


Article

A New Regional Distributed Photovoltaic Power Calculation Method Based on FCM-mRMR and nELM Model

Honglu Zhu ^{1,2,*} , Tingting Jiang ^{1,2}, Yahui Sun ^{1,2} and Shuang Sun ^{1,2}

¹ State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Beijing 102206, China

² School of New Energy, North China Electric Power University, Beijing 102206, China

* Correspondence: hongluzhu@ncepu.edu.cn

Abstract: As the proportion of distributed photovoltaic (DP) increases, improving the accuracy of regional distributed photovoltaic power calculation is crucial to making full use of PV and ensuring the safety of the power system. The calculation of regional power generation is the key to power prediction, performance evaluation, and fault diagnosis. Distributed photovoltaic plants (DPP) are characterized by scattered distribution and small installed capacity, lots of DPPs are not fully monitored, and their real-time output power is difficult to obtain. Therefore, to improve the observability of DPPs and increase the accuracy of calculation, a new method that combines with fuzzy c-means (FCM), Max-Relevance and Min-Redundancy (mRMR) and Extreme Learning Machine (ELM), which can calculate the regional DPP output power without meteorological data is proposed, and validated using actual operational data of regional DPPs in China. The calculations results show good robustness in different months. The innovation of this study is the combination of the benchmark DPP selection method FCM-mRMR and the power calculation method nELM, and the mean absolute error (MAPE) of the proposed method is 0.198 and the coefficient of determination (R²) is 0.996.

Keywords: distributed photovoltaic; power calculation; Extreme Learning Machine; Max-Relevance and Min-Redundancy



Citation: Zhu, H.; Jiang, T.; Sun, Y.; Sun, S. A New Regional Distributed Photovoltaic Power Calculation Method Based on FCM-mRMR and nELM Model. *Sustainability* **2022**, *14*, 13880. <https://doi.org/10.3390/su142113880>

Academic Editor: Giuseppe Todde

Received: 29 August 2022

Accepted: 21 October 2022

Published: 26 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Against the greenhouse effect and the energy crisis, PV power has developed rapidly as a sustainable, green renewable energy resource (RES). By the end of 2021, worldwide PV installed capacity increased by 145 TWh to nearly 1000 TWh [1]. DP has the advantage of local consumption, and DP-based power systems can reduce the cost of electrification in rural and suburban communities [2,3], making DP one of the main options for developing RES. China has the largest PV market globally, with 5488 GW of new PV systems in 2021, of which 2927.9 GW is DP [4]. As DP grows, the power system becomes more complex and uncertain. Since the output power of most DPPs is not monitored, the operating status is unknown to the power system, so the power calculation of DPPs is crucial for the safe operation of DPPs. At the same time, regional distributed photovoltaic power calculation (DPPC) can provide practical information to support online functions of the distribution network, such as optimal power dispatching and fault diagnosis [5], improve the observability of the DP, and contribute to the safe operation of the distribution network [6].

The methods for PV power calculation fall into three main categories: the methods based on state estimation methods [7–9], physical modeling methods, and data-driven methods. State estimation methods mainly use AC voltage and node power data in the measurement points, Shen and Liang [10] present an improved algorithm based on PV branch current and PV branch power for the unmonitored PV state estimation model. Fang et al. [11] propose a state estimation algorithm with insufficient data. The physical

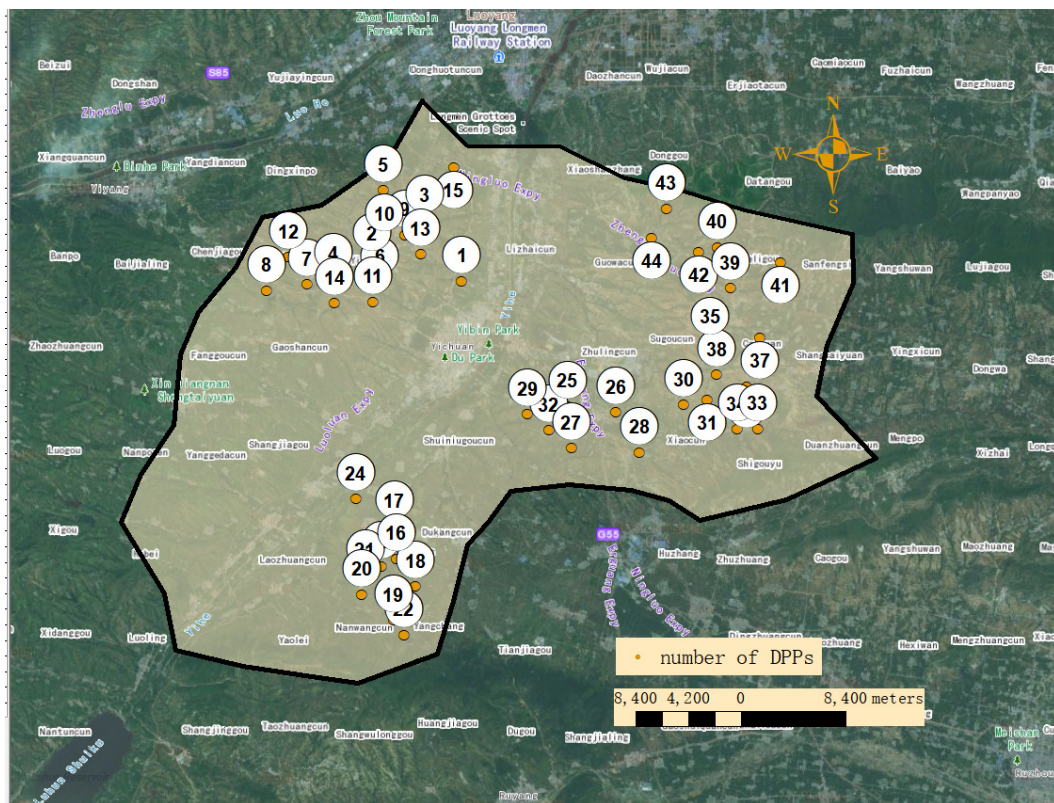
modeling methods model the distributed PV output based on the physical mechanism, Yun et al. [12] model the internal physical characteristics and external environmental characteristics of the PV system separately. Farzana et al. [13], Wendell et al. [14] use weather data, netload data, etc., to model the output power for PV arrays. Data-driven methods analyze the relationship between weather forecast data, PV historical data, and PV power, using algorithms such as neural networks, support vector machine [15,16], and random forest [17,18].

The above research mainly focuses on a single PV station or a few PV nodes in the distribution network. The DPPs in a region are usually massive and widely distributed in the actual situation. Real-time, high-precision monitoring of PV power generation requires a higher sampling frequency and additional measurement sites, which raises the cost of DPPs constructing. Therefore, calculating the PV output power of the region with a small amount of measurement data is an important research direction [19,20]. Xin et al. [21] combined the K-means method and linear estimation model to obtain the total output power of regional distributed PV. Wu et al. [22] applied the K-medoids algorithm to cluster the distributed PV plants with installed capacity. Shaker et al. [23] proposed a hybrid method based on K-means clustering and principal component analysis to calculate the output of unmonitored PV plants using benchmark plants. Most power calculation method for DPPs require weather information such as wind speed, temperature, and irradiance, those are often missing in practical situations. Since neighboring regions have similar meteorological patterns, up-scaling the power data of benchmark stations allows for calculating PV power for the whole area. In [24,25], the Australian Photovoltaic Institute (APIV) calculates the regional PV output using measured PV output and regional installed capacity. Saint-Drenan et al. [26], to increase the accuracy of regional PV power calculation, develop a method based on satellite and measurement data. However, the previous work did not investigate the power output characteristics of PV plants in the region, and the analysis of PV power output characteristics is the foundation for calculating regional PV power generation.

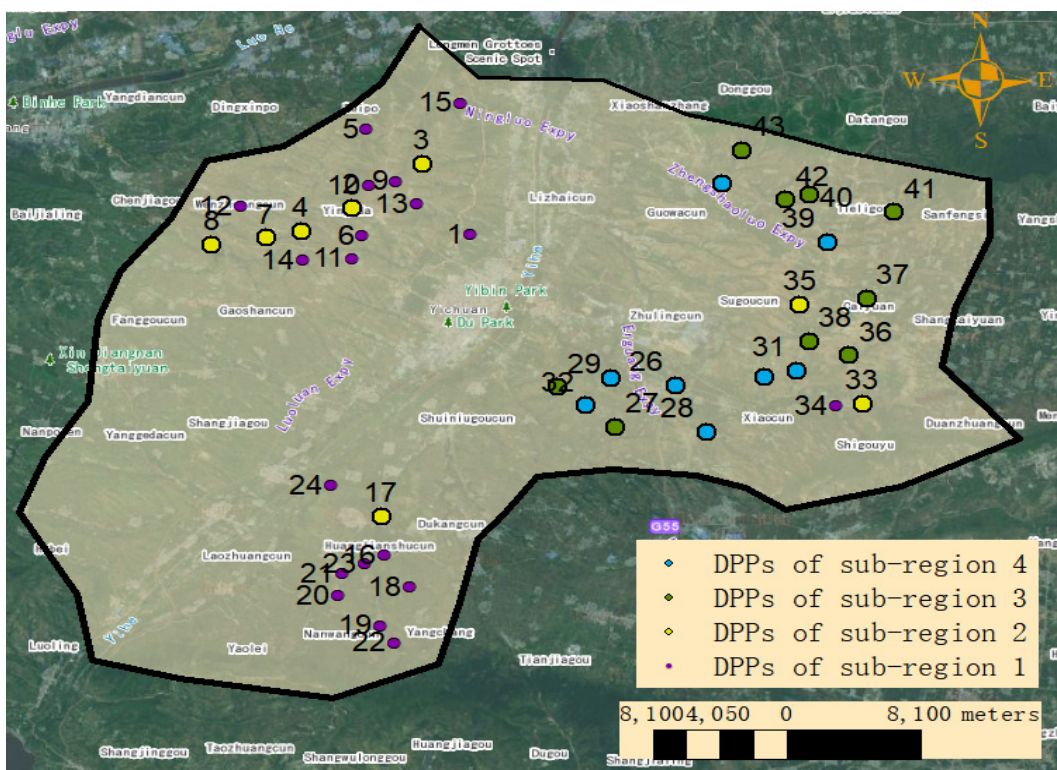
Based on the above analysis, the paper conducted the study on regional DPPC. For regional DPPs with large coverage areas and without measured meteorological information, dividing regions, and selecting benchmark DPPs for different sub-region can effectively improve the accuracy of power calculation. Due to the fast speed of the Extreme Learning Machine (ELM), the parallel ELM model optimized by Particle swarm optimization (PSO) is used for power calculation [27]. Based on existing modeling techniques [28,29], the fuzzy c-means (FCM) and Max-Relevance and Min-Redundancy (mRMR) algorithms were used for benchmark power plant selection. The improved algorithm reduces the data dependence compared to existing studies while improving the computational accuracy. The innovation of this study mainly lies in selecting benchmark power plants based on the output feature of PV and employing parallel ELM to increase the accuracy of the calculation. Accordingly, the paper is organized as follows: In Section 2, data sources and output characteristics of regional DPPs were studied. In Section 3, the principle of the algorithm and the proposed method are explained. In Section 4, the validity of the proposed model for region division and benchmark power plant selection is verified. In Section 5, the calculation results are shown and compared with other algorithms. Finally, Section 6 concludes the paper and suggests future work.

