IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 July 2024, accepted 11 August 2024, date of publication 21 August 2024, date of current version 2 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3446790

RESEARCH ARTICLE

Data-Driven Modeling for Photovoltaic Power Output of Small-Scale Distributed Plants at the 1-s Time Scale

JIA WEI, (Student Member, IEEE), WEIJIA YANG, (Member, IEEE), XUDONG LI, (Member, IEEE), AND JUNSONG WANG

State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China

Corresponding author: Weijia Yang (weijia.yang@whu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 52079096, and in part by the Natural Science Foundation of Hubei Province of China under Grant 2024AFA058.

ABSTRACT Under the condition of a small time scale (e.g. second), distributed photovoltaic (PV) power generation output has the problems of strongly fluctuating and difficult to accurately simulate. It affects the control strategy and operation mode of hybrid energy systems. To address this problem, a data-driven small-scale distributed PV plant power output model on a 1-second time scale is proposed for the generation of second-by-second PV power output scenarios in hybrid energy systems. Firstly, this work analyzes the characteristics of PV power output at the 1-second time scale based on the probability distribution of power output fluctuations. Secondly, an index system that characterizes the PV power output fluctuation characteristics at the 1-second time scale is constructed. Then, using the data-driven method, a BP neural network model is constructed to simulate the PV power output at the 1-second time scale. Finally, a simulation is performed using the measured data from the PV plant. The findings demonstrate that compared to PV power output models in seconds based on Pearson systematic random numbers: (1) The correlation coefficient (r) of the proposed model is more than 0.8, in a higher degree of fit; (2) The root mean square error (RMSE) of the proposed model achieves 0.005, generally representing a 37.12% reduction. Overall, both the time scale and model accuracy of this model have deep potential value in PV power output modeling and system regulation.

INDEX TERMS Photovoltaic, data-driven, second-time scale, fluctuation characteristics, hybrid energy system.

I. INTRODUCTION

A. BACKGROUND

In recent years, as the threat posed by human-induced climate change has intensified, the advantages of clean energy generation have become apparent [1]. China has set the goal of achieving a ''carbon peak'' by 2030 and put forward the vision of ''carbon neutrality'' by 2060 [2]. The country has also made great efforts to promote the integration of source-grid-load-storage and the development of multi-energy complementation [3]. Photovoltaic (PV) power generation, as a form of clean energy generation, has been developing rapidly in recent years [4]. PV has been widely used currently, and it will have a profound impact on the operation of power systems.

Currently, there are two types of prevalent PV power generation methods: distributed and centralized. Distributed photovoltaic power generation is smaller in scope than centralized photovoltaic power generation; part of the power generated is for self-consumption, and the remaining amount is connected to the power grid. When connected to the grid, distributed PV power output volatility affects scheduling, regulation, and system functioning [5].

PV power output is susceptible to volatility due to factors such as solar irradiance, temperature, and humidity. The pumped storage power station, a more developed type of

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

energy storage, exhibits excellent fast response and high adjustability, used in hybrid energy systems in tandem with PV power plants could lessen light abandonment and the effect that oscillations in PV power output have on the grid [6].

At present, there is are relatively few studies focusing on simulation modeling at the second time-scale in hybrid energy systems, and there still offers significant potential for improvement. In this work, our focus is on the second time-scale dynamics, particularly emphasizing the establishment of a PV power output model to generate scenarios for second-by-second PV power output in simulation studies of hybrid energy systems.

B. LITERATURE REVIEW

Aiming at the problem of joint operation of hybrid energy systems containing PV power generation, a systematic analysis of the active regulation of hybrid energy systems has been carried out by previous researchers. Hydropower is the largest capacity and most realistic way to regulate photovoltaic (PV) power fluctuation at present [7]. Riza et al. study the mechanisms of how the micro-hydro sub-systems work to compensate for the inconvenience found in the PV via a PV-micro-hydro hybrid system installed in Indonesia since 1989 [8], and it could be concluded that modeling the uncertainty of PV power is important for system stability. Some scholars also establish the model of hydro-PV hybrid energy system [9], [10], [11]. However, the operation mode of the hybrid energy system under the condition of the small-time scale still needs to be optimized. The development of an accurate PV power output model at the 1-second time scale is beneficial to enhancing the operating efficiency of PV power plants, the operating condition of the power system, and the economic operation of the hybrid system.

In the research on PV power output characteristics, some scholars use Gaussian distribution to analyze the fluctuation characteristics of PV power output and establish a PV power output fluctuation model accordingly. For example, Wang et al. model PV power output fluctuations based on t-location distribution functions [12], and Wang et al. model PV power output fluctuations based on mixed Gaussian distribution [13]. However, these models simply analyze and simulate the fluctuation features of PV power output, without translating them into PV power output curves, and they cannot be used for joint operation of grid-connected PV systems.

In the research on PV power output modeling, scholars have done a lot of studies on PV power output modeling, and some research results have been obtained. We distinguished here by modeling methods and time scales, there is also some overlap between the two. On the time scale with a resolution of 1h, scholars typically applied statistical methods to establish the PV power output model by statistical and curve fitting of historical data: Guan et al. established the PV power output model by multi-scale cluster analysis [14]; Sheng et al. propose a model employed the weighted Gaussian process regression approach [15]. On the time scale with a resolution of 5min-15min, Some scholars established PV power output models based on the physical causes of PV power output: Hummon, et al. established a PV power output model with an accuracy of 4s based on the temporal and spatial Correlation of PV power output [16]; Shi et al. model and analysis the stability of grid-connected PV system via variation of solar irradiance and cell temperature [17]; Yang et al. propose a novel multi time-scale data-driven PV power model that leverages both spatial and temporal correlations among neighboring solar site [18]. In recent years, with the development of artificial neural network algorithm technology based on deep learning, more scholars have applied deep learning algorithms to establish PV power output modeling on the time scale with a resolution of 5min-15min [19], and the accuracy has been improving. For example, Ospina et al. proposed a novel hybrid wavelet-based LSTM-DNN forecasting model designed to short-term forecast PV power [20]; Li et al. propose a hybrid deep learning model combining wavelet packet decomposition (WPD) and long short-term memory (LSTM) networks to utilize for one-hour-ahead PV power forecasting with five-minute intervals [21]; Zhang et al. propose a regional PV power prediction model based on the dynamic graph convolutional layer and bidirectional long short-term memory (LSTM) module [22]; Miao et al. utilize a backpropagation neural network (BPNN) model improve the accuracy of PV power output predictions [23]. However, the time scale of the majority of these models is clustered at 5min-1h, and there are fewer works on simulating PV power output on a 1-second time scale. Moreover, these models are largely focused on physical elements like weather to develop prediction models, and they rarely simulate PV output from data attributes.

The data-driven algorithm is an artificial intelligence algorithm based on data. Data-driven is a systematic approach to improve the data and model by deriving/adding features to address the problem identified during the iterative loop of forecasting model development [24]. In recent years, the application of data-driven approaches in PV power output modeling has been steadily increasing. González Ordiano et al. established a model for PV power output forecast based on the data-driven approach, and this model doesn't require any weather data [25]; Kang et al. utilized data-driven methods and Long Short-Term Memory (LSTM) networks to achieve one-minute-ahead prediction of PV power output [26]; Sonia et al. utilized data-driven methods present a time series based energy forecasting system using a solar home database [27].

