IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 22, 2021, accepted July 18, 2021, date of publication July 26, 2021, date of current version August 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3099169

Short-Term Photovoltaic Power Forecasting Based on VMD and ISSA-GRU

PENGYUN JIA[1], HAIBO ZHANG[1], (Senior Member, IEEE), XINMIAO LIU[2], AND XIANFU GONG[2]

[1]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China
[2]Power System Planning Research Center, Guangdong Power Grid Company, Ltd., Guangzhou 510080, China

Corresponding author: Haibo Zhang (zhb@ncepu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (No. 51777069) and the Science and Technology Program of China Southern Power Grid Company (037700KK52190010(GDKJXM20198272)).

ABSTRACT Photovoltaic (PV) power generation is affected by many meteorological factors and environmental factors, which has obvious intermittent, random, and volatile characteristics. To improve the accuracy of short-term PV power prediction, a hybrid model (VMD-ISSA-GRU) based on variational mode decomposition (VMD), improved sparrow search algorithm (ISSA) and gated recurrent unit (GRU) is proposed. First of all, the PV time series is decomposed into a series of different subsequences by VMD to reduce the non-stationarity of the original data. Then, the main factors affecting PV power generation are obtained by using the correlation coefficients of Spearman and Pearson, which reduces the computational complexity of the model. Finally, the GRU network optimized by ISSA is used to predict all the subsequences and residual error of VMD, and the prediction results are reconstructed. The results show that the hybrid VMD-ISSA-GRU model has stronger adaptability and higher accuracy than other traditional models. The mean absolute error (MAE) in the whole test set is 1.0128 kW, the root mean square error (RMSE) is 1.5511 kW, and the R^2_{adj} can reach 0.9993.

INDEX TERMS Short-term PV power forecast, variational mode decomposition, improved sparrow search algorithm, GRU neural network.

I. INTRODUCTION

With the rapid development of the global new energy power generation industry, solar energy has been widely used because of its several advantages of safety, efficiency, and wide distribution [1]–[3]. The process of PV power generation is random and unstable, which brings great challenges to the safe operation of power system [4]–[6]. Therefore, the way to improve the stability of power system operation after PV grid connected is very important. Accurate short-term PV power prediction can not only improve the operation efficiency of the PV power station, but also help the dispatching department to make more accurate real-time scheduling plans, and ensure the stable operation of large-scale PV power station after connecting to the grid.

The current short-term PV power prediction methods mainly include physical methods and statistical methods [7]. Physical methods use detailed meteorological conditions and environmental information to establish physical models for prediction [8]–[10]. Physical methods do not rely on a large amount of historical power generation data, and their prediction speeds are fast. However, physical models are complex and have poor anti-interference ability, so it is difficult to guarantee the prediction accuracy of the models. Statistical methods include traditional statistical methods and artificial intelligence methods. Traditional statistical methods mainly establish a simple mapping relationship between historical data and output power, such as regression analysis method [11], fuzzy theory [12], grey theory method [13], and so on. Traditional statistical methods have the advantages of simple model and fast prediction speed, but they can not fit the complex nonlinear relationship. Artificial intelligence methods use the intelligent algorithm to fully mine the internal characteristics and hidden change rules of the data, can better fit the complex nonlinear relationship, and realize the PV power prediction [14]–[16]. Liu *et al.* [17] proposed an improved PV power forecasting model with the Assistance of Aerosol Index Data. The experimental results have shown that the method has high prediction performance. William *et al.* [18] established the prediction model

The associate editor coordinating the review of this manuscript and approving it for publication was Guangya Yang.

of support vector machine (SVM) optimized by genetic algorithm (GA) to realize the short-term power prediction of residential scale PV system. Compared with other models, GA-SVM has better prediction performance. However, most of the shallow machine learning methods are difficult to accurately describe such complex nonlinear mapping relationship. Their algorithms are single, low robustness, and easy to fall into the local optimum [19].

In recent years, with the rapid development of artificial intelligence algorithms, deep learning algorithms have broken through the limitations of shallow machine learning models. The deep learning algorithm model represented by convolutional neural network (CNN) [20] and recurrent neural network (RNN) [21], [22] has been widely used in the field of short-term PV power prediction. RNN is mainly used to process time series data, but it is prone to long-term dependence. Long short term memory (LSTM) introduces the gate structure based on RNN structure, which realizes the function of selective memory of historical information and solves the problem of long-term dependence of RNN. Li *et al.* [23] proposed a hybrid model based on CNN and LSTM, which used CNN to deeply mine the nonlinear characteristics and invariant structures of data, and then combined LSTM for prediction. Mei *et al.* [24] used quantile regression averaging (QRA) and LSTM integrated model to obtain the probability prediction of PV output. The experimental results have shown that LSTM-QRA has a higher prediction performance. However, LSTM has complex structure, many parameters and long training time. K. Cho *et al.* [25] proposed gated recurrent unit network (GRU), which is another RNN gating architecture after LSTM. There are only two gates in it, which has less training parameters than LSTM. GRU not only ensures high prediction accuracy, but also alleviates the problem of LSTM over-fitting [26]–[28]. Compared with the shallow machine learning algorithm, the accuracy of the above methods has been greatly improved, but there are still problems that can not fully mine the local features and internal hidden information of historical PV data.

As PV power is affected by many factors, it has obvious nonlinear and non-stationary characteristics. The wavelet transform (WT) [29], the empirical mode decomposition (EMD) [30], the local mean decomposition (LMD) [31] and the variational mode decomposition (VMD) [32] have been widely used in the field of PV power prediction. WT is not adaptive, and it can not have high accuracy in time and frequency at a certain scale. EMD and LMD are recursive decomposition methods, which have the problems of modal aliasing, endpoint effect, sensitivity to noise and sampling, and difficulty to separate the near frequency components. VMD is an adaptive, completely non-recursive method of modal variation and signal processing, which solves the problem that EMD and LMD can not get the signal characteristic components under the background of bad noise. Wang *et al.* [33] used VMD to decompose the historical PV power, and then combined with the LSTM optimized by the improved particle swarm optimization (IPSO) algorithm to predict. The residual error of VMD is also very important to the prediction results, which has not been predicted and analyzed in [33]. The particle swarm optimization (PSO) has a fast convergence speed and strong universality, but it is easy to fall into the local optimum [34]. The IPSO based on non-linear inertia weight improves the convergence accuracy of PSO to a certain extent, but there is still the risk of falling into the local optimum. Moreover, there are too many internal parameters of LSTM, the long training time of the model, and prone to overfitting. Xue *et al.* [35] proposed a new swarm intelligence optimization algorithm, namely the sparrow search algorithm (SSA). The experimental results have shown that this algorithm has higher optimization accuracy, faster convergence speed, and better robustness than the gray wolf optimizer (GWO), gravitational search algorithm (GSA), and particle swarm optimization (PSO). Table 1 shows the specific prediction methods of some references.

TABLE 1. Overview of different PV power prediction models.

Ref.	Year	Model	Parameters	Error evaluated	Prediction time
[17]	2015	BP(AI)	3 layers	MAPE	24h
[18]	2019	GASVM	Gaussian kernel	MAPE, RMSE	1h
[20]	2019	PVPNet	3 conv layers 3 pooling layers	MAE, RMSE	24h
[21]	2019	AHPA-LSTM	18 layers	MAPE, THI RMSE	24h
[23]	2020	CNN-LSTM	-	MAE, R^2, RMSE	15min-180min
[24]	2020	LSTM-QRA	12 layers	MAPE, RMSE	24h
[33]	2020	OVMD-IPSO-LSTM	4 layers	MAPE, MRE, R^2	24h

Based on the previous research, this paper proposes a new hybrid model based on VMD, improved sparrow search algorithm (ISSA), and GRU to enhance the accuracy of short-term PV power prediction. First of all, the historical PV power data is adaptively decomposed into subsequences with different center frequencies by VMD. The optimal decomposition parameters are determined by using the correlation coefficient between the residual error of VMD and the original data, so that the historical PV power data can be fully decomposed. Secondly, Pearson and Spearman coefficients are used to determine the main influencing factors of PV power generation, to improve the calculation ability of the model. Then, the Levy flight strategy [36] and adaptive t-distribution are used to improve SSA to reduce the risk of falling into the local optimization. Finally, the decomposed subsequences and residual error combined with the reduced dimension meteorological data are input into the GRU model for prediction. All the results prediction by the GRU network are reconstructed. The structural parameters of GRU are optimized by ISSA.