2. Data Sources and Regional PV Plant Power Output Characteristics

In this work, 44 DPPs are selected for analysis in a county of China, numbered 1 to 44, and the locations of the DPPs are shown in Figure 1a. The data are real-time power data from 44 DPPs with a time resolution of 5 min. The output power curves of 44 DPPs in one day are shown in Figure 2, which shows that there are similarities and differences among different DPPs.

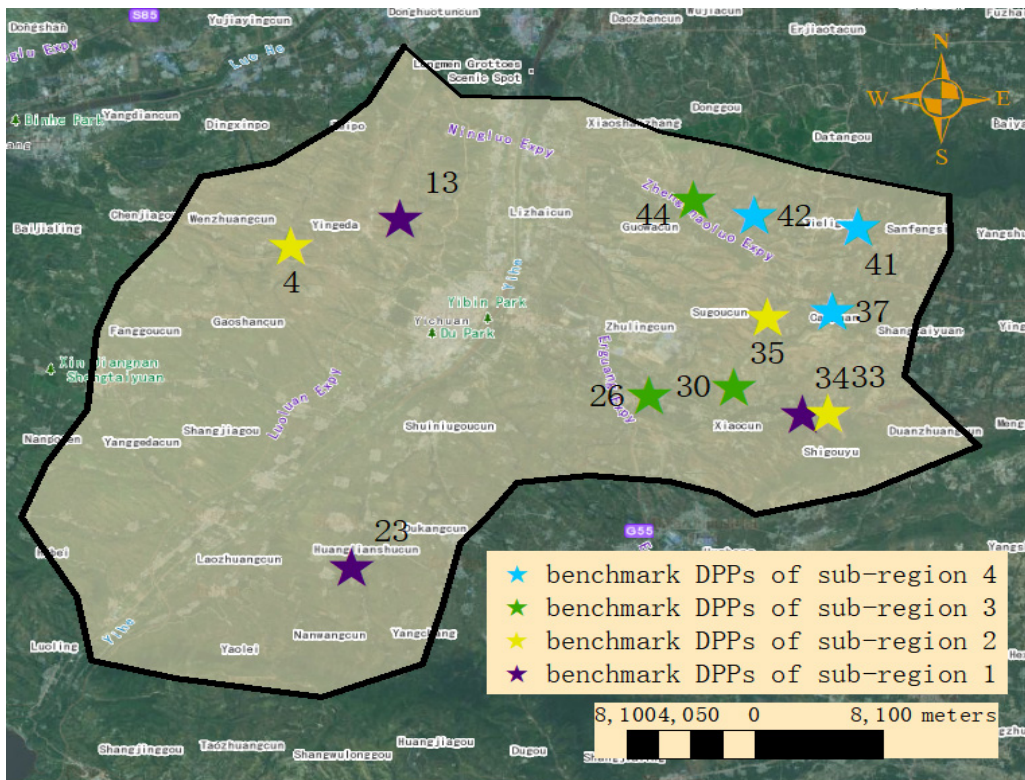


(a)



(b)

Figure 1. Cont.



(c)

Figure 1. Distribution of power stations in the satellite map. (a) Regional DPPs distribution. (b) sub-region clustering results. (c) benchmark plants of different sub-region.

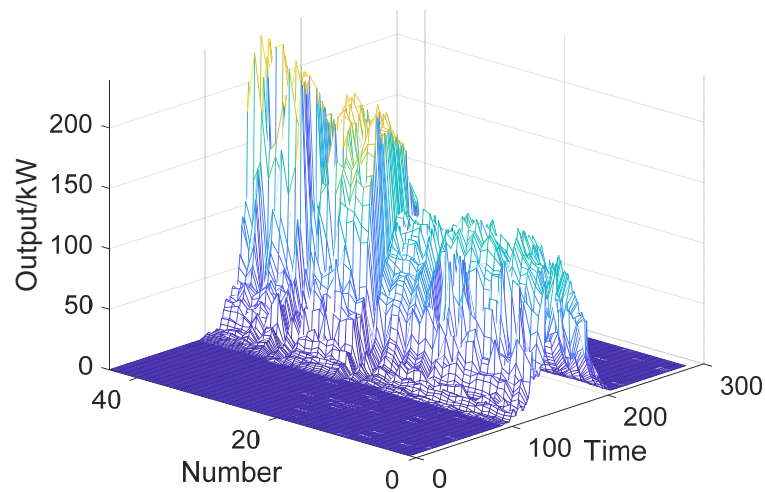


Figure 2. Output power of DPPs.

The Pearson coefficient between the power of DPPs and the entire region was analyzed. And ρ is a linear correlation coefficient that measures how closely two variables are related, it takes values between -1 and 1 , which can be calculated as Equations (1)–(3):

$$\rho = \text{Cov}(X, Y) / \sqrt{\text{Var}(X) \cdot \text{Var}(Y)} \quad (1)$$

$$\text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

$$\text{Cov}(X, Y) = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (3)$$

ρ between different DPPs are shown in Figure 3, where numbers 1–44 refer to DPPs, and number 45 refers to all DPPs in the region, and ρ have been normalized. In the correlation matrix diagram, the closer the correlation is to 1, the closer the color is to yellow; the closer the correlation is to 0, the closer the color is to blue. Parts 1, 2, and 3 represent three parts in Figure 3, and the correlation coefficients vary significantly among different parts. The single DPP and the region have the lowest correlation coefficient. Parts 1 and 3 have a higher correlation (0.7, 1), but part 2 has a lower correlation (0.4, 0.7). Therefore, the sub-region division can precisely separate the DPPs with similar output. DPPs with strong correlation can be used to calculate the output of others. For the case of large amount DPPs, the correlation among different power plants is different, and the sub-region division of the region can effectively improve the accuracy of calculation. Accordingly, this paper proposes a regional DPPC method, which can calculate the real-time total power of the region based on the selected benchmark plants in the absence of meteorological information.

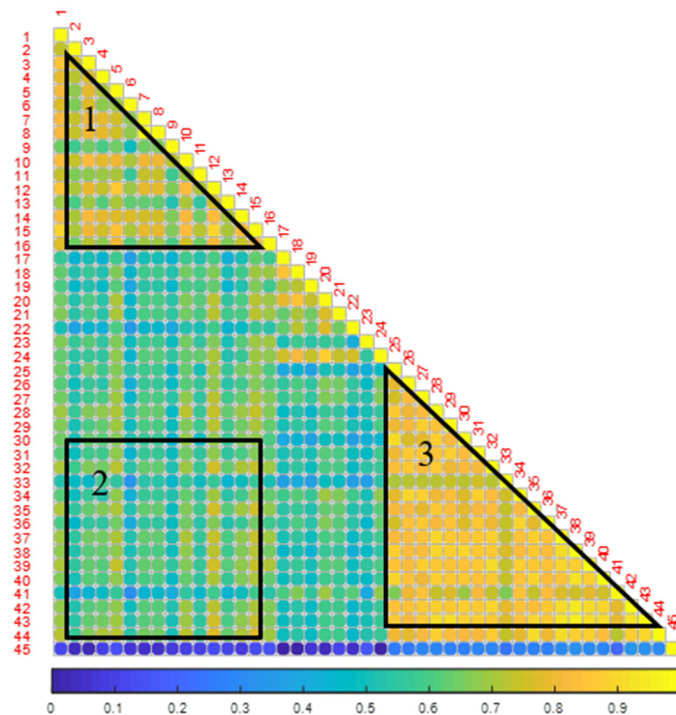


Figure 3. Pearson correlation coefficients between DPPs.

3. Principles of the Algorithm

In the paper, Fuzzy C Means (FCM) clustering algorithm is used to separate DPPs into different sub-regions. Then benchmark DPPs are selected by the Max-Relevance and Min-Redundancy (mRMR) method for different sub-regions. Finally, a combined PSO-ELM approach is used to realize the regional DPPC.

3.1. Improved FCM-mRMR Algorithm

FCM is used for sub-region division, and the aim is to have the samples grouped into the same cluster as similar as possible [30,31]. The objective function of the FCM is

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (4)$$

The objective function is minimized by taking derivatives for all variables, resulting in:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (5)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c (d_{kj})^{2/(m-1)}} \quad (6)$$

where, u_{ij} is the membership degree, which ranges from 0 to 1, reflects the degree to which each sample belongs to a class; c_i is the cluster center of the fuzzy group I and $d_{ij} = \|c_i - x_j\|$ is the Euclidean Distance between i th cluster center and j th data point. Then, the process of minimizing the objective function automatically divide the samples.

