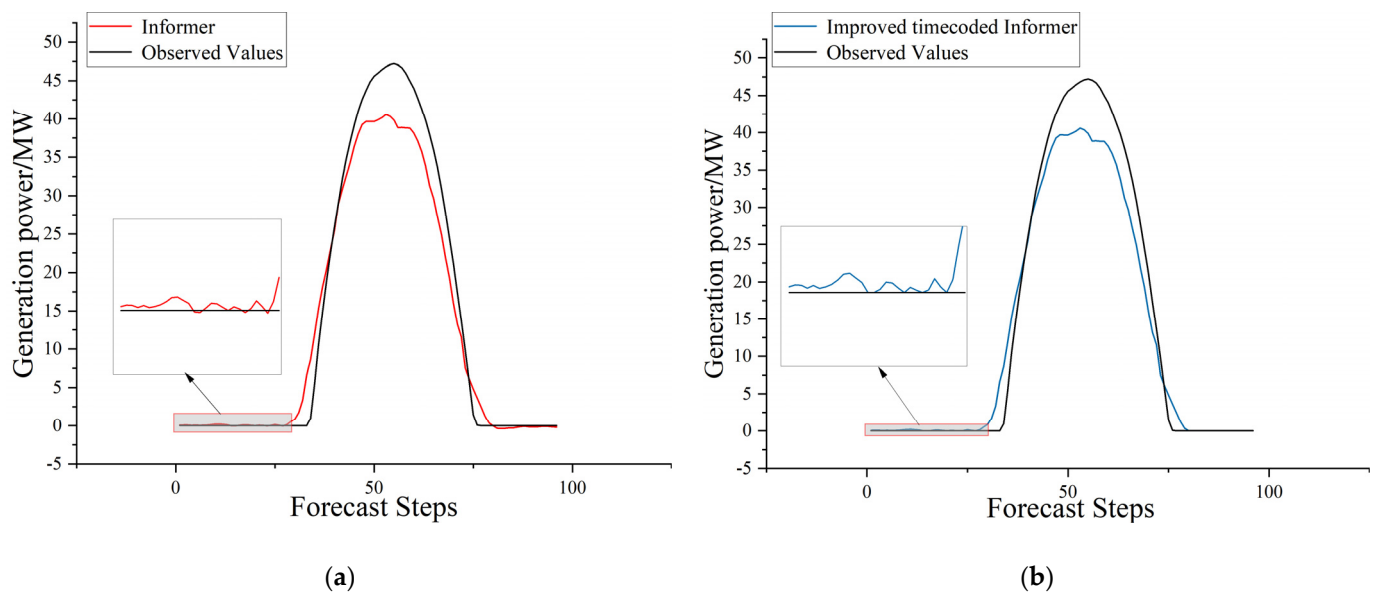


Figure 10. (a) Informer 1-day PV power forecast curves; (b) Improved time-coded Informer 1-day PV forecast curves.

4.4. VMD and SSA Ablation Study

In order to verify whether VMD as well as SSA can improve the capability of the Informer model, the Informer predictions, VMD-Informer predictions, and SSA-VMD-Informer predictions are compared with the true values.

From Figure 11 and Table 3, it can be seen that the RMSE of the model with VMD of the raw PV power generation data is reduced by 0.0167, MAE is reduced by 0.0941, SDE is reduced by 0.06, and R^2 is improved by 0.0004. The model with SSA-VMD decomposition of the raw PV power generation data has a reduction of RMSE by 0.718, MAE by 0.1743, and SDE by 0.29, compared to the model decomposed with only VMD, which has a reduction of RMSE by 1.342, MAE by 0.7748, SDE by 0.84, and R^2 by 0.0223. This suggests that VMD and SSA improve model performance.