The main contributions of the paper are further explained as follows:

(1) The optimal decomposition mode is determined by the correlation coefficient between the residual error of VMD and the original PV data. The local characteristics and internal hidden information of the original PV power data are fully mined to reduce its instability.
(2) According to the characteristics of the sparrow search algorithm, Levy flight strategy and adaptive t-distribution are used to improve the performance of the sparrow search algorithm and alleviate the problem of falling into the local optimum.
(3) Pearson and Spearman coefficients are used to analyze the main factors affecting the PV power to reduce the dimension of high-dimensional meteorological data.
(4) A new hybrid prediction model (VMD-ISSA-GRU) is proposed, which used the subsequences and residual error of VMD to analyze and predict. Compared with other models, the accuracy of PV power prediction has been significantly improved.

II. THE HYBRID VMD-ISSA-GRU MODEL

A. THE ARCHITECTURE OF THE PROPOSED MODEL

Since the PV power time series is affected by many factors, it is obviously unstable and nonlinear. VMD can effectively separate the signal and noise of the original data, fully mine the local features contained in the PV power data, and reduce the instability of the original data. To improve the accuracy of short-term PV power prediction, the historical PV power time series is decomposed into K different frequency subsequences by VMD. The optimal decomposition parameters are determined by using the correlation coefficient between the residual error of VMD and the original data, so that the PV power data can be fully decomposed. The mathematical theory of the VMD algorithm is defined in Section II.B.

When forecasting short-term PV power, the meteorological factors with weak correlation will seriously affect the training efficiency and prediction accuracy of the model. Therefore, the main meteorological factors affecting PV output are determined by Spearman and Pearson correlation coefficients to reduce the dimension of meteorological data.

The structure parameters of the GRU network are very important for the training and prediction of samples. The reasonable network structure parameters are the important premise of PV power prediction. This paper adopts the structure of the four-layer neural network, including an input layer, an output layer, and two hidden layers. The improved sparrow search algorithm (ISSA) and adaptive moment estimation algorithm (Adam) [37] are used to optimize the GRU network. ISSA mainly optimizes the structural parameters of the GRU network, including the number of neurons in the first hidden layer, the number of neurons in the second hidden layer, the number of iterations, and the learning rate. The number of GRU network parameters is taken as the dimension of sparrow search, and the GRU network parameters are encoded as the initial position vector of the sparrow individuals. Through many iterations, the global optimal point is searched, and the GRU optimal network structure parameters are obtained by decoding the position of the optimal solution. The mathematical theories of the ISSA and the GRU network are defined in Section II.C, and Section II.D, respectively.

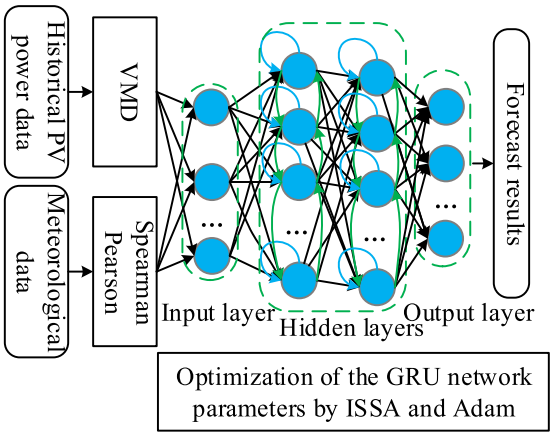


FIGURE 1. Structure of the hybrid VMD-ISSA-GRU model.

At the same time, Adam algorithm is used for local optimization of the GRU network, so that the training can adaptively calculate the learning efficiency of each parameter, reduce the influence of parameter selection on the accuracy of the model, and improve the convergence speed. The exponential decay rate of first-order moment estimation is set to 0.9, and the exponential decay rate of second-order moment estimation is set to 0.999 [37]. The subsequences and residual error of VMD are combined with the meteorological factors after dimension reduction. The combined new data sets are predicted through the GRU network optimized by ISSA. Fig. 1 shows the structure of the hybrid VMD-ISSA-GRU model.

B. VARIATIONAL MODE DECOMPOSITION

Variational mode decomposition (VMD) [32] is an adaptive, completely non-recursive method of mode variation and signal processing. The core of VMD is to construct and solve variational problems. According to the pre-determined modal number K, the relevant frequency scale can be determined adaptively and the corresponding modal function can be estimated, which reduces the complexity of nonlinear time series, and has strong robustness to sampling and noise. The working principle of the VMD is as follows:

Step1: Construct the objective function.

$$\begin{cases} \min \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + j/\pi t \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ \sum_{k=1}^{K} u_k = f(t) \end{cases} \tag{1}$$

where, $\{u_k\} = \{u_1, u_2, \ldots, u_K\}$ and $\{\omega_k\} = \{\omega_1, \omega_2, \ldots, \omega_K\}$ represent the set of all subsequences and respective center frequencies, $\delta(t)$ refers to the impulse function, j refers to the

imaginary part, K refers to the decomposed mode number, $f(t)$ refers to the historical PV power time series.

Step2: Solves variational problems.

The quadratic penalty α and Lagrange multipliers λ are introduced to turn the constrained variational problem into an unconstrained variational problem.

$$\begin{aligned} &L\left(\{u_k\},\{\omega_k\},\lambda\right) \\ &:=\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+j/\pi t\right)*u_k(t)\right]e^{-j\omega_k t}\right\|_2^2 \\ &\quad+\left\|f(t)-\sum\nolimits_{k=1}^{K}u_k(t)\right\|_2^2 \\ &\quad+\left\langle\lambda(t),f(t)-\sum\nolimits_{k=1}^{K}u_k(t)\right\rangle \end{aligned} \tag{2}$$

The alternating direction method of multiplier (ADMM) is used to optimize and update u_k and ω_k to find the optimal solution of the constrained variational model.

$$\begin{cases} \hat{u}_k^{n+1}(\omega) \\ \quad=\dfrac{\hat{f}(\omega)-\sum_{i<k}\hat{u}_i^{n+1}(\omega)-\sum_{i>k}\hat{u}_i^{n}(\omega)+\dfrac{\hat{\lambda}^n(\omega)}{2}}{1+2\alpha\left(\omega-\omega_k^n\right)^2} \\ \omega_k^{n+1}=\dfrac{\int_0^\infty\omega\left|\hat{u}_k^{n+1}(\omega)\right|^2d\omega}{\int_0^\infty\left|\hat{u}_k^{n+1}(\omega)\right|^2d\omega} \\ \hat{\lambda}^{n+1}(\omega)=\hat{\lambda}^n(\omega)+\tau\left(\hat{f}(\omega)-\sum\limits_k\hat{u}_k^{n+1}(\omega)\right) \\ \sum_k\left\|\hat{u}_k^{n+1}-\hat{u}_k^n\right\|_2^2\Big/\left\|\hat{u}_k^n\right\|_2^2<\varepsilon \end{cases} \tag{3}$$

where, $\hat{u}_k^n(\omega)$, $\hat{\lambda}_k^n(\omega)$ and $\hat{f}(\omega)$ are the Fourier transforms of $u_k^n(\omega)$, $\lambda_k^n(\omega)$ and $f(\omega)$ respectively.