It can be seen that fuzzy C-means clustering is a simple iterative process. The clustering steps of DPPs in the region are shown as follows.

- (1) Initialize the membership matrix U with random numbers, each matrix element is a random number from 0–1.
- (2) Calculate the clustering centers $c_i, i = 1, \dots, c$.
- (3) Calculate the objective function according to Equation (4). If it is less than a defined threshold, or if the change from the last value is less than the threshold, the process stops.
- (4) Compute the membership degree matrix using Equation (6). Then return to step (2).

In information theory, entropy can be used to measure how much information is carried. The entropy of a variable is defined as

$$H(X) = -\sum_{i=1}^n P(x_i) \log P(x_i) \quad (7)$$

Mutual information (MI) measures the amount of information that one variable contains about another variable. It is commonly used in feature selection. Metrics such as correlation coefficient and MI reflect the similarity between features and labels, but the selected features with best correlation in the dataset is not necessarily the optimal subset, and such strategy may lead to the missing of important information. If the correlation between features is high, there may be some duplicated information, resulting in increased redundancy in the selected subset. The mRMR algorithm ensures the minimum redundancy and the maximum correlation between subset features and labels.

The maximum correlation is defined as $\max\{D, S\}$,

$$D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c) \quad (8)$$

Minimum redundancy is defined as $\min R(S)$,

$$R = \frac{1}{|S^2|} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (9)$$

where S is the feature set; c is the category; $I(x_i; c)$ is the mutual information between feature i and the target category c ; $I(x_i; x_j)$ is the mutual information between feature i and feature j .

The feature selection criteria for MRMR are:

$$\max \phi(D, R), \phi = D - R \quad (10)$$

$$\max \phi_1(D, R), \phi_1 = D/R \quad (11)$$

The proposed sub-regions division and benchmark DPPs selection algorithm based on FCM and mRMR, divide the DPPs into different sub-regions by FCM clustering, then the benchmark DPPs within the sub-regions are selected using mRMR. The results of the FCM-MI and FCM-mRMR are shown in Table 1 and Figure 1c.

Table 1. Results of the FCM-mi and FCM-mRMR.

Sub-Region Number	FCM-MI	FCM-mRMR
1	13, 22, 34	34, 23, 13
2	4, 33, 35	33, 4, 35
3	29, 40, 42	26, 30, 44
4	26, 28, 44	37, 41, 42

3.2. nELM Algorithm

The Extreme Learning Machine (ELM) [32–34] is a single-layer neural network, with the advantages of few training parameters, fast learning speed, and strong generalization ability, it has been widely used in the field of PV power prediction in recent years. The input layer weights are randomly generated during the training process. The output layer bias is calculated by a generalized inverse matrix, the structure diagram of ELM is shown in Figure 4.

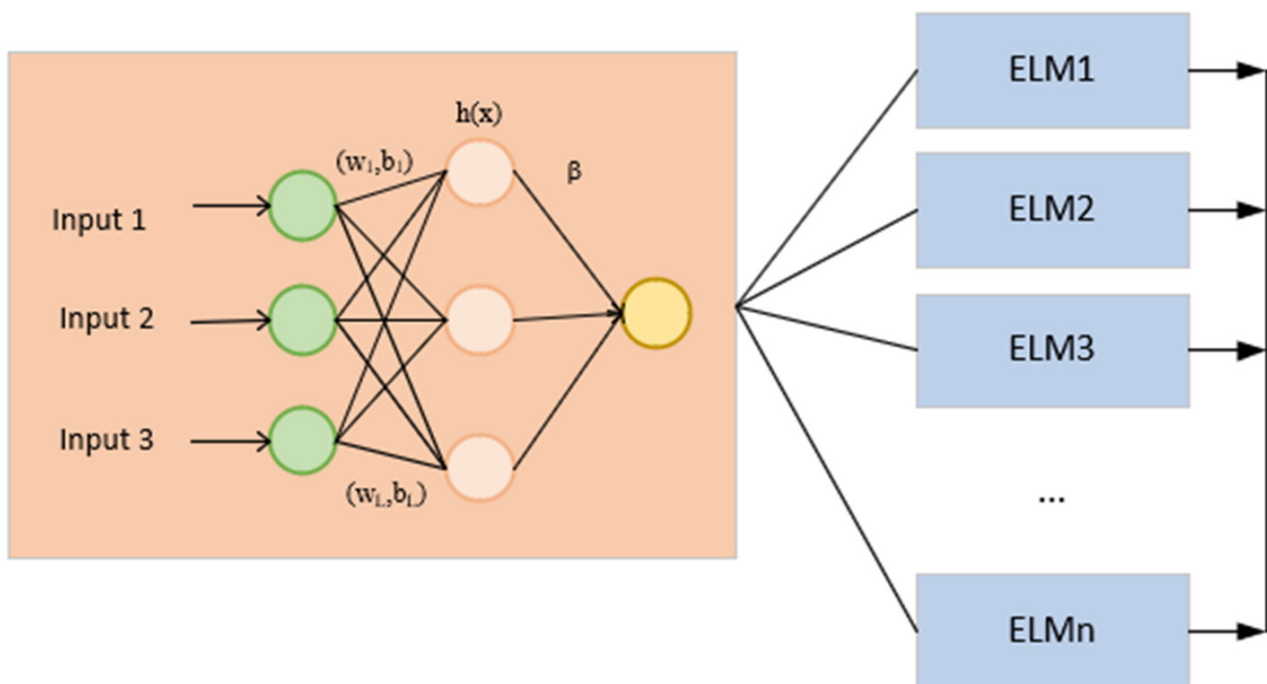


Figure 4. Model of nELM.

W is the connection weights matrix between the input layer and the hidden layer,

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1D} \\ w_{21} & w_{22} & \dots & w_{2D} \\ \dots & \dots & \dots & \dots \\ w_{L1} & w_{L2} & \dots & w_{LD} \end{bmatrix}_{L \times D} \quad (12)$$

where, w_{ji} is the connection weights of the i th neuron in the input layer and the j th neuron in the hidden layer.

$$t = [t_1 \quad t_2 \quad \dots \quad t_Q] \quad (13)$$

$$t_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \dots \\ t_{mj} \end{bmatrix}_{m \times 1} = \begin{bmatrix} \sum_{i=1}^l \beta_{i1} g(w_i x_j + b_i) \\ \sum_{i=1}^l \beta_{i2} g(w_i x_j + b_i) \\ \dots \\ \sum_{i=1}^l \beta_{im} g(w_i x_j + b_i) \end{bmatrix}_{m \times 1} \quad (j = 1, 2, \dots, Q) \quad (14)$$

β is the connection weight between the hidden layer and the output layer, b is the threshold of the neuron in the hidden layer, X is the input set, Y is and the output set. $g(x)$ is the hidden layer neuron activate function, and t is the network output.

Equations (15) and (16). can be expressed as

$$H\beta = T' \quad (15)$$

where T' is the transpose of the matrix T and H is the hidden layer output matrix.

The connection weights β are obtained by solving the least-squares solution.

$$\hat{\beta} = H^+ T' \quad (16)$$

where H^+ is the Moore-Penrose generalized inverse of the hidden layer output matrix H .

However, as the initial values of the parameters w and b are chosen randomly, ELM has an unstable performance. Therefore, a parallel framework that combines n ELMs is proposed to reduce the randomness of the model during training.

3.3. Regional DPPC Method

The following are some of the difficulties with regional DPPC:

1. Due to financial considerations, not all DPPs are equipped with meteorological stations, resulting in a paucity of meteorological data.
2. The number of DPPs in the region is considerable, making it difficult to extract useful information.

As a result, the FCM-mRMR method is proposed in this study to divide the sub-regions and select the benchmark power plants, with the PSO-ELM algorithm estimating the total power based on the real-time power of the benchmark power plants. The regional DPPC method proposed in the paper can be described as following steps:

Divide the sub-regions. Assuming there are N DPPs in the region, use the FCM algorithm to divide the power dataset of DPPs into k sub-region datasets, denoted as $\{x_1, x_2, \dots, x_k\}$.

Select m benchmark PV plants in each sub-region by mRMR, and the power data of the benchmark plants are used as input.

The regional DPPC is realized using a combined PSO-nELM method. In the nELM algorithm, the connection weights β and thresholds of the hidden layers b are determined by Particle Swarm Optimization (PSO).

Figure 5 shows the framework of the proposed method.

