

Article

Boosting the Development and Management of Wind Energy: Self-Organizing Map Neural Networks for Clustering Wind Power Outputs

Yanqian Li, Yanlai Zhou , Yuxuan Luo , Zhihao Ning  and Chong-Yu Xu 

State Key Laboratory of Water Resources Engineering and Management, Wuhan University, Wuhan 430072, China; yanqian.li@whu.edu.cn (Y.L.); luoyx0389@whu.edu.cn (Y.L.); zhihao.ning@whu.edu.cn (Z.N.); c.y.xu@geo.uio.no (C.-Y.X.)

* Correspondence: yanlai.zhou@whu.edu.cn

Abstract: Aimed at the information loss problem of using discrete indicators in wind power output characteristics analysis, a self-organizing map neural network-based clustering method is proposed in this study. By identifying the appropriate representativeness and topological structure of the competition layer, cluster analysis of the wind power output process in four seasons is realized. The output characteristics are evaluated through multiple evaluation indicators. Taking the wind power output of the Hunan power grid as a case study, the results underscore that the 1×3 -dimensional competition layer structure had the highest representativeness (72.9%), and the wind power output processes of each season were divided into three categories, with a robust and stable topology structure. Summer and winter were the most representative seasons. Summer had strong volatility and small wind power outputs, which required the utilization of other power sources to balance power supply and load demand. Winter featured low volatility and large wind power outputs, necessitating cooperation with peak-shaving power sources to enhance the power grid's absorbability to wind power. The seasonal clustering analysis of wind power outputs will be helpful to analyze the seasonality of wind power outputs and can provide scientific and technical support for guiding the power grid's operation and management.



Citation: Li, Y.; Zhou, Y.; Luo, Y.; Ning, Z.; Xu, C.-Y. Boosting the Development and Management of Wind Energy: Self-Organizing Map Neural Networks for Clustering Wind Power Outputs. *Energies* **2024**, *17*, 5485. <https://doi.org/10.3390/en17215485>

Academic Editor: Hua Li

Received: 26 September 2024

Revised: 24 October 2024

Accepted: 31 October 2024

Published: 1 November 2024



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

Keywords: wind power output; clustering analysis; self-organizing map neural network; data mining; Hunan grid

1. Introduction

With the advancement of the “carbon peaking and carbon neutrality” policy, the integration of new energy into the power system has been developed rapidly [1–3], and wind energy is currently one of the main forms of new energy utilization [4,5]. However, wind power output exhibits strong intermittency and volatility [6,7], posing significant challenges to the stable operation and programming of the power grid [8,9]. To effectively accomplish tasks such as power balance and operational control in the power grid, the clustering analysis of wind power output is crucial, as its accuracy directly impacts wind power operation [10,11]. Wind power clustering involves dividing a collection of high-dimensional wind power output data into multiple clusters composed of similar objects [12,13], providing a reference for wind power output prediction [14,15], and offering scientific decision-making support for wind power operation, which is of paramount importance for promoting the effective development and utilization of wind power [16,17].

Characteristic analysis is a fundamental method to study the law of wind power output [18]. It mainly analyzes the characteristic indicators of wind power output, which are usually expressed as extreme values and average values. Based on the extreme and average values, Han et al. [19] adopted the fluctuation ratio and ramp ratio to describe the volatility of wind power output; Wang et al. [20] utilized kurtosis and skewness as

two indicators to characterize the characteristics of wind power output curves; and Li et al. [21] employed fluctuation magnitude and rate indicators to reveal the fluctuation characteristics of wind power. These characteristic analyses use discrete indicators to describe the continuous output characteristics of wind power, resulting in a certain amount of information loss. When fewer indicators are selected, only local characteristics of the output can be revealed; on the contrary, when too many indicators are selected, the information similarity will be too high, increasing the uncertainty of characteristic analysis results. This problem currently exists in wind power characteristic analysis. Wind power characteristic indicators often use simple statistical indicators such as maximum output, minimum output, and daily peak-to-valley difference, which are all discrete characteristic indicators. The analysis results make it difficult to characterize the topological structure of wind power output data. As China implements “carbon peaking and carbon neutrality”, new energy such as wind power has received unprecedented attention [22–24]. If merely discrete indicators are used to analyze wind power characteristics, it will not be able to meet the needs of the new era for the refined management of new energy classification.

Clustering analysis provides important foundational support for the refined management of new energies such as wind power [25,26]. Commonly used clustering analysis methods include K-means [27], principal component analysis [28], projection pursuit [29], and the Self-Organizing Map (SOM) neural network [30]. Munshi [25] applied three clustering algorithms of two different unsupervised techniques to wind power output clustering and improved the K-means method; Sumathy et al. [31] evaluated the application of K-means and C-means algorithms to wind power data. An SOM can identify the entire process characteristics and topological structure of the data [32,33], which not only reduces information loss and clustering uncertainty but also overcomes the difficulty in clustering visualization caused by high input dimensions. Therefore, an SOM is more suitable for the clustering analysis of high-dimensional nonlinear distributed wind power output data [34,35].

Wind power output is intermittent and volatile and exhibits both diurnal and seasonal variations [36,37]. The wind power output time series with the daily minimum window has the strongest periodicity and has obvious daily and seasonal distribution characteristics. Therefore, clustering the daily wind power output processes by season can extract wind farm scenarios that simultaneously represent both diurnal and seasonal variations.

To address the issue of information loss in the analysis of wind power output characteristics based on discrete indicators, this study proposes a wind power output clustering method based on the Self-Organizing Map neural network to extract typical wind power scenarios and perform characteristic analysis. Using the wind power output process of the Hunan power grid, SOM networks with different dimensions of competitive layer structures are trained. Combining representativeness and topology analysis, the optimal dimension of the competitive layer structure (the number of clusters) was selected. Based on the clustering results of the obtained optimal dimension, the seasonal patterns of wind power output in the Hunan power grid were analyzed by considering the overall indicators, segmented indicators, and single-point indicators. The extraction of typical wind power output scenarios and characteristic analysis provide a basis for the refined management of wind power output classification, which is conducive to improving the grid’s absorption level of new energy and is of great significance to ensuring the stable operation of the power system.

2. Methods

This study employs data-mining algorithms to extract typical seasonal scenarios of wind power outputs to elevate wind power utilization rates and grid system stability. The proposed framework of clustering analysis of wind power outputs is illustrated in Figure 1 and comprises three components. Firstly, the typical seasonal scenarios of wind power outputs across various dimensions are determined by using the SOM to establish a foundation for characteristic analysis (Figure 1a). Subsequently, the competitive layer structure for

the best dimensions is identified by utilizing relevant evaluation indicators to determine the typical scenarios of wind power outputs (Figure 1b). Finally, the typical scenarios are analyzed by adopting characteristic indicators to obtain the seasonal characteristics of wind power outputs (Figure 1c). A concise overview of the adopted methods is provided below.

