

Article

Photovoltaic Power Forecasting Based on Variational Mode Decomposition and Long Short-Term Memory Neural Network

Zhijian Hou ¹, Yunhui Zhang ^{1,*}, Xuemei Cheng ^{1,2} and Xiaojiang Ye ^{2,*}¹ School of Mechanical and Electrical Engineering, Shenzhen Polytechnic University, Shenzhen 518055, China; zhijianhou@szpu.edu.cn (Z.H.); 22210010025@stu.wit.edu.cn (X.C.)² School of Optical Information and Energy Engineering, Wuhan Institute of Technology, Wuhan 430200, China

* Correspondence: zhangyunhui1106@szpu.edu.cn (Y.Z.); xjye@wit.edu.cn (X.Y.)

Abstract

The accurate forecasting of photovoltaic (PV) power is vital for grid stability. This paper presents a hybrid forecasting model that combines Variational Mode Decomposition (VMD) and Long Short-Term Memory (LSTM). The model uses VMD to decompose the PV power into modal components and residuals. These components are combined with meteorological variables and their first-order differences, and feature extraction techniques are used to generate multiple sets of feature vectors. These vectors are utilized as inputs for LSTM sub-models, which predict the modal components and residuals. Finally, the aggregation of prediction results is used to achieve the PV power prediction. Validated on Australia's 1.8 MW Yulara PV plant, the model surpasses 13 benchmark models, achieving an MAE of 63.480 kW, RMSE of 81.520 kW, and R^2 of 92.3%. Additionally, the results of a paired *t*-test showed that the mean differences in the MAE and RMSE were negative, and the 95% confidence intervals for the difference did not include zero, indicating statistical significance. To further evaluate the model's robustness, white noise with varying levels of signal-to-noise ratios was introduced to the photovoltaic power and global radiation signals. The results showed that the model exhibited higher prediction accuracy and better noise tolerance compared to other models.



Academic Editor: Santiago Silvestre

Received: 29 May 2025

Revised: 26 June 2025

Accepted: 4 July 2025

Published: 7 July 2025

Citation: Hou, Z.; Zhang, Y.; Cheng, X.; Ye, X. Photovoltaic Power Forecasting Based on Variational Mode Decomposition and Long Short-Term Memory Neural Network. *Energies* **2025**, *18*, 3572. <https://doi.org/10.3390/en18133572>

Copyright: © 2025 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/>).

Keywords: PV power forecasting; sample entropy; variational mode decomposition; long short-term memory; signal-to-noise ratio

1. Introduction

The increasing penetration of photovoltaic (PV) power into the electricity grid has made it crucial to forecast the output of PV systems accurately. Accurate and reliable forecasting methods for PV power generation can assist grid operators in effectively managing the variability and uncertainty of renewable energy affected by weather and other factors, improving grid stability and renewable energy integration efficiency [1].

Due to various factors such as meteorological conditions, equipment characteristics, data quality, and model technology, there are still challenges in predicting PV power generation. Physical models are based on the fundamental principles of physics and utilize equations to predict the output of PV systems based on the underlying physical and meteorological conditions [2]. They are interpretable and adaptable to extreme scenarios; however, they rely on high-precision parameters and are susceptible to data errors, complex calculations, and a limited ability to characterize complex systems [3]. Statistical

methods utilize historical data to predict power generation [4–7]. Although they are relatively straightforward to implement, they have limitations in capturing changing weather conditions and may lack good generalization ability due to the nonlinear relationship between PV power and its influencing factors [8]. Machine learning and deep learning techniques have gained significant popularity in recent years. Machine learning techniques can automatically learn complex relationships between input features and output variables from historical data, making them well-suited for handling the inherent uncertainty and non-linearity of PV power generation [9]. Common machine learning methods, such as traditional neural networks [10,11], support vector machine (SVM) [12], support vector regression (SVR) [13], and eXtreme Gradient Boosting (XGBoost) [14] are effective in accurately forecasting PV power generation. However, machine learning models often rely on shallow feature extraction and complex feature engineering, making it difficult to accurately capture the complex features of PV power generation, which limits the improvement of the prediction accuracy. Deep learning, a subset of machine learning, automatically extracts features of meteorological and PV power data through multi-layer networks, making it more efficient and accurate than traditional machine learning in nonlinear modeling and multi-source data fusion for PV power prediction. For example, Dhaked et al. [9] utilized the Long Short-Term Memory (LSTM) algorithm to forecast the PV production of a solar plant in Brazil, achieving superior prediction accuracy compared to Back Propagation Neural Network (BPNN) models. In another study by AitYusen et al. [15], three deep learning models, Bidirectional Long Short-Term Memory (Bi-LSTM), Convolutional Neural Network (CNN), and Gated Recurrent Unit (GRU), were employed to predict PV power generation based on historical data. The accuracy of the models was evaluated using the mean squared error, comparing scenarios with and without nighttime values. The daytime predictions achieved correlation coefficients ranging from 96.9% to 97.2%, indicating excellent accuracy. A study was conducted by Mellit [16], who utilized various deep learning neural networks, including LSTM, Bi-LSTM, GRU, and bidirectional GRU (Bi-GRU) to forecast PV power. The authors demonstrate that these methods offer very high accuracy. Another study was performed to predict PV power, utilizing vanilla Transformer, Informer, and Autoformer models. The authors suggest using vanilla Transformer with a 15 min interval for a 4 h prediction and a 60 min interval for a 24 h prediction [17].

Hybrid methods have gained popularity in recent years. The main idea behind these methods is to overcome the limitations of individual models and improve the overall accuracy and robustness of the forecasting system. For instance, Vandeventer [18] proposed a GASVM model to predict the short-term power output of residential PV systems. The SVM classifiers classify historical weather data, and GAs optimize the performance of these classifiers. The proposed model outperforms the traditional SVM model. Similarly, Pan et al. [19] used SVM to develop a PV forecasting model, with ant colony optimization (ACO) used to optimize the parameters of SVM. In another study, Lim [20] proposed a hybrid model combining a Convolutional Neural Network (CNN) for weather condition classification and Long Short-Term Memory (LSTM) to learn power generation patterns based on weather conditions. Tested on PV power output data from a Busan, South Korea power plant, this model achieved accurate power generation predictions. In a study by Zhen et al. [21], a Bi-LSTM model was proposed to forecast the output of a PV power plant, and a genetic algorithm was used to optimize the structure and parameters of Bi-LSTM to obtain the best performance. In another study by Niu et al. [22], the original power sequence was decomposed using the complete ensemble empirical mode decomposition (CEEMD) algorithm, and the forecasting accuracy of BPNN (Back Propagation Neural Network) was improved by the particle swarm optimization algorithm. This method was tested in a PV power plant and proved to be a promising approach for PV power

generation forecasting. Li et al. [23] proposed a hybrid model for the PV power prediction based on CNN-GRU, which combines the spatial feature extraction capability of CNN with the temporal dynamic modeling advantage of GRU to achieve the accurate prediction of PV power. In a study by Liu [24], the authors proposed a hybrid encoder–decoder architecture that combines a Transformer encoder and BiLSTM decoder, which not only leverages the advantages of Transformer in extracting multivariate spatiotemporal features but also enhances bidirectional temporal dynamic modeling through BiLSTM. Experiments using the Alice Springs dataset in Australia showed that this model outperformed methods like GRU-CNN, GRU, and LSTM across different seasons and weather scenarios. Abid et al. [25] proposed a model that combines a spatiotemporal attention mechanism with CNN and a Bi-LSTM network. This approach combines the spatial feature extraction of CNN with the temporal modeling ability of Bi-LSTM, enabling accurate predictions of spatiotemporal sequences. Experimental results show that this model outperforms three leading methods due to its effective integration of spatio-temporal features. Tang et al. [26] proposed a short-term PV hybrid prediction model, which generates smooth intrinsic modal function components through the complete ensemble empirical mode decomposition (CEEMDAN) algorithm, extracts features from meteorological data, and captures multivariate frequency information based on an iTransformer network with an attention mechanism. Validated with data from the Desert Knowledge Australia Solar Centre, the model showed good applicability.

Hybrid methods improve the prediction performance and have better robustness than individual models because they can better handle uncertainty and variability. They typically combine machine learning algorithms or integrate them with other approaches. However, hybrid methods still face challenges that highlight the potential for improvement. The choice of the hybrid method depends on the specific data and requirements of the application, as no single method fits all scenarios. Further research is essential to maximize their potential in PV power prediction.

This paper proposes a hybrid model that utilizes Variational Mode Decomposition (VMD) to tackle the non-stationarity of PV power. The model incorporates a Long Short-Term Memory (LSTM) neural network to develop a one-hour-ahead prediction model for PV power. The effectiveness and accuracy of the model are tested on the 1.8 MW Yulara PV power station in Australia and compared with other benchmark models through a simulation. The main contributions of the paper are as follows:

- Optimizing the modal number of VMDs using sample entropy to avoid over-decomposition or the insufficient decomposition of VMD signals.
- For the first time, we construct the residual components obtained from VMD decomposition as independent sub-models and form a multi-source signal aggregation architecture with modal sub-models, breaking through the neglect of residual information in traditional methods. It not only utilizes modal components to improve the accuracy of stationary signal prediction but also captures sequence trend features through residual sub-models.
- By fusing the features of modal components, meteorological variables, and first-order differences in meteorological variables, the input features of the sub model are constructed to enable the model to more accurately capture the dynamic changes of a PV power time series, breaking through the limitations of traditional feature selection models in characterizing nonlinear changes.
- The model validation quantifies errors through metrics such as MAE, RMSE, and R^2 and combines a paired t -test to verify statistical significance. This not only achieves a numerical description of the prediction accuracy but also overcomes the limita-

tion of traditional methods that cannot distinguish random errors and confirms the advantages of the proposed model from a statistical perspective.

- Evaluate the generalization ability of the PV power prediction model by analyzing the impact of different noise signals and signal-to-noise ratios.

The rest of the paper is organized as follows: Section 2 provides a detailed explanation of the dataset and systematically elaborates on the main theoretical foundations relevant to the proposed method. Section 3 is structured to report the results and discussion, while Section 4 concludes the paper.

2. Materials and Methods

2.1. Data

This study uses the publicly accessible 1.8 MW Yulara PV Power Station dataset from the Australian Renewable Energy Agency (ARENA) [27] to evaluate and validate the proposed model, ensuring research reproducibility. The goal is to predict PV power for one hour ahead using historical data recorded at five-minute intervals from 1 January 2017 to 31 December 2018. To ensure accuracy, the dataset was cleaned, removing variables with significant missing data. Missing values were inferred from similar observations during adjacent periods or the same weather types. If a considerable number of records was lost during a specific time period, those periods were excluded from the training dataset. Overall, the dataset contains relatively few instances of continuous missing records, which has a minimal impact on its size and availability. Outliers were treated as missing values.

PV power plants do not generate any electricity at night, and the inclusion of a large amount of zero-output-power data in the modeling process will have a significant impact on the accuracy of PV power prediction, making it unsuitable for practical use. To ensure practicality, this paper selects data between 7:00 a.m. and 7:00 p.m. for analysis and resamples the historical data at one-hour intervals.

The final dataset includes active power and meteorological data such as the weather temperature, weather relative humidity, global horizontal radiation, diffuse horizontal radiation, wind direction, wind speed, weather daily rainfall, air pressure, and hail accumulation.

2.2. Variational Mode Decomposition

PV power is affected by various factors such as solar radiation, temperature, and cloud cover changes, and its output power signal exhibits significant nonlinear and non-stationary characteristics. Directly using this raw data for modeling and forecasting is challenging. Therefore, it is essential to use appropriate data processing methods to obtain more helpful information.