C. IMPROVED SPARROW SEARCH ALGORITHM

1) LEVY FLIGHT STRATEGY

Levy flight strategy [36] is a kind of random behavior that simulates the foraging of organisms in nature. Its frequent short-distance search can improve the local search ability of intelligent optimization algorithm, while the accidental long-distance search can increase the search range, to enhance the global search ability of organisms and prevent intelligent optimization algorithm from falling into the local optimum. The method of generating levy random step size is as follows:

$$\begin{cases} s=\dfrac{\mu}{|\upsilon|^{1/\beta}},\ 0<\beta<2 \\ \upsilon\sim N\left(0,\sigma_\upsilon^2\right),\ \sigma_\upsilon=1 \\ \mu\sim N\left(0,\sigma_\mu^2\right),\ \sigma_\mu=\left\{\dfrac{\Gamma(1+\beta)\sin\left(\pi\beta/2\right)}{\Gamma\left[(1+\beta)/2\right]\beta 2^{(\beta-1)/2}}\right\}^{1/\beta} \end{cases} \tag{4}$$

where, s refers to the random step size, β refers to the control coefficient, μ and υ are random numbers with the normal distribution.

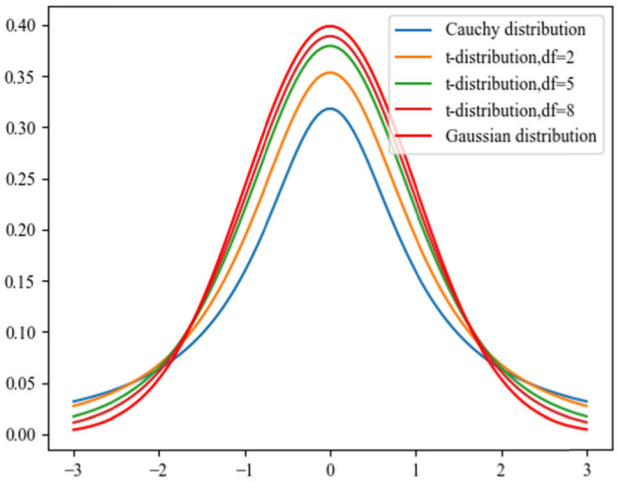


FIGURE 2. Comparison of Cauchy distribution, t-distribution, and Gaussian distribution.

2) T-DISTRIBUTION

The t-distribution is also called student distribution, and its distribution curve is related to the degree of freedom (df). The smaller the degree of freedom is, the smoother the t-distribution curve is. When the degree of freedom is 1, it is Cauchy distribution; when the degree of freedom tends to infinity, the distribution is closer to Gaussian distribution.

Adaptive t-distribution mutation uses the number of iterations as its degree of freedom parameter. At the beginning of the iteration, the adaptive t-distribution is close to Cauchy distribution and has strong global search ability. At the end of the iteration, with the increase of degrees of freedom, the adaptive t-distribution is closer to Gaussian distribution and has strong local search ability. Fig. 2 shows the comparison of Cauchy distribution, t-distribution, and Gaussian distribution.

3) IMPROVED SPARROW SEARCH ALGORITHM

Sparrow search algorithm (SSA) is a new swarm intelligence optimization algorithm. SSA mainly simulates the process of foraging behavior and anti-predation behavior of sparrows. The individuals who are easy to find food in sparrows are called discoverers, and other individuals are called followers. At the same time, a certain proportion of individuals in the sparrow population are selected as investigators to send out early warning information when danger is found. SSA has the advantages of fast convergence speed, high optimization accuracy, and strong robustness, but it also has the disadvantage of easily falling into the local optimum [38]. For details of SSA, please refer to [35].

In this paper, Levy flight strategy and adaptive t-distribution are used to improve the search accuracy of SSA and avoid falling into a local optimal solution.

Improvement to the location of discoverers is as follows:

$$x_{ij}^{m+1}=\begin{cases} x_{ij}^m\cdot\exp\left[-i/(rand\cdot M)\right], & R_2<ST \\ x_{ij}^m+x_{ij}^m\cdot\gamma\cdot Levy(\beta), & R_2\geq ST \end{cases} \tag{5}$$

where, M refers to the maximum number of iterations, *rand* refers to a random number between [0,1], R_2 and ST are warning value and safety value respectively, γ refers to the step coefficient, $\gamma = 0.01$. When $R_2 > ST$, discoverers need to lead other individuals to fly to other safe places quickly according to Levy flight strategy.

Improvement to the location of followers is as follows:

$$x_{ij}^{m+1} = \begin{cases} \gamma \cdot Levy(\beta) \cdot \exp\left(\frac{x_{wj}^m - x_{ij}^m}{i^2}\right), & i > NP/2 \\ x_{pj}^{m+1} + \left|x_{ij}^m - x_{pj}^{m+1}\right| \cdot A^+ \cdot L, & otherwise \end{cases} \quad (6)$$

where, x_{wj} refers to the current global worst position, x_{pj} refers to the best position occupied by discoverers, and A refers to a row of the multidimensional matrix whose elements are 1 or -1. When $i > NP/2$, the i-th participant with lower fitness did not get food and needed to fly to other places for food.

The location update formula of investigators is as follows:

$$x_{ij}^{m+1} = \begin{cases} x_{bj}^m + \gamma \cdot Levy(\beta) \cdot \left|x_{ij}^m - x_{bj}^m\right|, & f_i > f_g \\ x_{ij}^m + K\dfrac{\left|x_{ij}^m - x_{wj}^m\right|}{(f_i - f_w) + \zeta}, & f_i = f_g \end{cases} \quad (7)$$

where, x_{bj} is the global best position, f_i is the fitness value of the current individual, f_g and f_w are the global best and worst fitness values, ζ is a very small constant, and K is the step control coefficient, which is a random number of $[-1,1]$. When $f_i > f_g$, it means that the current individual is at the edge of the population and is vulnerable to predators. When $f_i = f_g$, it means that the individual in the middle of the population is aware of the danger and needs to be close to other sparrows to reduce the risk of predation.

At the same time, in SSA, the adaptive t-distribution mutation operation is added to carry out the adaptive mutation. When $rand < p$, the mutation is carried out to avoid the algorithm falling into the local optimal solution.

$$x_{ij}' = x_{ij} + x_{ij} \cdot t(m), \quad rand < p \quad (8)$$

where, $t(m)$ is the t-distribution with m degree of freedom.

D. GATED RECURRENT UNIT NETWORK

Gated recurrent unit network (GRU) [25] is a variant of long short-term memory network (LSTM). It uses reset gate and update gate to solve the problem of long-term dependence, and combines the state of the data unit and hidden layer to solve the problem of gradient disappearance. Therefore, GRU network has only two gates structure, with fewer training parameters and easy convergence, which alleviates the over-fitting problem of LSTM, greatly improves the learning efficiency, and maintains the prediction effect of LSTM. The GRU network in this paper adopts the structure of the four-layer neural network. Fig. 3 shows the structure diagram of GRU neural unit. Fig. 4 shows the data flow diagram of the GRU neural network.

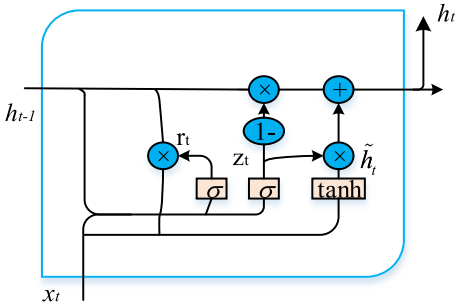


FIGURE 3. Neural unit structure of GRU network.

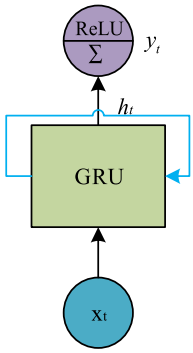


FIGURE 4. Data flow diagram of the GRU network.

Input layer: The subsequences (u_k) and residual error (ε_k) of VMD are combined with the main meteorological factors (Q) after dimension reduction as the input of the model (I).