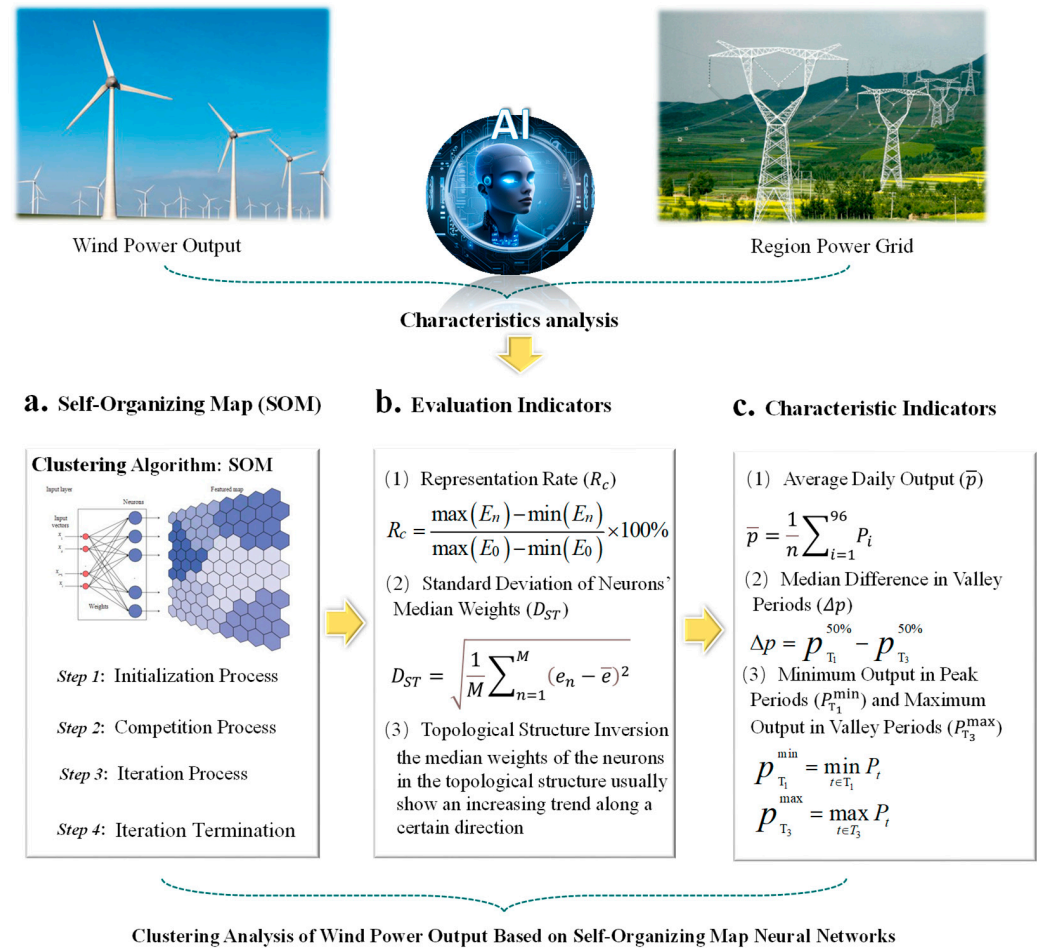


Figure 1. The proposed framework of clustering analysis of wind power outputs: (a) Self-Organizing Map; (b) evaluation indicators; (c) characteristic indicators.

2.1. Self-Organizing Map (SOM) Neural Network

The SOM, an unsupervised machine learning method, consists of an input layer and a competition layer. It can effectively preserve the topological structure and distribution characteristics of the original input data and is mainly used for the clustering analysis of high-dimensional data [34,35,38]. The SOM has been successfully applied in water quality analysis [39], flood process classification [40], and flood mapping classification [41], among others. However, there is no research on wind power output process clustering based on the SOM. The SOM can imitate the function of biological neurons and self-organize to mine the essential characteristics among sample data. Through competitive training of neurons in response to external stimuli [38], each high-dimensional wind power output process is mapped onto two dimensions while maintaining the original topological structure of the sample data. The principle and steps of clustering the wind power output process based on the SOM are as follows:

2.1.1. Initialization Process

1. Input a high-dimensional dataset of wind power outputs with N dimensions. Then, normalize and organize it into Q input vectors with N dimensions, X_q ($q = 1, 2, \dots, Q$), where Q is the number of samples in the dataset;

2. Design the competitive layer structure based on the estimated number of clusters, M , such as 1×2 , 1×3 , and 2×3 dimensions corresponding to 2, 3, and 6 clusters, respectively;
3. Set the competitive layer's neurons according to N and M , randomly assign initial values to the neurons' weights, and normalize them to obtain the initial weight vectors, $w = [w_n]^T$ ($n = 1, 2, \dots, N$), resulting in a weight matrix of $M \times N$ neurons;
4. Initialize the neighborhood radius $O_n(0)$, learning rate $v(0)$, learning rate threshold $v_{\min}$, and maximum number of iterations $D_{\max}$.

2.1.2. Competition Process

1. Calculate the distance between each input vector and the weight vector of each neuron using the Euclidean distance formula:

$$d_n = \sqrt{\sum_{q=1}^Q (x_q - w_{qn})^2} \quad (1)$$

where d_n is the distance between the weight vector of the competitive layer neuron and the input vector; x_q is the q -th input vector; and w_{qn} is the weight vector of the n -th neuron in the competitive layer corresponding to the q -th input variable.

2. Select the neuron with the smallest distance to each input vector as the winning neuron.

2.1.3. Iteration Process

1. Based on the winning neuron and the neighborhood radius, define the range of sample points within the neighborhood radius of the winning neuron, forming the winning neighborhood. The neighborhood radius shrinks as the number of iterations increases;
2. Update the weight vector of the winning neuron for the input vectors according to the idea that the closer to the winning neuron, the greater the update amplitude, with the purpose of making the winning neuron closer to the sample points within the winning neighborhood.

$$w_{nm}(t+1) = w_{nm}(t) + v(t)[x_{nq} - w_{nm}(t)] \quad (2)$$

where n is the index of the input variable corresponding to a neuron in the winning neighborhood; $w_{nm}(t)$ is the weight value of the m -th neuron corresponding to the n -th index at the t -th iteration; $v(t)$ is the learning rate at the t -th iteration; and x_{nq} is the q -th input vector corresponding to the n -th index.

2.1.4. Iteration Termination

If the learning rate $v(t)$ exceeds the threshold $v_{\min}$, or the number of iterations t does not reach the maximum number of iterations $D_{\max}$, continue the competition and iteration process; otherwise, stop the calculation and output the final clustering analysis results of wind power output data [42].

2.2. Evaluation Indicators for Identifying SOM Network Structure

1. Representation Rate (R_c): The ratio of neurons' median weight interval to samples' median interval can reflect the representativeness of clustering. The calculation formula is given as follows:

$$R_c = \frac{\max(E_n) - \min(E_n)}{\max(E_0) - \min(E_0)} \times 100\% \quad (3)$$

where E_n is the vector of neurons' median weights; E_0 is the vector of samples' medians; and max and min are the maximum and minimum elements of E_n , respectively. A larger R_c value indicates better clustering representativeness [41].

