

## Research Article

# Dynamic Equivalent Modeling of Wind Farm Based on Dominant Variable Hierarchical Clustering Algorithm

Wenbo Jiang <sup>1,2</sup> and Mingyue Zhong<sup>1,2</sup>

<sup>1</sup>School of Electrical Engineering and Electronic Information, Xihua University, Chengdu 610039, China

<sup>2</sup>Sichuan Provincial Key Laboratory of Signal and Information Processing, Xihua University, Chengdu 610039, China

Correspondence should be addressed to Wenbo Jiang; [caswenbojiang@gmail.com](mailto:caswenbojiang@gmail.com)

Received 23 September 2021; Accepted 9 November 2021; Published 29 November 2021

Academic Editor: Albert Alexander Stonier

Copyright © 2021 Wenbo Jiang and Mingyue Zhong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The actual operating state of the wind turbine group is influenced by the wake effect and control mode; however, the current models cannot describe the actual operating state very well. A dynamic equivalent modeling method for a doubly fed wind power generator is proposed on the basis of ensuring the accurate description of the wind turbine group. As the clustering index, dominant variables are used in the hierarchical clustering algorithm, which are extracted by principal component analysis. Three dynamic equivalent models of 24 wind turbines are established using PSCAD software platform, which use 13 state variables, wind speed, and dominant variables as clustering indexes, respectively. Furthermore, the active power and reactive power output curves of wind farm are simulated in the case of the three-phase short-circuit fault on the system side or wind speed fluctuation, respectively. The simulation results demonstrate that it is reasonable and effective to extract slip ratio and wind turbine torque as clustering index, and the maximal relative error between the dominant variable equivalent model and 13-state-variable model is only 9.9%, which is greatly lower than that of the wind speed model, K-means clustering model, neural network model, and support vector machine model. This model is easy to implement and has wider application prospect, especially for characteristics analysis of large-scale wind farm connected to power grid.

## 1. Introduction

Wind power generation is one of the most important renewable energies, which has attracted more and more attention from most of the countries for its mature technology and low cost in recent years [1]. However, wind energy has characteristics of random, intermittent, and instability. It is very difficult to study the operational characteristics of power grid with large-scale wind farm interconnection [2–4]. Therefore, it is necessary to establish an effective model of wind farm, which should be as close as possible to the actual operating state of wind farm. Literature [5] proposed a method of reduced-order model computation for planning grid connection of a large-scale wind farm, and the effectiveness of the proposed method is demonstrated and evaluated by an illustrative simple example wind farm and an example large-scale wind farm with 200 WTGs (wind

turbine generators). Literature [6] used the integrating field measurement method to characterize the monthly wind speed and wind direction distributions and investigate the wind characteristics in turbine wakes. The research work shows that good agreement is obtained for both mean wind speed and turbulence intensity, which also verifies the possibility of combining actual field measurements and high-fidelity simulations to describe the characterization of utility-scale wind farms. Literature [7] studied large-eddy simulations of coherent structures within and above different wind farm configurations in a neutral atmospheric boundary layer (ABL) using proper orthogonal decomposition (POD) to improve understanding of the flow structures in both physical and spectral space. The research work indicates that wind farm dynamics in the ABL are very complex. Literature [8] proposed a simplified floating offshore wind turbine model, which is applicable for the real-

time simulation of large-scale floating offshore wind farms. The real-time results of offshore wind farms show the feasibility of the proposed turbine models for the real-time model of large-scale offshore wind farms.

Generally speaking, the wind farm model can be divided into detailed models and equivalent models, where the detailed model is difficult to solve, does not have good adaptability to parameter changes, takes a lot of work to modify, and requires high computer performance [9,10]. Therefore, most of the current research studies choose the equivalent model to describe the wind farm.

The equivalent model can be further divided into single-machine equivalence and multi-machine equivalence, where the single-machine equivalent model uses a large-capacity wind turbine to replace the whole wind farm for reducing the complexity of the simulation system and improving the speed of calculation, and the multi-machine equivalent model divides the whole wind farm into many groups according to one or more indexes and then uses a wind turbine to replace the wind farm within the same group.