Hidden layer: Training and learning the input layer data, and introducing dropout weight constraint to prevent network overfitting.

$$\begin{cases} r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \\ z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases} \quad (9)$$

where, z_t stands for update gate, mainly for forgetting and memory; r_t is reset gate, which is used to determine whether to combine the current state with the previous information; x_t is the current input; h_{t-1} is the status information of the previous node, and $\tilde{h}_t$ represents the summary of x_t and h_{t-1}; W_z and W_r are connection weights of update gate and reset gate respectively; σ and tanh are activation functions.

Output layer: The full connection layer is adopted, and the ReLU function is selected as the activation function.

$$y_t^o = \mathrm{ReLU}(W_o \cdot h_t + b_o) \quad (10)$$

where, y_t^o is the output of the GRU network, W_o is the weight of the output layer, and b_o is the bias term.

III. CONSTRUCTION OF VMD-ISSA-GRU MODEL

A. OPTIMAL DECOMPOSITION OF VMD

The parameters of VMD need to be set in advance are modal number K, penalty factor α, fidelity coefficient τ, and stop

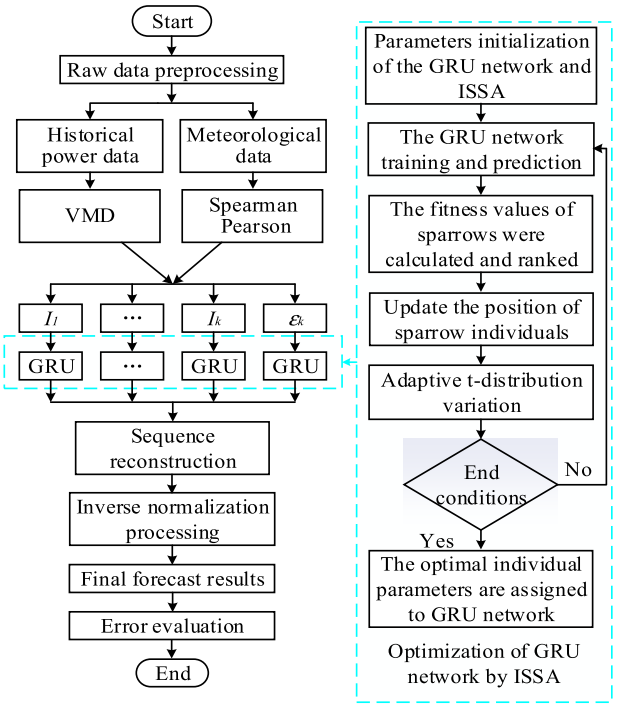


FIGURE 5. Flow chart of the hybrid VMD-ISSA-GRU model.

convergence condition ε. The setting of the mode number K has a great influence on the decomposition process. If the value of K is too large, it is easy to lead to over decomposition and increase the complexity of PV power prediction; if the value of K is too small, it is easy to lead to insufficient decomposition and can not achieve the purpose of reducing the instability of the original sequence. To fully decompose the PV power sequence, the correlation coefficient R_k between the residual of VMD and PV power sequence is used to determine the number of modes K. When the PV sequence is fully decomposed, the correlation between the residual and the original PV sequence should be minimal, and other parameters are set as default values. Starting from $K = 2$, the correlation coefficient between the residual after each decomposition and the original PV sequence is calculated. When R_K is minimum, K is the optimal number of decomposition modes.

$$\begin{cases} R_K = \rho_p \left(f\left(t\right), \varepsilon_K\left(t\right)\right) \\ \varepsilon_K\left(t\right) = f\left(t\right) - \sum_{i=1}^{K} u_{Ki}\left(t\right) \end{cases} \tag{11}$$

where, ρ_p is the Pearson correlation coefficient, u_{Ki} is the i-th modal function with modal number K, and ε_K is the residual of VMD.

B. THE CONSTRUCTION PROCESS OF HYBRID MODEL

The flow chart of the hybrid VMD-ISSA-GRU model is shown in Fig 5. The short-term PV prediction process based on a hybrid VMD-ISSA-GRU model is as follows:

Step1: Perform data cleaning on the original PV time series and multi-dimensional meteorological data, complete missing values, and eliminate outlier data. The mean interpolation [39] is a common method to make up for the missing values. The mean interpolation uses the mean value of the adjacent time data of the missing value to make up for the missing value. However, the PV power data often appear the phenomenon of multiple adjacent data missing. In this paper, the random weighted mean of the same time data of adjacent dates is used to make up for the missing value. Because of the different dimension levels between different features, the normalization method MinMaxScaler can effectively improve the data quality and the convergence speed of the model.

$$y_i' = \eta_1 y_i^{d-2} + \eta_2 y_i^{d-1} + \eta_3 y_i^{d+1} + \eta_4 y_i^{d+2} \tag{12}$$

$$\tilde{X}_i = \frac{\left(X_i - X_{\min}\right)}{\left(X_{\max} - X_{\min}\right)} \tag{13}$$

where, y_i' is the missing value, y_i^{d-2}, y_i^{d-1}, y_i^{d+1} and y_i^{d+2} are the measured value corresponding to the same time of four days before and after the missing value, η is the random weight, and $\eta_1+\eta_2+\eta_3+\eta_4 = 1$; $\tilde{X}_i$ is the normalized value, X_i is the value before normalization, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the data respectively.

Step2: The meteorological factors with high correlation $Q = (q_1, q_2, \ldots, q_l)$ are screened by Spearman and Pearson correlation coefficients. The larger the absolute value of the correlation coefficient is, the higher the correlation with PV power is.

$$\rho_s = 1 - \frac{6\sum_{i=1}^{N}\left(R\left(q_{il}\right) - R\left(y_i\right)\right)^2}{N\left(N^2-1\right)} \tag{14}$$

$$\rho_p = \frac{\sum_{i=1}^{N}\left(q_{il} - \bar{q}_l\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{N}\left(q_{il} - \bar{q}_l\right)^2 \sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2}} \tag{15}$$

where, ρ_s is Spearman rank correlation coefficient, ρ_p is Pearson correlation coefficient, q_{il} is the i-th value of the l-st dimension meteorological factor, y_i is the i-th measured power value of PV power data, $R(q_{il})$ and $R(y_i)$ are the ranking of q_{il} and y_i elements in their respective column vectors respectively, $\bar{q}_l$ is the average of meteorological elements in column l, $\bar{y}$ is the average of PV power data.

Step3: By analyzing the correlation between the residual and the original PV data, the optimal decomposition mode number K is obtained. The PV time series $f(t)$ is decomposed by VMD to obtain a series of subsequences with different center frequencies $\{u_k\} = \{u_1, u_2, \ldots, u_K\}$ and residual error ε_k. The subsequences and residual error of VMD are combined with the meteorological factors after dimension reduction respectively as the input vector $I_k = (u_k, q_1, q_2, \ldots, q_l)$, $I_\varepsilon = (u_\varepsilon, q_1, q_2, \ldots, q_l)$. At the same time, the training set and test set are divided.

Step4: ISSA parameters initialization, encoding GRU network structure parameters $pop = (pop_1, pop_2, pop_3, pop_4)$ to Sparrow's initial position vector $X_{ij} = (x_{i1}, x_{i2}, x_{i3}, x_{i4})$, pop_1-pop_4 correspond to the number of neurons in the first hidden layer, the number of neurons in the second hidden

layer, the number of iterations of GRU, and the learning rate respectively, X_{ij} represents the j-th position of the i-th sparrow in the 4-dimensional space.

Step5: Determine the fitness function of ISSA. The weighted average of the mean square error of training set and test set is used as the fitness function. After experiments, the weights of training set and test set account for 0.5 respectively.

$$F = \sum_{i=1}^{S} (\tilde{y}_i - y_i)^2 \Big/ 2S + \sum_{i=S+1}^{N} (\tilde{y}_i - y_i)^2 \Big/ 2(N-S) \tag{16}$$

where, N is the number of PV power data samples, S is the number of training set samples, $N - S$ is the number of test set samples, $\tilde{y}_i$ is the i-th predicted value of PV power data.