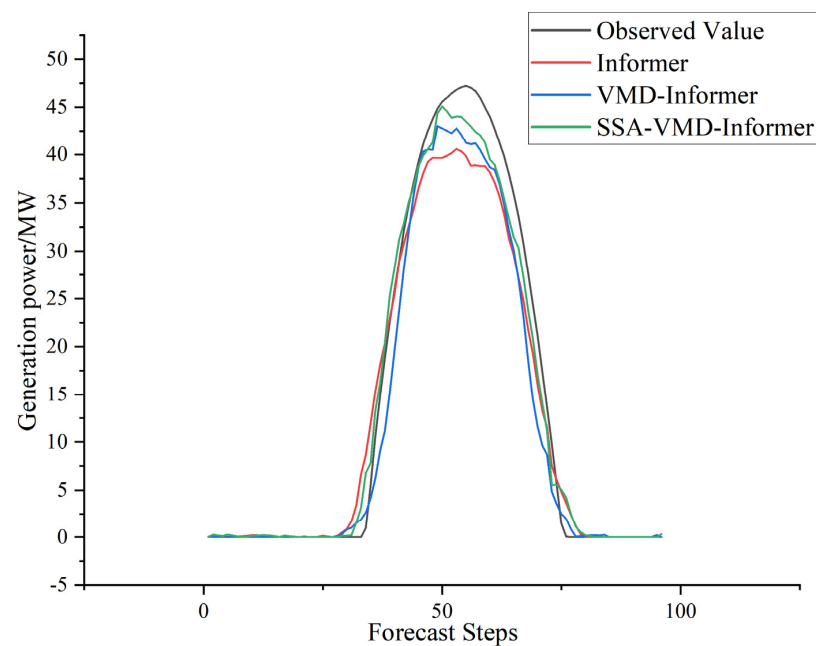


Figure 11. VMD and SSA ablation study forecast curves.

Table 3. VMD and SSA ablation study comparison of evaluation indicators.

Model	RMSE	MAE	SDE	R ²
Informer	3.5188	2.2589	3.30	0.9635
VMD-Informer	3.5021	2.1648	2.94	0.9639
SSA-VMD-Informer	2.1601	1.3900	2.10	0.9862

4.5. Comparative Experiments with Other Models

To validate the superior predictive ability of the SSA-VMD-Informer model, the commonly used classical prediction models LSTM, BiGRU-attention, and Transformer model were selected for prediction comparison experiments.

To verify the model's stability, we chose two random days for PV power prediction: 24 April in spring and 20 October in autumn. By comparing the model's predictions on these two days with the actual observations, we are able to evaluate the model's performance under different time conditions. Specifically, if the model's prediction results on these two different days are able to better match the actual situation, it shows that the model is highly stable and reliable. Such a validation method can effectively reveal the robustness of the model in the face of different input data and ensure its feasibility in practical applications.

As can be seen from the prediction curves in Figure 12, on 24 April and 20 October, the SSA-VMD-Informer model's prediction curves align significantly better with the actual value curves compared to other models. The smaller error between the predicted and observed values of the SSA-VMD-Informer model during these two different time periods indicates that the model has higher accuracy and stability in capturing the trend of PV power changes. In contrast, the prediction curves of the other models exhibit large fluctuations and deviations, which indicate their deficiencies in fitting the actual data. Further, through the comparative analysis in Tables 4 and 5, we can clearly see that the prediction results of the SSA-VMD-Informer model on 24 April and 20 October are excellent in all error indicators. Specifically, the RMSE, MAE, and SDE of the model for these two dates are lower than those of other comparative models, and R² is higher than those of other comparative models, which indicates that the SSA-VMD-Informer model possesses higher accuracy and reliability in short-term PV power prediction.