Literatures [11–14] proposed the single-machine equivalent model of wind farm, but the characteristics of wind turbines change greatly based on actual applications. Literature [15] proposed a systematic and simple method to model large-scale induction machine-based wind farms by a single WTG that contains an equivalent mechanical wind turbine and an equivalent electrical generator. The research results show that the proposed method is adequately accurate in both transients and steady-state responses and it can be readily used for modeling large-scale wind farms to reduce the overall computational burden of the system. In a sense, the single-machine equivalent model is not accurate enough.

Literatures [16–18] proposed the earlier research results on multi-machine equivalent models of wind farm, where the wind farm is mainly grouped according to the geographic location. This classification method needs a regular arrangement of wind turbines and does not have good adaptability. To overcome the above shortcoming, literatures [19–29] proposed some novel classification methods, which are more close to the actual operating state of wind turbines. Yan et al. [19] sorted the wind farm using the support vector machine (SVM) under the consideration of similar wind speed and the wake effect of the wind farm. This method cannot accurately characterize the actual operating state of wind farm, which will lead to large algorithm errors. Literatures [20–22] sorted the wind farm by clustering algorithm, which is based on the measured data of all the wind turbines, the wind speed model, and the steady-state model of wind farm. However, this method does not consider dynamic characteristics of wind turbines. The clustering index of [23–25] was the dominant variable during the operation, and the dominant variables were extracted by the feature analysis method. However, the influence of disturbance location and type on the operation characteristics of wind turbines has not been considered by the characteristic roots. Literatures [3, 5, 26, 27] sorted the wind farm considering 13 state variables of wind turbines during the operation. Literature [28] presented a method to develop computationally efficient

dynamic model of a wind farm suitable for large disturbance simulation. The method based on a trajectory piecewise linear (TPWL) approximation uses single and multiple training trajectory to develop a nonlinear reduced-order model (ROM). Simulation results demonstrated the effectiveness of the model in capturing dynamic behavior of wind farm following large disturbances. But this kind of classification method has the disadvantages of redundant data, high calculation complexity, low applicability, etc.

There are still some shortcomings in the above classification methods, such as large error, incomplete consideration of actual operating state, complicated calculation and low applicability, and so on. More importantly, the large-scale wind farm has the characteristics of multi-time-scale and strong nonlinearity on the time axis. With the increase of wind farm's scale, the shortcomings of existing modeling methods are more obvious. To overcome the above shortcomings, the multi-machine dynamic equivalent modeling of wind farm based on dominant variables hierarchical clustering algorithm is proposed in this paper. The essence of principal component analysis is a mathematical transformation method, which transforms a given set of related variables into another set of irrelevant variables by linear transformation. These new variables are arranged in descending order of variance. During the process of mathematical transformation, the total variance of variables should always remain unchanged, and the first variable has the largest variance, which is called the first principal component. The second variable has the second largest variance and is independent of the first variable, which is called the second principle component. Then, the third and the fourth variables are obtained in the same method. Compared with other parameter analysis methods (correlation analysis, feature analysis, factor analysis, and so on), principal component analysis can not only retain as much information as possible about the original variables but also achieve the goal of dimensionality reduction, thus reducing the computational complexity and the cost of the algorithm. Meanwhile, compared with other classification methods (support vector machine, neural network, density clustering, grid clustering, and so on), the hierarchical clustering method does not need to determine the number of classifications and has high classification efficiency, which is especially suitable for the classification of large data.

The specific research steps in this paper are as follows. Firstly, the 13 state variables of the wind turbines during actual operation are calculated. Secondly, the dominant variables are extracted from 13 state variables by principal component analysis. Then, the wind farm is clustered by hierarchical clustering algorithm, where the dominant variables are used as clustering index. Furthermore, three dynamic equivalent models of 24 wind turbines are established using PSCAD software, which use 13 state variables, wind speed, and dominant variables as clustering indexes, respectively. Finally, the active power and reactive power output curves of wind farm are simulated in the case of the three-phase short-circuit fault on the system side or wind speed fluctuation, respectively. The simulation results verify the effectiveness of dominant variables and the advantages of this proposed model.