Moreover, PV power output modeling on a small time scale has been gradually studied recently. Meng et al. propose a PV power output model based on ANN, and this model has a temporal resolution of 1 min [28]. The study of the impact of PV power production fluctuations on the system at small time scales can be advanced by modeling PV

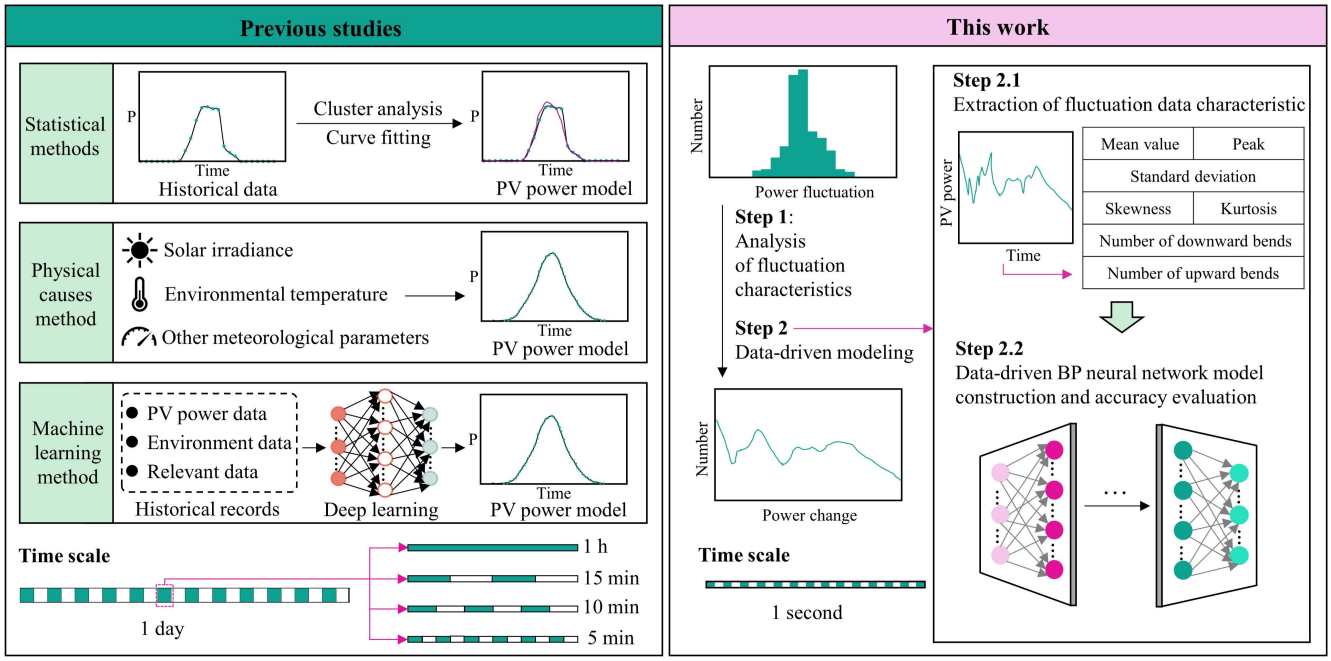


FIGURE 1. Comparison chart of technical routes.

power output on a 1-second time scale. Several research on the properties of PV power output at 1-second time scale have also been undertaken by scholars. Shedd et al. studied the characteristics of PV power output fluctuations on different small-time scales through numerical analysis and concluded that the amplitude of PV power output fluctuations increases with the increase of time scales [29]; Marcos et al. analyzed the fluctuations of PV power output on second-time scales for different PV power plant of different sizes of PV power plants, and obtained the conclusion that the trend of PV power output fluctuations is smoother the larger the PV power plant scale [30]; Yang et al. established a mathematical model to get 1-second time scale PV power output from the PV power output fluctuation characteristics using a statistical method based on Pearson system random numbers [31], but the model's accuracy needs to be improved.

In conclusion, the current PV power output modeling research has much space to improve in terms of time scale and model accuracy. An important highlight of this work is achieving higher model accuracy while shortening the time scale to 1 second. The comparison between previous research and this work is listed here, as shown in FIGURE 1.

C. THIS WORK

The main contributions of this work are summarized below:

(1) In terms of time scale, PV power output fluctuation characteristics of distributed PV power Plants in a 1-second time scale are investigated using measured PV power output data, and a fluctuation characteristics index system is proposed accordingly;

(2) In terms of model accuracy, data-driven and BP neural network methods are applied to realize PV power output modeling with input fluctuation characteristic index output 1-second time scale PV power output time series, and the accuracy is improved by 37.12% in general.

Moreover, a model of PV power output at 1-second time scale is proposed in this work, and it could generate scenarios for second-by-second PV power output in simulation studies of hybrid energy systems. Therefore, it has the potential to contribute to the study of control strategies under small time scale conditions for hybrid energy systems.

This work is based on the measured data of power output of a small-scale distributed PV power plant, the content is organized as follows.

Section II analyzes PV power output characteristics at 1-second time scale based on the probability distribution of PV power output fluctuation. Section III establishes the index system describing 1-second time scale fluctuation characteristics of PV power output based on the fluctuation characteristics, and establishes a BP neural network model to simulate the PV power output under second time scales by using the data-driven methodology. Section IV validates the validity and accuracy of this model by combining it with measured data and calculating examples.

II. FLUCTUATION CHARACTERISTIC AT 1-SECOND TIME SCALE

The data in this work is from a ground-mounted PV array located at the National Institute of Standards and Technology (NIST), in Gaithersburg, U.S.A., using two consecutive years of PV power output data with a temporal resolution of

1-second for a small-scale distributed PV plant with a rated power output of 271 kW, as shown in FIGURE 2 [32].

Define the amount of change in the PV generation system power output as the PV power output fluctuation, and its expression is

$$\Delta P_i = P_{i+1} - P_i \tag{1}$$

where: ΔP_i represents the amount of PV power output fluctuation, P_i represents the value of PV power output at the ith s, and P_{i+1} represents the value of PV power output at the $(i+1)$th s.

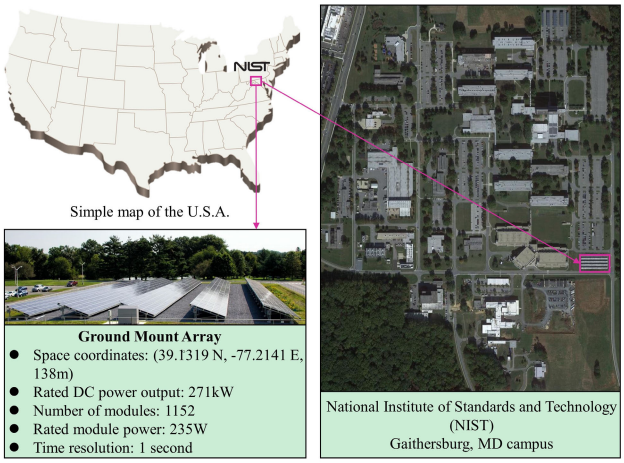


FIGURE 2. Location and schematic diagram of the distributed PV plants of this case in the U.S.A.

Calculate the PV power output fluctuation at 1-second resolution for all output times of all output periods for multiple days and use the histogram of statistical distribution to represent the probability distribution of the fluctuation, as shown in FIGURE 3.

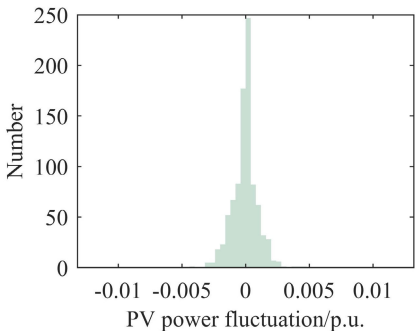


FIGURE 3. Statistical distribution of PV power output fluctuations at 1s resolution over multiple days.