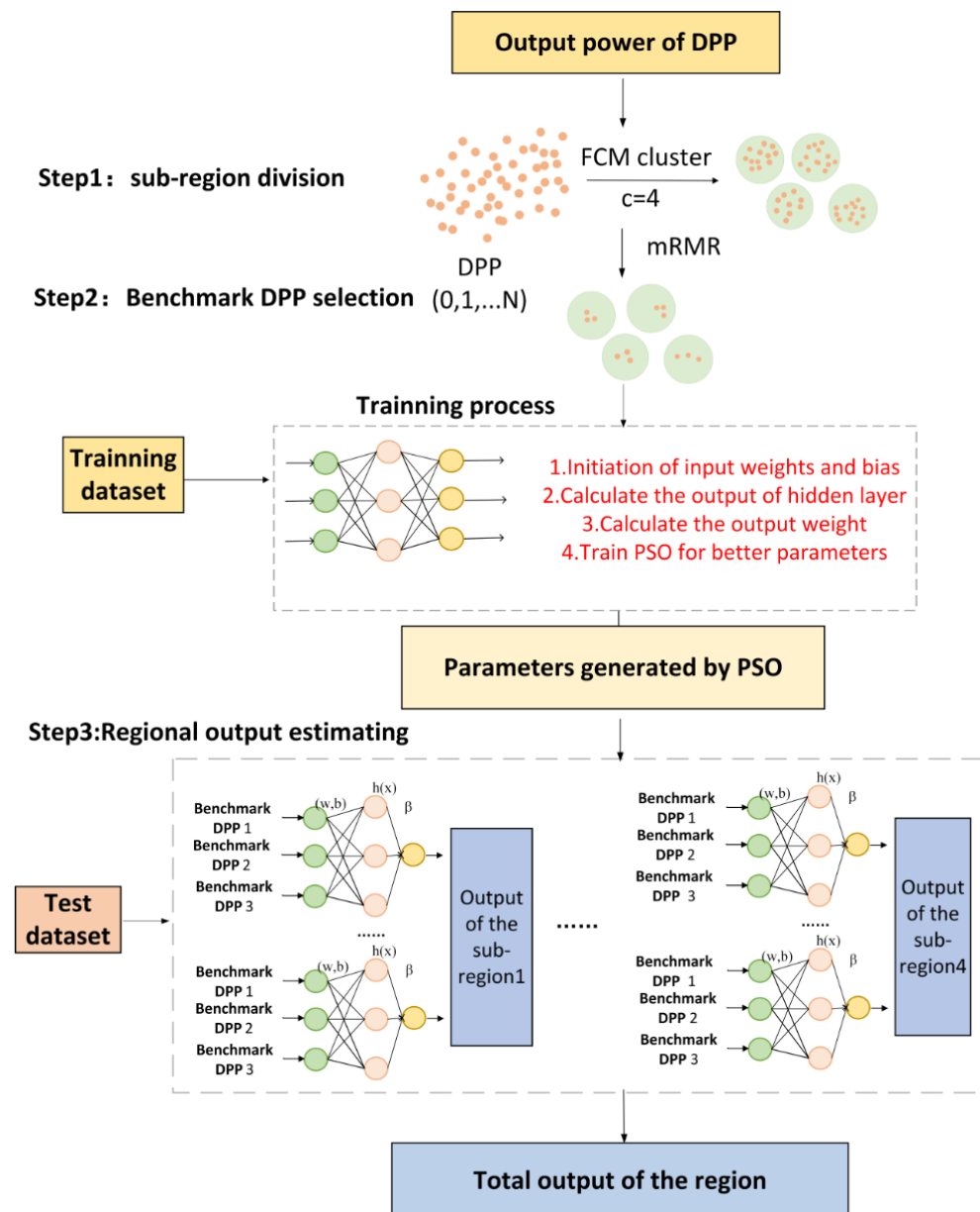


Figure 5. Framework of the proposed method.

4. Rationality Analysis

4.1. Experimental Scenarios Setting

All DPPs in the region have certain similarity due to the similar geographical location and certain difference due to the environmental changes. Sub-region division of DPPs in the region can extract the power plants that are similar to each other and separate the power plants that are different from each other. Therefore, the comparison scheme was designed to verify the effectiveness of sub-region division and benchmark power plant selection. There are 7 experimental scenarios numbered 1–7, which are shown in Table 2. The sub-region division methods include the FCM clustering method and the geographic location division method (marked as loc). The benchmark DPPs selection methods include the mRMR algorithm, MI ranking method, and random selection method (marked as rand). Where nc indicates no sub-regional division.

Table 2. Different regional division methods.

		Sub-Region Division			Benchmark Power Plant Selection		
		nc	FCM	loc	mRMR	mi	rd
1	nc-mi	✓				✓	
2	nc-mRMR	✓			✓		
3	loc-mi			✓		✓	
4	loc-mRMR			✓	✓		
5	FCM-mi		✓			✓	
6	FCM-mRMR		✓		✓		
7	FCM-rd		✓				✓

The estimation results were evaluated by the accuracy evaluation metrics of mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). Since the PV power is zero at night, all accuracy evaluations in this paper were the results after removing the nighttime data.

$$MAE = \frac{1}{m} \sum_{i=1}^m |Y_i - \hat{Y}_i| \quad (17)$$

$$R^2 = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (\bar{Y}_i - \hat{Y}_i)^2} \quad (18)$$

$$MAPE = \frac{100\%}{m} \sum_{i=1}^m \left| \frac{Y_i - \hat{Y}_i}{Y} \right| \quad (19)$$

4.2. Results Analysis

The training set is the DPPs' power data from 1 to 14 January 2020, while the test set is from 17 January to 30 April 2020.

Figure 6 depicts the power calculation results of different methods, Figure 7a and Table 3 provide the quantitative results. The MAPEs of nc-mi and nc-mRMR are larger than 30%, and the MAEs are greater than 75, according to the data. The MAPEs of loc-mi and loc-mRMR are less than 30%, while the MAEs are between 60% and 70%. The MAPEs of FCM-mi, FCM-mRMR, and FCM-rd are all less than 28%, with MAEs ranging from 50 to 70. The boxplot of the estimation error is shown in Figure 7b, where the error distribution is on both sides of 0. The upper and lower quartile intervals of nc-mi and nc-mRMR are the largest, distributed within (−39, 227) and (−37, 106); the upper and lower quartiles of loc-mi, loc-mRMR, FCM-mi, FCM-mRMR, FCM-rd intervals are (−79, 11), (−34, 79), (−93, 43), (−17, 118), (−7, 141), respectively, with FCM-mRMR having the fewest outlier points. Therefore, we can conclude that dividing sub-regions gives more accurate estimation results than those without division.

Table 3. Comparison of different regional division methods.

	nc-mi	nc-mRMR	loc-mi	loc-mRMR	FCM-mi	FCM-mRMR	FCM-rd
MAE	132.204	76.409	64.186	67.845	61.696	50.169	69.831
MAPE	0.331	0.501	0.153	0.300	0.173	0.272	0.128
R^2	0.993	0.997	0.997	0.996	0.997	0.998	0.997

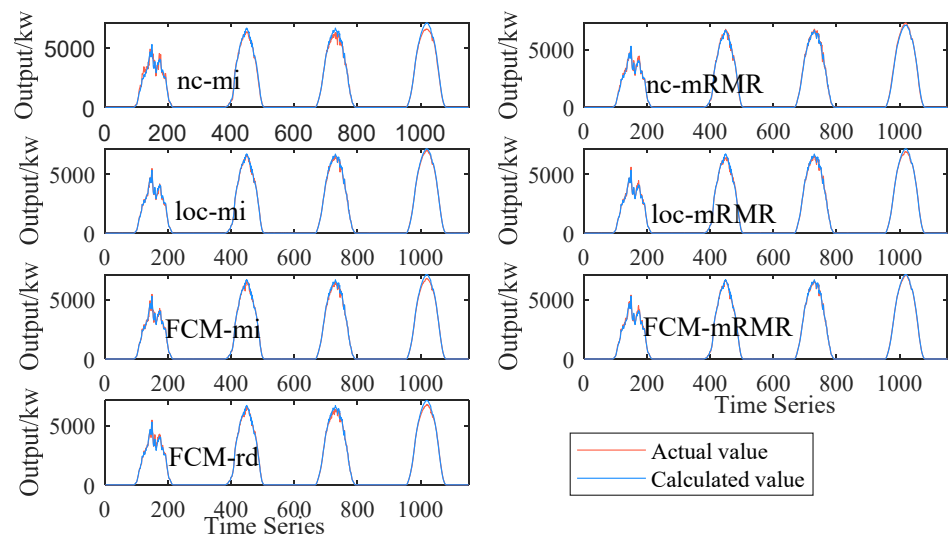


Figure 6. Comparison of different regional division methods.

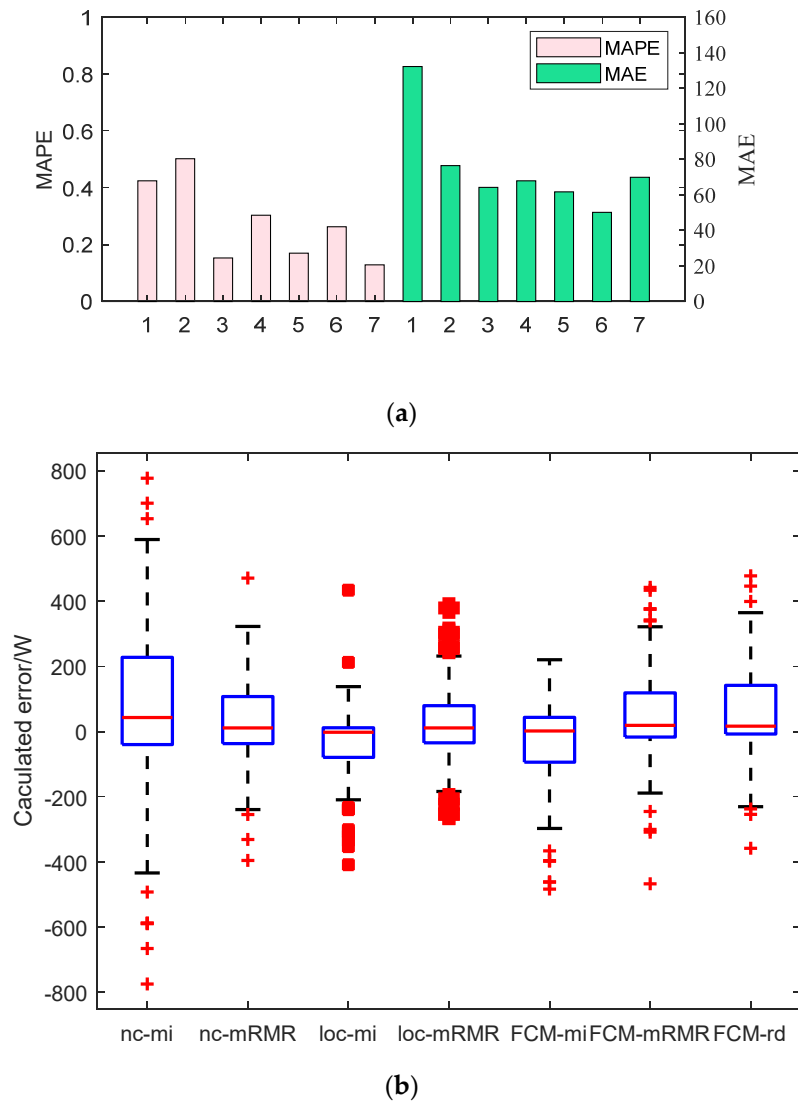


Figure 7. The quantitative results of different methods. (a) Evaluation of the calculation results. (b) Distribution of the calculation errors.