## 2. Mathematical Model of Doubly Fed Induction Generator

At present, the doubly fed induction generator (DFIG) is most widely used in wind farms all over the world. Its mathematical model mainly includes wind turbine module, induction motor module, transmission system module, back-to-back converter system control module, boost transformer module, R-L series components, etc. Some important modules will be described below [29–32].

**2.1. Wind Turbine Module.** The relationship between active power of the doubly fed induction generator and wind energy captured by wind turbines can be expressed as follows:

$$P_e = \frac{1}{2C_p\rho Av^3}, \quad (1)$$

where  $p_e$  is the active power of the generator,  $C_p$  is the utilization coefficient of wind energy,  $\rho$  is the air density,  $A$  is the sweeping area of the blade, and  $v$  is the input wind speed endured by the wind turbine.

In the case of a given wind speed, the captured wind power of the wind turbine mainly depends on the  $C_p$ , which is the function of the pitch angle ( $\beta$ ) and tip speed ratio ( $\lambda$ ).

$$\lambda = \frac{R\omega_r}{v}, \quad (2)$$

where  $\omega_r$  is the rotating speed of the wind turbine, which can be obtained from the power characteristic curve under the known active power of the generator ( $p_e$ ).

The characteristic curve of wind energy utilization coefficient is shown in Figure 1.

For variable-pitch wind turbines, the wind power of the impeller can be controlled by changing the pitch angle ( $\beta$ ). Usually, the wind energy utilization factor ( $C_p$ ) of the wind turbines reaches the maximum value when  $\beta = 0$ .

**2.2. Induction Motor Module.** In DQ coordinate, the flux-linkage equation and voltage equation of the doubly fed induction generator can be obtained from the characteristics of induction motor [23–25], which can be expressed as follows:

$$\begin{cases} \psi_{ds} = -L_{ss}I_{ds} + L_m I_{dr}, \\ \psi_{qs} = -L_{ss}I_{qs} + L_m I_{qr}, \\ \psi_{dr} = L_{rr}I_{dr} - L_m I_{ds}, \\ \psi_{qr} = L_{rr}I_{qr} - L_m I_{qs}, \end{cases} \quad (3)$$

$$\begin{cases} U_{ds} = -R_s I_{ds} + \frac{d\psi_{ds}}{dt} - \psi_{qs}, \\ U_{qs} = -R_s I_{qs} + \frac{d\psi_{qs}}{dt} + \psi_{ds}, \\ U_{dr} = R_r I_{dr} + \frac{d\psi_{dr}}{dt} - S\psi_{qs}, \\ U_{qr} = R_r I_{qr} + \frac{d\psi_{qr}}{dt} + S\psi_{ds}, \end{cases} \quad (4)$$

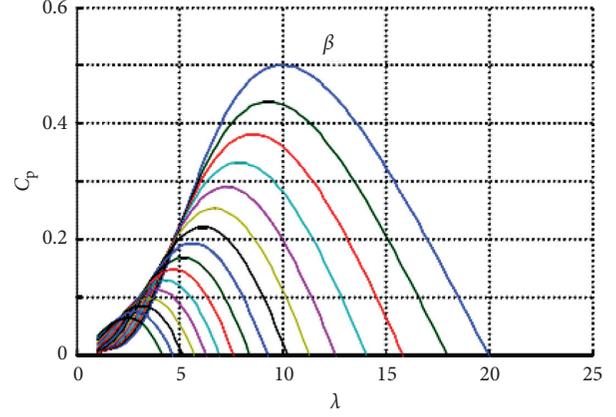


FIGURE 1: The characteristic curve of wind energy utilization coefficient.

where  $\psi$  and  $U$  represent flux linkage and the voltage, respectively; the subscripts  $d$  and  $q$  represent the direct-axis component and the quadrature-axis component, respectively; the subscripts  $s$  and  $r$  represent the stator and the rotor, respectively;  $L_{ss}$  and  $L_{rr}$  represent the self-inductance of the stator winding and rotor winding, respectively;  $L_m$  represents the mutual inductance between stator and rotor;  $R$  represents the resistance; and  $S$  represents the slip ratio.