According to FIGURE 3, the probability distribution of the PV power output fluctuation amount at 1s resolution, compared to the typical normal distribution, exhibits a larger peak and a thinner tail. Moreover, it shows the characteristic of obviously anomalous spiking. It indicates that the PV power output is more likely to exhibit small fluctuations and less likely to show large perturbations at the 1-second time scale.

However, it doesn't mean that the PV power output fluctuations all over the 1-second time scale, there are also momentary excessive amounts of fluctuations. It can be verified in the PV power output curves. FIGURE 4 shows the PV power output curves for a day with a total duration of 15 min. The significant fluctuations in photovoltaic output may be correlated with sudden changes in cloud cover and other related factors. Most of the curves show smooth power output, but large fluctuations can be observed. The amount of fluctuation in 1 s can even exceed 10% of the rated power output. In order to maintain the stability of the power system, other power generation in the grid needs to respond in time.

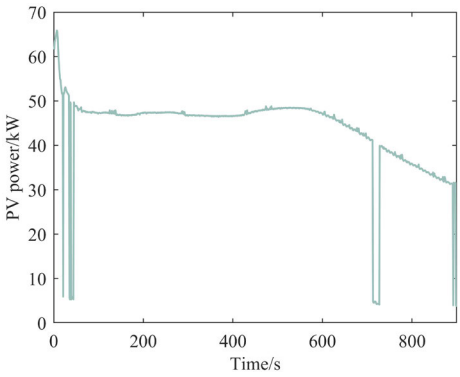


FIGURE 4. PV power output curve for a total duration of 15 minutes on a representative day.

The amount of fluctuation in the PV power output at the 1-second time scale for ten consecutive days is analyzed, as shown in Table 1. There are instances where there are more than 900 fluctuation points at the 1-second time scale, meaning that the variation in power output is greater than 5% of the rated power output. This highlights the significance of identifying and modeling PV power output fluctuations at 1-second time scale.

TABLE 1. Calculation of the amount of fluctuation in PV power output at 1-second time scale for ten consecutive days in 2017.

Date	09-29	09-30	10-01	10-02	10-03
Number of fluctuation points exceeding 5% of rated capacity	814	904	14	0	27
Number of fluctuation points exceeding 10% of rated capacity	87	69	0	0	16
Date	10-04	10-05	10-06	10-07	10-08
Number of fluctuation points exceeding 5% of rated capacity	0	32	16	312	0
Number of fluctuation points exceeding 10% of rated capacity	0	18	13	28	0

III. MODELING

A. METHODOLOGY

In this work, a data-driven model for PV power output at 1-second time scale is established based on the fluctuation characteristics. Data-driven approaches and BP neural network algorithms are the main methodologies used.

Data-driven is a method that uses systematic measurements as input and output data for experiments and creates models using data processing [33]. In recent years, data-driven artificial intelligence technology has been developing rapidly, gradually moving from shallow traditional machine learning to deep learning [34], and its application in the energy sector is gradually increasing [35]. Considering the high degree of stochasticity of PV power output, this work employs a data-driven modeling technique, with the goal of mining general relationships from a huge quantity of data to estimate PV power output.

BP neural network, one of the key deep learning techniques, is a multilayer forward network with adjustable error propagation in the backward direction and forward direction of signals [36]. It is quite flexible when it comes to tackling highly stochastic nonlinear issues. The primary idea of the BP neural network algorithm is learning by stimulation with a vast amount of sample data. During training, the feedback signals are continuously adjusted to the weights and thresholds, and training is repeated repeatedly to produce the intended results [37].

The BP neural network is composed of input, hidden, and output layers, with the algorithm flow shown in FIGURE 5. The network may learn intricate correlations between inputs and outputs and process information via numerous levels thanks to this architecture.

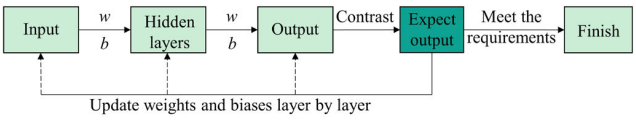


FIGURE 5. Flowchart for BP neural network algorithm.

B. FRAMEWORK

Applying the research methodology mentioned in section III-A, The process of PV power output data-driven modeling at 1-second time scale is shown in FIGURE 6.

The research framework of this model is specified as follows:

(1) Choose the PV power output data at 1-second time scale from the entire data set, and then perform data preprocessing, including normalizing the data and removing data sets with outliers and voids.

(2) Divide the data set into training samples and test samples. Extract the fluctuation characteristic of the PV power output dataset as the input variable and the PV power output dataset as the output variable.

(3) Input the training samples into the BP neural network for training.

(4) Enter the test samples into the trained BP neural network after training is finished, and assess whether the output PV power output at 1-second time scale satisfies the accuracy requirement.

(5) If the accuracy requirement is not met, adjust the parameters and re-do the model construction until the accuracy requirement is met.

C. FLUCTUATION CHARACTERISTIC INDEX SYSTEM

To realize the simulation of PV power output by the fluctuation characteristics of PV output data, it is necessary to establish a fluctuation characteristic index system. This index system mainly describes the fluctuation of PV power output by characterizing the probability distribution of PV power output fluctuation at 1-second time scale, including the mean, standard deviation, kurtosis, skewness, kurtosis of the probability distribution of PV power output fluctuation, and the number of times of bends of PV power output, as shown in the FIGURE 7.

(1) PC_1: Mean value

The mean value (μ) is the average of the fluctuation amount of PV power output, reflecting the central tendency of the fluctuation data. The formula is

$$PC_1 = \mu = \frac{\sum_{i=1}^{n} \Delta P_i}{n} \tag{2}$$

where: ΔP_i represents the amount of PV power output fluctuation, n represents the number of samples, and the following symbols have the same meaning as before.

(2) PC_2: Standard deviation

The standard deviation (σ) is the arithmetic average of the squared differences between the PV power output fluctuation and its mean value, reflecting the degree of dispersion of the fluctuation data. The formula is

$$PC_2 = \sigma = \sqrt{\frac{\sum_{i=1}^{n} (\Delta P_i - \mu)^2}{n}} \tag{3}$$

(3) PC_3: Peak

The peak value (P_k) is the quantity of fluctuation data corresponding to the center of the probability distribution graph of the PV power output. It represents the maximum value of the distribution and reflects the concentration of the fluctuation data. A higher peak value indicates a larger smooth region in the PV power output.

(4) PC_4: Skewness

Skewness (S) is used to measure the direction and degree of deviation in the distribution of fluctuation data, reflecting the asymmetry of the data distribution. The larger the absolute value of the skewness, the greater the degree of skewness in the data distribution. The formula is

$$PC_4 = S = \frac{1}{n} \sum_{i=1}^{n} \left[\left(\frac{P_i - \mu}{\sigma} \right)^3 \right] \tag{4}$$