Variational mode decomposition (VMD), as an efficient processing method for nonlinear and non-stationary signals [28], can effectively separate different frequency components in signals, extract fluctuation features at different time scales, and capture the changing patterns in PV power. This provides clearer and more representative features for subsequent prediction models, thereby improving the accuracy and stability of predictions.

The core idea of VMD is to construct and solve variational problems [29]. Assuming that the original signal $f(t)$ is decomposed into K components $u_k(t)$, the following conditions must be met: (1) $u_k(t)$ is the modal component with a finite bandwidth and a central frequency, and the sum of estimated bandwidths for each modality is minimized; (2) the sum of all modes is equal to the original signal.

By introducing α as the penalty parameter and $\lambda(t)$ as the Lagrange multiplier, VMD transforms the constrained variational problem into an unconstrained variational problem. It can be expressed as follows:

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

where $L(u_k, \omega_k, \lambda)$ is an augmented Lagrangian function, u_k and ω_k represent the k -th mode component and center frequency, $f(t)$ is the original signal; $\lambda(t)$ is the Lagrange multiplier, k is the index of mode component, K is the number of modal components to be decomposed, α is a penalty factor, $\sum_{k=1}^K$ represents the summation from $k = 1$ to K , $\|\cdot\|_2$ represents the L_2 norm, $\langle \cdot, \cdot \rangle$ represents the inner product, $\delta(t)$ is the dirac delta function, ∂_t is the time derivative, j is the imaginary unit, and $*$ represents the convolution operation.

Using the Alternative Direction Method of Multipliers (ADMM) algorithm to solve Equation (1), for all $\omega \geq 0$, update and iterate each modal component $\hat{u}_k$, and center frequency ω_k and $\hat{\lambda}$ using Equations (2)–(4) until the iteration termination condition is met.

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i=1, i < k}^K \hat{u}_i^{n+1}(\omega) - \sum_{i=1, i > k}^K \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \quad (2)$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega} \quad (3)$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau \left[\hat{f}(\omega) - \sum_{k=1}^K \hat{u}_k^{n+1}(\omega) \right] \quad (4)$$

where n is the number of iterations. Initially, initialization starts from $n = 0$, and the value of n increases by 1 after each round of variable update; $\hat{u}_k(\omega)$, $\hat{f}(\omega)$, and $\hat{\lambda}^n(\omega)$ denote the Fourier transforms of $\hat{u}_i(t)$, $\hat{f}(t)$, $\hat{\lambda}(t)$, respectively. $(\cdot)^{n+1}$ represents the updated result after the n th iteration, while $(\cdot)^n$ corresponds to the current value before the update. As a noise tolerance, τ is used to control the convergence speed of constraint conditions during the iteration process.

The iteration termination condition is as follows:

$$\frac{\left\| \hat{u}_d^{n+1} - \hat{u}_d^n \right\|_2^2}{\left\| \hat{u}_d^n \right\|_2^2} < \varepsilon \quad (5)$$

where ε is the convergence value for the iteration.

2.3. Sample Entropy

Sample entropy is a metric that quantifies the complexity of a time series and was introduced by Richman and Moorman [30]. A smaller sample entropy value indicates a less complex time series with higher self-similarity. Conversely, a larger entropy value indicates a more complex signal with a greater probability of generating new patterns. The VMD method decomposes a time series into multiple intrinsic mode functions (IMFs) with different frequency scales and relative stationarity. These IMF components reflect the signal's different frequency and amplitude characteristics. However, the values of modal number K and penalty factor α can influence the decomposition effect. A lower value of K may cause some modal components to remain unrecognized, affecting the predictions' accuracy. Conversely, a higher value of K can increase the computational complexity and make the center frequencies of adjacent modal components too close.

This study introduces sample entropy into the VMD decomposition of PV power sequences. The main idea is that the multi-scale fluctuations in PV power require the decomposition method to capture the modal characteristics of a single physical process while avoiding blind parameter optimization. This method uses sample entropy to quantify the complexity of IMF. When the VMD parameter K is appropriate, the IMF corresponding to a single physical process exhibits low complexity (small sample entropy value), while multi-scale signal aliasing can lead to abnormally high sample entropy. This correlation provides a quantitative basis for the modal decomposition of PV power signals.

Compared to EMD's modal aliasing and black box optimization of the grid search, the advantage of this method lies in the following: (1) dynamically adjusting the K value through sample entropy to achieve the adaptive separation of physical processes such as irradiance fluctuations and cloud cover; (2) transforming parameter selection from "blind trial and error" to process decoupling; (3) the sensitivity of sample entropy to non-stationary signals naturally matches the time-varying characteristics of PV power, making the decomposition results both physically interpretable and predictive effective.

For a time series $\{x(1), x(2), \dots, x(N)\}$ composed of N data points, the algorithm for sample entropy is as follows [31]:

1. Sequence segmentation: form an m -dimensional vector based on the sequence number, i.e.,

$$X(i) = [x(i), x(i+1), \dots, x(i+m-1)], i = 1, 2, \dots, N-m+1 \quad (6)$$

where N is the length of the time series, i is the vector index, and m is the embedding dimension, generally chosen as 2.

2. Define the absolute value of the maximum difference between the elements corresponding to vectors $X(i)$ and $X(j)$ as the maximum distance between them, denoted as $d[X[i], X[j]]$:

$$d[X[i], X[j]] = \max\{|x(i+k) - x(j+k)|\}, i, j = 1, 2, \dots, N-m+1, k = 0, \dots, m-1, i \neq j \quad (7)$$

where j is the vector index, and k represents the offset of the elements within the vector, used to traverse the elements at each corresponding position in the vector.

3. Given a similarity tolerance $r = (0.1 \sim 0.25) \times SD$ to measure similarity, where SD is the standard deviation of the sequence, calculate the number of times $d[X[i], X[j]]$ is less than r for each i value, and calculate the ratio of this number to the total distance, denoted as follows:

$$C_i^m(r) = \frac{1}{N-m} \{\text{The number of } d[X(i), X(j)] < r\} \quad i = 1, 2, N-m+1, i \neq j \quad (8)$$

Calculate the average value of $C_i^m(r)$, to obtain the global similarity probability in m dimensions, denoted as $C^m(r)$:

$$C^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} C_i^m(r) \quad (9)$$

$C^m(r)$ represents the average probability that the distance between any two vectors in m -dimensional space is less than or equal to r , reflecting the self-similarity of the sequence in low dimensional space.

Increase the embedding dimension from m to $m+1$, and repeat the above steps to obtain $C_i^{m+1}(r)$ and $C^{m+1}(r)$.

4. Calculate sample entropy

To quantify the complexity differences in the time series at different scales, sample entropy is defined as follows:

$$\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left[-\ln \frac{C^{m+1}(r)}{C^m(r)} \right] \quad (10)$$

where $\ln$ represents the natural logarithm. But in reality, N cannot be infinite, but a finite value, so the estimated value of sample entropy is as follows:

$$\text{SampEn}(m, r) = -\ln \frac{C^{m+1}(r)}{C^m(r)} \quad (11)$$

From Equation (11), it can be seen that sample entropy quantifies the complexity and irregularity of a time series by measuring the probability ratio of similarity at different embedding dimensions m and $(m + 1)$.

2.4. Long Short-Term Memory (LSTM) Neural Network

Long Short-Term Memory (LSTM) is a special kind of Recurrent Neural Network (RNN) model that is mainly used to solve the vanishing and exploding gradient problems during long sequence training, and it can perform better in longer sequences than ordinary RNNs [32,33]. Therefore, it is widely used in prediction models [34,35]. The PV power sequence is nonlinear, time-varying, and uncertain. Using LSTM for PV power prediction can effectively capture the complex patterns and dynamic characteristics in the PV time series, improve prediction accuracy, and provide strong support for ensuring the stable operation of the power system.

The LSTM neural network is illustrated in Figure 1. The top part shows the time extension chain of the LSTM cell (represented by S), with each cell associated with a time instance. At the bottom of Figure 1, it can be seen that LSTM cells are controlled by three gates, namely the forget gate, input gate, and output gate, to control the cell state. These three gates determine the degree of retention of long-term memory flow, the degree to which new input information updates the cell state, and the degree to which short-term memory is presented.

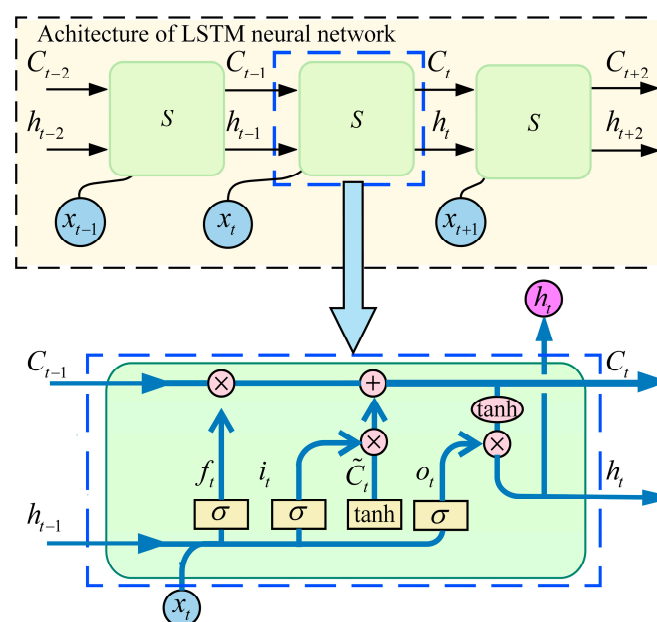


Figure 1. A generic LSTM neural network architecture.

The input of LSTM includes a single time step input vector x_t at the current time step, the previous hidden state h_{t-1} , and cell state C_{t-1} from the previous time step. The outputs are the hidden state h_t for the current time step (used as the prediction result or passed to the next layer) and the updated cell state C_t (passed to the next time step).

By introducing gating mechanisms and the cell state, LSTM can achieve the long-distance transmission of information [36]. The core process can be summarized as follows: firstly, the forget gate uses the sigmoid function to filter the effective information of the current input x_t and the previous hidden state h_{t-1} and generates the forget weight f_t to determine the retention ratio of the previous cell state C_{t-1} ; at the same time, the input gate generates the input weight i_t in the same way, which is multiplied by the candidate state $\tilde{c}_t$ generated by the hyperbolic tangent function (tanh) to control how much new information needs to be added to the cell state C_{t-1} in the current candidate state $\tilde{c}_t$. Next, by combining the outputs of the forget gate and input gate, the previous cell state C_{t-1} is updated to the new cell state C_t . Finally, the updated cell state C_t is normalized by tanh and then filtered by the output gate o_t to generate the hidden state h_t of the current time step as the output or input for the next time step.

2.5. Proposed Forecasting Strategy

This study proposes a hybrid method for PV power prediction that combines VMD and LSTM networks. This method effectively captures the temporal patterns of PV power and the influence of external meteorological factors, improving the prediction accuracy and providing a more reliable solution for predicting PV power.