**2.3. Transmission System Module.** By incorporating the inertia of gearbox into the inertia of generator rotor, a two-mass model of transmission device can be obtained [33], which can be expressed as follows:

$$\begin{cases} \frac{dT_t}{dt} = \frac{1}{T_d} (T_m - T_t), \\ \frac{dT_e}{dt} = \frac{1}{T_j} (T_e - T_t), \end{cases} \quad (5)$$

where  $T_m$  represents the wind turbine torque,  $T_t$  represents rotor mechanical torque,  $T_e$  represents the electromagnetic torque, and  $T_d$  and  $T_j$  represent the inertia time constant of the transmission system and the generator, respectively.

The electromagnetic power is the output power of the stator, and the expression of the electromagnetic torque ( $T_e$ ) can be deduced from the electromagnetic power, which can be expressed as follows:

$$T_e = L_m (I_{dr} I_{qs} - I_{ds} I_{qr}). \quad (6)$$

**2.4. Converter Control System Module.** The rotor of the doubly fed induction motor is connected to the power grid through back-to-back converter. Active power and reactive power of the generator consist of stator side and rotor side, where  $P_s$  and  $Q_s$  represent the active power and reactive power on stator side, while  $P_r$  and  $Q_r$  represent the active power and reactive power on rotor side [29–32]. In the steady-state process, the stator copper consumption is

neglected, the differential component is 0, and the direct-axis component of the stator voltage is 0. Combining with the stator side active power expression, the active power and reactive power expression of the stator side can be obtained, which can be expressed as follows:

$$\begin{cases} P_s = U_s I_{qs}, \\ Q_s = -U_s I_{ds}. \end{cases} \quad (7)$$

From the above formula, it can be seen that under the DQ coordinate, assuming that the direction of the stator flux linkage is consistent with that of the generator, the active power and reactive power of the stator of the doubly fed induction generator can be decoupled by adjusting the direct-axis component and quadrature-axis component, respectively.

When the stator flux-linkage orientation is used, the stator flux linkage coincides with the synchronous rotating coordinate system on the direct axis, and the quadrature-axis flux-linkage component is zero. Combining with the stator flux-linkage equation, the relationship between stator current and rotor current can be obtained, which can be expressed as follows:

$$\begin{cases} I_{ds} = \frac{U_s}{L_{ss}} - \frac{I_{dr} L_m}{L_{ss}}, \\ I_{qs} = -\frac{I_{qr} L_m}{L_{ss}}. \end{cases} \quad (8)$$

The relationship between the rotor voltage and the rotor current can be obtained from the aforementioned simultaneous flux-linkage equation and the stator-rotor voltage equation.

$$\begin{cases} U_{dr} = R_s I_{ds} - S \left( L_{rr} - \frac{(L_m)^2}{L_{ss}} \right) I_{qr}, \\ U_{qr} = R_s I_{qr} - S \left( L_{rr} - \frac{(L_m)^2}{L_{ss}} \right) I_{dr} + S \frac{U_s L_m}{L_{ss}}, \end{cases} \quad (9)$$

where  $R$  represents the resistance and  $S$  represents the slip ratio.

The relationship between the rotor voltage and the stator current is established based on formulas (8) and (9). The control of active power and reactive power on the stator side can be realized by controlling the rotor voltage by the control system.

**2.5. Transient Model.** When using the pulse width modulation (PWM) converter, the power factor of the PWM converter can be controlled to be close to 1.0. That is to say, the active power of the doubly fed wind turbine is approximately equal to the reactive power, which can be expressed as

$$\begin{cases} P_e = (1 - s)P_s, \\ Q_e = Q_s. \end{cases} \quad (10)$$

If the wind turbine works under the status of constant power factor, it can be expressed as

$$\frac{Q_s}{P_s} = C, \quad (C \text{ is a constant}). \quad (11)$$

The relationship between the power of the doubly fed wind turbine and the active power on the stator side can be obtained by equations (3)–(11), which can be expressed as

$$\begin{cases} P_e = \frac{R_r (L_{ss})^2}{(L_m)^2 (U_s)^2} (1 + C^2) (P_s)^2 + \left( 1 - S + \frac{2CL_{ss}R_r}{(L_m)^2} \right) P_s, \\ Q_e = CP_s. \end{cases} \quad (12)$$