The error distributions of FCM-rd, FCM-mi, and FCM-mRMR are shown in Figure 8, with the horizontal coordinates being the errors and the vertical coordinates being the probabilities of the error distributions. The probability distribution curve of FCM-mRMR is narrower and is mainly concentrated in $(-200, 200)$. The distribution curve of FCM-mi is concentrated in the interval $(-400, 200)$, and the curve of FCM-rd is mainly concentrated in the interval $(-800, 400)$. Therefore, the proposed FCM-mRMR method for benchmark power plants selection reduces the error of power calculation.

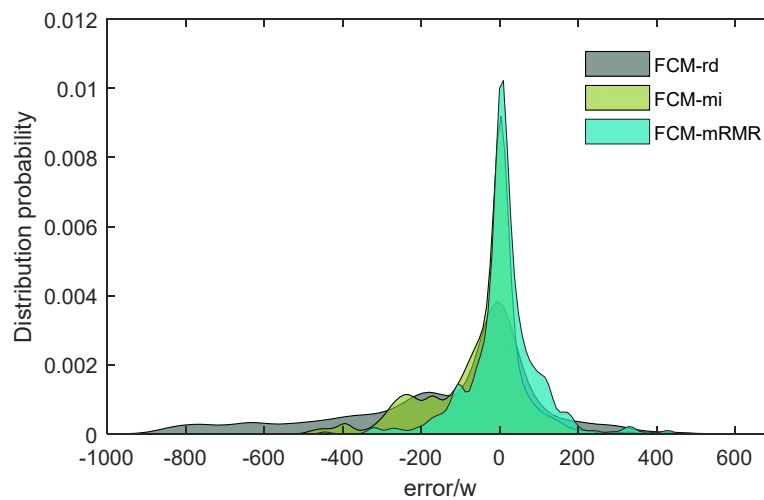


Figure 8. Error distribution of FCM-rd, FCM-mi, and FCM-mRMR.

5. Results Comparison

To get the optimal ELM parameter of the proposed method, the results of different setting of hidden neurons and parallel ELM models were compared. Then two optimization algorithms, the self-adaptive difference evolution algorithm (SaDE) and PSO were compared. Finally, the proposed method was compared with other calculation methods, including FCM-mRMR-BPANN and FCM-mRMR-PSO-SVM.

5.1. ELM Parameter Optimization

For ELM, the connection weights β and thresholds of the hidden layers b are set randomly, the number of neurons in the hidden layer needs to be optimized. Figure 9. shows the computational accuracy of different numbers of hidden layer neurons for FCM-mi and FCM-mRMR. Where, x coordinate is the number of hidden layer neurons, y coordinate is the number of program executions, and z denotes the MAE. We can conclude that the optimal number of neurons is $L = 2$ for FCM-mi and $L = 3$ for FCM-mRMR, and the accuracy is higher with the FCM-mRMR selecting method.

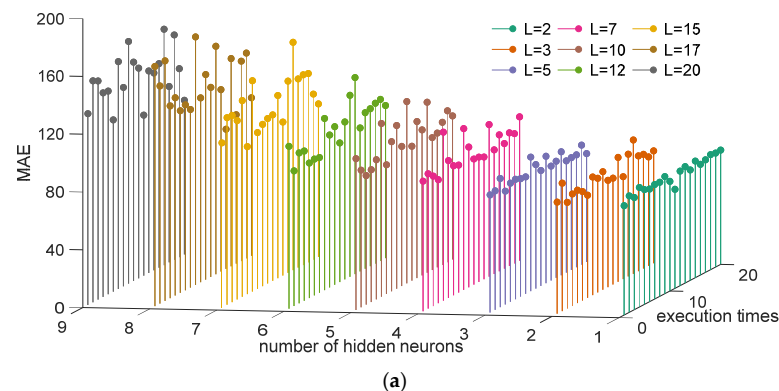


Figure 9. Cont.

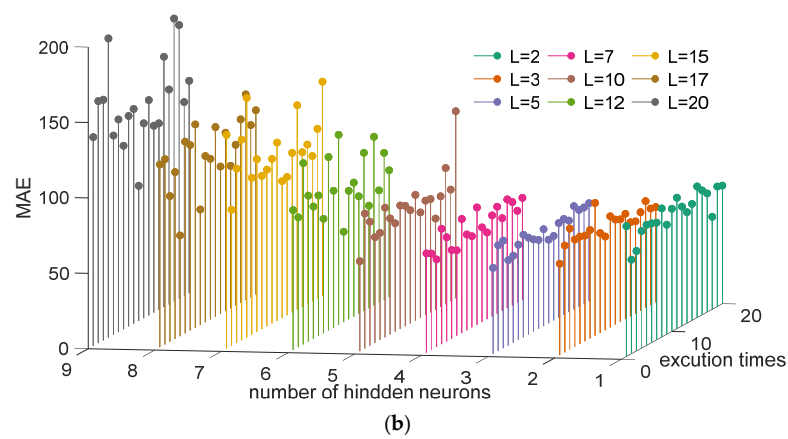


Figure 9. (a) Comparison of FCM-mi with the different number of hidden neurons; (b) Comparison of FCM-mRMR with the different number of hidden neurons.

The output results for ELM are unstable, to reduce the model’s randomness during training, a parallel framework combining n ELMs is presented. The average output of all ELMs is the final predicted value. Therefore, the determining of the number for parallel ELM models (N) is vital. Figure 10 shows the results of FCM-mi and FCM-mRMR with different N . It can be obtained that the operation results tend to be stable when $L = 2\sim 7$ so that $N = 4$, and when $L = 10\sim 20$ so that $N = 10$.

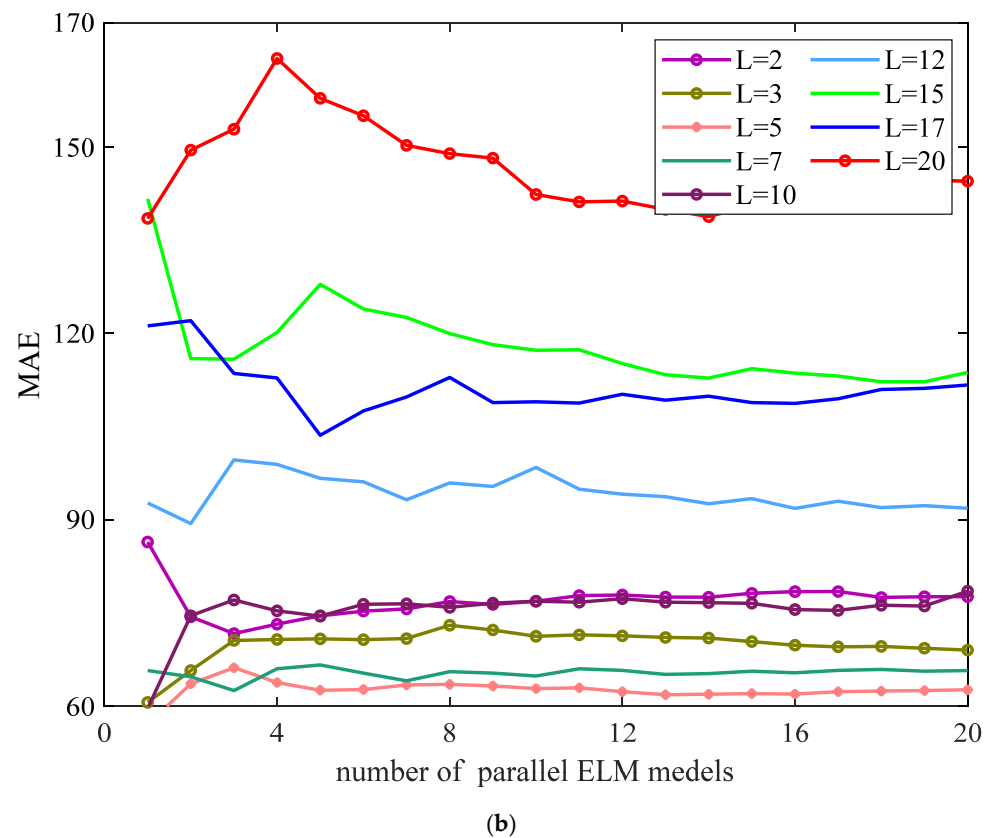


Figure 10. Cont.