2. Standard Deviation of Neurons' Median Weights (D_{ST}): The standard deviation of neurons' median weights can reflect the diversity of clusters. The calculation formula is given as follows:

$$D_{ST} = \sqrt{\frac{1}{M} \sum_{n=1}^M (e_n - \bar{e})^2} \quad (4)$$

where M is the number of clusters, e_n ($n = 1, 2, \dots, M$) is the median weight of each neuron, and $\bar{e}$ is the mean of e_n . A larger D_{ST} value indicates greater differences between clusters.

3. Topological Structure Inversion: After training, the median weights of the neurons in the topological structure usually show an increasing trend along a certain direction [24]. If the median weights of two adjacent neurons do not reflect such status, it indicates that the neurons at these two positions in the topological structure are inverted [41], meaning that the clustering results are unstable.

2.3. Characteristic Indicators of Wind Power Outputs

The wind power characteristic indicators selected in this study are not used in the input stage of the SOM algorithm but in the post-evaluation stage of clustering results. The peak and valley periods are divided according to the user load conditions in Hunan Province. The specific division results are shown in Table 1. Wind power characteristic indicators include overall indicators, segmented indicators, and single-point indicators.

1. Overall Indicators, including Average Daily Output ($\bar{p}$): The average wind power output of a day can be intuitively identified as "high wind" or "low wind" on certain days.

$$\bar{p} = \frac{1}{n} \sum_{i=1}^{96} P_i \quad (5)$$

where n is the total number of data points and P_i is the wind power output of the i -th data point.

2. Segmented Indicators, including Median Difference in Valley Periods (Δp): The difference between the median of peak periods and the median of valley periods, which can characterize the impact of wind power output on the power grid's regulation capacity.

$$\Delta p = P_{T_1}^{50\%} - P_{T_3}^{50\%} \quad (6)$$

where $P_{T_1}^{50\%}$ and $P_{T_3}^{50\%}$ represent the medians of peak and valley periods, respectively, and the calculation formula is given as follows:

$$P_{T_0}^{50\%} = \begin{cases} p'_{0.5(n+1)} & n \text{ is an odd number} \\ 0.5(p'_{0.5n} + p'_{0.5n+1}) & n \text{ is an even number} \end{cases} \quad (7)$$

where $o = 1, 2$, and 3 , and p'_i ($i = 1, 2, \dots, n < 96$) is the data point sequence sorted by wind power outputs.

3. Single-point Indicators, including Minimum Output in Peak Periods ($P_{T_1}^{\min}$) and Maximum Output in Valley Periods ($P_{T_3}^{\max}$): The former reflects the minimum contribution of wind power to the power grid during peak demand periods, while the latter reveals the maximum challenge to the power grid's stability during off-peak periods.

$$P_{T_1}^{\min} = \min_{t \in T_1} P_t \quad (8)$$

$$P_{T_3}^{\max} = \max_{t \in T_3} P_t \quad (9)$$

where P_t is the wind power output at time t .

Table 1. User load division in Hunan province.

Period	Symbolic Representation	Specific Time Periods
Peak period	T_1	11:00–14:00, 18:00–23:00
Off-peak period	T_2	7:00–11:00, 14:00–18:00
Valley period	T_3	23:00–7:00 the next day

3. Study Area and Data Materials

In recent years, the Hunan power grid has formed a relatively complete provincial-level power grid structure. The 500 kV level power network has built a backbone for “West-to-East power transmission, North-South mutual aid”, and has constructed a ring network operation pattern around the “two receiving ends” of eastern and southern Hunan Province. The 220 kV level power grid has been continuously improved, and the power supply service scope has expanded to 14 cities (prefectures) in the province, covering 96% of the province’s area and serving 98% of the province’s population, basically forming a provincial power supply pattern. The Hunan power grid, as a crucial infrastructure for energy transmission and distribution, shoulders the important responsibility of ensuring the power supply and reliability of Hunan Province.

With the advancement of the country’s “dual carbon” goals and the construction of a new power system with new energy as the main body, new energy has ushered in large-scale development opportunities [1,24]. The development of hydropower resources in Hunan Province has reached saturation. In the future, the development of clean energy will mainly focus on new energy fields such as wind power [2,3]. As of the end of 2017, Hunan power grid’s installed wind power capacity reached 2.635 million kW, accounting for about 6% of the total installed capacity; as of the end of March 2022, Hunan Province has built 120 wind farms, and Hunan power grid’s installed wind power capacity reached 9 million kW, accounting for about 15% of the total installed capacity. The proportion of wind power generation has been increasing continuously and has become one of the important energy sources in Hunan. Under the trend of low-carbon energy development, new energy sources such as wind power in Hunan will be connected to the power grid on a large scale. The peak of power generation and the peak of power consumption often do not match, resulting in a serious lack of peak load regulation during peak power consumption. In addition, Hunan’s wind power resource utilization efficiency is not high, belonging to the fourth category of wind energy resources, and the number of wind power utilization hours is only about 70% and 50% of those in Yunnan and East Mongolia, respectively. The cluster analysis of wind power outputs in the Hunan power grid has important reference value for power supply during peak power consumption, optimal resource allocation, optimal grid dispatching, and the “wind, light, water, fire, and storage” regulation and operation system, thereby helping Hunan Province strengthen the cross-provincial balance of power surplus and shortage.

Data on the entire daily wind power output process of the Hunan power grid from 2017 to 2022 (6 years) were collected, including 210,240 pieces of 15 min wind power output data ($=6 \text{ years} \times 365 \text{ days} \times 96 \text{-time steps}$) and a total of 2190 daily wind power output samples ($=6 \text{ years} \times 365 \text{ days}$), all of which are complete daily wind power output processes. This dataset not only contains information such as the minimum output during peak periods, the maximum output during off-peak periods, and fluctuations in output but also preserves the characteristics of the entire wind power process, providing reliable input for wind power cluster analysis. Approximately 549 daily wind power output scenarios ($=6 \text{ years} \times 3 \text{ months} \times 30.5 \text{ days}$) were used for each season to extract seasonal representatives of daily wind power output events. Additionally, 12 seasonal representatives of daily wind power output events for each of the four seasons ($=3 \text{ scenarios} \times 4 \text{ seasons}$) were used for clustering analysis of Hunan Province’s wind power output characteristics.

4. Results

Since the dimension of the entire wind power output process is high (96 points), it is necessary to train different competitive layer structures of the SOM to determine the appropriate number of clusters. Firstly, three sizes of an SOM, 1×2 , 1×3 , and 2×2 dimensions, are used to train the wind power output data of spring, summer, autumn, and winter for 10,000 iterations, and the wind power output processes of the Hunan power grid from 2017 to 2022 are clustered into 2, 3, and 4 categories. Secondly, the representation rate and topological structure indicators are used to optimize the competitive layer structure. Finally, clustering analysis is performed based on the selected competitive layer structure, and the reliability of the clustering results is evaluated using overall indicators, segmented indicators, and single-point indicators.