If the flux linkage of the stator is neglected, the rotor transient model can be expressed as

$$\begin{cases} \frac{dE'_d}{dt} = -\frac{R_r}{L_{rr}} \left( E'_d - \omega_s I_{qs} \frac{(L_m)^2}{L_{ss}} \right) + S\omega_s E'_q - \omega_s U'_{qr}, \\ \frac{dE'_q}{dt} = -\frac{R_r}{L_{rr}} \left( E'_q + \omega_s I_{ds} \frac{(L_m)^2}{L_{ss}} \right) + S\omega_s E'_d + \omega_s U'_{dr}. \end{cases} \quad (13)$$

From the above detailed deduction and analysis, one can see that these 13 state variables (slip ratio  $S$  (no unit, just a ratio), pitch angle  $\beta$  (unit: 0), wind turbine torque  $T_m$  (unit: N·m), rotor mechanical torque  $T_t$  (unit: N·m), electromagnetic torque  $T_e$  (unit: N·m), stator direct-axis current  $I_{ds}$  (unit: A), stator quadrature-axis current  $I_{qs}$  (unit: A), rotor direct-axis current  $I_{dr}$  (unit: A), rotor quadrature-axis current  $I_{qr}$  (unit: A), rotor direct-axis voltage  $U_{dr}$  (unit: V), rotor quadrature-axis voltage  $U_{qr}$  (unit: V), direct-axis transient potential  $E_d$  (unit: V), and quadrature-axis transient potential  $E_q$  (unit: V)) can describe the actual operation process of DFIGs. When some special faults or wind speed disturbances occur, the control system of DFIGs can be controlled according to the initial values and the variations of these 13 state variables, whose initial values can be used to describe the initial operating state of DFIGs.

### 3. Dominant Variable Hierarchical Clustering Algorithm

Based on the theoretical analysis and formula deduction in Section 2, 13 state variables of wind turbines during operation can be calculated only under the consideration of few parameters, such as wind speed, power factors, nameplate parameters and power characteristic curves of the wind turbines, and so on.

In practical application, the analysis and calculation of 13 state variables will be difficult for huge computational complexity, especially for large-scale wind farm. Therefore, this paper extracts the dominant variables using principal component analysis and considers them as clustering indexes, which can effectively reduce the computational

complexity. This method can not only reduce the data dimension but also ensure the information completeness, so it has better performance than analyzing the original data of wind farm [34–38].

The detailed modeling process can be expressed as follows:

- (i) Step 1: according to the mathematical model of doubly fed wind turbines, 13 complete state variables representing the actual operation state of wind turbines are calculated.
- (ii) Step 2: based on the principal component analysis method, the dominant variables representing the main oscillation characteristics of the 13 complete state variables are extracted.
- (iii) Step 3: considering dominant variables as clustering index, the wind farm of doubly fed wind turbines are clustered by hierarchical clustering algorithm, where the hierarchical clustering algorithm includes the following steps:
  - Step 3.1: each sample point is defined as a class.
  - Step 3.2: calculate the distance between sample points within different classes using Euclidean distance formula.
  - Step 3.3: according to the above distance, the two classes with the shortest distance are merged into one class.
  - Step 3.4: the distance between the merged class and other classes is recalculated using Euclidean distance formula.
  - Step 3.5: repeat the step 3.3.
  - Step 3.6: the algorithm ends until all samples are merged into one class.
- (iv) Step 4: according to the clustering results of dominant variables, wind farm is divided into several groups.
- (v) Step 5: according to the partition results, the equivalent model of wind farm is constructed.

## 4. Simulations

**4.1. Simulation Tool.** PSCAD is an electromagnetic simulation software platform, which has been widely used in many research fields. It can use the time-domain analysis method to solve the complete power system, and the operation results are very accurate. It also has an abundant component library and complex control modules, allowing users to flexibly establish circuit models for simulation analysis, such as inverter, transformer, rectifier, wind source, wind governor, wind turbine component, synchronous generator component, speed control component, and so on. In addition, it also provides a large number of simulation cases for learning. Therefore, PSCAD software platform is especially suitable for model establishment and performance analysis of large-scale wind farm.

**4.2. Initial Parameters.** A wind farm model with 24 DFIGs is established using PSCAD software platform. The terminal

voltage of wind turbine is 690 V, which is boosted to 33 KV by one-machine one-variable connection. Every 4 transformers are connected by overhead lines to public connection points, and then one boost transformer is connected to the infinite bus system.