Figure 2 shows the prediction process of the hybrid model, as proposed in this paper. It includes several steps:

1. Data cleaning: remove duplicate items and outliers from the raw data to ensure accuracy;
2. Feature screening: utilize the Random Forest method to evaluate the contribution of various meteorological data to PV output power, screen high-contribution key features, eliminate low-contribution features, and reduce the dimensionality of the input explanatory variables for the predictive model;
3. Data decomposition: the VMD method decomposes PV power into n intrinsic mode function (IMF) components and a residual (Res) component. Determine the optimal number of VMD modes based on the sample entropy of the summed modal components to avoid signal over-decomposition or insufficient decomposition;
4. Restructure the feature dataset: constructing a new feature dataset, including the components decomposed by VMD at time t , as well as the four meteorological parameters extracted from the second step, which are $GR(t)$, $DR(t)$, $WT(t)$, and $WR(t)$. To consider the impact of these parameters changing over time on the output, their first-order differences are added to the input parameters, which are $dif_GR(t)$, $dif_DR(t)$, $dif_WT(t)$, and $dif_WR(t)$;
5. Feature filtering: use the Random Forest method to identify the primary features that affect each modal component and residual at time $t + 1$ from the feature dataset established in the previous step. After extraction, the primary features are denoted as RF_1 , RF_2 , $\dots$, RF_n , and RF_{n+1} , respectively;
6. Model construction: use LSTM to establish multiple sub-models to forecast the modal components and residual of PV power. Each sub-model will take the primary features obtained in the previous step as input and output the forecast values of $IMF'_1(t + 1)$, $IMF'_2(t + 1)$, $\dots$, $IMF'_n(t + 1)$, and $Res'(t + 1)$, respectively;
7. Output reconstruction: the forecast results of each sub-model are obtained and added to obtain the final forecast value of PV power;

8. Performance evaluation: analyze the predictive performance of the model, including evaluating its performance on the validation dataset and evaluating its generalization ability using the test dataset.

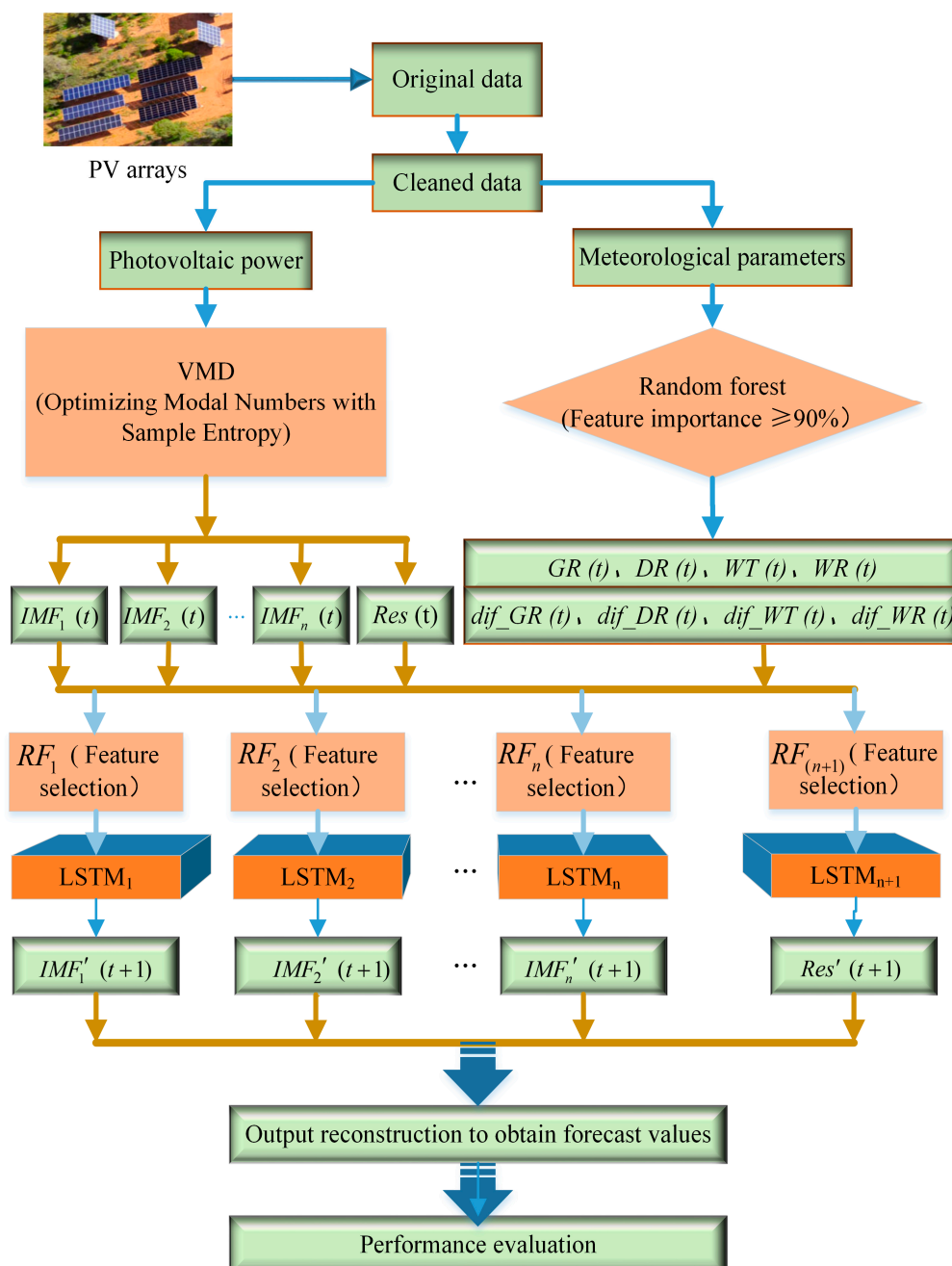


Figure 2. Flow diagram of the proposed model for PV power forecasting.

2.6. Performance Evaluation

The mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2) are important metrics used to evaluate the accuracy and reliability of PV power prediction models. The MAE measures the distance between observed and forecasted values, and the RMSE measures the degree of residual distribution. The MAE and RMSE have the same unit as the predicted variable (PV power), i.e., kilowatt (kW). In contrast, R^2 is a

dimensionless metric, ranging from 0 to 1, which measures the degree to which the model fits the observed values.

$$\left\{ \begin{array}{l} \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|^2} \\ \text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \\ \text{R}^2 = 1 - \frac{\sum_{i=1}^N |y_i - \hat{y}_i|^2}{\sum_{i=1}^N |y_i - \bar{y}|^2} \end{array} \right. \quad (12)$$

where N represents the number of observed samples; y_i and $\hat{y}_i$ represent the observed and forecasted values of the i -th sample, respectively; and $\bar{y}$ is the average of all observed values.

The MAE, RMSE, and R^2 , as quantitative indicators of model error, can only reflect the average error magnitude and dispersion and cannot explain the statistical significance of performance differences. The paired t -test analyzes the distribution of error differences from multiple training sessions to determine whether the performance differences of the model are caused by random factors, providing a statistical basis for the performance comparison [37]. This study aims to use the proposed model as the target model and multiple other models as benchmark models. Error metrics such as the MSE and RMSE are calculated on the same test set, and paired sample t -tests are conducted to verify the significance of performance differences between the two types of models.

The core of the paired sample t -test is to evaluate the significance of these differences by calculating the difference values between paired samples. Its null hypothesis (H_0) sets the mean difference between two paired samples to 0; that is, there is no significant difference in the mean of the two measurements. The alternative hypothesis (H_1) assumes that the mean difference between two paired samples is not 0; that is, there is a significant difference in the mean of the two measurements caused by real factors rather than sampling errors or accidental factors.

For each benchmark model and target model, their error metrics form paired samples. The target model error sequence is $\{e_{k,j}\}_{j=1}^n$, and the benchmark model error sequence is $\{e_{i,j}\}_{j=1}^n$ ($i \neq k$), where j represents the j -th sample, $e_{k,j}$ is the error of the target model (indexed by k), and $e_{i,j}$ is the error of the benchmark model (indexed by i , $i \neq k$). The calculation process of the paired t -test is briefly expressed as follows:

1. For each of the n paired observations, calculate the error difference between the target model and the benchmark model:

$$d_j = e_{k,j} - e_{i,j}, \quad j = 1, 2, \dots, n, \quad (13)$$

where d_j is the error difference for the j -th sample;

2. Calculate the mean and standard deviation of the error difference:

$$\left\{ \begin{array}{l} \bar{d} = \frac{1}{n} \sum_{j=1}^n d_j \\ s_d = \frac{\sum_{j=1}^n (d_j - \bar{d})^2}{n-1} \end{array} \right. \quad (14)$$

where $\bar{d}$ is the mean of the error difference; s_d is the standard error of the mean difference;

3. Calculate the t -statistic:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}. \quad (15)$$

4. Compare the t -statistic with the t -distribution to derive the p -value, and then, assess the significance of the performance difference between the target model and the benchmark model. If the p -value is less than the preset significance level (such as 0.05), the null hypothesis H_0 is rejected, indicating that there is a significant difference in the mean between the two sets of data; that is, the performance difference between the models is not caused by sampling errors or accidental factors.

3. Results and Discussion

3.1. Experimental Condition

The simulation is implemented in Python 3.11 on a PC system with an Intel(R) Core(TM) i7-10700 CPU @ 2.90 GHz and 8 GB RAM. Python libraries are employed as essential tools to support the implementation and analysis processes [38–40].

3.2. Feature Screening

The meteorological dataset included various features, such as the weather temperature (WT), weather relative humidity (WR), global horizontal radiation (GR), diffuse horizontal radiation (DR), wind direction (WD), wind speed (WS), daily rainfall, air pressure, and hail accumulation. However, some of these features might be considered to have a minimal impact on PV power or could be deemed redundant.

To reduce the dimensionality of explanatory variables and identify the most informative features, Random Forest is employed for the feature extraction of meteorological variables, as it can automatically assess feature importance. Due to the relatively small variations in parameters such as daily rainfall, air pressure, and hail accumulation, these features are not considered for selection during feature screening. During the initial feature selection phase, the aim is to find a balance between model performance and complexity and to minimize the risk of overfitting. To achieve this, the importance ranking of various meteorological variables is quantified using the Random Forest algorithm. Then, an optimal subset of features is constructed based on a selection strategy that ensures an accumulated contribution rate of 90%.

Figure 3 presents the contribution of each feature to PV power through the feature importance ranking. The vertical axis in the figure displays the features being evaluated, while the horizontal axis quantifies each feature's contribution to PV power. A higher value indicates that the feature plays a more important role in predicting PV power, while a lower value indicates that the feature is less important. The result shows that GR is the most critical feature, with an importance score of 0.74, highlighting its crucial role in the model's predictions. WT, with an importance of 0.156, though not as high as GR, is still a relatively important feature, indicating that it also provides valuable information for the model's prediction. DR, with a score of 0.042, contributes modestly compared to others. Lower scores indicate minor contributions from the remaining features. The global horizontal radiation, weather temperature, and diffuse horizontal radiation, which ranked highest and have an accumulated contribution rate exceeding 90%, are identified as important characteristics that influence PV power. Thus, they are selected to form the optimal feature subset.

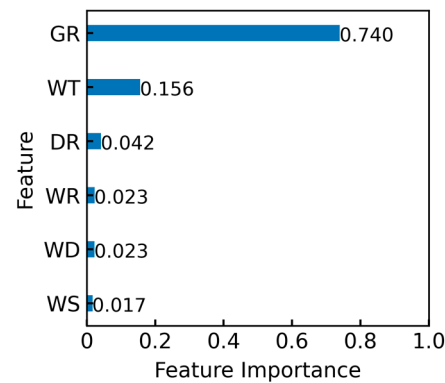


Figure 3. Ranking of feature importance in Random Forests.

After initially screening features using the Random Forest method, only those with significant contributions to the predictions were selected and retained, thereby significantly reducing the dimensionality of the dataset. This, in turn, reduces the complexity of subsequent processing. On this basis, the reduced feature set will be used for the selection of input variables to train the final prediction model, which can improve the computational efficiency and interpretability without sacrificing model performance.