4.1. Training SOM Neural Networks

Figure 2 reveals the variation of each neuron's median weight corresponding to training frequency for different SOM dimensions. The topology diagram of the competitive layer is shown in the upper left corner, with each neuron identified by different colors and circles. The right upper corner displays the R_c values and D_{ST} values at the completion of training. As shown in Figure 2, the distribution of the median weight values after training is more dispersed than the initial values. The 1×3 dimension exhibits the largest value for each period, with the maximum values being 61.2%, 72.9%, 66.5%, and 52.2% for spring, summer, autumn, and winter, respectively. The SOM neurons are not uniformly distributed between maximum and minimum values but tend to be located in the intervals where sample data are more densely grouped, ensuring that neurons are trained to move closer to the sample points in their respective winning neighborhoods. The R_c values indicate that the 1×2 -dimensional competitive layer structure has poor representativeness for wind power output clustering and is therefore not recommended. The 2×2 -dimensional structure still has neurons that nearly overlap each other, with unclear category boundaries, so it is not recommended either. Further identification of the appropriate number of clusters will be conducted using the competitive layer topology diagram.

Figure 3 provides a comparison diagram of the competitive layer topology in different dimensions of the SOM. The median weight value of each neuron is represented by the size and color depth of the circles, with the specific size reference on the right side. The number inside each circle indicates the specific weight value of a neuron, and the information in square brackets represents the position of a neuron in the competitive layer structure, with arrows indicating the direction from small to large weight values. After training, the median weight values of the neurons in all seasons increase in a certain direction without inversion between neurons, indicating that the topological structure in each dimension can reflect the original topological structure of the data. The 1×3 -dimensional structure has the highest representativeness, and the representativeness of the 2×2 -dimensional structure is similar to that of the 1×3 -dimensional structure. However, too many categories in wind power output clustering will blur the boundaries of each category; therefore the 1×3 -dimensional structure was selected as the competitive layer structure for the classification of wind power output in the Hunan power grid.

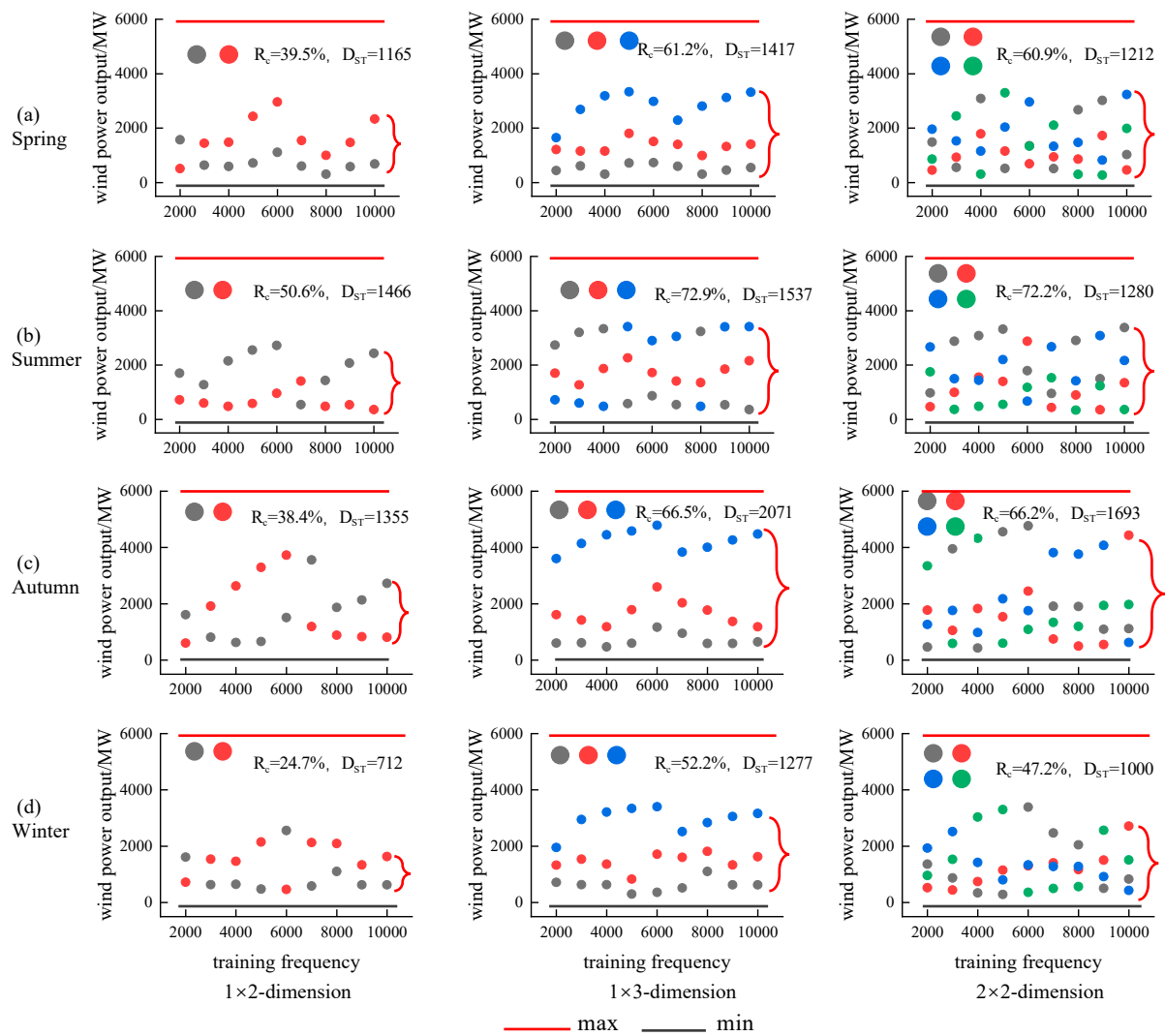


Figure 2. The variation of each neuron's median neuron weight value corresponding to training frequency in different stages of wind power generation.

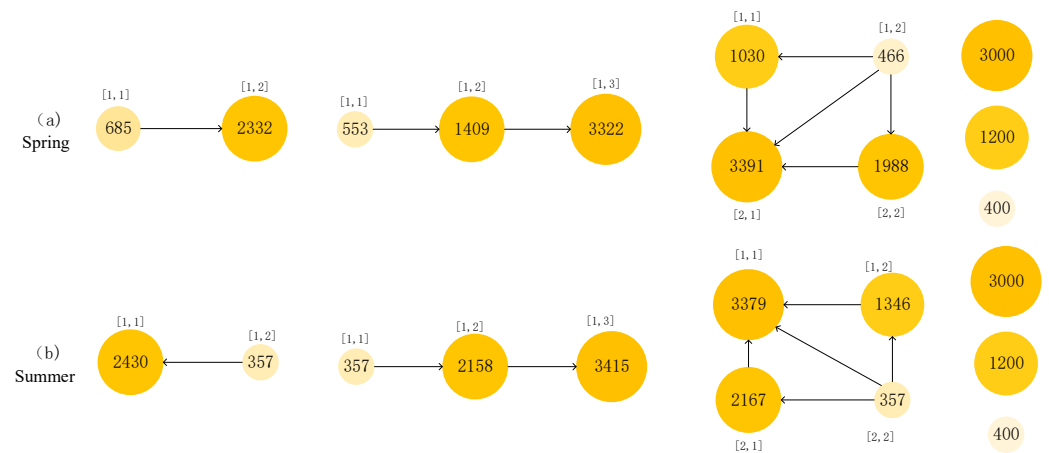


Figure 3. Cont.