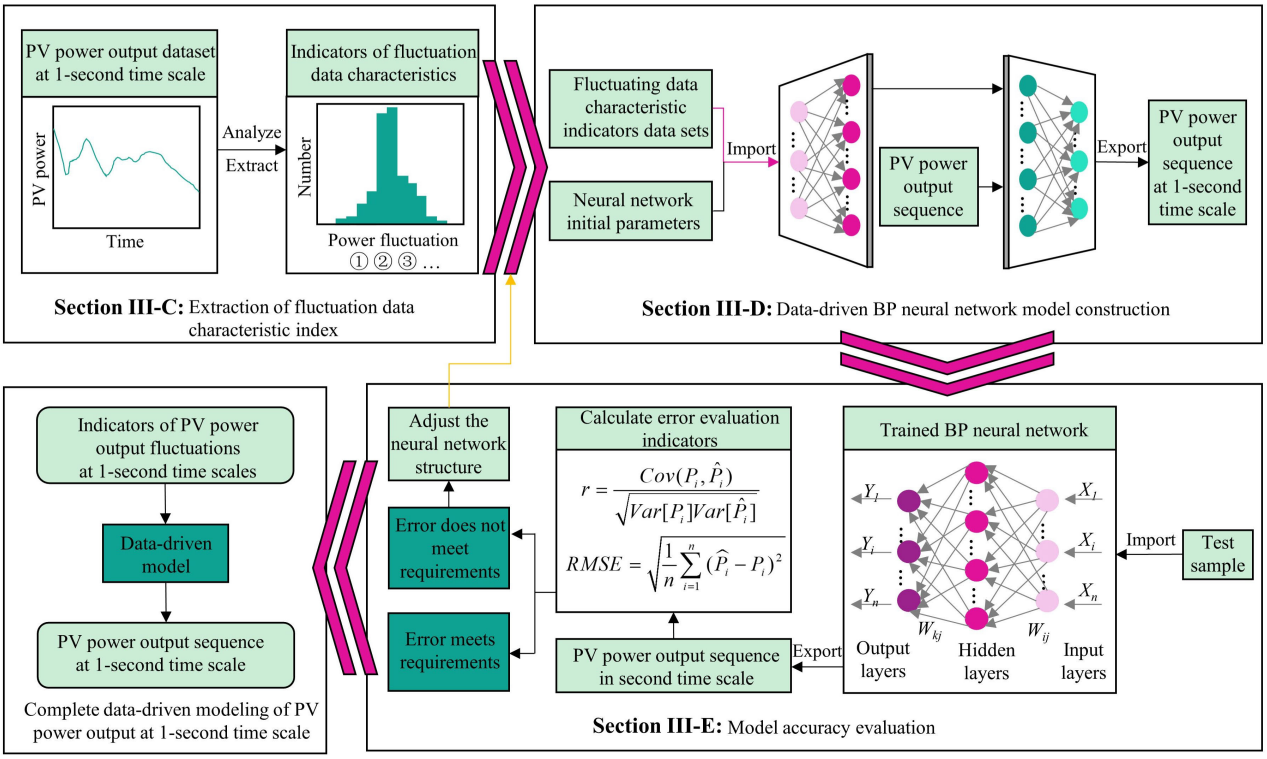
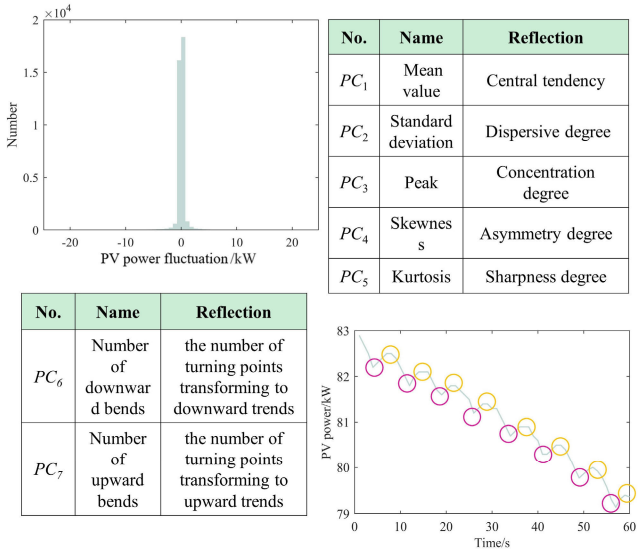


FIGURE 6. Flowchart for modeling of PV power output data-driven at 1-second time scale.



No.	Name	Reflection
PC_1	Mean value	Central tendency
PC_2	Standard deviation	Dispersive degree
PC_3	Peak	Concentration degree
PC_4	Skewness	Asymmetry degree
PC_5	Kurtosis	Sharpness degree

No.	Name	Reflection
PC_6	Number of downward bends	the number of turning points transforming to downward trends
PC_7	Number of upward bends	the number of turning points transforming to upward trends

FIGURE 7. Flowchart for modeling of PV power output data-driven at 1-second time scale.

(5) PC_5: Kurtosis

Kurtosis (K) is used to measure the steepness of the distribution of fluctuation data. Intuitively, kurtosis reflects the sharpness of the peak. The larger the kurtosis, the steeper the distribution. The formula is

$$PC_5 = K = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{P_i - \mu}{\sigma}\right)^4\right] \tag{5}$$

(6) PC_6: Number of downward bends

The number of downward bends r_l is the number of turning points where PV power output changes from an upward trend to a downward trend, expressed as the number of local maximum points in the PV power output curve for a given period.

$$PC_6 = r_l \tag{6}$$

(7) PC_7: Number of upward bends

The number of upward bends r_h is the number of turning points where PV power output changes from a downward trend to an uptrend, expressed as the number of local minimum points in the PV power output curve for a given period.

$$PC_7 = r_h \tag{7}$$

D. BP NEURAL NETWORK CONSTRUCTION

The neural network is shown in FIGURE 8.

1) INPUT LAYER

In order to examine the impact of PV power output fluctuations at 1-second time scale, this work focuses on seven fluctuation data characterization metrics, as presented in section III-C. Therefore, utilize these seven indicators as input variables.

2) OUTPUT LAYER

The model investigates the effect of different fluctuation characteristics on PV power output. Considering that a

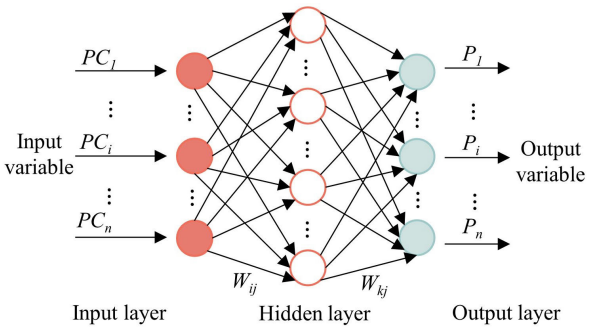


FIGURE 8. Flowchart for modeling of PV power output data-driven at 1-second time scale.

minimum time resolution of 15 minutes is required for ultra-short-term prediction of PV power production and ultra-short-term dispatch of electric power systems, PV power output data at a 1-second time scale with a total duration of 15 minutes is identified as the model output variable. Therefore, the output layer variable is 900 continuous PV power output data with a time resolution of 1 s.

3) HIDDEN LAYER

The number of hidden layers and nodes directly affects the accuracy and size of the model. The number of nodes in the output layer of the BP neural network used in this study is high. To strengthen the stability of the model, according to the least mean square (LMS) [38], the formula to calculate the number of nodes in the hidden layer is as follows:

$$s = \sqrt{0.43mn + 0.12n^2 + 2.54m + 0.77n + 0.35} + 0.51 \tag{8}$$

where: s represents the number of hidden layer nodes, m represents the number of input layer nodes, and n represents the number of output layer nodes.

E. EVALUATION INDICATORS

To verify the accuracy of the model, the correlation coefficient (r) and the Root Mean Square Error (RMSE) are used as the evaluation indicators of the accuracy of the model. The formula is

$$\mathrm{r} = \frac{Cov(P_i, \hat{P}_i)}{\sqrt{Var[P_i]Var[\hat{P}_i]}} \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{P}_i - P_i)^2} \tag{10}$$

where: n represents the number of samples; P_i represents the actual value of PV power output; $\hat{P}_i$ represents the simulated value of PV power output; $Cov(P_i, \hat{P}_i)$ represents the covariance between the real value and the simulated value; $Var[P_i]$ represents the variance.

The r describes the degree of linear correlation between the fitted values and the true values [39], in the range $[-1, 1]$. The closer the r is to 1, the stronger the model correlation. The RMSE describes the deviation of each data from the actual value [40], in the range $[0, +\infty)$. The closer the RMSE is to 1, the smaller the model error.