Actually, the initial wind speed data of 24 wind turbines have a great influence on the calculation cost of the algorithm. Literature [26] uses a K-means clustering algorithm to sort the wind farm, which is one of the state-of-the-art clustering methods. Meanwhile, the wake effect and topographic differences have also been considered in this literature, and the initial wind speed data of each wind turbine have been set as different values. To compare and quantitatively analyze the advantages and disadvantages of hierarchical clustering algorithm and K-means clustering algorithm, we use the same initial wind speed data as literature [26], as shown in Table 1, where all the data retain one significant digit after decimal point.

The 13 state variables of 24 DFIGs are calculated according to initial wind speed data in Table 1 and equations (1)–(13) described in Section 2, and the calculation results are shown in Table 2, where all the data retain two significant digits after decimal point.

Because the transient time constant of motor stator is very small,  $I_{ds}$  is ignored because its value is close to 0. Meanwhile, to obtain maximal value of wind energy utilization factor ( $C_p$ ) of wind turbines, the pitch angle ( $\beta$ ) should be equal to 0. Moreover, assuming that the wind speed is a fixed value, principal component analysis is carried out based on the above 13 state variables of 24 wind turbines. In general, according to the algorithm rules of principal component analysis, these variables can be considered as dominant variables as long as they can explain or express more than 85% variance. The analysis results are shown in Table 3.

As shown in Table 3, 94.331% variance can be expressed or explained by extracting only two factors, such as slip rate ( $S$ ) and wind turbine torque ( $T_m$ ). It demonstrates that these two variables can represent the actual operating state of wind turbines, so slip rate ( $S$ ) and wind turbine torque ( $T_m$ ) are regarded as dominant variables.

**4.3. Numerical Simulations.** To verify the effectiveness of dominant variables, three clustering trees are obtained using hierarchical clustering algorithm, where 13 state variables, dominant variables, and wind speed are regarded as clustering indexes, respectively. Figure 2 shows three clustering trees.

References [3, 26, 27] indicate that it is most reasonable to divide 24 wind turbines into four groups. According to Figure 2, the clustering results can be obtained, as shown in Table 4.

Figures 2(a), 2(b), and 2(c) correspond to (a), (b), and (c) in Table 4, respectively. From (a) and (b) in Table 4, one can see that clustering results are consistent when the 13 state variables and dominant variables are regarded as clustering indexes, respectively. From (a) and (c) in Table 4, one can see that the clustering results are quite different when the 13

TABLE 1: Initial wind speed data of 24 wind turbines.

No	1	2	3	4	5	6	7	8
Initial wind speed (m/s)	10.6	9.7	8.7	7.7	10.6	9.7	8.7	7.7
No	9	10	11	12	13	14	15	16
Initial wind speed (m/s)	10.3	9.3	8.2	6.9	10.3	9.3	8.2	6.9
No	17	18	19	20	21	22	23	24
Initial wind speed (m/s)	8.2	6.9	8.2	6.9	10.3	9.3	8.2	6.9

TABLE 2: Calculation results of 13 state variables of 24 wind turbines.