3.3. Data Decomposition

VMD decomposes a time series into multiple intrinsic mode functions (IMFs) with different frequency scales and relative stationarity. These IMF components reflect the signal's different frequency and amplitude characteristics. However, the values of modal number K and penalty factor α can influence the decomposition effect. A lower value of K may cause some modal components to remain unrecognized, affecting the predictions' accuracy. Conversely, a higher value of K can increase computational complexity and make the center frequencies of adjacent modal components too close.

As mentioned in Section 2.3, the increasing sample entropy represents the improvement in data complexity and unpredictability, which enables a direct quantification of the complexity of modal components. Therefore, sample entropy is used to evaluate the effectiveness of modal decomposition in this study. Specifically, it is achieved by calculating the sample entropy of the reconstructed signal obtained during VMD at different modal numbers and determining the optimal number of modes based on the inflection point of the sample entropy curve. This method combines computational efficiency and physical interpretability, effectively avoiding excessive or insufficient decomposition.

In this study, the original signal of PV power generation is first decomposed using VMD, resulting in a series of modal components $\{IMF_k(t)\}$ and a residual ($Res(t)$), where $k = 1, 2, \dots, K$. Subsequently, these modal components are summed as an aggregated signal, and the sample entropy of the aggregated signal is calculated to quantify its complexity and irregularity. The process of calculating sample entropy is outlined as follows:

1. Initialize parameters: set the embedding dimension $m = 2$ and similarity tolerance $r = 0.15 \times SD$ to measure similarity;
2. Form subsequences: form subsequences of length m using Equation (6);
3. Calculate the distance between vectors using Equation (7);
4. Calculate the probability that the distance between vectors is less than r according to Equations (8) and (9);
5. Increase the embedding dimension to $m + 1$ and repeat the above steps;
6. Calculate the sample entropy using Equation (11).

This study achieves the adaptive optimization of the optimal mode number K for VMD by constructing different combinations of K and α , based on the stabilization of the

sample entropy of the summed modal components $\{IMF_k(t)\}$. The range of K values is set from 1 to 15, and the α parameter is varied within the range of 500 to 9200, with a step size of 300. For each K value, the calculation steps are as follows:

1. VMD is performed on the original signal, sum up all the modal components obtained, and obtain an aggregated signal;
2. Calculate the sample entropy of the aggregated signal corresponding to different α values;
3. Select the minimum value from the sample entropy calculation results corresponding to different α values.

This process is repeated across all K values, yielding a series of minimum sample entropy values.

Figure 4 illustrates the trend of these minimum sample entropy values for different mode numbers K , where the horizontal axis represents the mode number of PV power time series decomposed via VMD, and the vertical axis corresponds to the minimum sample entropy value (dimensionless) of the aggregated signal. Data points represent minimum values of each K after scanning different α values. Initially, when K is small, the signal's frequency components and features may not be adequately captured through the VMD decomposition, resulting in a lower sample entropy. As K increases, the modal components can more comprehensively represent the signal's characteristics, leading to an increase in the minimum sample entropy. However, after a certain point ($K = 7$), additional modal components contribute less to the signal's overall complexity, causing the minimum sample entropy tending to stabilize.

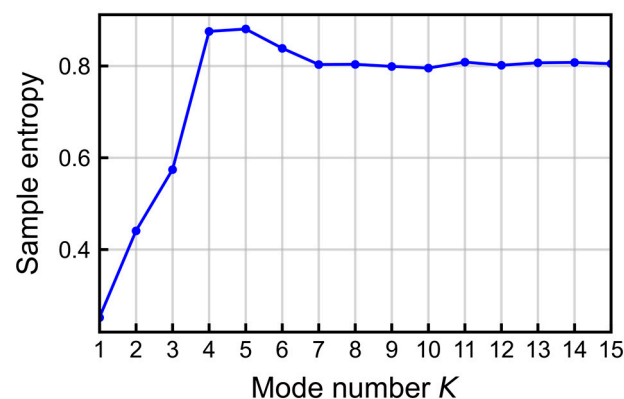


Figure 4. Sample entropy vs. mode number plot.

The optimal K value is determined as the one where the minimum sample entropy stabilizes, indicating a balance between capturing the signal's essential features and avoiding excessive decomposition. Then, determine the penalty parameter α corresponding to the minimum sample entropy associated with the optimal K . The optimal modal number $K = 7$ and penalty parameter $\alpha = 1700$ are determined through the above process. This combination of parameters ensures that the VMD decomposition retains the detailed features of the signal while avoiding unnecessary interference from noise or nonlinear components.

Apply the parameter combination ($K = 7$, $\alpha = 1700$) to perform VMD on the PV power time series in 2017. To present this more clearly, Figure 5 shows the VMD results from 7:00 a.m. on 1 October, to 7:00 p.m. on 31 December 2017 (a total of 1183 sample points). From top to bottom, they are the time-domain waveform of the original signal, each modal component subgraph, and the residual diagram in sequence. The horizontal axis represents the sample point index, and the vertical axis represents the amplitude, with the unit of kW. Each modal component subgraph corresponds to an IMF component, with the first

subgraph for IMF1, the second for IMF2, and so on. The residual diagram displays the difference between the original signal and the sum of all modal components, reflecting the residual information not captured via VMD.

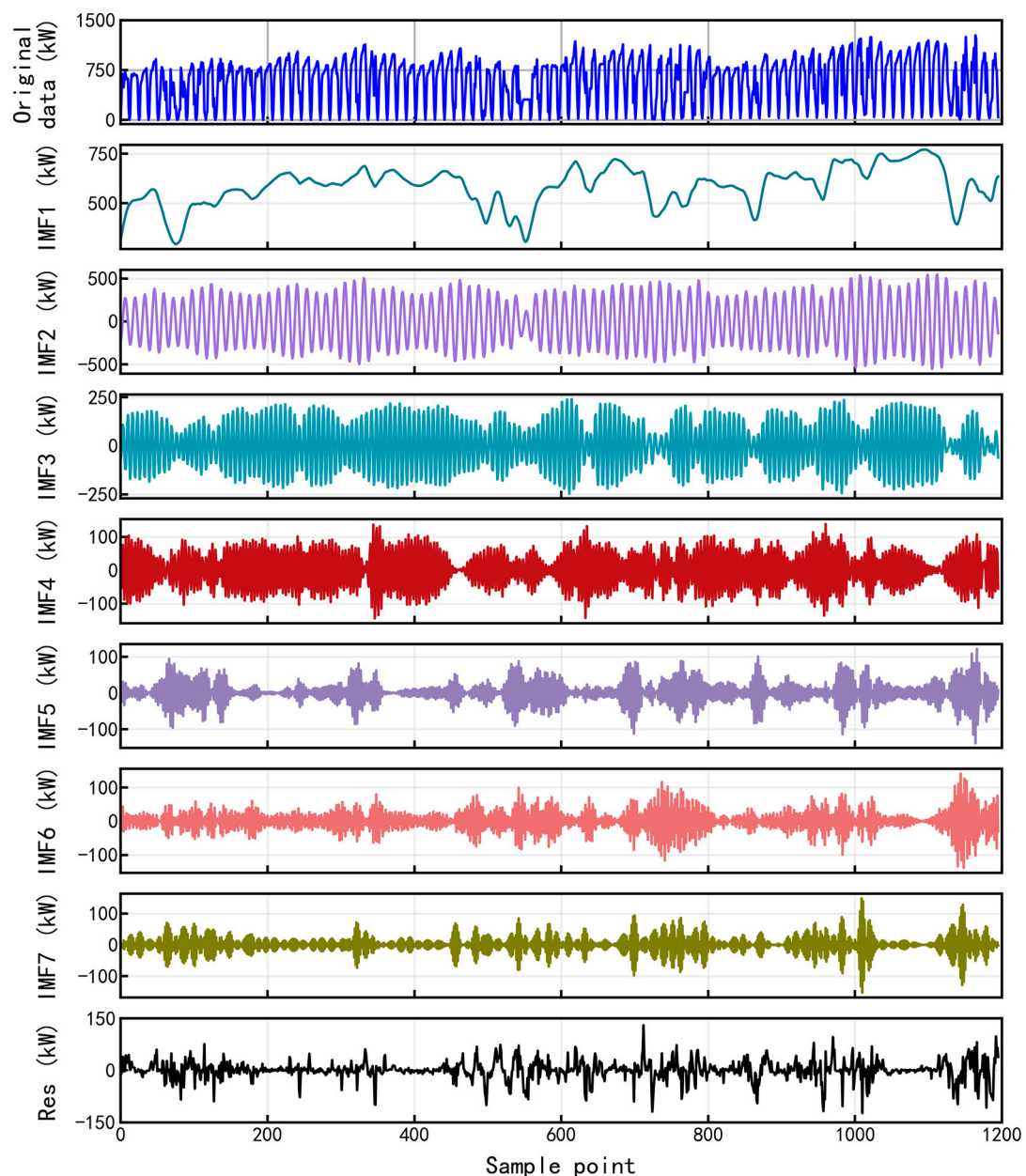


Figure 5. VMD decomposition results in 2017 (from 1 October 2017 to 31 December 2017).

As can be seen from Figure 5, the IMF1 component changes gently and represents the trend component, reflecting the overall trend of the PV power time series; IMF2 and IMF3 components have reasonable regularity and apparent periodicity. The components IMF4–IMF7 represent the fluctuation component, reflecting the random fluctuation details of the curve. The decomposed sequence presents better stationarity and more concentrated similar details compared to the original sequence, which can reduce the difficulty of modeling.

3.4. Input and Output of the LSTM Model

The performance of the forecast models greatly relies on its input and output. If all the information in a time series is mapped into an output vector nonlinearly, it couples

different scales and dimensions of information, potentially leading to noise sensitivity and affecting the forecast performance.

In Section 3.3, the optimal number of modes for the VMD of PV power signals has been determined to be 7. After VMD of the PV power signal at time t , the modal components and residuals at that time are obtained, denoted as $IMF_1(t)$, $IMF_2(t)$, $\dots$, $IMF_7(t)$, and $Res(t)$, respectively. Then, LSTM was used to construct submodels for predicting modal components ($IMF_1, IMF_2, \dots, IMF_7$) and residuals (Res) at time $t + 1$. The outputs of each submodel correspond to the values of each component at time $t + 1$.

The components obtained using the VMD technique may influence each other, so it is crucial to consider these interactions while selecting input variables for the submodels. Specifically, the following steps can be followed to determine the input variables for each submodel:

1. Reconstruct the feature dataset

The main meteorological parameters influencing PV power were obtained through the feature screening described in Section 3.2. These meteorological parameters, together with their first-order differential signals and the VMD components of PV power, were combined to form a new feature set. The first-order differences in these meteorological parameters indicate the rate of change of these parameters over time.

A more specific description of the new feature variables is shown in Table 1.

Table 1. Feature variable description.