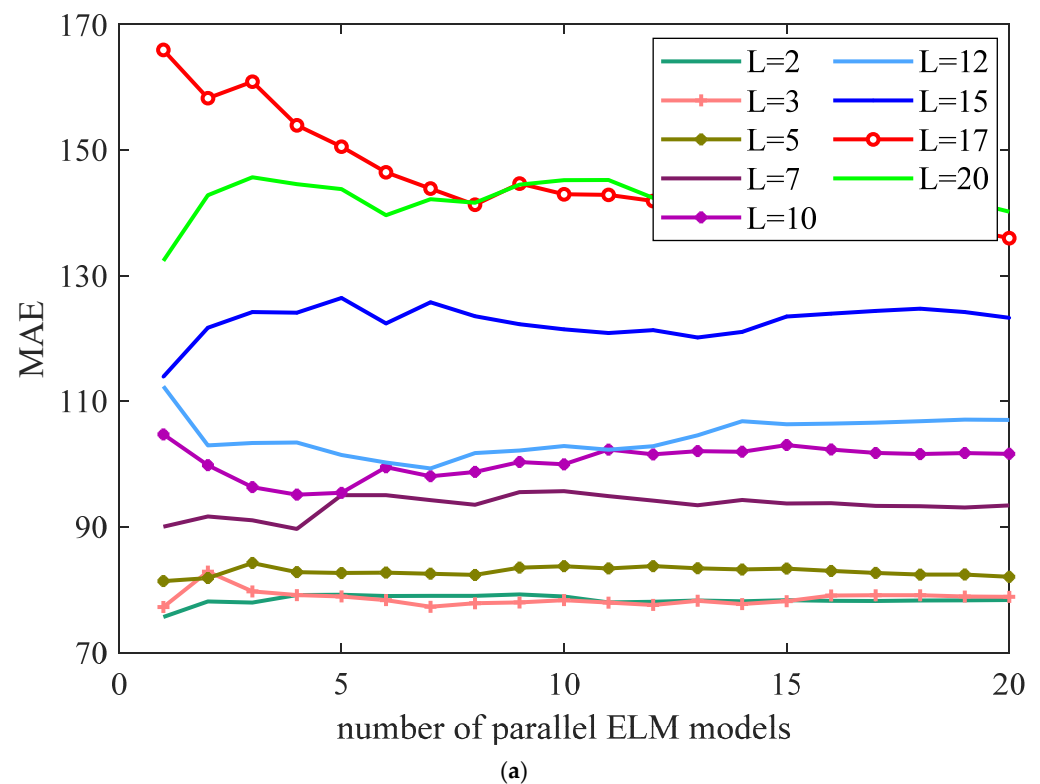


Figure 10. Comparison of ELM parameter. (a) Comparison of FCM-mi with the different number of parallel ELMs. (b) Distribution of the calculation errors.

5.2. Comparison of Different Optimization Algorithms

PSO and SaDE were used to optimize the parameters of proposed FCM-mRMR-nELM model. PSO is based on the idea of a massless particle that behaves like a flock of birds. Each particle has two properties: velocity and position, and based on its position at a certain velocity, it seeks out the optimal solution. Comparing the optimal solutions generates the global optimal solution. The optimal solution as well as the global optimal solution leads all particles to modify their velocity and position.

SaDE is a population-based parallel search approach for parameter optimization. Two individuals in the population are randomly selected, their difference vector is combined with the third individual to generate variant individuals. Crossover is a process in which variable individuals are combined with target individuals to create experimental individuals. If the experimental individuals' fitness is higher than that of the target individuals, the target individuals will be replaced by experimental individuals in the next generation, this process is known as selection. During each generation of an evolutionary process, continuously iterative computation retains the good individuals and eliminates the bad ones, guiding the search process to approach the global optimal solution. The optimization results of PSO and SaDE are shown in Table 4, it indicates that the PSO has a better optimization performance.

Table 4. Results of different optimization algorithms.

	MAPE	MAE	R^2
PSO-nELM-FCM-mRMR	0.198	63.800	0.996
SaDE-nELM-FCM-mRMR	0.316	132.543	0.991

The proposed FCM-mRMR-PSO-nELM method can obtain the real-time power of the region based on the benchmark DPPs. Figure 11 displays the results of the sub-region

calculation and the real power curves, while Figure 12 displays the results of the region calculation and the actual power curves. The correlation scatters diagram is shown in Figures 13 and 14. The red dots in represent the calculated values with significant errors. We can see that fewer points deviate from the actual value using the proposed region calculation method, and there is less error deviation.

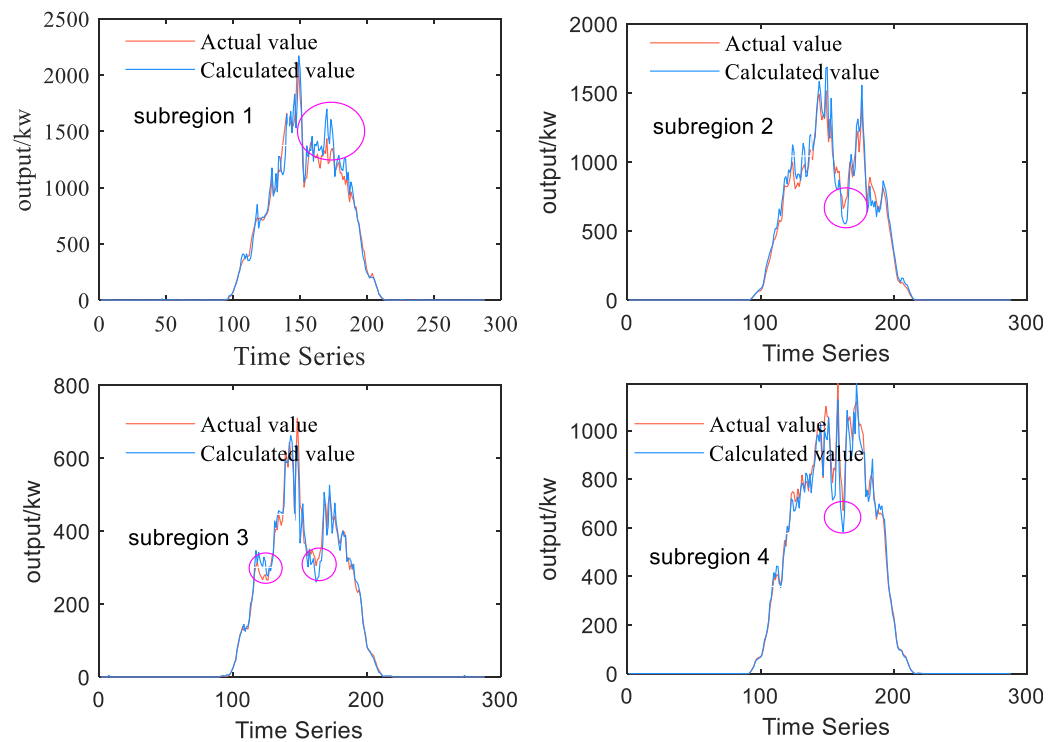


Figure 11. Actual and calculated power curves for the sub-region.

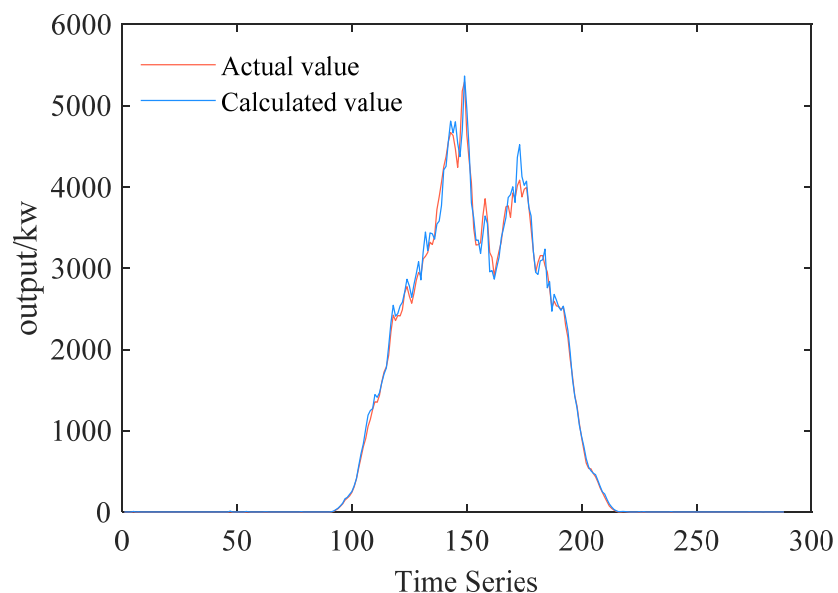


Figure 12. Actual and calculated power curves for the entire region.

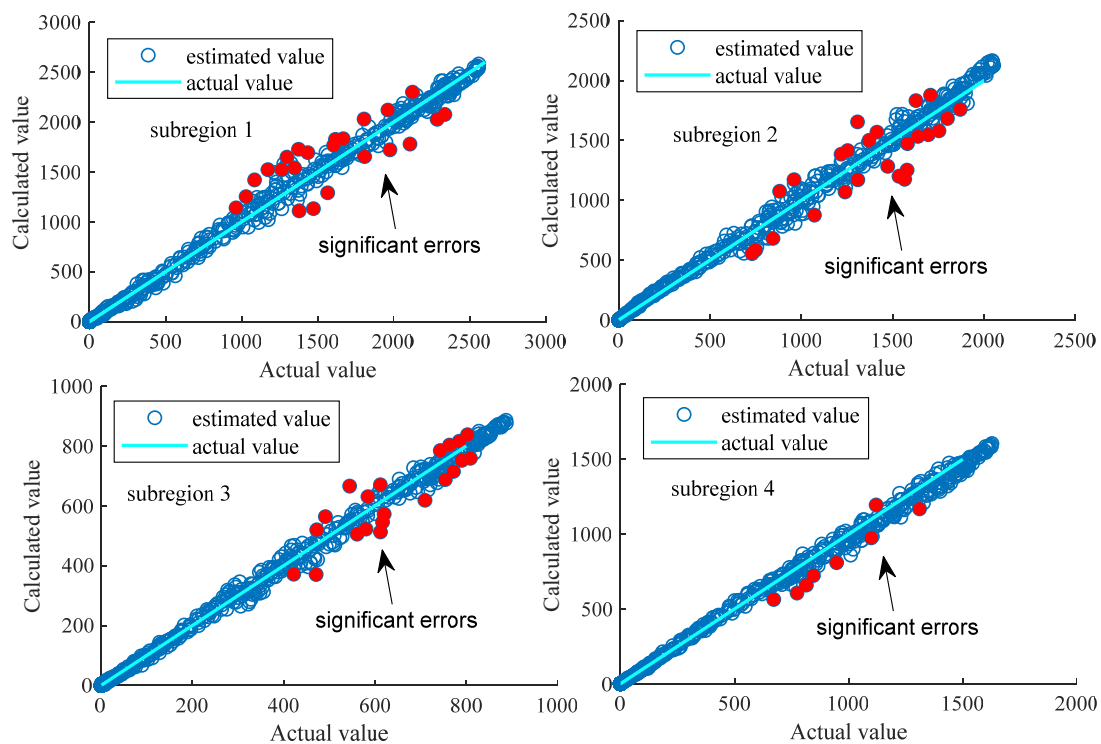


Figure 13. Scatter plot of actual and calculated power for the sub-region.