IV. CASE STUDY

A. SCENARIO SETTING

To verify the accuracy of the model in this work, the model-building process above is programmed in the PyCharm platform. Firstly, carry out the data processing, select the PV power plant during 2017-2018 a total of 730 days, sampling time interval of 1-second PV power output measured data. To improve the training accuracy, the PV power output during the period of relatively high daily light irradiance was selected. According to the temporal and spatial correlation of PV power output, this simulation finally takes the daily PV power output data at 12:00-13:00 and divides it into a total length of 15 minutes.

Divide the first 511d as the training set and the last 189d as the test set. Eliminate all problem datasets containing outliers and vacancies, and finalize 1900 sets of training set data and 209 sets of validation set data. To eliminate the influence of the scale and other physical quantities, all the output data are normalized as follows:

$$P'_t = \frac{P_t}{P_R} \tag{11}$$

where: P'_t represents a moment of PV power output per unit, p.u.; P_t represents a moment of PV power plant power, kW; P_R represents the rated power output of PV power plant, kW.

According to section III-D, the number of nodes in the input layer of the BP neural network established for this simulation is 7 and the number of nodes in the output layer is 900. Based on formula (6) yields, the number of nodes in the hidden layer is approximately 317. After model training and validation, the number of hidden layers of the BP neural network used in this model is finally determined to be 3 and the number of nodes is 320.

Set the learning rate of the BP neural network algorithm to 0.001, and establish the BP neural model as shown in FIGURE 9.

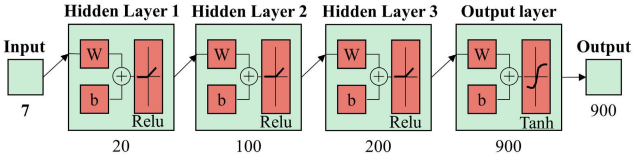


FIGURE 9. Flowchart for modeling of PV power output data-driven at 1-second time scale.

B. RESULT AND DISCUSSION

1) FLUCTUATION CHARACTERISTICS ANALYSIS

Two sets of data were randomly selected and the amount of force fluctuation was calculated and represented in the frequency histogram as shown in FIGURE 10.

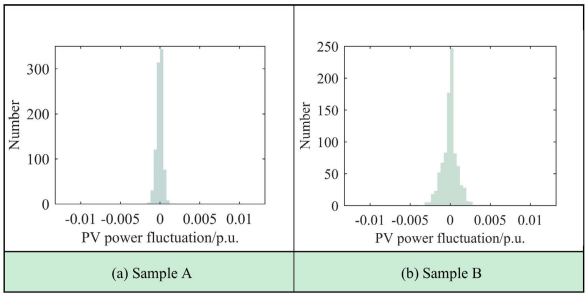


FIGURE 10. Distribution of PV power output fluctuations at the 1-second time scale of samples.

According to the frequency distribution histograms of the PV power output fluctuations, both of them show high peaks and thin tails with obvious anomalous spikes, in line with the fluctuation characteristics of PV power output at the 1-second time scale proposed in Section II.

Calculate the fluctuation characteristic index based on the contents of Section III-C, as shown in Table 2. Compare the fluctuation characteristic index of Sample A and Sample B:

(1) The means are approximately equal, indicating that the data concentration trend is similar and the amount of fluctuation is small overall.
(2) The standard deviation of Sample B is larger, indicating that the Sample B fluctuation data is more discrete.
(3) The peak of sample A is larger, indicating that the fluctuation data of sample A is more concentrated. It also means that the power output of sample A is smoother.
(4) The skewness and kurtosis of both samples are similar, indicating that the direction and degree of the steepness and deviation are similar.
(5) The number of bends the sample A is greater, indicating that the overall change trend of sample A is more.

TABLE 2. PV power output fluctuation characteristic index at the 1-second time scale of the samples.

No.	Fluctuation characteristic	Sample A	Sample B
PC_1	Mean value	6.0×10^{-5}	6.65×10^{-5}
PC_2	Standard deviation	0.000382	0.00103
PC_3	Peak	347	245
PC_4	Skewness	4.90	4.364553
PC_5	Kurtosis	25.78	24.46744
PC_6	Number of downward bends	174	76
PC_7	Number of upward bends	175	77

2) TIME DOMAIN CHARACTERISTICS ANALYSIS

After model training, the PV power output model established is used to realize the simulation of PV power output from the fluctuation characteristics. The simulated PV power output curves of the model in this work and the model based on Pearson random numbers are compared with the measured values at the 1-second time scale, as shown in FIGURE 11.

In terms of the resultant error of the model output power, the RMSE of the PV power output in the time domain of the data-driven model proposed in this work reaches 0.0053 and 0.0206, respectively, reduced by 86.4% and 47.8% compared with the model based on Pearson's random numbers, respectively.

In terms of the correlation between the results of the model power output and the measured data in the time domain, the r of the model proposed in this work are all more than 0.8, indicating a strong correlation between them. Moreover, the degree of correlation is greatly improved compared with the model based on Pearson random numbers, indicating that the model proposed in this work is well-fitted.

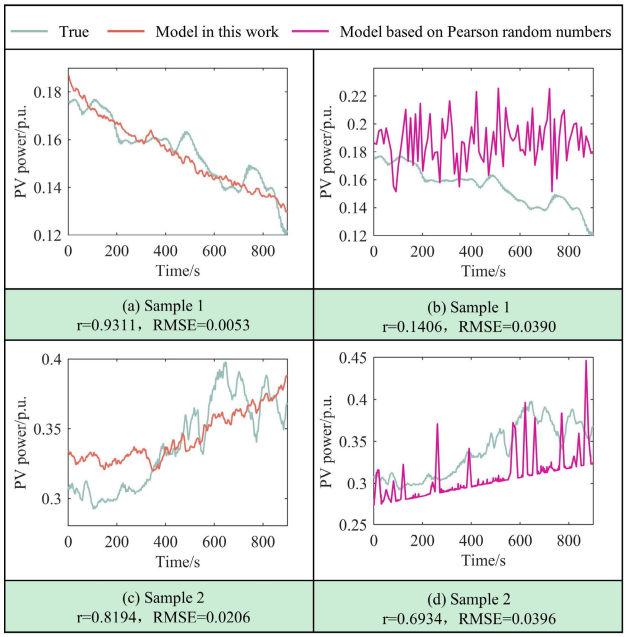


FIGURE 11. Comparison of PV power output model power output in time domain.

3) FREQUENCY DOMAIN CHARACTERISTICS ANALYSIS

Considering the role distributed PV power plants play in the grid, the PV power output from the two models is converted in the frequency domain and compared with the frequency domain curves of the actual power output, as shown in FIGURE 12.

In terms of resultant error of the model power output in the frequency domain, the RMSE of PV power output in the frequency domain from the data-driven model proposed in this work reaches 0.2645, reduced by 65% to 90% compared to the model based on Pearson random numbers.

In terms of correlation between the results of the model power output and the measured data in the frequency domain, the r of both two models exceeds 0.999, but the r of the data-driven model proposed in this work is still higher than the model based on Pearson random numbers.