No	S	$\beta$	$T_m$	$T_t$	$T_e$	$I_{ds}$	$I_{qs}$	$I_{dr}$	$I_{qr}$	$U_{dr}$	$U_{qr}$	$E_d$	$E_q$
1	-0.20	0	0.72	0.72	0.72	0	0.67	0.59	-0.70	-0.02	-0.23	0.14	1.08
2	-0.20	0	0.58	0.58	0.58	0	0.54	0.59	-0.57	-0.02	-0.23	0.11	1.08
3	-0.20	0	0.44	0.44	0.44	0	0.41	0.59	-0.43	-0.01	-0.23	0.09	1.08
4	-0.11	0	0.33	0.33	0.33	0	0.31	0.59	-0.32	0.00	-0.13	0.06	1.08
5	-0.20	0	0.72	0.72	0.72	0	0.67	0.59	-0.70	-0.02	-0.23	0.14	1.08
6	-0.20	0	0.58	0.58	0.58	0	0.54	0.59	-0.57	-0.02	-0.23	0.11	1.08
7	-0.20	0	0.44	0.44	0.44	0	0.41	0.59	-0.43	-0.01	-0.23	0.09	1.08
8	-0.11	0	0.33	0.33	0.33	0	0.31	0.59	-0.32	0.00	-0.13	0.06	1.08
9	-0.20	0	0.68	0.68	0.68	0	0.63	0.59	-0.67	-0.02	-0.23	0.13	1.08
10	-0.20	0	0.53	0.53	0.53	0	0.49	0.59	-0.52	-0.02	-0.23	0.10	1.08
11	-1.97	0	0.38	0.38	0.38	0	0.35	0.59	-0.38	-0.01	-0.23	0.07	1.08
12	-0.01	0	0.27	0.27	0.27	0	0.25	0.59	-0.27	0.00	-0.01	0.05	1.08
13	-0.20	0	0.68	0.68	0.68	0	0.63	0.59	-0.67	-0.02	-0.23	0.13	1.08
14	-0.20	0	0.53	0.53	0.53	0	0.49	0.59	-0.52	-0.02	-0.23	0.10	1.08
15	-0.19	0	0.38	0.38	0.38	0	0.35	0.59	-0.38	-0.01	-0.23	0.07	1.08
16	-0.01	0	0.27	0.27	0.27	0	0.25	0.59	-0.27	0.00	-0.12	0.05	1.08
17	-0.20	0	0.68	0.68	0.68	0	0.63	0.59	-0.67	-0.02	-0.23	0.13	1.08
18	-0.20	0	0.53	0.53	0.53	0	0.49	0.59	-0.52	-0.02	-0.23	0.10	1.08
19	-1.96	0	0.38	0.38	0.38	0	0.35	0.59	-0.38	-0.01	-0.23	0.07	1.08
20	-0.01	0	0.27	0.27	0.27	0	0.25	0.59	-0.27	0.00	-0.12	0.05	1.08
21	-0.20	0	0.68	0.68	0.68	0	0.63	0.59	-0.67	-0.02	-0.23	0.13	1.08
22	-0.20	0	0.53	0.53	0.53	0	0.49	0.59	-0.52	-0.02	-0.23	0.10	1.08
23	-0.19	0	0.38	0.38	0.38	0	0.35	0.59	-0.38	-0.01	-0.23	0.07	1.08
24	-0.01	0	0.27	0.27	0.27	0	0.25	0.59	-0.27	0.00	-0.12	0.05	1.08

TABLE 3: Results of principal component analysis.

Components	Variations (%)	Total (%)
S	82.918	82.918
$T_m$	11.413	94.331
$T_t$	3.224	97.555
$T_e$	2.309	99.864
$I_{qs}$	0.130	99.994
$I_{dr}$	0.006	100
$I_{qr}$	$3.281 \times 10^{-6}$	100
$U_{dr}$	$3.332 \times 10^{-8}$	100
$U_{qr}$	$1.309 \times 10^{-24}$	100
$E_d$	$1.008 \times 10^{-25}$	100
$E_q$	$2.131 \times 10^{-26}$	100

state variables and wind speed are regarded as clustering indexes, respectively. Therefore, it is proved that it is effective and feasible to extract dominant variables as clustering indexes by principal component analysis.

Based on the above results, three models are established using PSCAD software platform, which use 13 state

variables, dominant variables, and wind speed as clustering indexes, respectively. In order to compare the advantages and disadvantages of the above three models, the active power and reactive power output curves of wind farm are simulated in the case of the three-phase short-circuit fault on the system side or wind speed fluctuation, respectively. Because the exact data of large-scale wind farm are not readily available in real time, we set the wind speed of each wind turbine as a non-zero constant value, and the values of each wind turbine are not exactly equal during simulation process, and the detailed value is shown in Table 1. Therefore, the active power and reactive power output curves of wind farm are continuous but relatively flat, where the X-axis is running time of wind turbine (unit: s). Meanwhile, the faults or wind speed fluctuations are superimposed when the wind turbine is running, not stationary. If this method is applied to real measurements, just only reset the initial wind speed data of each wind turbine and change the wind speed from a constant value to a variable value during the actual operating process. This method is also valid, but the computational cost will be largely increased. Moreover, the output curve is no longer so flat but fluctuates a lot.