Step6: Calculate the fitness value of the sparrow, update the individual position of the sparrow, and compare with the historical optimal value. If the stop condition is satisfied, the iteration will be terminated. Otherwise, the iteration will continue.

Step7: Decode and transform the optimal solution optimized by ISSA to get the optimal parameters of the GRU network, and assign the optimal parameters to the GRU network. The optimized GRU network is used to train and predict the input vector I_K and I_ε. The final prediction result $P = Y_\varepsilon + \sum Y_k$ is obtained by reconstructing the prediction results of each subsequence and the prediction result of residual, and inversely normalizing them. The mean squared error (MSE) is used as the loss function of the GRU network. MSE can evaluate the change degree of the square error between the predicted value and the real value. The smaller the value of MSE, the higher the accuracy of GRU network prediction.

$$\sigma_{MSE} = \frac{1}{N} \sum_{1}^{N} (y_i - \tilde{y}_i)^2 \tag{17}$$

Step8: Error evaluation [40]. The root mean square error (RMSE), the mean absolute error (MAE), and the adjusted R-square (R^2_{adj}) are used to evaluate the prediction performance of the model.

$$\sigma_{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\tilde{y}_i - y_i)^2}{N}} \tag{18}$$

$$\sigma_{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \tilde{y}_i| \tag{19}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\tilde{y}_i - y_i)^2}{\sum_{i=1}^{N} (\bar{y} - y_i)^2} \tag{20}$$

$$R^2_{adj} = 1 - \frac{\left(1 - R^2\right)(N-1)}{N - p - 1} \tag{21}$$

IV. CASE STUDY AND ANALYSIS

A. DATA PREPROCESSING

The data set of PV power generation is from a PV power station in Australia, including the daily PV power data from August 1, 2016 to April 30, 2018, and the collection time is 5minutes. Meteorological data includes relative humidity, temperature, global horizontal radiation, diffuse horizontal radiation, wind speed, wind direction, and rainfall. After data preprocessing, there are 629 days of data, the first 600 days of data as the training set, the next 29 days of data as the test set. The collection time is adjusted from 5 minutes to 15 minutes, and samples are collected once for every three records. The daily prediction points of PV power are reduced appropriately to improve the prediction efficiency of the model.

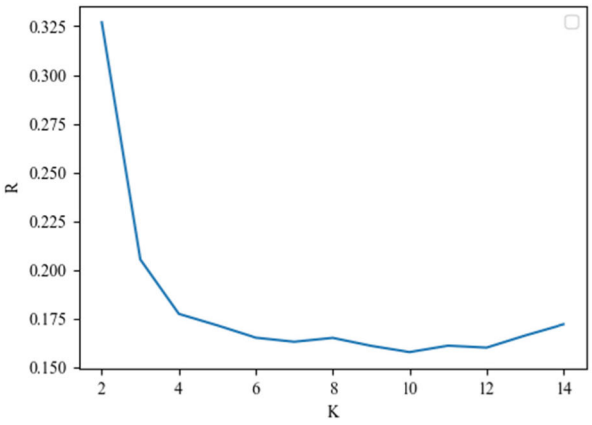


FIGURE 6. Trend chart of the residual correlation coefficient.

The historical PV power data is decomposed into some subsequences by VMD. It can be seen from Fig. 6 that the correlation coefficient between the residual corresponding to different mode number K and the historical PV power. When $K < 10$, the residual correlation coefficient R_K decreases as a whole, and the decomposition of PV power data is incomplete; when $K > 10$, the residual correlation coefficient R_K begins to show an upward trend, and the PV power data is over decomposed; when $K = 10$, the residual correlation coefficient R_K is the minimum, and the PV power data is completely decomposed. Fig. 7 shows the results of the optimal VMD. u_1 is the trend component, reflecting the overall trend of the original PV power; u_2 to u_9 are the detail components, reflecting the changing trend of the original PV power in different details; u_{10} is a random component, reflecting the random characteristics of the original PV power.

The Spearman and Pearson correlation coefficients of meteorological data were analyzed. It can be seen from Table 2 that the correlation coefficient of rainfall, relative humidity, and wind direction with PV power is less than 0, which belongs to negative correlation. PV power will decrease with the increase of negative correlation factors. The correlation coefficient of temperature, wind speed, global horizontal radiation, and diffuse horizontal radiation with PV power is greater than 0, which is positive correlation. PV power will increase with the increase of positive correlation factors. The absolute value of wind direction, wind speed, and rainfall correlation with PV power are less than 0.1, which belongs to weak correlation. Therefore, this paper

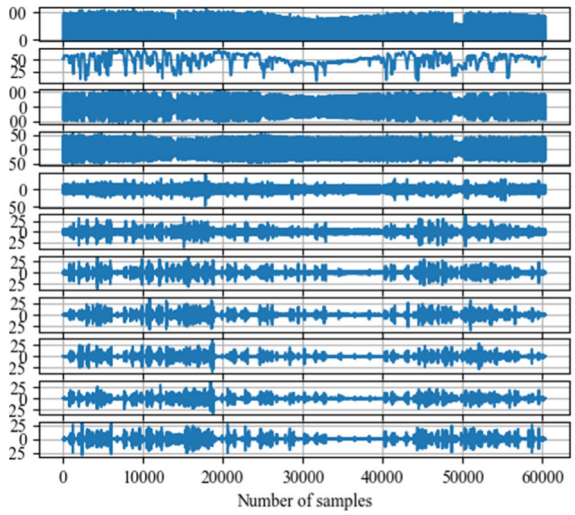


FIGURE 7. Results of VMD.

TABLE 2. Spearman and pearson correlation coefficient.

	Spearman	Pearson
Temperature	0.4845	0.4179
Wind speed	0.0187	0.0059
Wind direction	-0.0807	-0.0450
Rainfall	-0.0354	-0.0435
Relative humidity	-0.4680	-0.4292
Global horizontal radiation	0.9226	0.9759
Diffuse horizontal radiation	0.8461	0.5056

takes temperature, relative humidity, global horizontal radiation, and diffuse horizontal radiation as the main factors affecting PV output.

B. RESULT ANALYSIS

To verify the superiority of the model used in this paper, 12 different models are used to predict the PV power of the test set for 29 days. All the models are completed in python3.7, TensorFlow2.0, Intel (R) Core (TM) i5-9500 CPU@3.00GHZ, 8.0GB RAM, and 64-bit operating system personal computer. The IPSO algorithm used in the comparison model is referred to [32]. Table 3 shows the algorithm parameters of the hybrid VMD-ISSA-GRU model; Table 4 shows the comparison of the GRU network prediction errors with different values of dropout; Table 5 shows the comparison results of RMSE, MAE, and R^2_{adj} of each model from 7:00 to 19:00 every day under different weather conditions; Table 6 shows the comparison results of RMSE, MAE, and R^2_{adj} of each model in the whole test set; Fig. 8-10 show the prediction curves of some models under different weather conditions; Fig. 11 shows the time taken by each model; Fig. 12 shows the mean absolute error of each model in the test set.

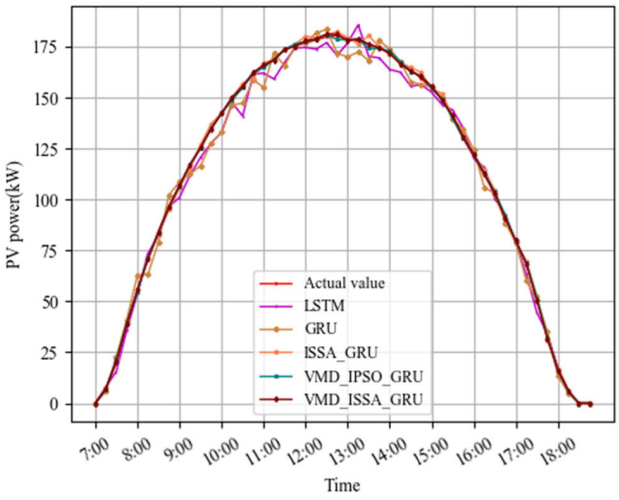


FIGURE 8. Prediction results of some models with relatively stable weather conditions.