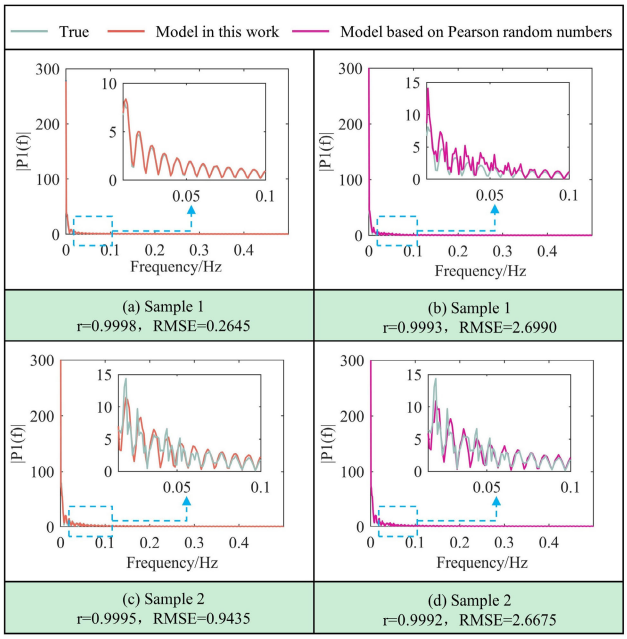


FIGURE 12. Comparison of PV power output model power output in frequency domain.

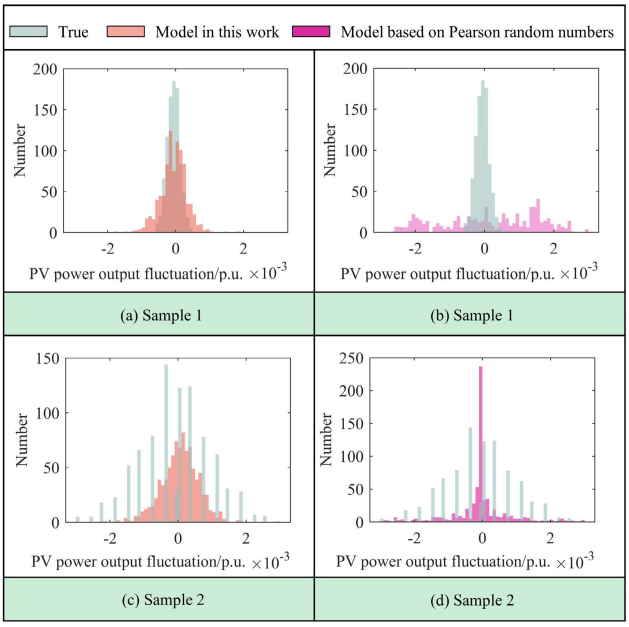


FIGURE 13. Comparison of PV power output model power output in fluctuation distribution.

Moreover, it indicates in terms of the effect on the stability and stabilization of the system, the data-driven model proposed in this work is more similar to the real power output.

4) FLUCTUATION DISTRIBUTION DOMAIN CHARACTERISTICS ANALYSIS

The amount of PV power output fluctuation is calculated and represented via Gaussian distribution, as shown in FIGURE 13.

Comparing the histograms of power output fluctuations of the data-driven model proposed in this work and the model based on Pearson random numbers with the actual real power output, it can be found that the power output fluctuation of the model based on Pearson random numbers more in line with the characteristics of anomalous spiking at 1-second time scale in sample 2. However, the data-driven model proposed in this work has a better fit with the real power fits real event data, making it more appropriate for simulating PV power output simulations at the 1-second time scale.

To quantitatively analyze the advantages and disadvantages of the two models in terms of fluctuation distribution, calculate the fluctuation characteristic index of the PV output generated by both models, as shown in Table 3 and Table 4. For the sake of simplicity, we use M1 to represent the data-driven model proposed in this work and M2 to represent the model based on Pearson random numbers. It is evident that the PV power output model proposed in this work (M1) is closer to the actual values in terms of fluctuation distribution characteristics, particularly in terms of mean value, standard deviation, kurtosis, and skewness. This also confirms that this model is more suitable for generating PV power output scenarios at the 1-second time scale.

TABLE 3. Table of fluctuation characteristic index for Sample 1 from two models.

No.	Fluctuation characteristic	True	M1	M2
PC_1	Mean value	-6.02×10^{-5}	-6.31×10^{-5}	-0.81×10^{-5}
PC_2	Standard deviation	3.82×10^{-4}	1.91×10^{-4}	24×10^{-4}
PC_3	Peak	124	185	41
PC_4	Skewness	2.2546	3.0607	1.1253
PC_5	Kurtosis	7.0948	11.2696	4.4994
PC_6	Number of downward bends	181	85	32
PC_7	Number of upward bends	182	86	32

5) SENSITIVITY ANALYSIS

To verify the sensitivity of the data-driven model proposed in this work, 30 groups of random scene samples are randomly selected and analyzed for error. In this part, we still use M1 to represent the data-driven model proposed in this work and M2 to represent the model based on Pearson random numbers. Compare the RMSE and r of the output results of the two models in 30 groups of random scenes, as shown in FIGURE 14 and FIGURE 15. M1 has a smaller RMSE compared to M2 overall, indicating lower error. Additionally, the r of M1 is closer to 1 than that of M2, suggesting a better fit overall.

Calculate the average RMSE and the r values for results from different models, and list them in Table 5. The average

TABLE 4. Table of fluctuation characteristic index for Sample 2 from two models.

No.	Fluctuation characteristic	True	M1	M2
PC_1	Mean value	6.64×10^{-5}	5.91×10^{-5}	5.72×10^{-5}
PC_2	Standard deviation	9.53×10^{-4}	5.39×10^{-4}	35×10^{-4}
PC_3	Peak	152	82	238
PC_4	Skewness	2.6619	1.4805	6.6080
PC_5	Kurtosis	903601	3.8964	48.1725
PC_6	Number of downward bends	76	92	118
PC_7	Number of upward bends	77	92	118

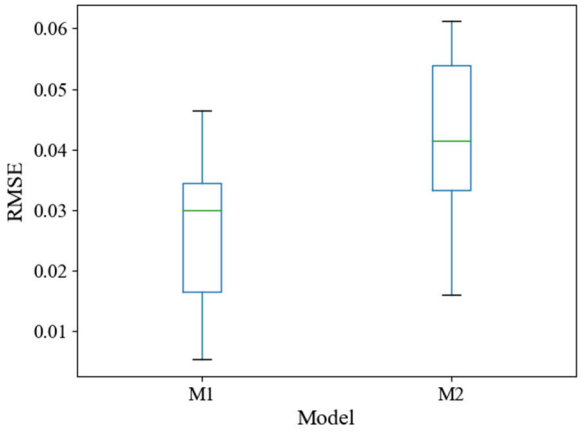


FIGURE 14. Comparison of box plot of RMSE of the model.

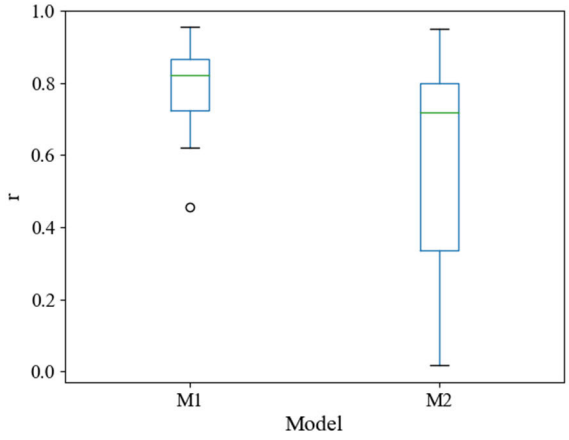


FIGURE 15. Comparison of box plot of the r of the model.

RMSE of M1 decreased by 37.12% compared to M2, while the r increased by 32.07% compared to M2. This also indicates that M1 has a lower error and better fit compared to M2.

TABLE 5. Comparison table of evaluation indicators for different models.

Evaluation indicators	M1	M2
RMSE	0.0271	0.0431
r	0.7931	0.6005

Based on the above, the data-driven model proposed in this study (M1) demonstrates better fitting capability and stability.