Number	Symbol	Description
1	$IMF_1(t), IMF_1(t+1)$	Modal component IMF1 of PV power at time t and time $(t+1)$
2	$IMF_2(t), IMF_2(t+1)$	Modal component IMF2 of PV power at time t and time $(t+1)$
3	$IMF_3(t), IMF_3(t+1)$	Modal component IMF3 of PV power at time t and time $(t+1)$
4	$IMF_4(t), IMF_4(t+1)$	Modal component IMF4 of PV power at time t and time $(t+1)$
5	$IMF_5(t), IMF_5(t+1)$	Modal component IMF5 of PV power at time t and time $(t+1)$
6	$IMF_6(t), IMF_6(t+1)$	Modal component IMF6 of PV power at time t and time $(t+1)$
7	$IMF_7(t), IMF_7(t+1)$	Modal component IMF7 of PV power at time t and time $(t+1)$
8	$Res(t), Res(t+1)$	Residual of PV power at time t and $(t+1)$
9	$GR(t)$	Global horizontal radiation at time t
10	$DR(t)$	Diffuse horizontal radiation at time t
11	$WT(t)$	Weather temperature at time t
12	$Dif_GR(t)$	The first-order difference of global horizontal radiation at time t
13	$Dif_DR(t)$	The first-order difference of diffuse horizontal radiation at time t
14	$Dif_WT(t)$	The first-order difference of weather temperature at time t

2. Feature Filtering

Using the Random Forest method, important features with high contribution to the target variable are selected from the feature set generated in the previous step, achieving feature dimensionality reduction and improving the model efficiency and generalization ability. If the total importance of the top-ranked features is over 90%, the bottom-ranked features are excluded. Thus, the optimal feature sets for each LSTM submodel are constructed, represented as $RF_1, RF_2, \dots, RF_7$, and RF_8 , respectively.

3. Feature scaling

Data normalization is an essential method of feature scaling used for data preprocessing. Normalization aims to remove the effect of dimensionality between various features and ensure comparability between them. This study implements the Z-score normalization

method to normalize all the data. This method maps the data to a distribution with a mean of 0 and a standard deviation of 1. Equation (16) is used to achieve Z-score standardization:

$$x_{scale} = \frac{x - \mu}{\sigma}, \quad (16)$$

where x denotes the original data, x_{scale} is the corresponding normalized value, μ is the mean of the original data, and σ is the standard deviation of the original data.

3.5. Comparison of Forecast Performance with Benchmark Models

3.5.1. Dataset Splitting

The data from 2017 will be used for both model training and validation purposes, with 2/3 allocated as training samples and the remaining 1/3 as testing samples. The training set is used to train the model by adjusting the model's parameters to fit the data. The validation set is used to evaluate the model's performance during the training process, aiding in the adjustment of hyperparameters and the prevention of overfitting. The data collected between 20 July 2018 07:00 and 17 October 2018 19:00 will be used as the test dataset to evaluate the performance of the trained model on unseen data, providing an assessment of its generalization capability. The time window is shown in Table 2.

Table 2. Dataset for the proposed method.

Dataset	Time Window
Training Dataset	1 January 2017 07:00–2 September 2017 19:00
Validation Dataset	3 September 2017 07:00–31 December 2017 19:00
Test Dataset	20 July 2018 07:00–17 October 2018 19:00

3.5.2. Model Overview

To verify the effectiveness of the proposed hybrid model, the results of its one-hour-ahead forecast of PV power will be compared with different benchmark models. The benchmark models include the Persistence model, traditional machine learning models (SVR, XGboost, BP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and its variants (LSTM, GRU, Bi-LSTM), the self-attention mechanism model (Transformer), or hybrid models that combine the advantages of different models, such as CNN-GRU, CNN-Transformer, and GRU-Transformer. Each model traverses hyperparameter values through the grid search method to select the optimal combination of hyperparameters and ensure the effectiveness of the simulation.

Each benchmark model's input comprises the meteorological parameters GR, DR, and WT. The proposed model includes multiple sub-models, which undergo 200 training iterations with a time step of 13 and a batch size of 32. The input variables of each sub-model were obtained through feature reconstruction and filtering, as described in Section 3.4.

Given that the training model utilizes nearly a year's diverse seasonal data for VMD decomposition, it offers comprehensive signal information while precisely capturing local signal characteristics. This ensures that the decomposition results accurately reflect the signal's true nature. Consequently, during the model testing phase, the VMD process is repeated using the consistent time window, the number of modes, and the penalty parameter. This entire VMD procedure is repeated at each online time step. As new data points are collected, the time window slides forward, ensuring that the modal components decomposed via VMD are continuously updated. This dynamic adaptation allows the prediction model to respond effectively to changes in the data over time.

3.5.3. Performance Testing

This section will compare the proposed model with 13 benchmark models. Table 3 compares the forecast accuracy of different forecast models by evaluating three evaluation metrics: MAE, RMSE, and R^2 for the test dataset. The benchmark models were labeled with IDs (Model 1 to Model 13), and the proposed is labeled as “Target Model”.

Table 3. Comparison of the forecast accuracy for different forecast models.

ID	Model	MAE (kW)	RMSE (kW)	R^2
Model 1	Persistence	139.847	198.897	0.544
Model 2	BP	87.478	129.500	0.807
Model 3	SVR	101.779	142.386	0.731
Model 4	XGBoost	84.656	125.518	0.818
Model 5	CNN	87.915	129.816	0.806
Model 6	RNN	92.178	132.319	0.798
Model 7	GRU	86.760	132.452	0.798
Model 8	LSTM	86.203	132.479	0.798
Model 9	Bi-LSTM	93.386	133.296	0.795
Model 10	CNN-GRU	72.418	110.599	0.838
Model 11	Transformer	70.459	110.352	0.839
Model 12	CNN-Transformer	65.851	104.29	0.856
Model 13	GRU-Transformer	68.745	105.926	0.851
Target Model	Proposed	63.480	81.520	0.923

In Table 3, the Persistence model performs poorly in performance metrics such as the MAE, RMSE, and R^2 and may not meet the demand for high-precision prediction. The forecast performance metrics of RNN, LSTM, GRU, and Bi LSTM are relatively close. Compared to other benchmark models, the proposed model demonstrates significant advantages in prediction performance, characterized by lower MAE values (63.480 kW for the test dataset) and RMSE values (81.520 kW for the test dataset), indicating that its predictions are close to the actual data. Furthermore, the high R^2 values (0.923 for the test dataset) suggest that the model captures a significant proportion of the data’s variability, demonstrating its ability to effectively extract information from time series data and provide more accurate prediction results.

Compared with the single LSTM model, the proposed model has significantly better evaluation metrics and can achieve better prediction results. For the test dataset, the MAE decreased by 26.4%, the RMSE decreased by 38.5%, and R^2 increased by 15.7%. This indicates that the proposed model combines the advantages of modal decomposition and LSTM, utilizing the first-order differential processing of meteorological data to provide better input features for LSTM. This enables LSTM to capture complex patterns in the time series more effectively, thereby improving the accuracy and stability of predictions. In addition, this combination method performs well in terms of generalization ability.

Although the MAE and RMSE can quantify the magnitude of prediction bias and R^2 can reflect the explanatory power of the model based on the data, these metrics cannot reveal the statistical significance of observed differences. To address this, we validated performance disparities via paired t -tests: first, optimal hyperparameters for each model were determined through a grid search; then, models with optimal hyperparameters were independently trained 50 times using distinct random seeds to ensure independent initialization. Error metrics (MAE and RMSE) were collected from the same test set, forming a paired t -test dataset with 50 samples. Increasing training repetitions reduces randomness, enabling a more accurate estimation of mean errors and narrower confidence intervals to better reflect true error levels and model discrepancies.

Figure 6 shows the performance differences between the target model and benchmark models (Model 1 to Model 13) in terms of MSE and RMSE metrics. The left subgraph illustrates the mean of MSE differences (target model-benchmark models) across multiple trials, while the right subgraph shows the corresponding mean differences for the RMSE. The vertical axis of the figure shows the benchmark model numbers, which correspond to paired samples of error metrics (such as MAE, RMSE) between each benchmark model and the target model. The horizontal axis represents the mean difference in these error metrics, visually illustrating the magnitude of the error difference between the benchmark models and the target model. This mean difference is defined as the difference between the target model error and the benchmark model error. The negative value area (on the left) indicates that the target model is better. The error bars indicate the confidence intervals for the differences.

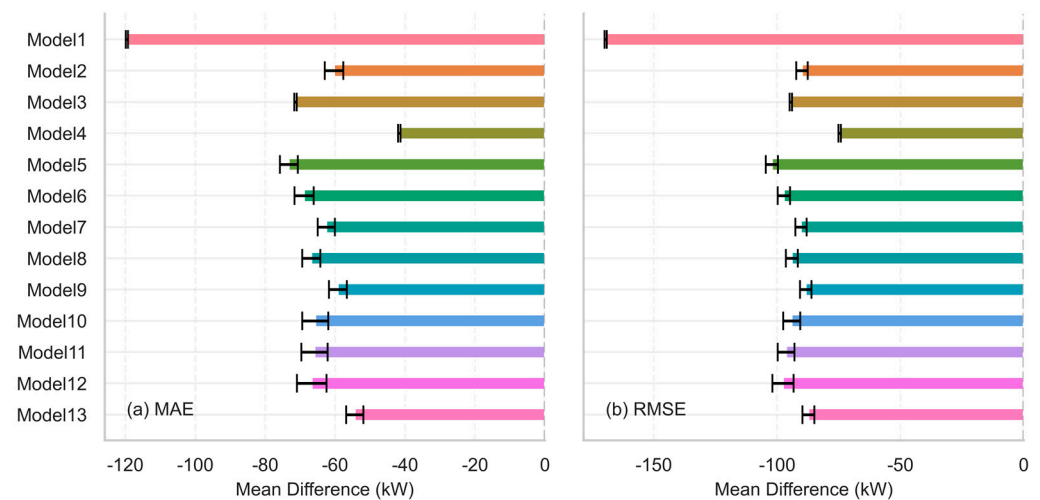


Figure 6. Comparison of model differences under different metrics.

The comparative results between the target model proposed and benchmark models showed that the mean differences were all negative for MAE and RMSE metrics, indicating that the target model had smaller errors and more accurate predictions. Moreover, all error bars were positioned to the left of the zero line, and the 95% confidence intervals of the differences did not include zero, which corresponds to a p -value of less than 0.05 in the paired t -test. This statistical evidence confirms that the target model's advantages were statistically significant. The performance superiority of the target model was not attributable to sampling errors or random fluctuations but truly reflected its substantial improvements in prediction accuracy and data fitting capability, demonstrating that the target model's average errors were significantly lower than those of all benchmark models ($p < 0.05$). The length of the bar reflects the magnitude of the error difference. The longer the bar, the greater the difference in error between the target and benchmark models.

From Figure 6, it can be seen that the mean difference range of MAE is from -119.556 kW to -41.66 kW, and the mean difference range of RMSE is from -169.515 kW to -74.638 kW, indicating that the prediction accuracy difference between the target model and each benchmark model is quite significant. Among them, the bar chart length of Model 1 is the longest in both metrics, and its MAE and RMSE differences significantly deviate from zero, reflecting that the benchmark model has a significant performance disadvantage compared to the target model. Although the original error values of Model 4 may not be optimal, the paired t -test results indicate that its bar length is the shortest and the confidence interval for error differences is narrow, indicating that the performance difference

between this model and the target model is relatively small and less affected by random fluctuations, with high statistical significance.

The error bars of models 1 and model 3 are relatively short, indicating that they have small error fluctuations, strong prediction consistency, and are less affected by random factors in repeated experiments; a narrow confidence interval means that the true error estimate is more accurate, and the statistical significance of paired *t*-tests is more reliable; Although the original error is not optimal, the stable performance suggests that its differences from the target model are more likely due to algorithm characteristics rather than accidental fluctuations.

To clarify the superiority of the proposed model more clearly, data from 7:00 a.m. on 2 October 2018 to 7:00 p.m. on 5 October 2018 were selected to compare the results of the benchmark model and the proposed models. The selected data covers several scenarios of PV power changes, including stable changes, certain fluctuations, and significant fluctuations. Figure 7 visually displays the changes in the forecasted and observed values of different models, and the proposed model has significantly better forecast performance than other models. Due to the poor metrics of the Persistence model and SVR model, Figure 7 only shows the predictive performance of other models.