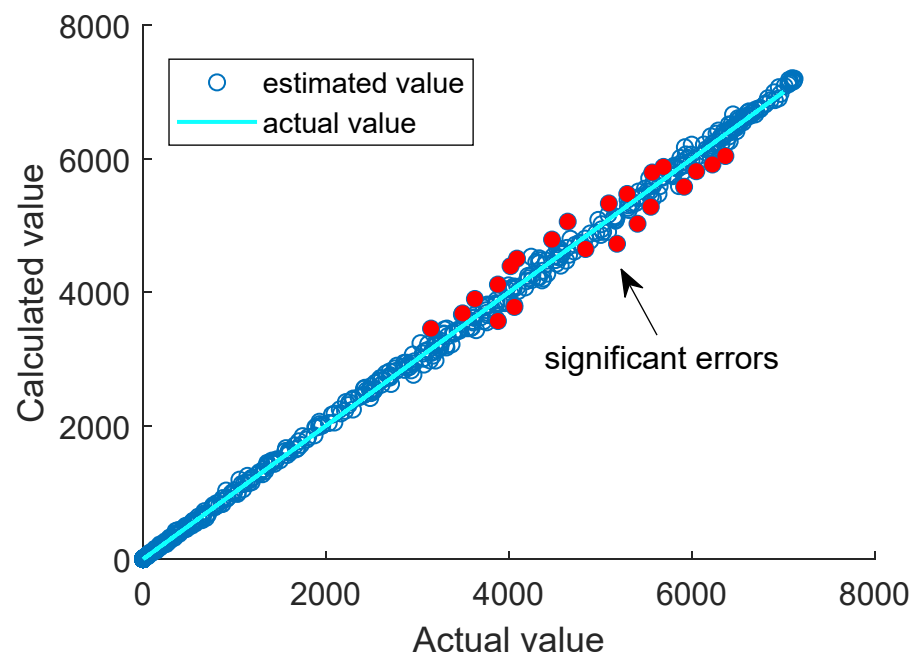


Figure 14. Scatter plot of actual and calculated power for the entire region.

5.3. Performance Comparison for Different Modeling Algorithms

The calculated model fundamentally affects the calculated results, this paper compared the results of back propagation artificial neural network (BPANN), support vector machine (SVM) and ELM applied to DPPC. Therefore, the FCM-mRMR-PSO-nELM method proposed has been compared with FCM-mRMR-BPANN and FCM-mRMR-PSO-SVM methods, and the results are shown in Table 5. The BPANN computational model is set to a three-layer structure, with the number of hidden neurons being 15 and 20, the maximum number of training times being 5000, the training accuracy being 0.05, and the learning rate

being 0.05. The SVM model has two essential parameters, c and g . The model parameters of the SVM are optimized by the PSO algorithm with parameters $c = 81$ and $g = 0.7$.

Table 5. Value of parameters of each algorithm.

SVM		BPANN	
Parameter	Value	Parameter	Value
c	81	layer	3
g	0.7	hidden neurons	20
		hidden neurons	15
		maximum training times	5000
		training accuracy	0.05
		learning rate	0.05

The train data consists of 15 days of measured data with 5 min temporal resolution. The results are shown in Figure 15, where the horizontal coordinate is the sub-region number, and the vertical coordinates are the MAE and MAPE. Table 6 shows the calculation performance of different methods. The MAE of the FCM-mRMR-BPANN and FCM-mRMR-PSO-nELM are in (15, 55), while the MAE of the FCM-mRMR-PSO-SVM is in (20, 150). The MAPE for sub-regions 1 and 4 are more than 1, which indicates that the calculated results for points at lower power are too large. We conclude that the proposed method for calculating sub-region power has lower MAE and MAPE. Through the above analysis, the ELM method has better computational performance.

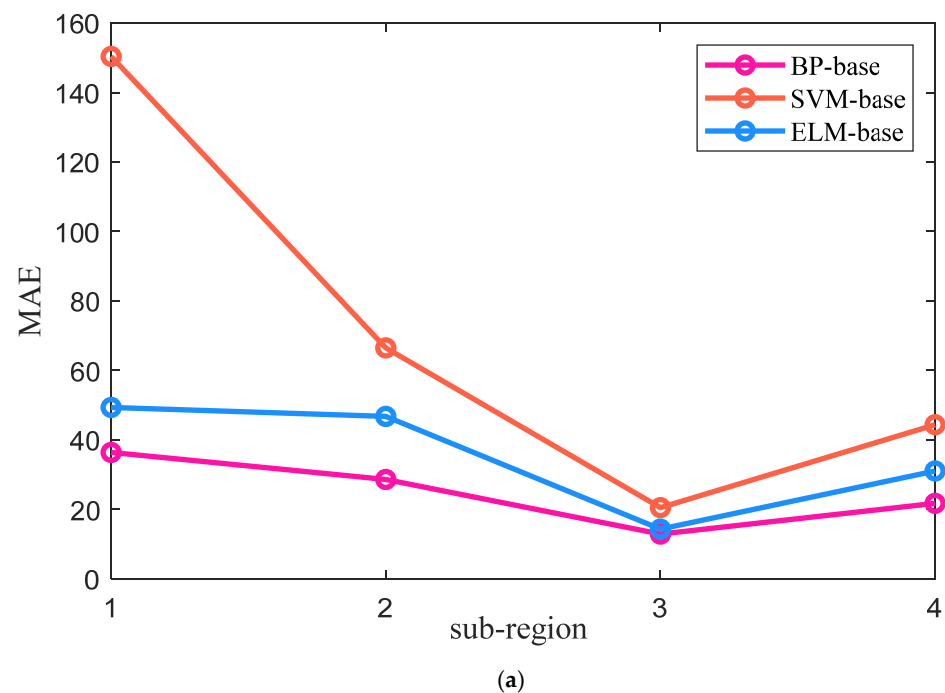


Figure 15. Cont.

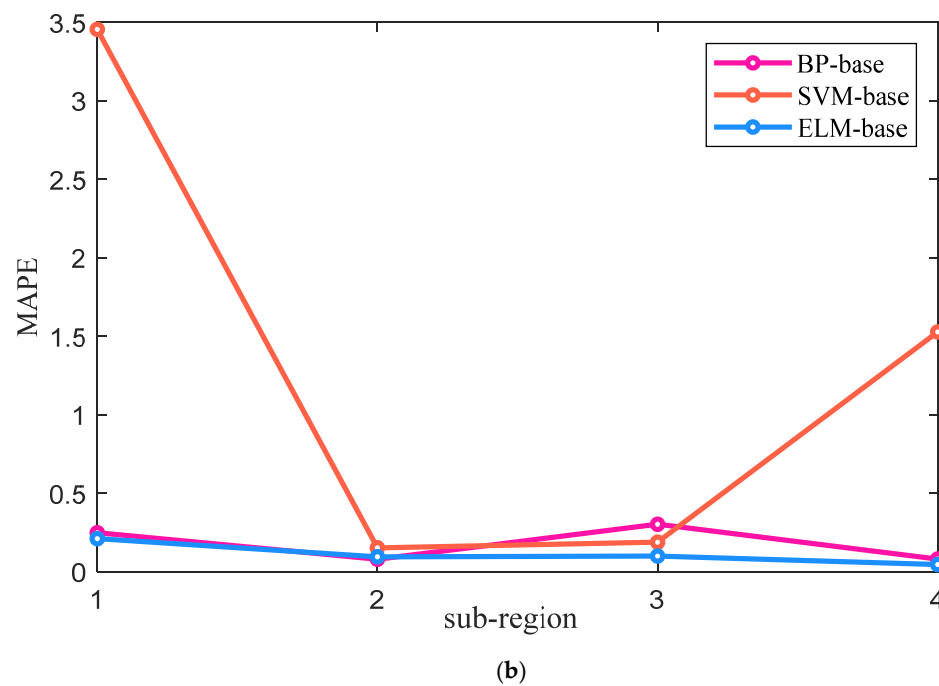


Figure 15. Outcome of sub-region calculation. (a) MAEs of sub-region calculation. (b) MAPEs of sub-region calculation.

Table 6. Comparison of model calculation results.

Model	FCM-mRMR-BPANN		FCM-mRMR-PSO-SVM		FCM-mRMR-PSO-nELM	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
January	81.918	0.663	81.186	0.255	61.348	0.224
February	130.153	0.418	185.142	0.219	65.371	0.175
March	148.445	0.371	162.261	10.098	122.070	0.213

6. Conclusions

For real-time DP monitoring and enhancing the secure and stable operation of distribution networks, accurate region power calculation is essential. Lack of meteorological data and poor calculation accuracy are issues with practical applications. This study's main contribution is to improve the Regional DPPC method. First, the region is divided into sub-region using the FCM clustering algorithm. Next, the sub-region benchmark plant is obtained using the mRMR algorithm. The power of the benchmark power plant is then inputted to compute the sub-region's power using PSO-nELM, and the total power is obtained by weighting the power of the sub-region. This study compares various methods for dividing regions, selecting benchmark power plants, and power calculating to simulate the uncertainty of the regression. The optimization algorithm and model parameters are also discussed. It was shown through empirical study that the proposed FCM-mRMR-PSO-nELM approach increases DPPC accuracy. The main innovation of the paper is to present a new DPPC method FCM-mRMR-PSO-nELM, the calculation results with night data removed were able to reach an MAPE of 0.198 and a R^2 of 0.996, reducing data dependence and computational complexity. The future research direction is to verify the generalizability of the method on a larger scale, and to calculate the regional PV power in combination with numerical weather prediction (NWP).

Author Contributions: Methodology and writing—review & editing, T.J.; project administration, H.Z.; data curation, H.Z.; writing—original draft preparation, Y.S.; investigation S.S.; conceptualization and funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ownership problem of the data.