V. CONCLUSION

Based on observed PV power plant output data, this work studies the PV power output characteristics and proposes a data-driven, small-scale distributed PV power plant output model, all at the 1-second time scale. Moreover, the accuracy of this model is verified by comparing it with the observed data and the model developed based on Pearson random numbers, proposed by previous scholars. The main conclusions are as follows:

(1) In terms of time scale, a fluctuation characteristic index system is established via the analysis of the observed data, and the data-driven model proposed in this work simulates the PV power output at the 1-second time scale via the fluctuation characteristics.

(2) In terms of model accuracy, the results of the simulation show that the PV power output model proposed in this work at 1-second time scale is feasible. The model output correlation coefficient (r) exceeds 0.8, and the RMSE can be as high as 0.005, generally representing a 37.12% reduction compared to the model based on Pearson random numbers.

However, there are still some limitations and suggested future work for this paper:

(1) In terms of model accuracy, more representative indicators of fluctuation characteristics could be mined and used in this model. At the same time, this study did not gather more measured PV power output data at 1-second time scales, and obtaining more measured data would equally enhance the accuracy of the model.

(2) In terms of model generalizability, more calculations are required to evaluate the validity of this method, particularly conducting comparisons across different time scales.

(3) In terms of hybrid energy systems, based on this work for generating scenarios for second-by-second PV power output, this model can be combined with pumped storage and other power generation system models. This combination allows for research on the control strategy of hybrid energy systems under small time scale conditions, ultimately promoting the development of hybrid energy systems.

REFERENCES

[1] P. Norrlund, L. Saarinen, A. Witt, B. Smith, J. Yang, and U. Lundin, "Burden on hydropower units for short-term balancing of renewable power systems," *Nature Commun.*, vol. 9, Jul. 2018, Art. no. 2633, doi: 10.1038/s41467-018-05060-4.

[2] X. Guo, X. Chen, X. Chen, P. Sherman, J. Wen, and M. McElroy, "Grid integration feasibility and investment planning of offshore wind power under carbon-neutral transition in China," *Nature Commun.*, vol. 14, no. 1, Apr. 2023, Art. no. 2447, doi: 10.1038/s41467-023-37536-3.

[3] N. Zhang, Z. Hu, B. Shen, S. Dang, J. Zhang, and Y. Zhou, "A source–grid–load coordinated power planning model considering the integration of wind power generation," *Appl. Energy*, vol. 168, pp. 13–24, Apr. 2016, doi: 10.1016/j.apenergy.2016.01.086.

[4] A. Kannan and V. K. Tumuluru, "Shared investment in PV panels and battery storage for residential building," *Energy, Ecol. Environ.*, vol. 7, no. 3, pp. 236–249, Jun. 2022, doi: 10.1007/s40974-021-00235-0.

[5] B. Zakeri and S. Syri, "Electrical energy storage systems: A comparative life cycle cost analysis," *Renew. Sustain. Energy Rev.*, vol. 42, pp. 569–596, Feb. 2015, doi: 10.1016/j.rser.2014.10.011.

[6] X. Li, W. Yang, Z. Zhao, R. Wang, X. Yin, and P. Liu, "Risk-averse energy management of hydro/thermal/pumped storage complementarily operating with wind/solar: Balancing risk, cost and carbon emission," *Sustain. Energy Technol. Assessments*, vol. 60, Dec. 2023, Art. no. 103534, doi: 10.1016/j.seta.2023.103534.

[7] C. Zhang, W. Cao, T. Xie, C. Wang, C. Shen, X. Wen, and C. Mao, "Operational characteristics and optimization of hydro-PV power hybrid electricity system," *Renew. Energy*, vol. 200, pp. 601–613, Nov. 2022, doi: 10.1016/j.renene.2022.10.005.

[8] M. Riza, M. Aman, S. Wahyu, P. Minwon, and M. Kenji, "The 10 years operation of a PV-micro-hydro hybrid system in Taratak, Indonesia," *Solar Energy Mater. Solar Cells*, vol. 67, nos. 1–4, pp. 621–627, Mar. 2001, doi: 10.1016/S0927-0248(00)00334-2.

[9] Y. Ren, X. Yao, D. Liu, R. Qiao, L. Zhang, K. Zhang, K. Jin, H. Li, Y. Ran, and F. Li, "Optimal design of hydro-wind-PV multi-energy complementary systems considering smooth power output," *Sustain. Energy Technol. Assessments*, vol. 50, Mar. 2022, Art. no. 101832, doi: 10.1016/j.seta.2021.101832.

[10] K. Lei, J. Chang, X. Wang, A. Guo, Y. Wang, and C. Ren, "Peak shaving and short-term economic operation of hydro-wind-PV hybrid system considering the uncertainty of wind and PV power," *Renew. Energy*, vol. 215, Oct. 2023, Art. no. 118903, doi: 10.1016/j.renene.2023.118903.

[11] J. Yang, J. Liu, G. Qiu, J. Liu, S. Jawad, and S. Zhang, "A spatio-temporality-enabled parallel multi-agent-based real-time dynamic dispatch for hydro-PV-PHS integrated power system," *Energy*, vol. 278, Sep. 2023, Art. no. 127915, doi: 10.1016/j.energy.2023.127915.

[12] H. Wang, J. Zhao, Q. Sun, and H. Zhu, "Probability modeling for PV array output interval and its application in fault diagnosis," *Energy*, vol. 189, Dec. 2019, Art. no. 116248, doi: 10.1016/j.energy.2019.116248.

[13] Z. Wang, J. Kang, L. Cheng, Z. Pei, C. Dong, and Z. Liang, "Mixed Gaussian models for modeling fluctuation process characteristics of photovoltaic outputs," *Frontiers Energy Res.*, vol. 7, p. 76, Aug. 2019, doi: 10.3389/fenrg.2019.00076.

[14] G. U. A. N. Lin, Z. H. A. O. Qi, Z. H. O. U. Baorong, and L. Y. U. Yaotang, "Multi-scale clustering analysis based modeling of photovoltaic power characteristics and its application in prediction," *Automat. Electr. Power Syst.*, vol. 42, no. 15, pp. 24–30, 2018, doi: 10.7500/AEPS20171027003.

[15] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 300–308, Jan. 2018, doi: 10.1109/TIE.2017.2714127.

[16] M. Hummon, A. Weekley, K. Searight, and K. Clark, "Downscaling solar power output to 4-seconds for use in integration studies," in *Proc. 3rd Int. Workshop Integr. Sol. Power Power Syst.*, London, U.K., 2013, pp. 1–7. [Online]. Available: https://www.nrel.gov/docs/fy14osti/60336.pdf

[17] Y. Shi, Y. Sun, J. Liu, and X. Du, "Model and stability analysis of grid-connected PV system considering the variation of solar irradiance and cell temperature," *Int. J. Electr. Power Energy Syst.*, vol. 132, Nov. 2021, Art. no. 107155, doi: 10.1016/j.ijepes.2021.107155.

[18] C. Yang, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 104–112, Jan. 2015, doi: 10.1109/TSTE.2014.2359974.

[19] W. Zhang, Q. Li, and Q. He, "Application of machine learning methods in photovoltaic output power prediction: A review," *J. Renew. Sustain. Energy*, vol. 14, no. 2, Mar. 2022, Art. no. 22701, doi: 10.1063/5.0082629.

[20] J. Ospina, A. Newaz, and M. O. Faruque, "Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model," *IET Renew. Power Gener.*, vol. 13, no. 7, pp. 1087–1095, May 2019, doi: 10.1049/iet-rpg.2018.5779.

[21] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114216, doi: 10.1016/j.apenergy.2019.114216.