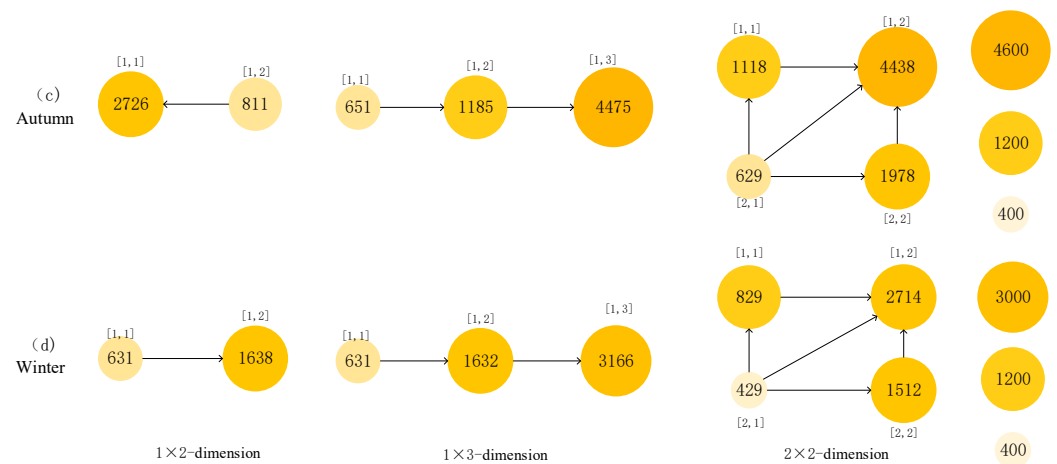


Figure 3. Comparison diagram of topological structures in different dimensions of wind power in different periods.

4.2. Clustering Analysis of Wind Power Output Processes

Based on the 1×3 -dimensional competitive layer structure, the wind power output clustering results were further analyzed. First, the clustered wind power output process diagram was drawn according to the competitive layer topology diagram of the neurons (Figure 4), where the upper part of the figure labels the number of wind power output samples in each category, and the lower part labels the position of the neuron in the 1×3 competitive layer structure (e.g., [1,3]).

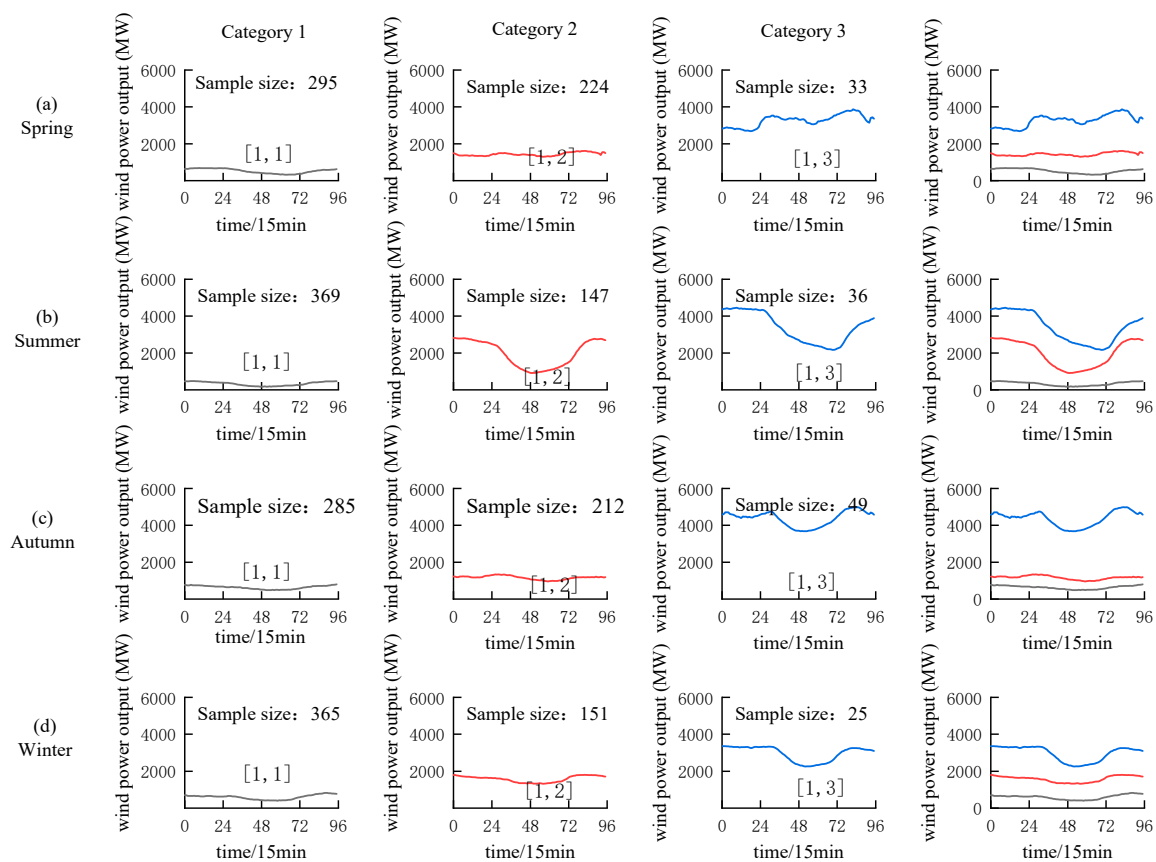


Figure 4. Wind power output process diagram for each period based on a 1×3 -dimensional competitive layer structure. The lines colored in gray, red, and blue represent the small, medium, and large power outputs.

Table 2 presents the characteristic values corresponding to each category of wind power outputs. Combining Figure 4 and Table 2, it can be observed that the two most distinct types of wind power outputs (Category 1 and Category 3) are distributed at the ends of the competitive layer. Considering the topology of the SOM, the three indicators of Category 1 are described as Low/Low/Low and those of Category 3 are described as High/High/High, while Category 2 is described as Medium/Low/High. All categories have positive Δp values, indicating positive peak shaving, and vice versa for negative peak shaving. The wind power output samples in each category are named according to season and category, such as Spring Category 1, Spring Category 2, and Spring Category 3, etc. Category 3 has the fewest wind power output samples, which is consistent with the general law of wind power outputs.

Combining the percentages in Table 2 and the overall indicator information, it is noted that both the probability and $\bar{p}$ value of Category 3 in winter are the lowest. Moreover, the trend of the typical winter scenario is the most stable, indicating that wind power output is relatively stable in winter compared to those of the other seasons. Although both the probability and $\bar{p}$ value of Category 3 in autumn are the highest, considering the trend of the typical autumn scenario, the fluctuation in wind power output is relatively stable in autumn compared to those of spring and summer. Figure 4 shows that wind power output fluctuations are greatest in summer, followed by spring.