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

Nomenclature

PV	Photovoltaics
DP	Distributed Photovoltaic
DPP	Distributed Photovoltaic Plants
FCM	Fuzzy C-Means
mRMR	Max-Relevance And Min-Redundancy
ELM	Extreme Learning Machine
RES	Renewable Energy Resource
TWh	Tera Watt Hour
PSO	Particle Swarm Optimization
DPPC	Distributed Photovoltaic Power Calculation
SaDE	Self-Adaptive Difference Evolution Algorithm
MI	Mutual Information

References

- Global Energy Review. 2021. Available online: <https://www.iea.org/reports/global-energy-review-2021> (accessed on 1 January 2022).
- Liang, M.; Xiao, Y.; Shaobo, Y.; Wen, Z.; Xuekai, H.; Can, S. The Influence of High Permeability Distributed Photovoltaic Access on Power Grid. In Proceedings of the 2020 4th International Conference on HVDC, HVDC 2020, Xi'an, China, 6–9 November 2020; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2020; pp. 449–453.
- Zahedi, R.; Zahedi, A.; Ahmadi, A. Strategic Study for Renewable Energy Policy, Optimizations and Sustainability in Iran. *Sustainability* **2022**, *14*, 2418. [[CrossRef](#)]
- Statistic of Solar PV Generation and Operation. Available online: http://www.nea.gov.cn/2022-03/09/c_1310508114.htm (accessed on 1 January 2022).
- Han, L.; Han, X.S.; Chen, F.; Zha, H. An Effective Hybrid Approach for Dynamic State Estimation in Power System. In Proceedings of the 2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, Nanjing, China, 6–9 April 2008.
- Sebastian, M.; Devaux, O.; Huet, O. Description and Benefits of a Situation Awareness Tool Based on a Distribution State Estimator and Adapted to Smart Grids. In Proceedings of the CIRED Seminar 2008: SmartGrids for Distribution, Frankfurt, Germany, 23–24 June 2008.
- Ishigame, A.; Matsuda, M.; Genji, T. A State Estimation Method for Photovoltaic Power Generation Using Independent Component Analysis. In Proceedings of the 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS), Seoul, Korea, 7–10 August 2011.
- Li, C.; Yuan, S.; Wu, C.; Chen, N.; Gao, B. Research on State Estimation of Power System with Large-Scale Photovoltaic Plant. In Proceedings of the 6th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems, IEEE-CYBER 2016, Chengdu, China, 19–22 June 2016; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2016; pp. 205–209.
- Niknam, T.; Bahmani-Firouzi, B. A Practical Algorithm for Distribution State Estimation Including Renewable Energy Sources. *Renew. Energy* **2009**, *34*, 2309–2316. [[CrossRef](#)]
- Shen, D.; Liang, H. Research on Distribution System State Estimation with Photovoltaic Generation. In Proceedings of the IEEE Region 10 Annual International Conference, Proceedings/TENCON, Xi'an, China, 22–25 October 2013.
- Fang, Z.; Lin, Y.; Song, S.; Li, C.; Lin, X.; Chen, Y. State Estimation for Situational Awareness of Active Distribution System with Photovoltaic Power Plants. *IEEE Trans. Smart Grid* **2021**, *12*, 239–250. [[CrossRef](#)]

12. Yun, T.; Zuohao; Chunlai, L.; Qian, H. Research on Modeling of Integrated Information System for Photovoltaic Power Plant. In Proceedings of the 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China, 9–10 October 2017; pp. 417–419.
13. Kabir, F.; Yu, N.; Yao, W.; Yang, R.; Zhang, Y. Joint Estimation of Behind-the-Meter Solar Generation in a Community. *IEEE Trans. Sustain. Energy* **2021**, *12*, 682–694. [[CrossRef](#)]
14. Stainsby, W.; Zimmerle, D.; Duggan, G.P. A Method to Estimate Residential PV Generation from Net-Metered Load Data and System Install Date. *Appl. Energy* **2020**, *267*, 114895. [[CrossRef](#)]
15. Mo, H.; Zhang, Y.; Xian, Z.; Wang, H. Photovoltaic (PV) Power Prediction Based on ABC—SVM. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *199*, 052031. [[CrossRef](#)]
16. Fan, S.; Cao, S.; Zhang, Y. Temperature Prediction of Photovoltaic Panels Based on Support Vector Machine with Pigeon-Inspired Optimization. *Complexity* **2020**, *2020*, 9278162. [[CrossRef](#)]
17. Assouline, D.; Mohajeri, N.; Scartezzini, J.L. Large-Scale Rooftop Solar Photovoltaic Technical Potential Estimation Using Random Forests. *Appl. Energy* **2018**, *217*, 189–211. [[CrossRef](#)]
18. Yang, M.; Zhao, M.; Liu, D.; Ma, M.; Su, X. Improved Random Forest Method for Ultra-Short-Term Prediction of the Output Power of a Photovoltaic Cluster. *Front. Energy Res.* **2021**, *9*, 749367. [[CrossRef](#)]
19. Tanomura, K.; Ogita, Y.; Kaneshige, Y.; Ishii, J.; Arai, J. Calculation of Distribution System Voltage and Power Flow State Using Measured Values. *Electr. Eng. Jpn. (Engl. Transl. Denki Gakkai Ronbunshi)* **2008**, *164*, 33–42. [[CrossRef](#)]
20. Kamono, K.; Ueda, Y. Real Time Estimation of PV Output in Distribution Systems Based on Smart Meters and Irradiance Measurement. In Proceedings of the 2015 IEEE 42nd Photovoltaic Specialist Conference, PVSC 2015, New Orleans, LA, USA, 14–19 June 2015; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2015.
21. Chen, X.; Li, S.; Wang, F.; Li, J.; Tang, C. Power Estimation Method of Low-Voltage Distributed Photovoltaic Generation Based on Similarity Aggregation. *Energy Rep.* **2021**, *7*, 1344–1351. [[CrossRef](#)]
22. Ji, W.; Xu, C.; Xiang, Z.; Hai, Z.; Fang, C. Research on Estimation of Regional Distributed Photovoltaic Output Based on K-Medoids Algorithm. In Proceedings of the 2019 IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia), Chengdu, China, 21–24 May 2019.
23. Shaker, H.; Zareipour, H.; Wood, D. A Data-Driven Approach for Estimating the Power Generation of Invisible Solar Sites. *IEEE Trans. Smart Grid* **2016**, *7*, 2466–2476. [[CrossRef](#)]
24. Haghdadi, N.; Bruce, A.; MacGill, I. Assessing the Representativeness of “Live” Distributed PV Data for Upscaled PV Generation Estimates. In Proceedings of the Asia-Pacific Power and Energy Engineering Conference, APPEEC, Suzhou, China, 15–17 April 2016; IEEE Computer Society: Washington, DC, USA, 2016.
25. Haghdadi, N.; Dennis, J.; Bruce, A.; Macgill, I. Real Time Generation Mapping of Distributed PV for Network Planning and Operations. In Proceedings of the Asia-Pacific Power and Energy Engineering Conference, APPEEC, Brisbane, Australia, 15–18 November 2015; IEEE Computer Society: Washington, DC, USA, 2016.
26. Saint-Drenan, Y.M.; Bofinger, S.; Ernst, B.; Landgraf, T.; Rohrig, K. Regional Nowcasting of the Solar Power Production with PV-Plant Measurements and Satellite Images. In Proceedings of the 30th ISES Biennial Solar World Congress 2011, SWC 2011, Kassel, Germany, 28 August–2 September 2011; Volume 2, pp. 1464–1474.
27. Liu, J.; Xu, M. Kernelized Fuzzy Attribute C-Means Clustering Algorithm. *Fuzzy Sets Syst.* **2008**, *159*, 2428–2445. [[CrossRef](#)]
28. Rajamoorthy, R.; Arunachalam, G.; Kasinathan, P.; Devendiran, R.; Ahmadi, P.; Pandiyan, S.; Muthusamy, S.; Panchal, H.; Kazem, H.A.; Sharma, P. A Novel Intelligent Transport System Charging Scheduling for Electric Vehicles Using Grey Wolf Optimizer and Sail Fish Optimization Algorithms. *Energy Sources Part A Recover. Util. Environ. Eff.* **2022**, *44*, 3555–3575. [[CrossRef](#)]
29. Worku, M.Y. Recent Advances in Energy Storage Systems for Renewable Source Grid Integration: A Comprehensive Review. *Sustainability* **2022**, *14*, 5985. [[CrossRef](#)]
30. Bora, B.J.; Dai Tran, T.; Prasad Shadangi, K.; Sharma, P.; Said, Z.; Kalita, P.; Buradi, A.; Nhanh Nguyen, V.; Niyas, H.; Tuan Pham, M.; et al. Improving Combustion and Emission Characteristics of a Biogas/Biodiesel-Powered Dual-Fuel Diesel Engine through Trade-off Analysis of Operation Parameters Using Response Surface Methodology. *Sustain. Energy Technol. Assess.* **2022**, *53*, 102455. [[CrossRef](#)]
31. Wang, Z. Comparison of Four Kinds of Fuzzy C-Means Clustering Methods. In Proceedings of the 2010 Third International Symposium on Information Processing, Qingdao, China, 15–17 October 2010; pp. 563–566. [[CrossRef](#)]
32. Lan, Y.; Soh, Y.C.; Huang, G. Bin Ensemble of Online Sequential Extreme Learning Machine. *Neurocomputing* **2009**, *72*, 3391–3395. [[CrossRef](#)]
33. Zhao, J.; Wang, Z.; Park, D.S. Online sequential extreme learning machine with forgetting mechanism. *Neurocomputing* **2012**, *87*, 79–89. [[CrossRef](#)]
34. Wang, X.; Han, M. Online Sequential Extreme Learning Machine with Kernels for Nonstationary Time Series Prediction. *Neurocomputing* **2014**, *145*, 90–97. [[CrossRef](#)]