[22] X. Zhang, R. Gao, C. Zhu, C. Liu, and S. Mei, "Ultra-short-term prediction of regional photovoltaic power based on dynamic graph convolutional neural network," *Electr. Power Syst. Res.*, vol. 226, Jan. 2024, Art. no. 109965, doi: 10.1016/j.epsr.2023.109965.

[23] Y. Miao, S. S. Y. Lau, K. K. N. Lo, Y. Song, H. Lai, J. Zhang, Y. Tao, and Y. Fan, "Harnessing climate variables for predicting PV power output: A backpropagation neural network analysis in a subtropical climate region," *Sol. Energy*, vol. 264, Nov. 2023, Art. no. 111979, doi: 10.1016/j.solener.2023.111979.

[24] R. Trivedi, S. Patra, and S. Khadem, "A data-driven short-term PV generation and load forecasting approach for microgrid applications," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 3, no. 4, pp. 911–919, Oct. 2022, doi: 10.1109/JESTIE.2022.3179961.

[25] J. Á. González Ordiano, S. Waczowicz, M. Reischl, R. Mikut, and V. Hagenmeyer, "Photovoltaic power forecasting using simple data-driven models without weather data," *Comput. Sci.-Res. Develop.*, vol. 32, nos. 1–2, pp. 237–246, Mar. 2017, doi: 10.1007/s00450-016-0316-5.

[26] J. Kang, J. Lee, and S. Lee, "Data-driven minute-ahead forecast of PV generation with adjacent PV sector information," *Energies*, vol. 16, no. 13, p. 4905, Jun. 2023, doi: 10.3390/en16134905.

[27] S. Sonia, C. Asmae, and T. Mohamed, "Data-driven prediction models of photovoltaic energy for smart grid applications," *Energy Rep.*, vol. 9, no. 9, pp. 90–105, 2023, doi: 10.1016/j.egyr.2023.05.237.

[28] X. Meng, X. Shi, W. Wang, Y. Zhang, and F. Gao, "An upscaling minute-level regional photovoltaic power forecasting scheme," *Int. J. Electr. Power Energy Syst.*, vol. 155, Jan. 2024, Art. no. 109609, doi: 10.1016/j.ijepes.2023.109609.

[29] S. Shedd, B. M. Hodge, A. Florita, and K. Orwig, "Statistical characterization of solar photovoltaic power variability at small timescales," in *Proc. 2nd Annu. Int. Workshop Integr. Sol. Power into Power Syst. Conf.*, Lisbon, Portugal, 2012, pp. 1–8. [Online]. Available: https://www.nrel.gov/docs/fy12osti/56165.pdf

[30] J. Marcos, L. Marroyo, E. Lorenzo, D. Alvira, and E. Izco, "From irradiance to output power fluctuations: The PV plant as a low pass filter," *Prog. Photovolt., Res. Appl.*, vol. 19, no. 5, pp. 505–510, Aug. 2011, doi: 10.1002/pip.1063.

[31] W. Yang, J. Yang, Z. Zhao, and D. Li, "Preliminary study on dynamic performance of variable speed pump-turbine unit for hybrid photovoltaic-pumped storage power system," in *Proc. 11th Int. Conf. Appl. Energy (ICAE)*, Västerås, Sweden, 2020, p. 951, doi: 10.46855/energy-proceedings-1844.

[32] M. Boyd, T. Chen, and B. Dougherty, "NIST campus photovoltaic (PV) arrays and weather station data sets," Dept. Commerce, Nat. Inst. Standards Technol., Gaithersburg, MA, USA, 2017, doi: 10.18434/M3S67G.

[33] Y. Jin, H. Wang, T. Chugh, D. Guo, and K. Miettinen, "Data-driven evolutionary optimization: An overview and case studies," *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 442–458, Jun. 2019, doi: 10.1109/TEVC.2018.2869001.

[34] H. Bourenane, A. Berkani, K. Negadi, F. Marignetti, and K. Hebri, "Artificial neural networks based power management for a battery/supercapacitor and integrated photovoltaic hybrid storage system for electric vehicles," *J. Européen des Systèmes Automatisés*, vol. 56, no. 1, pp. 139–151, Feb. 2023, doi: 10.18280/jesa.560118.

[35] H. Tian, H. Zhao, H. Li, X. Huang, X. Qian, and X. Huang, "Digital twins of multiple energy networks based on real-time simulation using holomorphic embedding method. Part II: Data-driven simulation," *Int. J. Electr. Power Energy Syst.*, vol. 153, Nov. 2023, Art. no. 109325, doi: 10.1016/j.ijepes.2023.109325.

[36] P. Baldi, P. Sadowski, and Z. Lu, "Learning in the machine: Random backpropagation and the deep learning channel," *Artif. Intell.*, vol. 260, pp. 1–35, Jul. 2018, doi: 10.1016/j.artint.2018.03.003.

[37] G. Chen, B. Tang, X. Zeng, P. Zhou, P. Kang, and H. Long, "Short-term wind speed forecasting based on long short-term memory and improved BP neural network," *Int. J. Electr. Power Energy Syst.*, vol. 134, Jan. 2022, Art. no. 107365, doi: 10.1016/j.ijepes.2021.107365.

[38] M.-L. Huang, Y.-H. Hung, and W.-Y. Chen, "Neural network classifier with entropy based feature selection on breast cancer diagnosis," *J. Med. Syst.*, vol. 34, no. 5, pp. 865–873, Oct. 2010, doi: 10.1007/s10916-009-9301-x.

[39] X. Li, W. Yang, Y. Liao, S. Zhang, Y. Zheng, Z. Zhao, M. Tang, Y. Cheng, and P. Liu, "Short-term risk-management for hydro-wind-solar hybrid energy system considering hydropower part-load operating characteristics," *Appl. Energy*, vol. 360, Apr. 2024, Art. no. 122818, doi: 10.1016/j.apenergy.2024.122818.

[40] A. Inteha, Nahid-Al-Masood, F. Hussain, and I. A. Khan, "A data driven approach for day ahead short term load forecasting," *IEEE Access*, vol. 10, pp. 84227–84243, 2022, doi: 10.1109/ACCESS.2022.3197609.

JIA WEI (Student Member, IEEE) received the B.S. degree from the School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China, in 2024, where he is currently pursuing the M.S. degree with the State Key Laboratory of Water Resources Engineering and Management. His research interests include modeling and regulation of hybrid energy system dominated by renewable energy, for example, hydro/wind/PV/storage hybrid system.

WEIJIA YANG (Member, IEEE) received the B.S. and M.S. degrees from the School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China, in 2011 and 2013, respectively, and the Ph.D. degree from the Division of Electricity, Department of Engineering Sciences, Uppsala University, Uppsala, Sweden, in 2017.

He is currently an Associate Professor with the State Key Laboratory of Water Resources Engineering and Management, Wuhan University. He main on the dynamics and control of hydropower systems, in the interdisciplinary field regarding hydraulics, mechanics, and electrical engineering. His current research interests include large-scale hydropower plants, hydro/wind/PV/storage hybrid systems, and variable speed pumped storage. He is a member of IEC/TC4/WG 36 and IAHR.

XUDONG LI (Member, IEEE) received the B.S. degree from the School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China, in 2021. He is currently pursuing the M.S. degree with the State Key Laboratory of Water Resources Engineering and Management, Wuhan University. His research interests include optimization and management of hybrid energy system dominated by renewable energy. He is also a Young Professional Member of IAHR.

JUNSONG WANG received the B.S. degree from the School of Water Resources and Hydropower Engineering, Wuhan University, Wuhan, China, in 2024, where he is currently pursuing the M.S. degree with the State Key Laboratory of Water Resources Engineering and Management. His research interests include construction of water conservancy and hydropower projects.