According to the percentages in Table 2 and the segmented and single-point indicators, it can be seen that the frequencies of Category 1 and Category 2 in winter are positive peak shaving, and the probability of Category 3 being negative peak shaving in winter is lower than that of the other seasons. Moreover, the fluctuation is smaller in winter, indicating that winter is relatively conducive to the stable operation of the power grid. In summer, all categories are negative peak shaving, the two single-point indicators differ greatly, and there is significant volatility in summer, indicating that summer wind power output poses a certain threat to the stability of the power grid.

The overall and single-point indicators of Category 3 at [1, 3] in autumn are much higher than those of the other categories of wind power outputs, indicating that the largest wind power output in a year most likely occurs in autumn. The overall and single-point indicators of Category 1 at [1, 1] in summer are the lowest among all categories, and the probability of occurrence is the highest, indicating that the probability of the smallest wind power output in summer is the highest. By overall control of the occurrence probabilities of different categories in each season, the output is relatively small in summer but relatively large in winter.

Combining Figure 4 and Table 2, it can be concluded that summer and winter are the two most representative seasons. In summer, wind power and hydropower complement each other in seasonal fluctuations [43]. During summer, wind power output is relatively small and exhibits greater volatility. The proportion of wind power in the power system is unstable, especially when the widespread use of air-conditioning equipment leads to a significant increase in electricity demand, yet the load will fluctuate significantly. Therefore, it is necessary to be prepared to use backup power sources such as pumped storage power plants to cope with potential power shortages. Compared with thermal power units and conventional hydropower units, pumped storage power stations not only have more flexible regulation capabilities but also have a greater regulatory effect on the rapid fluctuations in wind power [44,45]. They can store energy during high power output (Category 3) and release energy during low power output (Category 1 and Category 2) or when demand increases, which not only balances supply and demand but also improves energy utilization and reduces reliance on thermal power units.

Table 2. Eigenvalues of different clusters of wind power output in different seasons.

Period	Category	Location	Sample Size	Percentage	Overall /MW	Segmented/MW	Single-Point /MW		Category Representation *	
					$\bar{p}$	Δp	$p_{T_1}^{\min}$	$p_{T_3}^{\max}$		
Spring	1	[1, 1]	295	53.40%	527	−228	370	687	1 March 2017	2 March 2017
	2	[1, 2]	224	40.60%	1435	168	1317	1555	30 March 2017	31 March 2017
	3	[1, 3]	33	6.00%	3264	716	3107	3440	12 March 2022	13 March 2022
Summer	1	[1, 1]	369	66.80%	333	−124	180	476	1 June 2017	2 June 2017
	2	[1, 2]	147	26.60%	1969	−844	912	2813	5 June 2019	6 June 2019
	3	[1, 3]	36	6.60%	3374	−1589	2239	4449	12 July 2021	13 July 2021
Autumn	1	[1, 1]	285	52.20%	636	−89	492	801	1 September 2017	2 September 2017
	2	[1, 2]	212	38.80%	1165	−56	1012	1335	11 October 2017	12 October 2017
	3	[1, 3]	49	9.00%	4375	121	3676	4710	7 October 2021	8 October 2021
Winter	1	[1, 1]	365	67.50%	597	2	414	790	1 January 2017	2 January 2017
	2	[1, 2]	151	27.90%	1581	73	1326	1797	19 December 2017	20 December 2017
	3	[1, 3]	25	4.60%	2962	−229	2258	3362	17 December 2021	18 December 2021

* Note: Classification representation involves selecting two consecutive days or a continuous period that consistently falls into the same category as the representative.

In winter, wind power has good regulation performance because it has relatively small output fluctuation and relatively stable output. Therefore, wind power generation can be better integrated with hydropower generation. When wind power output is high (Category 3), hydropower stations can reduce power generation and use the excess electricity for energy storage in pumped storage power stations. When wind power output is low (Category 1 and Category 2), hydropower stations can cooperate with pumped storage power stations to increase power generation, supplement the power demand, and effectively improve wind power absorption capacity.

Observing the two consecutive samples representing Category 3 wind power output in each cluster, it was found that both occurred after 2021. In July 2021, Hunan Province was hit by Typhoon “Inferno”, and in December 2021, Typhoon “Rai” occurred. This demonstrates the reliability of the SOM in clustering wind power output.

5. Discussion

Usually, the analysis of wind power output characteristics often adopts indicators similar to those used for load characteristics. However, unlike the stable characteristics of the load curves, wind power output has obvious intermittent and fluctuating characteristics. This indicator method not only fails to accurately reflect the characteristics of wind power outputs but also overlooks crucial information. It does not provide guidance for the operation of the power grid system. The clustering analysis of wind power outputs based on the SOM conducted in this study can complete the data mining of high-dimensional and long-sequence wind power outputs. Munshi [25] leveraged ant colony swarm clustering, bat clustering, and K-means clustering techniques to analyze wind power output. The evaluation of these unsupervised clustering algorithms was conducted using eight validity indices, which involved a complex process to determine the optimal number of clusters. Alternatively, Ou [46] employed the K-means algorithm to categorize the daily output curves of offshore wind power into 20 distinct categories. The SOM clustering method adopted in this paper is concise and clear while ensuring the accuracy of the clustering results. Based on the wind power output clustering rules, future research can focus on short-term wind power output prediction, peak shaving scheduling, and new energy management. For instance, when forecasting wind power outputs, subsequent wind power output processes can be predicted by comparing the current characteristics of wind power outputs on the forecast day with those of various types of wind power outputs from the corresponding period. During peak shaving scheduling, procedures can be developed for different categories of wind power outputs. When similar wind power outputs occur, the established procedures for peak shaving scheduling can be followed, ensuring stable operation of the power grid. In new energy management, optimization scheduling plans can be developed based on the characteristics of different categories of wind power outputs to reduce the abandonment rate of new energy and enhance power supply security.

6. Conclusions and Recommendations

This study developed a seasonal wind power output clustering analysis method based on the SOM, analyzed the representativeness and topological structure of different competitive layer structures, selected the wind power output clustering results based on the optimal competitive layer structure, realized the topological relationship mining of the entire process of wind power output, and conducted a reasonable evaluation on SOM results using overall, segmented, and single-point indicators. The main conclusions are presented as follows:

First, the 1×2 -dimensional competitive layer structure had significantly lower representativeness compared to other dimensions. The 2×2 -dimensional competitive layer structure had unclear classification boundaries. In contrast, the 1×3 -dimensional competitive layer structure had the highest representativeness, there is no topological structure inversion phenomenon, the number of categories is appropriate, and it can effectively identify the clustering characteristics of the wind power output process in Hunan Province.