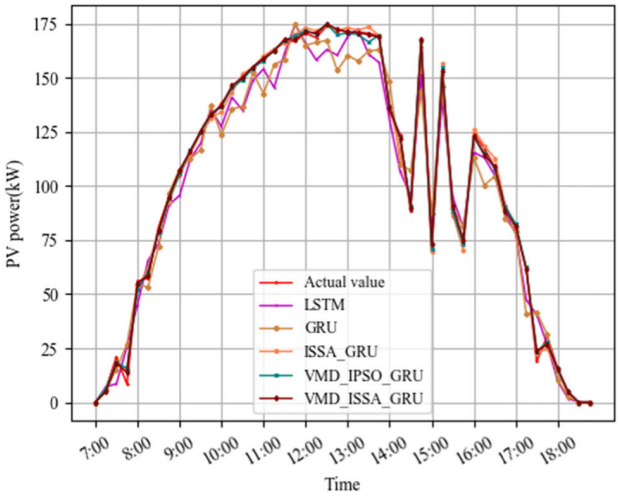


FIGURE 9. Prediction results of some models with obvious fluctuation of weather conditions.

TABLE 3. Parameters setting of model.

Name	Value	Name	Value
Maximum iterations of ISSA	10	Variation probability	0.3
Population number	5	The first hidden layer	[0,200]
Optimization dimension	4	The second hidden layer	[0,200]
Proportion of Discoverers	0.2	Maximum iterations of GRU	[100,1000]
Safety value	0.8	Learning rate	[0.001,0.01]

The dropout weight is introduced into the GRU network structure to prevent the GRU network from over-fitting. Through repeated experiments, the value of dropout was determined to be 0.2. When dropout is 0.2, the RMSE and

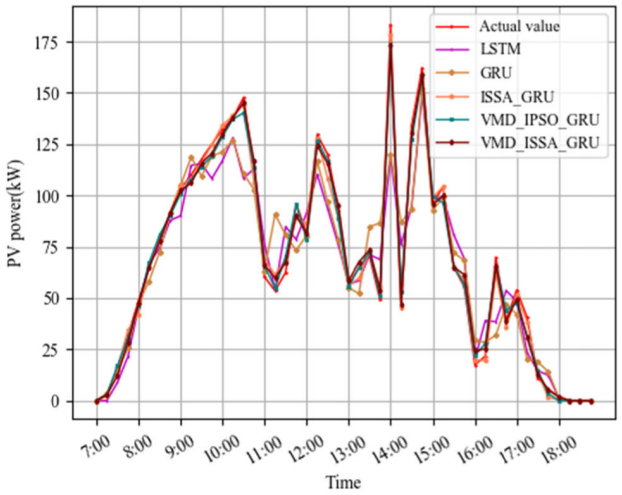


FIGURE 10. Prediction results of some models with severe fluctuation of weather conditions.

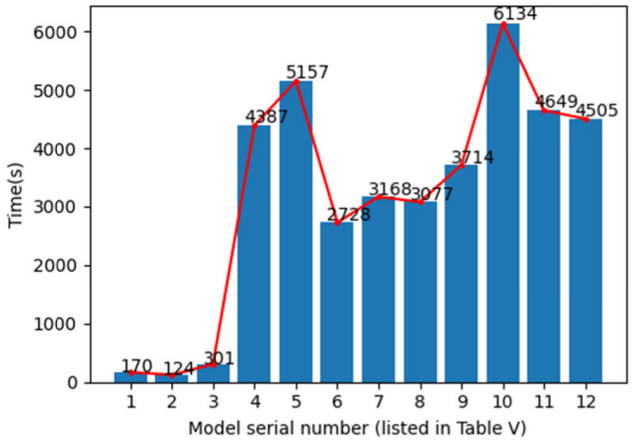


FIGURE 11. Prediction time of different models.

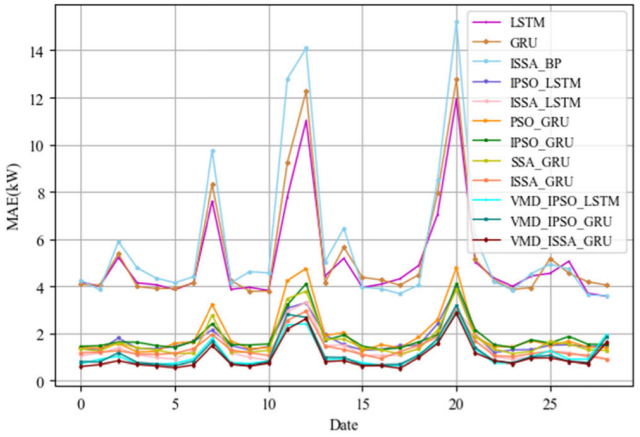


FIGURE 12. The MAE of each model in the test set.

MAE of GRU network are the smallest, and the R^2_{adj} is the highest.

From Table 5 and Fig. 8, we can see that the weather conditions are relatively stable, and each model can predict the trend of PV output. VMD-ISSA-GRU has the smallest RMSE and MAE, which are 0.6898 kW, 0.5409 kW. Compared with VMD-IPSO-GRU, ISSA-GRU and GRU, the MAE of VMD-ISSA-GRU is reduced by 13.95%, 52.76% and 86.13% respectively, and the R^2_{adj} can reach 0.9999, which indicates that the VMD-ISSA-GRU model can better fit the PV output when the weather conditions are relatively stable.

TABLE 4. Comparison of GRU prediction errors under different values of dropout.

Dropout	RMSE(kW)	MAE(kW)	R^2_{adj}
0.2	7.9934	5.3039	0.9821
0.4	8.2042	5.5002	0.9811
0.6	8.0248	5.5598	0.9819
0.8	8.4570	5.7937	0.9799
-	8.6474	6.6022	0.9790

From Table 5 and Fig. 9, we can see that the weather conditions fluctuate obviously, LSTM, GRU, and ISSA-BP can not well fit the fluctuation of PV output, while other models can well fit the fluctuation of PV output. The RMSE and MAE of VMD-ISSA-GRU are the lowest, and the R^2_{adj} is the highest. Compared with VMD-IPSO-GRU, ISSA-GRU and GRU, the MAE of VMD-ISSA-GRU is reduced by 10.54%, 24.32% and 82.31% respectively, which indicates that the VMD-ISSA-GRU model has higher prediction accuracy and more stability than other models in the case of obvious fluctuation of weather conditions

From Table 5 and Fig. 10, we can see that the weather conditions fluctuate violently and the PV output is extremely unstable. LSTM, GRU, and ISSA-BP can not fit the fluctuation of PV output at all, and the overall performance is poor, while other models can fit the changing trend of PV output. Compared with VMD-IPSO-GRU, ISSA-GRU and GRU, the prediction accuracy of VMD-ISSA-GRU is still the highest, and the MAE can reach 2.8565 kW, which is reduced by 10.69%, 4.79% and 77.62% respectively. The RMSE is also the smallest, and the R^2_{adj} can reach 0.9930, which is better than other models.

From Fig. 11 and Fig. 12, we can know that the prediction time of LSTM, GRU, and ISSA-BP is short, but the prediction accuracy is low, which can not well fit the fluctuation of PV output. The overall prediction accuracy of LSTM and GRU is close, but the prediction time of LSTM is longer, the prediction time of GRU is 27.06% shorter than LSTM, and the prediction time of ISSA-GRU is 27.98% shorter than ISSA-LSTM. The time used by the VMD-ISSA-GRU model includes the prediction time of VMD-GRU and the time of ISSA optimizing GRU. It takes 3714s for ISSA to optimize the GRU model, which is shorter than manual parameter adjustment, and the optimized parameters are more accurate. Using the GRU model optimized by ISSA, the actual prediction time of VMD-GRU is 791s. With the improvement of

TABLE 5. The comparison results of RMSE, MAE, and R^2_{adj} of each model under different weather conditions.