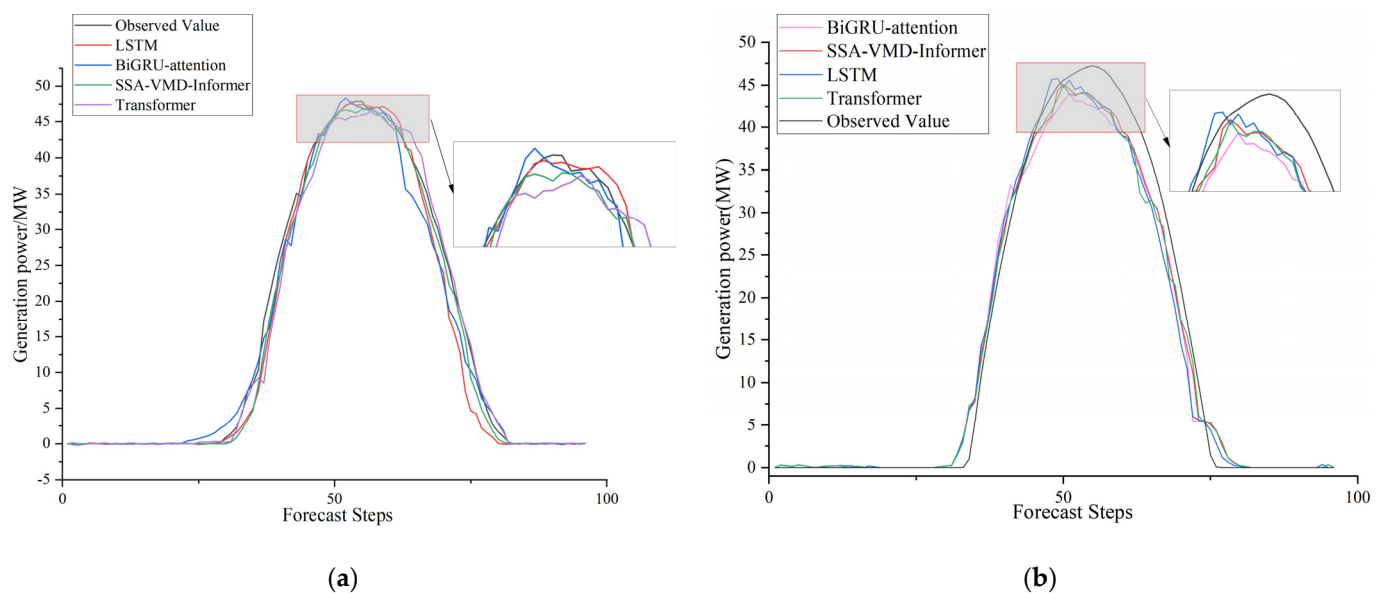
**Figure 12.** (a) Forecast curve for 24 April; (b) Forecast curve for 20 October.

Table 4. Assessment indicators for different models on 24 April.

Model	RMSE	MAE	SDE	R ²
SSA-VMD-Informer	1.3882	0.8310	1.14	0.9944
BiGRU-attention	2.0200	1.2202	1.90	0.9882
LSTM	2.3605	1.2592	2.11	0.9839
Transformer	1.7933	0.9642	1.75	0.9907

Table 5. Assessment indicators for different models on 20 October.

Model	RMSE	MAE	SDE	R ²
SSA-VMD-Informer	2.1601	1.3900	2.10	0.9862
BiGRU-attention	2.7036	1.7757	2.60	0.9785
LSTM	2.7031	1.6528	2.63	0.9785
Transformer	2.3627	1.4499	2.28	0.9835

5. Conclusions

To enhance the accuracy of short-term PV power predictions using neural network models, this paper introduces the SSA-VMD-Informer model. Validation experiments conducted on measured PV power data from a Xinjiang PV power plant over one year in 2019 show that the SSA-VMD-Informer model achieves higher prediction accuracy, leading to the following conclusions:

1. The Informer model in the SSA-VMD-Informer model is more effective for PV power prediction than time series models like LSTM and Transformer.
2. VMD optimised with SSA is fused with the Informer model to make the model more robust and adaptable to the complexities of real applications.
3. Adding seasonal and night-day factors to the Informer location coding can improve the accuracy of short-term PV power prediction.
4. Through comparative simulation verification, the RMSE, MAE, SDE, and R² of the PV power prediction results of this paper's model are better than those of other models, which effectively improves the PV power prediction accuracy.

Author Contributions: Conceptualisation, W.X. and D.L.; methodology, D.L.; software, W.X. and D.L.; validation, W.X., D.L. and W.D.; formal analysis, D.L.; investigation, D.L.; resources, W.X.; data curation, D.L.; writing—original draft preparation, D.L.; writing—review and editing, Q.W. and W.X.; visualisation, D.L.; supervision, W.X.; project administration, D.L.; funding acquisition, W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the National Natural Science Foundation of China (U1802271).