Second, the wind power output process in Hunan Province could be divided into four periods, spring, summer, autumn, and winter, each of which could be clustered into three categories. The characteristics of wind power output in different categories varied, and the topological structure of the corresponding neurons could effectively represent the relationship and differences among the wind power output samples. This indicated that SOM clustering could effectively extract wind power output characteristics and the clustering results were reliable.

Third, summer was the most volatile for wind power output, while winter was the most stable; summer posed the greatest threat to grid operation, while winter was relatively conducive to grid stability. The largest and smallest wind power outputs in a year most likely occurred in the autumn and summer, respectively. Summer and winter were the most representative seasons. Summer required the use of backup water sources, such as pumped storage power stations, to balance power supply and demand during high and low output, respectively. In winter, hydropower stations were needed to cooperate with pumped storage power stations, reducing power generation when wind power output was high and increasing hydropower generation when wind power output was low, to improve wind power absorption capacity.

The seasonal clustering analysis of wind power outputs using the SOM helps to analyze seasonal wind power output patterns, which is of great significance for the stable operation of the power grid. However, this study requires further examination in the selection of indicators and lacks a comprehensive description of how the outcomes of cluster analysis are applied to wind power systems. Future research will focus on the short-term prediction of wind power outputs based on clustering analysis, providing a foundation for the refined management of wind power output classification, and enhancing the grid's ability to absorb new energy.

Author Contributions: Conceptualization, Y.Z. and C.-Y.X.; data curation, Y.L. (Yanqian Li) and Y.L. (Yuxuan Luo); formal analysis, Y.L. (Yanqian Li), Y.L. (Yuxuan Luo), and Z.N.; investigation, Y.L. (Yanqian Li), Y.Z., and C.-Y.X.; methodology, Y.L. (Yanqian Li), Y.Z., and C.-Y.X.; project administration, Y.Z. and C.-Y.X.; resources, Y.Z.; software, Y.L. (Yanqian Li) and Y.L. (Yuxuan Luo); supervision, Y.Z. and C.-Y.X.; validation, Y.L. (Yuxuan Luo) and Z.N.; visualization, Y.L. (Yanqian Li), Y.L. (Yuxuan Luo), and Z.N.; writing—original draft, Y.L. (Yanqian Li) and Y.Z.; writing—review and editing, C.-Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Key Research and Development Program of China (No. 2021YFC3200303).

Data Availability Statement: The data are not publicly available due to a legal agreement between the authors and Hunan Provincial Grid.

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

References

1. Gao, J.; Pan, L. A System Dynamic Analysis of Urban Development Paths under Carbon Peaking and Carbon Neutrality Targets: A Case Study of Shanghai. *Sustainability* **2022**, *14*, 15045. [\[CrossRef\]](#)
2. Shu, H.; Song, W.; Song, Z.; Guo, H.; Li, C.; Wang, Y. Multistep short-term wind speed prediction with rank pooling and fast Fourier transformation. *Wind. Energy* **2024**, *27*, 667–694. [\[CrossRef\]](#)
3. Fan, W.; Cai, L.; Li, M.; Zhou, M.; Cao, J.; Wang, J.; Fan, Y. Investigation of Upward Leader Development Characteristics by Two Kinds of Charge Density Models in Wind Turbines. *Wind. Energy* **2024**, *27*, 979–993. [\[CrossRef\]](#)
4. Roga, S.; Bardhan, S.; Kumar, Y.; Dubey, S.K. Recent technology and challenges of wind energy generation: A review. *Sustain. Energy Technol. Assess.* **2022**, *52*, 102239. [\[CrossRef\]](#)
5. Superchi, F.; Giovannini, N.; Moustakis, A.; Pechlivanoglou, G.; Bianchini, A. Optimization of the power output scheduling of a renewables-based hybrid power station using MILP approach: The case of Tilos island. *Renew. Energy* **2024**, *220*, 119685. [\[CrossRef\]](#)
6. Chen, L.; Li, J.; Zhang, L.; Gao, W.; Ai, C.; Ding, B. A novel collaborative control algorithm for maximum power point tracking of wind energy hydraulic conversion system. *Wind. Energy* **2023**, *26*, 1299–1317. [\[CrossRef\]](#)
7. Wang, P.; Zhang, Z.; Huang, Q.; Lee, W.-J. Wind Farm Dynamic Equivalent Modeling Method for Power System Probabilistic Stability Assessment. *IEEE Trans. Ind. Appl.* **2020**, *56*, 2273–2280. [\[CrossRef\]](#)