No.	Model	The weather conditions are relatively stable			The weather conditions fluctuate obviously.			The weather conditions fluctuated violently		
		RMSE (kW)	MAE (kW)	R^2_{adj}	RMSE (kW)	MAE (kW)	R^2_{adj}	RMSE (kW)	MAE (kW)	R^2_{adj}
1	LSTM	4.9669	3.8476	0.9924	9.2025	7.6064	0.9725	17.4440	11.9388	0.8513
2	GRU	4.9791	3.9001	0.9923	10.5367	8.3149	0.9640	18.8791	12.7621	0.8258
3	ISSA-BP	5.3852	4.1426	0.9910	14.3050	9.7492	0.9336	22.7449	15.2158	0.7472
4	IPSO-LSTM	1.9145	1.4363	0.9989	2.8637	2.1302	0.9973	5.2217	3.9260	0.9867
5	ISSA-LSTM	1.2026	0.8973	0.9996	2.9737	2.3833	0.9971	4.1467	3.0050	0.9916
6	PSO-GRU	1.8537	1.5746	0.9989	4.1273	3.2175	0.9945	6.2039	4.7834	0.9812
7	IPSO-GRU	1.7970	1.3927	0.9990	3.3015	2.4148	0.9965	5.3760	4.1191	0.9859
8	SSA-GRU	1.5841	1.1327	0.9992	3.5842	2.7680	0.9958	5.4615	3.8602	0.9854
9	ISSA-GRU	1.4707	1.1450	0.9993	2.8150	1.9440	0.9974	4.0245	3.0002	0.9921
10	VMD-IPSO-LSTM	0.9528	0.7195	0.9997	2.4213	1.7787	0.9981	3.8310	2.9339	0.9928
11	VMD-IPSO-GRU	0.8087	0.6286	0.9998	2.1615	1.6444	0.9985	4.1312	3.1983	0.9917
12	VMD-ISSA-GRU	0.6898	0.5409	0.9999	1.9933	1.4711	0.9987	3.7858	2.8565	0.9930

computer hardware, the time in practical application will be even shorter.

From Table 6 and Fig.12, we can know that the average error of VMD-ISSA-GRU in the whole test set is the smallest, and the R^2_{adj} is the highest, the RMSE can reach 1.5511 kW, and the MAE is 1.0128 kW. Compared with GRU, PSO-GRU, IPSO-GRU, and SSA-GRU, the RMSE and MAE of ISSA-GRU are the smallest, the R^2_{adj} is the highest, and the MAE is reduced by 73.56%, 27.54%, 24.10%, and 16.36%, respectively. Compared with LSTM and IPSO-LSTM, the RMSE and MAE of ISSA-LSTM are the smallest, the R^2_{adj} is the highest, and the MAE is reduced by 72.25% and 17.34%, respectively. As a result, the optimization performance of ISSA is obviously better than that of SSA, IPSO, and PSO. Compared with ISSA-LSTM, the MAE of ISSA-GRU is smaller, the RMSE of ISSA-LSTM is smaller, but the prediction time of ISSA-GRU is 27.98% shorter than that of ISSA-LSTM. It can be concluded that the prediction accuracy of ISSA-GRU and ISSA-LSTM is close, but the prediction time of ISSA-GRU is obviously better than that of ISSA-LSTM. Compared with VMD-IPSO-GRU, VMD-IPSO-LSTM and ISSA-GRU, the RMSE and MAE of VMD-ISSA-LSTM are the smallest, the R^2_{adj} is the highest, and the MAE is reduced by 11.60%, 10.12% and 27.78% respectively. As a result, the average error of VMD-ISSA-GRU in the whole test set is obviously lower than that of other models.

To sum up, the results show that optimization performance of ISSA is better than that of SSA, PSO and IPSO. Compared with other models in this paper, VMD-ISSA-GRU has the smallest RMSE and MAE in the whole test set, and the R^2_{adj} is the highest. Moreover, whether the weather conditions are stable or violent fluctuations, the VMD-ISSA-GRU model can well fit the PV output.

TABLE 6. The comparison results of MAE, RMSE, and R^2_{adj} of each model in the whole test set.

No.	Model	RMSE (kW)	MAE (kW)	R^2_{adj}
1	LSTM	7.4590	5.1063	0.9844
2	GRU	7.9934	5.3039	0.9821
3	ISSA-BP	9.5248	5.7890	0.9745
4	IPSO-LSTM	2.4945	1.7144	0.9982
5	ISSA-LSTM	2.0753	1.4172	0.9988
6	PSO-GRU	2.8207	1.9352	0.9978
7	IPSO-GRU	2.6427	1.8475	0.9980
8	SSA-GRU	2.5412	1.6765	0.9982
9	ISSA-GRU	2.1300	1.4023	0.9987
10	VMD-IPSO-LSTM	1.7073	1.1268	0.9992
11	VMD-IPSO-GRU	1.7606	1.1457	0.9991
12	VMD-ISSA-GRU	1.5511	1.0128	0.9993

C. DISCUSSION

The prediction results show that the hybrid VMD-ISSA-GRU can reduce the instability of the original PV power data, and the prediction accuracy is greatly improved compared with

other traditional models. However, the model still has some limitations. (1) This paper only verifies that the model has achieved certain advantages in the field of short-term PV prediction, and does not verify the prediction effect in other fields. (2) The optimization time of the model is long and the efficiency is not very high. To improve the efficiency of the model, this paper uses ISSA to optimize the parameters of GRU network based on the original PV power data, and does not optimize all the subsequences. If all subsequences of VMD are optimized, the prediction accuracy of the model can be improved to a certain extent, but the running time of the model will be greatly increased. (3) Due to the influence of data, this paper only analyzes the common meteorological factors, but PV power is affected by many meteorological factors, such as atmospheric cloud, air quality, energy storage configuration, and so on.

Therefore, the future work will promote the operation efficiency of the model, enhance the applicability of the model in different situations, and further improve the accuracy of PV power prediction.

V. CONCLUSION

PV power is affected by many factors, which has obvious volatility and intermittence. Most models can not mine the internal characteristics of PV power data, and can not reduce the instability of PV data. Therefore, a hybrid model for short-term PV power prediction based on VMD-ISSA-GRU is proposed. From the forecasting results, it can be concluded that: (1) Through the correlation analysis between the residual error of VMD and the original PV power data, the optimal number of decomposition modes is determined, which can effectively avoid under decomposition and over decomposition. (2) The sparrow search algorithm improved by Levy flight strategy and adaptive t-distribution has better performance than SSA, PSO and IPSO. (3) Compared with the LSTM and BP optimized by ISSA, the GRU network optimized by ISSA not only ensures the prediction accuracy, but also has higher prediction efficiency. (4) Whether the weather conditions are stable or fluctuating, the hybrid VMD-ISSA-GRU model can better fit the PV output.

REFERENCES

[1] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Wei, and G. Sun, "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *Int. J. Electr. Power Energy Syst.*, vol. 118, Jun. 2020, Art. no. 105790, doi: 10.1016/j.ijepes.2019.105790.

[2] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113315, doi: 10.1016/j.apenergy.2019.113315.

[3] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 912–928, Jan. 2018, doi: 10.1016/j.rser.2017.08.017.

[4] M. Yang and X. Huang, "Ultra-short-term prediction of photovoltaic power based on periodic extraction of PV energy and LSH algorithm," *IEEE Access*, vol. 6, pp. 51200–51205, Sep. 2018, doi: 10.1109/ACCESS.2018.2868478.

[5] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Eng. Sci. Technol., Int. J.*, vol. 21, no. 3, pp. 428–438, Jun. 2018, doi: 10.1016/j.jestch.2018.04.013.

[6] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, Jun. 2020, Art. no. 115023, doi: 10.1016/j.apenergy.2020.115023.

[7] F. Xu, J. Tong, and S. Cai, "Model cloud modeling for distributed ultra-short-term PV power forecasting," *J. Sol. Energy*, vol. 37, pp. 1748–1755, Jul. 2016.