Data Availability Statement: The data analysed in this study are subject to the following licenses/restrictions: the raw data supporting the conclusions of this article will be made available by the authors, without undue reservation. Requests to access these datasets should be directed to D.L., 15774697105@163.com.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kabir, E.; Kumar, P.; Kumar, S.; Adelodun, A.A.; Kim, K.-H. Solar energy: Potential and future prospects. *Renew. Sustain. Energy Rev.* **2018**, *82*, 894–900. [\[CrossRef\]](#)
2. Stein, G.; Letcher, T.M. 15—Integration of PV Generated Electricity into National Grids. In *A Comprehensive Guide to Solar Energy Systems*; Letcher, T.M., Fthenakis, V.M., Eds.; Academic Press: Cambridge, MA, USA, 2018; pp. 321–332.
3. Wang, D. Combination Forecasting of Medium and Long-Term Power Load Based on Entropy Weight Method. Doctoral Dissertation, North China Electric Power University, Beijing, China, 2007.
4. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
5. Zhang, Q.; Ma, Y.; Li, G.; Ma, J.; Ding, J. Application of Frequency Domain Decomposition and Deep Learning Algorithms in Short-term Load and Photovoltaic Power Prediction. In Proceedings of the CSEE, Rome, Italy, 7–9 April 2019; Volume 39, pp. 2221–2230.

6. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. *ACM SIGKDD Explor. Newsl.* **2016**, *17*, 49–57.
7. Lang, Q.; Wang, X.; Wang, Y. Research on PV power output prediction based on GBDT and SVM. *J. Shanghai Electr. Power Univ.* **2023**, *39*, 275–280.
8. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
9. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using rnn encoder–decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078, 1724–1734.
10. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-Based Learning Applied to Document Recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
11. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
12. Geoffrey, E.H.; Simon, O.; Yee-Whye, T. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* **2006**, *18*, 1527–1554.
13. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 2–9 February 2021; pp. 11106–11115.
14. Wen, Q.; Zhou, T.; Zhang, C.; Chen, W.; Ma, Z.; Yan, J.; Sun, L. Transformers in time series: A survey. *arXiv* **2022**, arXiv:2202.07125.
15. Zhang, Y.; Wu, R.; Dascalu, S.M.; Harris, F.C. Multi-scale Transformer Pyramid Networks for Multivariate Time Series Forecasting. *arXiv* **2023**, arXiv:2308.11946. [[CrossRef](#)]
16. Cao, H.; Yang, J.; Zhao, X.; Yao, T.; Wang, J.; He, H.; Wang, Y. Dual-Encoder Transformer for Short-Term Photovoltaic Power Prediction Using Satellite Remote-Sensing Data. *Appl. Sci.* **2023**, *13*, 1908. [[CrossRef](#)]
17. Tao, K.; Zhao, J.; Wang, N.; Tao, Y.; Tian, Y. Short-term photovoltaic power forecasting using parameter-optimized variational mode decomposition and attention-based neural network. *Energy Sources Part A Recovery Util. Environ. Eff.* **2024**, *46*, 3807–3824. [[CrossRef](#)]
18. Wenjin, C.; Feng, Z.; Tongyan, Z.; Jun, Z.; Fengming, Z.; Dong, X.; Wei, R.; Meiya, S.; Qiang, S. A Photovoltaic Power Prediction Method Based on AFSA-BP Neural Network. *Zhejiang Electr. Power* **2022**, *41*, 7.
19. Li, Z.; Xu, R.; Luo, X.; Cao, X.; Du, S.; Sun, H. Short-term photovoltaic power prediction based on modal reconstruction and hybrid deep learning model. *Energy Rep.* **2022**, *8*, 9919–9932. [[CrossRef](#)]
20. Korkmaz, D.; Acikgoz, H.; Yildiz, C. A Novel Short-Term Photovoltaic Power Forecasting Approach based on Deep Convolutional Neural Network. *Int. J. Green Energy* **2021**, *18*, 525–539. [[CrossRef](#)]
21. Huang, Y.; Wu, Y. Short-Term Photovoltaic Power Forecasting Based on a Novel Autoformer Model. *Symmetry* **2023**, *15*, 238. [[CrossRef](#)]
22. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
23. Huang, N.E.; Wu, Z. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Rev. Geophys.* **2014**, *46*, 965. [[CrossRef](#)]
24. Xue, J.; Shen, B. A novel swarm intelligence optimization approach: Sparrow search algorithm. *Syst. Sci. Control Eng.* **2020**, *8*, 22–34. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.