8. Lang, J.; Peng, X.; Li, W.; Cai, T.; Gan, Z.; Duan, S.; Li, C. A novel two-stage interval prediction method based on minimal gated memory network for clustered wind power forecasting. *Wind. Energy* **2021**, *24*, 450–464. [\[CrossRef\]](#)
9. Ma, G.; Li, Y.; Yang, J.; Wu, B. Cyber-physical Modeling Technique based Dynamic Aggregation of Wind Farm Considering LVRT Characteristic. *CES Trans. Electr. Mach. Syst.* **2023**, *7*, 229–236. [\[CrossRef\]](#)
10. Song, F.; Cui, J.; Yu, Y. Dynamic volatility spillover effects between wind and solar power generations: Implications for hedging strategies and a sustainable power sector. *Econ. Model.* **2022**, *116*, 106036. [\[CrossRef\]](#)
11. Hajou, A.; El Mghouchi, Y.; Chaoui, M. Novel approaches for wind speed evaluating and solar-wind complementarity assessing. *Renew. Energy Focus* **2024**, *48*, 100547. [\[CrossRef\]](#)
12. Yang, Z.; Peng, X.; Zhang, X.; Song, J.; Wang, B.; Liu, C. Short-Term Offshore Wind Power Prediction Based on Significant Weather Process Classification and Multitask Learning Considering Neighboring Powers. *Wind. Energy* **2024**, *27*, e2490. [\[CrossRef\]](#)
13. Yang, M.; Shi, C.; Liu, H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy* **2021**, *218*, 119515. [\[CrossRef\]](#)
14. Wang, Y.; Wang, D.; Tang, Y. Clustered Hybrid Wind Power Prediction Model Based on ARMA, PSO-SVM, and Clustering Methods. *IEEE Access* **2020**, *8*, 17071–17079. [\[CrossRef\]](#)
15. Xu, A.; Yang, T.; Ji, J.; Gao, Y.; Gu, C. Application of cluster analysis in short-term wind power forecasting model. *J. Eng.* **2019**, *2019*, 5423–5426. [\[CrossRef\]](#)
16. Gutierrez-Osorio, C.; Pedraza, C. Modern data sources and techniques for analysis and forecast of road accidents: A review. *J. Traffic Transp. Eng. (Engl. Ed.)* **2020**, *7*, 432–446. [\[CrossRef\]](#)
17. Zhao, M.; Wang, Y.; Wang, X.; Chang, J.; Zhou, Y.; Liu, T. Modeling and Simulation of Large-Scale Wind Power Base Output Considering the Clustering Characteristics and Correlation of Wind Farms. *Front. Energy Res.* **2022**, *10*, 810082. [\[CrossRef\]](#)
18. Li, Y.; Huang, X.; Tee, K.F.; Li, Q.; Wu, X.-P. Comparative study of onshore and offshore wind characteristics and wind energy potentials: A case study for southeast coastal region of China. *Sustain. Energy Technol. Assess.* **2020**, *39*, 100711. [\[CrossRef\]](#)
19. Han, S.; Zhang, L.-N.; Liu, Y.-Q.; Zhang, H.; Yan, J.; Li, L.; Lei, X.-H.; Wang, X. Quantitative evaluation method for the complementarity of wind–solar–hydro power and optimization of wind–solar ratio. *Appl. Energy* **2019**, *236*, 973–984. [\[CrossRef\]](#)
20. Wang, Y.; Hu, Q.; Li, L.; Foley, A.M.; Srinivasan, D. Approaches to wind power curve modeling: A review and discussion. *Renew. Sustain. Energy Rev.* **2019**, *116*, 109422. [\[CrossRef\]](#)
21. Li, L.; Li, Y.; Zhou, B.; Wu, Q.; Shen, X.; Liu, H.; Gong, Z. An adaptive time-resolution method for ultra-short-term wind power prediction. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105814. [\[CrossRef\]](#)
22. Zhang, Y. Analysis of China's energy efficiency and influencing factors under carbon peaking and carbon neutrality goals. *J. Clean. Prod.* **2022**, *370*, 133604. [\[CrossRef\]](#)
23. Wang, S.; Chen, F.; Wang, Y. Low-carbon technological innovation of coal production and utilization impetus carbon neutrality in China. *Environ. Sci. Pollut. Res.* **2023**, *30*, 80916–80930. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Ghoushchi, S.J.; Manjili, S.; Mardani, A.; Saraji, M.K. An extended new approach for forecasting short-term wind power using modified fuzzy wavelet neural network: A case study in wind power plant. *Energy* **2021**, *223*, 120052. [\[CrossRef\]](#)
25. Munshi, A.A. Clustering of Wind Power Patterns Based on Partitional and Swarm Algorithms. *IEEE Access* **2020**, *8*, 111913–111930. [\[CrossRef\]](#)
26. Wang, Z.H.K. Clustering analysis to improve the reliability and maintainability of wind turbines with self-organizing map neural network. *Int. J. Perform. Eng.* **2013**, *9*, 245. [\[CrossRef\]](#)
27. Krishna, K.; Murty, M.N. Genetic K-means algorithm. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **1999**, *29*, 433–439. [\[CrossRef\]](#)
28. Greenacre, M.; Groenen, P.J.; Hastie, T.; d'Enza, A.I.; Markos, A.; Tuzhilina, E. Principal component analysis. *Nat. Rev. Methods Primers* **2022**, *2*, 100. [\[CrossRef\]](#)
29. Friedman, J.; Tukey, J. A Projection Pursuit Algorithm for Exploratory Data Analysis. *IEEE Trans. Comput.* **1974**, *100*, 881–890. [\[CrossRef\]](#)
30. Corona, F.; Mulas, M.; Baratti, R.; Romagnoli, J.A. On the topological modeling and analysis of industrial process data using the SOM. *Comput. Chem. Eng.* **2010**, *34*, 2022–2032. [\[CrossRef\]](#)
31. Sumathy, P.; Senthilkumar, P. Comparative Analysis of Clustering Algorithm for Wind Power. *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.* **2018**, *3*, 565–574.
32. Kalth, A.M.; Hjorth, P.; Berndtsson, R. Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environ. Model. Softw.* **2008**, *23*, 835–845. [\[CrossRef\]](#)
33. Chon, T.-S. Self-Organizing Maps applied to ecological sciences. *Ecol. Inform.* **2011**, *6*, 50–61. [\[CrossRef\]](#)
34. Liu, Y.; Wang, Y.; Wang, Q.; Zhang, K.; Qiang, W.; Wen, Q.H. Recent advances in data-driven prediction for wind power. *Front. Energy Res.* **2023**, *11*, 1204343. [\[CrossRef\]](#)
35. Ahmad, A.U.; Starkey, A. Application of feature selection methods for automated clustering analysis: A review on synthetic datasets. *Neural Comput. Appl.* **2018**, *29*, 317–328. [\[CrossRef\]](#)
36. Graabak, I.; Korpås, M. Variability Characteristics of European Wind and Solar Power Resources—A Review. *Energies* **2016**, *9*, 449. [\[CrossRef\]](#)
37. Li, J.; Chen, S.; Wu, Y.; Wang, Q.; Liu, X.; Qi, L.; Lu, X.; Gao, L. How to make better use of intermittent and variable energy? A review of wind and photovoltaic power consumption in China. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110626. [\[CrossRef\]](#)
38. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69. [\[CrossRef\]](#)

39. Li, H.Y.; Li, Z.X.; Wang, C.J.; Han, Y.F.; Zhao, D.; He, J.C. An Evaluation Method for Water Quality Based on the Improved SOM Neural Network. *Appl. Mech. Mater.* **2013**, 373–375, 1220–1223. [[CrossRef](#)]
40. Kumar, V.; Azamathulla, H.M.; Sharma, K.V.; Mehta, D.J.; Maharaj, K.T. The State of the Art in Deep Learning Applications, Challenges, and Future Prospects: A Comprehensive Review of Flood Forecasting and Management. *Sustainability* **2023**, *15*, 10543. [[CrossRef](#)]
41. Chang, L.-C.; Wang, W.-H.; Chang, F.-J. Explore training self-organizing map methods for clustering high-dimensional flood inundation maps. *J. Hydrol.* **2021**, 595, 125655. [[CrossRef](#)]
42. Chang, F.-J.; Chang, L.-C.; Kao, H.-S.; Wu, G.-R. Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network. *J. Hydrol.* **2010**, 384, 118–129. [[CrossRef](#)]
43. de Jong, P.; Sánchez, A.; Esquerre, K.; Kalid, R.; Torres, E. Solar and wind energy production in relation to the electricity load curve and hydroelectricity in the northeast region of Brazil. *Renew. Sustain. Energy Rev.* **2013**, *23*, 526–535. [[CrossRef](#)]
44. Shahzad, S.; Jasińska, E. Renewable Revolution: A Review of Strategic Flexibility in Future Power Systems. *Sustainability* **2024**, *16*, 5454. [[CrossRef](#)]
45. Hunt, J.D.; Zakeri, B.; Lopes, R.; Barbosa, P.S.F.; Nascimento, A.; de Castro, N.J.; Brandão, R.; Schneider, P.S.; Wada, Y. Existing and new arrangements of pumped-hydro storage plants. *Renew. Sustain. Energy Rev.* **2020**, *129*, 109914. [[CrossRef](#)]
46. Ou, Z.; Lan, W.; Zhang, L.; Tong, Z.; Liu, D.; Liu, Z. Comparative Analysis of Offshore Wind Power Prediction Models and Clustering-Based Daily Output Classification. In Proceedings of the 2024 IEEE 2nd International Conference on Power Science and Technology, Yunnan, China, 11–19 May 2024; pp. 1482–1487. [[CrossRef](#)]

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