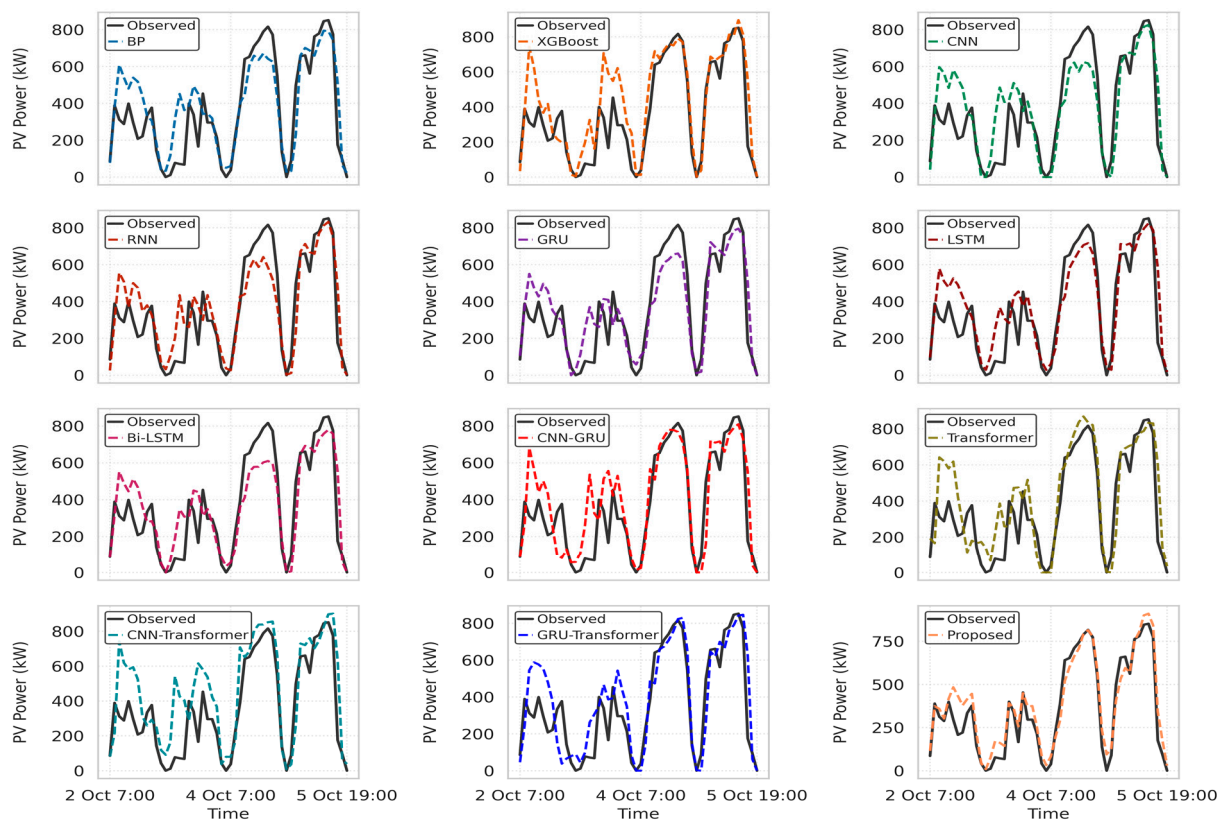


Figure 7. Comparison of forecast results for different models.

As shown in Figure 7, when the PV power changes steadily (from 7:00 a.m. on 4 October 2018 to 7:00 a.m. on 5 October 2018), almost all models effectively track the observed values, accurately reflecting variations in PV power. Whether it is stable or fluctuating PV power, the RNN, LSTM, GRU, and Bi-LSTM models achieve similar tracking performance. However, when the PV power fluctuates significantly, there is a notable discrepancy between the predicted values of various benchmark models and the observed value. Compared with other models, the proposed model has high consistency between forecasted and observed values, and it is robust to changes and noise in the test dataset. It

performs better in capturing the trends and peaks of PV power changes. This is because, after VMD decomposes the original time series, different frequency sub-sequences can be obtained, which reduces the adverse effects of noise on the original data, minimizes prediction errors caused by data anomalies or noise pollution, and ultimately improves the prediction accuracy. Meanwhile, the model introduces meteorological parameters and their differential signals as key explanatory variables, comprehensively reflecting the real-time status and trend in the meteorological system, making the model robust and able to maintain stable predictive performance under different meteorological conditions.

3.6. Robustness Testing of the Models

3.6.1. Robustness Testing Method Description

In the practical operation of PV power stations, sensor data are susceptible to environmental noise, posing challenges to models' reliability and generalization capability. A robustness testing approach has been adopted to comprehensively evaluate the models' performance under noisy conditions. Robustness testing is conducted to assess a model's ability to handle unknown or abnormal data by simulating the noisy environment encountered in actual operation, thereby judging its reliability and generalization capability in real-world applications. Specifically, noise is artificially added to the test data, and changes in model performance are observed.

The implementation steps of testing are as follows:

1. Baseline performance testing

The performance of the trained and validated model is evaluated using the original test data as a benchmark.

2. Noise data generation

White noise is added to critical signals in PV power station operation, such as PV power signals and global horizontal radiation signals; the intensity of the noise is adjusted to simulate different sources and levels of noise interference.

3. Model performance test and evaluation

Noise added to PV power signals: the model's performance in power prediction is evaluated by incorporating noise into the PV power signals. If the model maintains high prediction accuracy based on noisy data, it indicates strong robustness to noise.

Noise added to global horizontal radiation signals: similarly, noise is added to global horizontal radiation signals to evaluate the model's performance in power prediction.

Variation in signal-to-noise ratio (SNR): the SNR is a measure of the strength of a signal compared to the background noise, which quantifies the ratio of signal power to noise power. The higher the SNR, the smaller the noise, usually expressed in decibels (dB). Mathematically, it is defined as follows:

$$\text{SNR} = 10 \log_{10} \left(\frac{P_s}{P_n} \right), \quad (17)$$

where P_s represents the power of the signal, while P_n denotes the power of the noise. The logarithm is calculated to the base 10.

In robustness testing, varying the SNR allows for the simulation of different noise environments, which helps observe the model performance. If the model's performance gradually declines with increasing noise intensity but the decline is small, it indicates high tolerance to noise; conversely, it suggests the model is more sensitive to noise.

In summary, through the robustness testing approach, the performance of models can be comprehensively evaluated under various sources and intensities of noise, providing a solid foundation for their application in real-world PV power station operations.

3.6.2. Robustness Test Against PV Power Noise

The experiment involved adding various signal-to-noise ratios (SNRs), such as 10 dB, 30 dB, and 50 dB, to PV power signals. Noise was specifically introduced during two time periods: from 7:00 a.m. on 21 July 2018 to 7:00 p.m. 23 July 2018 and from 7:00 a.m. on 10 October 2018 to 7:00 p.m. on 12 October 2018. During these times, Gaussian white noise was tested on PV power signals. The performance metrics of the PV output prediction models were then evaluated for the periods from 7:00 a.m. on 10 October 2018 to 7:00 p.m. on 18 October 2018.

Table 4 presents a comparison of the performance metrics of the models when different levels of noise are added to the PV power signal.

Table 4. Performance comparison of model with different noise levels based on PV power signal.

Model	Metrics	SNR			Rate of Change (%) (from 30 dB to 10 dB)	Rate of Change (%) (from 50 dB to 30 dB)
		10 dB	30 dB	50 dB		
Persistence	MAE (kW)	189.623	178.833	178.353	6.033	0.269
	RMSE (kW)	253.826	243.344	243.09	4.308	0.104
	R ²	0.388	0.388	0.384	0.000	1.061
BP	MAE (kW)	108.276	93.588	93.384	15.694	0.218
	RMSE (kW)	157.916	135.126	135.173	16.866	−0.035
	R ²	0.763	0.811	0.809	−5.921	0.221
SVR	MAE (kW)	113.615	87.594	86.319	29.706	1.477
	RMSE (kW)	174.583	133.236	132.216	31.033	0.772
	R ²	0.711	0.816	0.818	−12.972	−0.149
XGBoost	MAE (kW)	117.179	94.177	92.997	24.424	1.269
	RMSE (kW)	173.949	136.546	134.876	27.392	1.238
	R ²	0.713	0.807	0.81	−11.716	−0.375
CNN	MAE (kW)	112.907	96.255	95.661	17.300	0.620
	RMSE (kW)	160.555	138.052	138.088	16.300	−0.026
	R ²	0.755	0.803	0.801	−5.946	0.228
RNN	MAE (kW)	105.353	94.609	95.356	11.356	−0.783
	RMSE (kW)	154.023	137.23	137.778	12.237	−0.398
	R ²	0.775	0.805	0.802	−3.796	0.409
GRU	MAE (kW)	105.951	89.168	88.986	18.822	0.204
	RMSE (kW)	161.348	136.368	136.312	18.318	0.041
	R ²	0.753	0.808	0.806	−6.803	0.189
LSTM	MAE (kW)	102.954	85.492	85.407	20.424	0.100
	RMSE (kW)	157.61	130.402	129.874	20.865	0.407
	R ²	0.764	0.824	0.824	−7.290	0.013
Bi-LSTM	MAE (kW)	112.627	96.912	96.544	16.216	0.381
	RMSE (kW)	159.944	137.624	137.448	16.218	0.128
	R ²	0.757	0.804	0.803	−5.857	0.150
CNN-GRU	MAE (kW)	80.435	78.299	78.192	2.728	0.137
	RMSE (kW)	120.622	117.711	117.637	2.473	0.063
	R ²	0.809	0.817	0.817	−0.979	0.000
Transformer	MAE (kW)	91.798	91.069	91.14	0.800	−0.078
	RMSE (kW)	129.681	128.517	128.609	0.906	−0.072
	R ²	0.779	0.781	0.781	−0.256	0.000

Table 4. Cont.

Model	Metrics	SNR			Rate of Change (%) (from 30 dB to 10 dB)	Rate of Change (%) (from 50 dB to 30 dB)
		10 dB	30 dB	50 dB		
CNN-Transformer	MAE (kW)	74.826	72.985	72.83	2.522	0.213
	RMSE (kW)	118.695	116.303	116.226	2.057	0.066
	R ²	0.815	0.821	0.821	−0.731	0.000
GRU-Transformer	MAE (kW)	95.245	92.742	92.681	2.699	0.066
	RMSE (kW)	136.996	133.814	133.731	2.378	0.062
	R ²	0.753	0.763	0.763	−1.311	0.000
Proposed	MAE (kW)	71.468	70.703	70.888	1.082	−0.261
	RMSE (kW)	91.671	87.035	87.218	5.326	−0.210
	R ²	0.92	0.922	0.921	−0.216	−0.111

The results showed that all models showed significant improvements in the MAE, RMSE, and R² values as the SNR increased from 10 dB to 30 dB. This trend suggests that as the noise level decreases, the models can predict PV output more accurately. This phenomenon is because the reduction in noise enables the models to more effectively capture the underlying patterns in the data, thereby optimizing the prediction performance. Conversely, as the SNR decreases from SNR = 30 dB to SNR = 10 dB, the MAE of the SVR model increased by 29.706%, the RMSE increased by 31.033%, and R² decreased by 12.972%. This indicates that when the noise level increases, the performance of the SVR model drops significantly, showing its high sensitivity to noise and low tolerance.