[8] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manage.*, vol. 156, pp. 459–497, Jan. 2018, doi: 10.1016/j.enconman.2017.11.019.

[9] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Appl. Energy*, vol. 283, Feb. 2021, Art. no. 116239, doi: 10.1016/j.apenergy.2020.116239.

[10] A. A. Cardenas, M. Carrasco, F. Mancilla-David, A. Street, and R. Cardenas, "Experimental parameter extraction in the single-diode photovoltaic model via a reduced-space search," *IEEE Trans. Ind. Electron.*, vol. 64, no. 2, pp. 1468–1476, Feb. 2017.

[11] Y. Li, Y. Su, and L. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renew. Energy*, vol. 66, pp. 78–89, Jun. 2014.

[12] M. R. Douiri, "Particle swarm optimized neuro-fuzzy system for photovoltaic power forecasting model," *Sol. Energy*, vol. 184, pp. 91–104, May 2019, doi: 10.1016/j.solener.2019.03.098.

[13] S. Ding, R. Li, and Z. Tao, "A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting," *Energy Convers. Manage.*, vol. 227, Jan. 2021, Art. no. 113644, doi: 10.1016/j.enconman.2020.113644.

[14] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, and F. Qin, "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *J. Cleaner Prod.*, vol. 277, Dec. 2020, Art. no. 123948, doi: 10.1016/j.jclepro.2020.123948.

[15] A. Asrari, T. X. Wu, and B. Ramos, "A hybrid algorithm for short-term solar power prediction—Sunshine state case study," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 582–591, Apr. 2017, doi: 10.1109/TSTE.2016.2613962.

[16] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. D. Monache, "Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble," *Renew. Energy*, vol. 108, pp. 274–286, Aug. 2017, doi: 10.1016/j.renene.2017.02.052.

[17] J. Liu, W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 434–442, Apr. 2015, doi: 10.1109/TSTE.2014.2381224.

[18] V. D. William, E. Jamei, G. S. Thirunavukkarasu, M. Seyedmahmoudian, T. K. Soon, B. Horan, S. Mekhilef, and A. Stojcevski, "Short-term PV power forecasting using hybrid GASVM technique," *Renew. Energy*, vol. 140, pp. 367–379, Feb. 2019, doi: 10.1016/j.renene.2019.02.087.

[19] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renew. Power Gener.*, vol. 13, no. 7, pp. 1009–1023, May 2019, doi: 10.1049/iet-rpg.2018.5649.

[20] C.-J. Huang and P.-H. Kuo, "Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting," *IEEE Access*, vol. 7, pp. 74822–74834, 2019, doi: 10.1109/ACCESS.2019.2921238.

[21] M. Chai, F. Xia, S. Hao, D. Peng, C. Cui, and W. Liu, "PV power prediction based on LSTM with adaptive hyperparameter adjustment," *IEEE Access*, vol. 7, pp. 115473–115486, 2019, doi: 10.1109/ACCESS.2019.2936597.

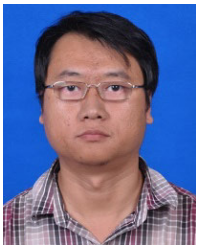
[22] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019, doi: 10.1109/ACCESS.2019.2946057.

[23] G. Li, S. Xie, B. Wang, J. Xin, Y. Li, and S. Du, "Photovoltaic power forecasting with a hybrid deep learning approach," *IEEE Access*, vol. 8, pp. 175871–175880, 2020, doi: 10.1109/ACCESS.2020.3025860.

[24] F. Mei, J. Gu, J. Lu, J. Lu, J. Zhang, Y. Jiang, T. Shi, and J. Zheng, "Day-ahead nonparametric probabilistic forecasting of photovoltaic power generation based on the LSTM-QRA ensemble model," *IEEE Access*, vol. 8, pp. 166138–166149, 2020, doi: 10.1109/ACCESS.2020.3021581.

[25] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Sep. 2014, pp. 1724–1734.
[26] R. Wang, C. Li, W. Fu, and G. Tang, "Deep learning method based on gated recurrent unit and variational mode decomposition for short-term wind power interval prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 3814–3827, Oct. 2020, doi: 10.1109/TNNLS.2019.2946414.
[27] B. Liu, C. Fu, A. Bielefield, and Y. Liu, "Forecasting of Chinese primary energy consumption in 2021 with GRU artificial neural network," *Energies*, vol. 10, no. 10, p. 1453, Sep. 2017, doi: 10.3390/en10101453.
[28] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600, doi: 10.1109/MWSCAS.2017.8053243.
[29] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information," *Renew. Energy*, vol. 118, pp. 357–367, Apr. 2018, doi: 10.1016/j.renene.2017.11.011.
[30] N. Nayak and A. K. Pani, "Short term PV power forecasting using empirical mode decomposition based orthogonal extreme learning machine technique," *Indian J. Public Health Res. Develop.*, vol. 9, no. 11, pp. 2170–2182, 2018.
[31] D. Zhang, C. Li, D. Luo, C. Wang, L. Zeng, B. Yan, and P. Wang, "Prediction of PV output based on local mean decomposition under limited information," in *Proc. 43rd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2017, pp. 5945–5950, doi: 10.1109/IECON.2017.8217031.
[32] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014, doi: 10.1109/TSP.2013.2288675.
[33] L. Wang, Y. Liu, T. Li, X. Xie, and C. Chang, "Short-term PV power prediction based on optimized VMD and LSTM," *IEEE Access*, vol. 8, pp. 165849–165862, 2020, doi: 10.1109/ACCESS.2020.3022246.
[34] M. Seyedmahmoudian, E. Jamei, G. Thirunavukkarasu, T. Soon, M. Mortimer, B. Horan, A. Stojcevski, and S. Mekhilef, "Short-term forecasting of the output power of a building-integrated photovoltaic system using a metaheuristic approach," *Energies*, vol. 11, no. 5, p. 1260, May 2018, doi: 10.3390/en11051260.
[35] J. Xue and B. Shen, "A novel swarm intelligence optimization approach: Sparrow search algorithm," *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, Jan. 2020, doi: 10.1080/21642583.2019.1708830.
[36] G. Iacca, V. C. dos Santos Junior, and V. V. de Melo, "An improved Jaya optimization algorithm with Lévy flight," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113902, doi: 10.1016/j.eswa.2020.113902.
[37] Z. Chang, Y. Zhang, and W. Chen, "Electricity price prediction based on hybrid model of Adam optimized LSTM neural network and wavelet transform," *Energy*, vol. 187, Nov. 2019, Art. no. 115804, doi: 10.1016/j.energy.2019.07.134.
[38] C. Zhang and S. Ding, "A stochastic configuration network based on chaotic sparrow search algorithm," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106924, doi: 10.1016/j.knosys.2021.106924.
[39] U. Garciarena and R. Santana, "An extensive analysis of the interaction between missing data types, imputation methods, and supervised classifiers," *Expert Syst. Appl.*, vol. 89, pp. 52–65, Dec. 2017, doi: 10.1016/j.eswa.2017.07.026.
[40] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access J. Power Energy*, vol. 7, pp. 376–388, 2020, doi: 10.1109/OAJPE.2020.3029979.

PENGYUN JIA was born in Shanxi, China, in 1993. He is currently pursuing the master's degree with North China Electric Power University. His research interests include new energy generation and grid-connected operation.



HAIBO ZHANG (Senior Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 2005. He is currently a Professor with North China Electric Power University, Beijing. His research interests include energy management systems (EMS), and new energy power system planning, operation, and control.

XINMIAO LIU is currently a Senior Engineer with the Power System Planning Research Center, Guangdong Power Grid Company Ltd., Guangzhou, China. His research interests include power system planning and energy storage application planning.

XIANFU GONG is currently a Senior Engineer with the Power System Planning Research Center, Guangdong Power Grid Company Ltd., Guangzhou, China. His research interests include energy and power development planning and energy storage application planning.