Compared with SVR, XGBoost has a slightly better adaptation to noise, but its performance changes are still relatively large compared to other models. During the process of SNR variation, the performance metrics such as the MAE, RMSE, and R² of GRU, LSTM, and BiLSTM exhibit significant synchronicity. Compared to the high sensitivity of SVR models to noise, GRU, LSTM, and BiLSTM models exhibit similar moderate tolerance characteristics to noise disturbances, indicating relatively stable predictive ability in complex noise environments.

When the SNR is further increased from 30 dB to 50 dB, the Persistence model, SVR model, and XGBoost model exhibit minor improvements, while others show inconsistent changes, with some even experiencing slight declines in performance. This suggests that different models have varying tolerances to noise, and further reductions in noise beyond a certain threshold are difficult to continuously optimize model performance.

When the SNR changes, CNN-GRU, Transformer, CNN-Transformer, and GRU-Transformer exhibit relatively stable performance, with small fluctuations in their MAE, RMSE, and R². However, the MAE, RMSE, and R² of these three models are still significantly higher than those of the proposed model.

Compared to other models, the proposed model exhibits the most stable performance metrics when SNR varies. From SNR = 30 dB to SNR = 10 dB, the MAE increased by only 1.082%, RMSE increased by 5.326%, and R² is decreased by 0.216%. When the SNR changed from 50 dB to 30 dB, the performance metrics also showed slight changes, with the MAE and RMSE slightly decreasing and R² slightly increasing. This demonstrates that the proposed model maintains good stability and adaptability at different noise levels, exhibiting a high tolerance to noise.

In summary, the proposed model possesses stronger generalization ability and maintains good performance under different noise levels.

3.6.3. Robustness Test Against Global Horizontal Radiation Noise

The robustness test for global radiation signals adopts testing conditions identical to those for PV power signals: adding Gaussian white noise with the same SNR during the same time period, and completing model performance testing during the same evaluation period.

Table 5 displays a comparison of the model's performance indicators when different levels of noise are added to the global horizontal radiation signal.

The results show that the performance of the models measured based on the MAE, RMSE, and R^2 will vary with the SNR of global horizontal radiation noise.

Under different levels of global horizontal radiation noise, the proposed model exhibits excellent stability and adaptive capability. Whether it is subtle interference in low-SNR environments or complex disturbances in high-SNR environments, its performance metrics, such as the MAE, RMSE, and R^2 , always fluctuate within a very small range. From SNR = 30 dB to SNR = 10 dB, the MAE increased by 0.271%, the RMSE increased by 0.338%, and the R^2 decreased by 0.058%. When the SNR changed from 50 to 30, there was almost no change in performance metrics. This indicates that the proposed model has a high tolerance to noise and can continuously output stable and reliable prediction results under non-ideal data conditions.

Table 5. Performance comparison of model with different noise levels based on global horizontal radiation signal.

Model	Metrics	SNR			Rate of Change (%) (from 30 dB to 10 dB)	Rate of Change (%) (from 50 dB to 30 dB)
		10 dB	30 dB	50 dB		
Persistence	MAE (kW)	157.911	157.911	157.911	0.000	0.000
	RMSE (kW)	217.325	217.325	217.325	0.000	0.000
	R^2	0.507	0.507	0.507	0.000	0.000
BP	MAE (kW)	92.816	93.333	93.409	−0.555	−0.081
	RMSE (kW)	135.027	135.216	135.219	−0.140	−0.002
	R^2	0.81	0.809	0.809	0.066	0.001
SVR	MAE (kW)	88.388	85.492	86.116	3.388	−0.724
	RMSE (kW)	133.275	131.692	132.104	1.202	−0.312
	R^2	0.815	0.819	0.818	−0.535	0.139
XGBoost	MAE (kW)	93.843	90.835	92.299	3.312	−1.587
	RMSE (kW)	130.866	132.237	133.786	−1.037	−1.158
	R^2	0.821	0.817	0.813	0.461	0.529
CNN	MAE (kW)	95.087	95.445	95.587	−0.375	−0.149
	RMSE (kW)	137.738	138.015	138.121	−0.200	−0.077
	R^2	0.802	0.801	0.801	0.099	0.038
RNN	MAE (kW)	96.284	95.383	95.455	0.944	−0.076
	RMSE (kW)	139.171	137.879	137.879	0.937	0.001
	R^2	0.798	0.801	0.802	−0.466	0.000
GRU	MAE (kW)	90.278	89.112	89.025	1.309	0.097
	RMSE (kW)	137.166	136.414	136.355	0.551	0.044
	R^2	0.804	0.806	0.806	−0.267	−0.021
LSTM	MAE (kW)	87.738	85.589	85.43	2.511	0.186
	RMSE (kW)	132.712	130.059	129.876	2.040	0.141
	R^2	0.816	0.823	0.824	−0.884	−0.060

Table 5. Cont.

Model	Metrics	SNR			Rate of Change (%) (from 30 dB to 10 dB)	Rate of Change (%) (from 50 dB to 30 dB)
		10 dB	30 dB	50 dB		
Bi-LSTM	MAE (kW)	97.531	96.566	96.522	0.999	0.046
	RMSE (kW)	138.59	137.529	137.474	0.771	0.040
	R ²	0.799	0.803	0.803	−0.381	−0.020
CNN-GRU	MAE (kW)	78.5	78.195	78.185	0.390	0.013
	RMSE (kW)	118.216	117.648	117.633	0.483	0.013
	R ²	0.815	0.817	0.817	−0.245	0.000
Transformer	MAE (kW)	92.572	91.225	91.158	1.477	0.073
	RMSE (kW)	131.465	128.73	128.63	2.125	0.078
	R ²	0.771	0.78	0.781	−1.154	−0.128
CNN-Transformer	MAE (kW)	73.112	72.854	72.819	0.354	0.048
	RMSE (kW)	116.457	116.219	116.218	0.205	0.001
	R ²	0.82	0.821	0.821	−0.122	0.000
GRU-Transformer	MAE (kW)	93.214	92.694	92.677	0.561	0.018
	RMSE (kW)	134.247	133.728	133.725	0.388	0.002
	R ²	0.761	0.763	0.763	−0.262	0.000
Proposed	MAE (kW)	71.124	70.932	70.932	0.271	0.000
	RMSE (kW)	87.58	87.285	87.285	0.338	0.000
	R ²	0.92	0.92	0.92	−0.058	0.000

When comparing the performance trends of models in change rates across different SNR ranges, the Persistence model's performance metrics (MAE, RMSE, R²) remain relatively consistent across all SNR levels. This consistency is due to the model's lack of responsiveness to noise variations. As the SNR decreased, the performance of models BP and CNN showed a nonlinear improvement. When the SNR decreases from 30 dB to 10 dB, the MAE of the CNN model decreases by 0.375%, the RMSE decreases by 0.2%, and R² increases by 0.099%; when the SNR decreases from 50 dB to 30 dB, the MAE decreases by 0.149%, the RMSE decreases by 0.077%, and R² also increases by 0.038%. This improvement can be attributed to the strong nonlinear correlation between noise and the signal, which allows the models to distinguish between the two more effectively through nonlinear transformations. In high-SNR scenarios, the presence of strong signals can lead to “saturation bias” during feature extraction. This results in an excessive reliance on a single feature, while neglecting complementary information from the noise. As a result, the models may better decouple mixed features in low-SNR environments.

Most models are more sensitive to noise variations at lower SNR levels. When the SNR is at a low level, changes in noise can cause fluctuations in metrics such as the MAE, RMSE, and R². As the noise levels increase, the performance metrics of the model show a trend of an increasing MAE and RMSE and decreasing R². For example, when the SNR decreases from 30 dB to 10 dB, the MAE of the SVR model increases by 3.388%, the RMSE increases by 1.202%, and the R² decreases by 0.535%. This may be due to noise interference weakening the model's ability to extract effective data features, resulting in a decrease in the prediction accuracy, an increase in overall deviation and extreme differences between predicted and actual values, and a decrease in data correlation. The MAE of the XGBoost model increases by 3.312%, the RMSE decreases by 1.037%, and R² increases by 0.461%, indicating that although it is affected by noise, the correlation between its predictions and observations has been enhanced, and improvements have been made in handling abnormal errors, and it is still sensitive to noise; the MAE of the LSTM model increases by 2.511%, the RMSE

increases by 2.040%, and R^2 decreases by 0.884%. This indicates that with increasing noise, the prediction accuracy of LSTM will decrease, and the correlation between predicted and observed values will also weaken; Other models exhibit intermediate levels of performance variation and varying sensitivities to noise.

4. Conclusions

With the increasing importance of renewable energy, PV power generation, as a clean and sustainable form of energy, is receiving increasing attention. However, the instability and intermittent nature of PV power generation pose particular challenges to the safe and stable operation of the power grid. Therefore, the accurate prediction of PV power is crucial for grid scheduling, energy planning, and the operation and management of PV power systems.

This study combines the advantages of VMD in signal multi-scale decomposition and LSTM in sequence modeling to construct a one-hour-ahead prediction model for PV power. This research is the first to independently construct VMD residual components as sub-models, forming a multi-source aggregation architecture with modal sub-models to address the traditional neglect of residuals. This design not only improves the prediction accuracy of stationary signals through modal components but also captures sequence trend features through residual sub-models. The feature set of the sub-model introduces the first-order differential signals of meteorological parameters to capture the dynamic changes in the PV power time series, enabling the model to maintain good adaptability in weather fluctuation scenarios. The superiority of the proposed model was validated through a thorough comparison with 13 benchmark models. The findings of the research are as follows:

- Compared with other benchmark models, the proposed model exhibits excellent environmental adaptability under different weather fluctuations and can capture real-time data trends of photovoltaic power generation. The test dataset shows that its MAE and RMSE are significantly reduced compared to the benchmark model, and the R^2 value is significantly improved, verifying the dual advantages of the model in prediction accuracy and robustness. The proposed model achieved a mean absolute error (MAE) of 63.480 kW, a root mean square error (RMSE) of 81.520 kW, and a coefficient of determination (R^2) of 0.923 based on the test dataset;
- Statistical validation of model superiority via a paired t -test, beyond traditional accuracy metrics. Using the proposed model as the target model and multiple other models as benchmarks, calculate the MAE and RMSE on the same test set; by conducting the paired t -test analysis on the MAE and RMSE between the target model and the benchmark model, the results showed that the mean differences in the MAE and RMSE were negative, and all error bars were located on the left side of the zero line. The 95% confidence interval does not contain zero, and the p -value of the paired t -test is less than 0.05, confirming that the average error of the target model is significantly lower than that of each benchmark model, and the accuracy advantage is significant;
- Through the investigation of the impact of different noise signals and SNR on PV power prediction models, it has been found that noise significantly affects their performance. As the SNR decreases, the prediction errors of most models tend to rise, as measured based on the MAE and RMSE. Furthermore, R^2 also tends to decrease with a lower SNR, suggesting a weaker correlation between the predicted and actual PV power outputs.
- Different models exhibit varying degrees of sensitivity to noise. The proposed model maintains stable predictive performance under various noise sources and different SNR levels, and its noise tolerance is relatively better than other methods. This feature

demonstrates the excellent robustness and strong environmental adaptability of the model when facing new data scenarios.

Although the superiority of the proposed model has been fully demonstrated, future research directions to enhance its practical application value are as follows:

- Study the dynamic VMD strategy and dynamic feature extraction during data updates to avoid information leakage and ensure the accuracy and robustness of model predictions;
- Expand the types and intensities of noise interference sources, conduct robustness testing against complex noise such as sensor drift and weather-induced fluctuations, and enhance the model's adaptability to complex scenarios by designing adaptive anti-interference mechanisms, further improving the model's generalization performance to real-world application environments.

Author Contributions: Z.H. was responsible for the conceptualization, methodology, software, formal analysis, data curation, visualization, funding acquisition, and writing—the original draft. Y.Z. was responsible for writing—reviewing and editing, methodology, software, data curation, validation, and visualization. X.C. was responsible for software, formal analysis, data curation, and visualization. X.Y. was responsible for data curation, software, investigation and writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the article processing charge (APC) were funded by the Department of Education of Guangdong Province, China, under Grant No. 2021ZDZX1106 for the Key Field Special Project in Higher Education (New Generation Information Technology).

Data Availability Statement: The data presented in this study are available on request from the corresponding authors.

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

References

1. Li, J.; Rao, C.; Gao, M.; Xiao, X.; Goh, M. Efficient Calculation of Distributed Photovoltaic Power Generation Power Prediction via Deep Learning. *Renew. Energy* **2025**, *246*, 122901. [\[CrossRef\]](#)
2. Zhi, Y.; Sun, T.; Yang, X. A Physical Model with Meteorological Forecasting for Hourly Rooftop Photovoltaic Power Prediction. *J. Build. Eng.* **2023**, *75*, 106997. [\[CrossRef\]](#)
3. Mayer, M.J. Benefits of Physical and Machine Learning Hybridization for Photovoltaic Power Forecasting. *Renew. Sustain. Energy Rev.* **2022**, *168*, 112772. [\[CrossRef\]](#)
4. Fara, L.; Diaconu, A.; Craciunescu, D.; Fara, S. Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models. *Int. J. Photoenergy* **2021**, *2021*, 6777488. [\[CrossRef\]](#)
5. Islam, M.S.S.; Ghosh, P.; Faruque, M.O.; Islam, M.R.; Hossain, M.A.; Alam, M.S.; Islam Sheikh, M.R. Optimizing Short-Term Photovoltaic Power Forecasting: A Novel Approach with Gaussian Process Regression and Bayesian Hyperparameter Tuning. *Processes* **2024**, *12*, 546. [\[CrossRef\]](#)
6. Miao, Z.; Shao, J.; Lin, B.; Hu, K. Prediction of Stable Power Generation Period of Photovoltaic Power Station Based on Time Series ARIMA Model. *J. Phys. Conf. Ser.* **2025**, *2993*, 012003. [\[CrossRef\]](#)
7. Yang, X.; Yang, Y.; Meng, L.; Zhao, Y. Spatio-Temporal PV Power Forecasting Considering the Time-Shift Correction and the Information Fusion Strategy of Multi-Stations. *ISA Trans.* **2023**, *139*, 376–390. [\[CrossRef\]](#)
8. Fu, J.; Sun, Y.; Li, Y.; Wang, W.; Wei, W.; Ren, J.; Han, S.; Di, H. An Investigation of Photovoltaic Power Forecasting in Buildings Considering Shadow Effects: Modeling Approach and SHAP Analysis. *Renew. Energy* **2025**, *245*, 122821. [\[CrossRef\]](#)
9. Dhaked, D.K.; Dadhich, S.; Birla, D. Power Output Forecasting of Solar Photovoltaic Plant Using LSTM. *Green Energy Intell. Transp.* **2023**, *2*, 100113. [\[CrossRef\]](#)
10. Lin, P.; Peng, Z.; Lai, Y.; Cheng, S.; Chen, Z.; Wu, L. Short-Term Power Prediction for Photovoltaic Power Plants Using a Hybrid Improved Kmeans-GRA-Elman Model Based on Multivariate Meteorological Factors and Historical Power Datasets. *Energy Convers. Manag.* **2018**, *177*, 704–717. [\[CrossRef\]](#)
11. Feng, Z.; Liu, P.; Niu, W. Solar Power Generation Prediction Using Radial Basis Function Neural Network with Mode Decomposition and Cooperation Search Algorithm. *Appl. Soft Comput.* **2025**, *180*, 113418. [\[CrossRef\]](#)

12. Monteiro, R.V.A.; Guimarães, G.C.; Moura, F.A.M.; Albertini, M.R.M.C.; Albertini, M.K. Estimating Photovoltaic Power Generation: Performance Analysis of Artificial Neural Networks, Support Vector Machine and Kalman Filter. *Electr. Power Syst. Res.* **2017**, *143*, 643–656. [\[CrossRef\]](#)
13. Ledmaoui, Y.; El Maghraoui, A.; El Aroussi, M.; Saadane, R.; Chebak, A.; Chehri, A. Forecasting Solar Energy Production: A Comparative Study of Machine Learning Algorithms. *Energy Rep.* **2023**, *10*, 1004–1012. [\[CrossRef\]](#)
14. Polo, J.; Martín-Chivelet, N.; Alonso-Abella, M.; Sanz-Saiz, C.; Cuenca, J.; De La Cruz, M. Exploring the PV Power Forecasting at Building Façades Using Gradient Boosting Methods. *Energies* **2023**, *16*, 1495. [\[CrossRef\]](#)
15. Ait Mansour, A.; Tilioua, A.; Touzani, M. Bi-LSTM, GRU and 1D-CNN Models for Short-Term Photovoltaic Panel Efficiency Forecasting Case Amorphous Silicon Grid-Connected PV System. *Results Eng.* **2024**, *21*, 101886. [\[CrossRef\]](#)
16. Mellit, A.; Pavan, A.M.; Lughi, V. Deep Learning Neural Networks for Short-Term Photovoltaic Power Forecasting. *Renew. Energy* **2021**, *172*, 276–288. [\[CrossRef\]](#)
17. Wu, J.; Zhao, Y.; Zhang, R.; Li, X.; Wu, Y. Application of Three Transformer Neural Networks for Short-Term Photovoltaic Power Prediction: A Case Study. *Sol. Compass* **2024**, *12*, 100089. [\[CrossRef\]](#)
18. VanDeventer, W.; Jamei, E.; Thirunavukkarasu, G.S.; Seyedmahmoudian, M.; Soon, T.K.; Horan, B.; Mekhilef, S.; Stojcevski, A. Short-Term PV Power Forecasting Using Hybrid GASVM Technique. *Renew. Energy* **2019**, *140*, 367–379. [\[CrossRef\]](#)
19. Pan, M.; Li, C.; Gao, R.; Huang, Y.; You, H.; Gu, T.; Qin, F. Photovoltaic Power Forecasting Based on a Support Vector Machine with Improved Ant Colony Optimization. *J. Clean. Prod.* **2020**, *277*, 123948. [\[CrossRef\]](#)
20. Lim, S.-C.; Huh, J.-H.; Hong, S.-H.; Park, C.-Y.; Kim, J.-C. Solar Power Forecasting Using CNN-LSTM Hybrid Model. *Energies* **2022**, *15*, 8233. [\[CrossRef\]](#)
21. Zhen, H.; Niu, D.; Wang, K.; Shi, Y.; Ji, Z.; Xu, X. Photovoltaic Power Forecasting Based on GA Improved Bi-LSTM in Microgrid without Meteorological Information. *Energy* **2021**, *231*, 120908. [\[CrossRef\]](#)
22. Niu, D.; Wang, K.; Sun, L.; Wu, J.; Xu, X. Short-Term Photovoltaic Power Generation Forecasting Based on Random Forest Feature Selection and CEEMD: A Case Study. *Appl. Soft Comput.* **2020**, *93*, 106389. [\[CrossRef\]](#)
23. Zhang, C.; Peng, T.; Nazir, M.S. A Novel Integrated Photovoltaic Power Forecasting Model Based on Variational Mode Decomposition and CNN-BiGRU Considering Meteorological Variables. *Electr. Power Syst. Res.* **2022**, *213*, 108796. [\[CrossRef\]](#)
24. Liu, M.; Rao, S.; Huang, M.; Deng, S. Short-Term Photovoltaic Power Forecasting Based on Improved Transformer with Feature Enhancement. *Sustain. Energy Grids Netw.* **2025**, *43*, 101759. [\[CrossRef\]](#)
25. Abid, M.S.; Ahshan, R.; Al-Abri, M.; Al Abri, R. Spatiotemporal Forecasting of Solar and Wind Energy Production: A Robust Deep Learning Model with Attention Framework. *Energy Convers. Manag. X* **2025**, *26*, 100919. [\[CrossRef\]](#)
26. Tang, H.; Kang, F.; Li, X.; Sun, Y. Short-Term Photovoltaic Power Prediction Model Based on Feature Construction and Improved Transformer. *Energy* **2025**, *320*, 135213. [\[CrossRef\]](#)
27. Data Download | DKA Solar Centre. Available online: <https://dkasolarcentre.com.au/download?location=yulara> (accessed on 14 June 2025).
28. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
29. Sun, J.; Zhang, X.; Tang, Q.; Wang, Y.; Li, Y. Knock Recognition of Knock Sensor Signal Based on Wavelet Transform and Variational Mode Decomposition Algorithm. *Energy Convers. Manag.* **2023**, *287*, 117062. [\[CrossRef\]](#)
30. Richman, J.S.; Moorman, J.R. Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy. *Am. J. Physiol.-Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [\[CrossRef\]](#)
31. Zhu, J.; Bai, W.; Zhao, J.; Zuo, L.; Zhou, T.; Li, K. Variational Mode Decomposition and Sample Entropy Optimization Based Transformer Framework for Cloud Resource Load Prediction. *Knowl.-Based Syst.* **2023**, *280*, 111042. [\[CrossRef\]](#)
32. Bao, G.; Liu, X.; Zou, B.; Yang, K.; Zhao, J.; Zhang, L.; Chen, M.; Qiao, Y.; Wang, W.; Tan, R.; et al. Collaborative Framework of Transformer and LSTM for Enhanced State-of-Charge Estimation in Lithium-Ion Batteries. *Energy* **2025**, *322*, 135548. [\[CrossRef\]](#)
33. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Peng, F.-L.; Qiao, Y.-K.; Yang, C. A LSTM-RNN Based Intelligent Control Approach for Temperature and Humidity Environment of Urban Utility Tunnels. *Heliyon* **2023**, *9*, e13182. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Alizamir, M.; Shiri, J.; Fard, A.F.; Kim, S.; Gorgij, A.D.; Heddami, S.; Singh, V.P. Improving the Accuracy of Daily Solar Radiation Prediction by Climatic Data Using an Efficient Hybrid Deep Learning Model: Long Short-Term Memory (LSTM) Network Coupled with Wavelet Transform. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106199. [\[CrossRef\]](#)
36. Guo, M.; Hao, Y.; Lu, Y.; Sun, L. Long-Short Term Memory Modeling and Performance Analysis of Extended-State Kalman Filter-Based Energy-Saving Model Predictive Control for Supercritical Unit. *Appl. Therm. Eng.* **2025**, *272*, 126202. [\[CrossRef\]](#)
37. Standaert, F.-X.; Oswald, E. (Eds.) *Constructive Side-Channel Analysis and Secure Design: 7th International Workshop, COSADE 2016, Graz, Austria, 14–15 April 2016, Revised Selected Papers*; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 9689, ISBN 978-3-319-43282-3.
38. Team, K. Keras Documentation: Developer Guides. Available online: <https://keras.io/guides/> (accessed on 14 June 2025).

39. Pandas Documentation—Pandas 2.3.0 Documentation. Available online: <https://pandas.pydata.org/docs/> (accessed on 14 June 2025).
40. Scikit-Learn: Machine Learning in Python—Scikit-Learn 1.7.0 Documentation. Available online: <https://scikit-learn.org/stable/> (accessed on 14 June 2025).

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.