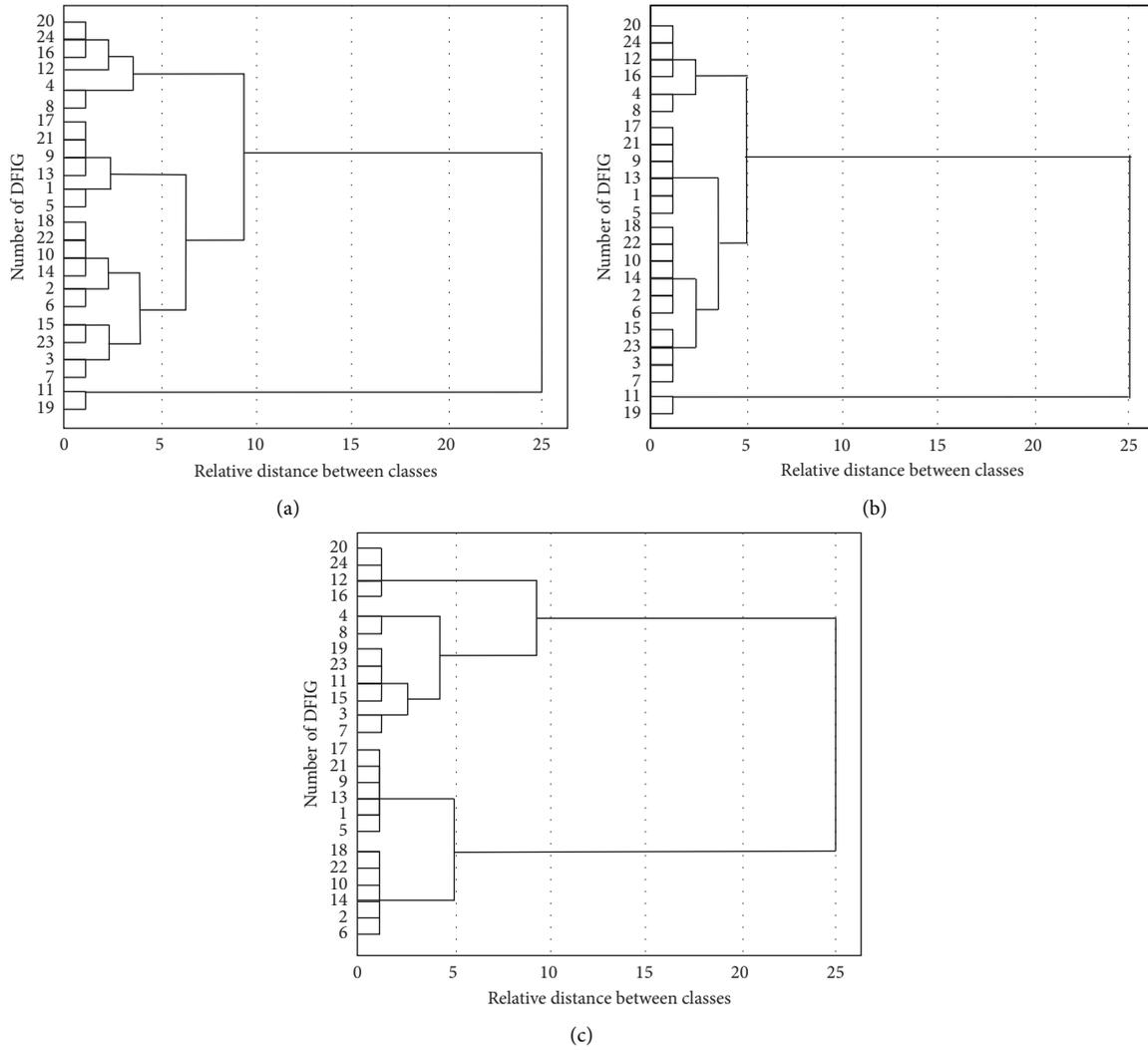


FIGURE 2: Clustering tree graphs under different indexes. (a) Clustering result using 13 state variables as clustering indexes. (b) Clustering result using dominant variables as clustering indexes. (c) Clustering result using wind speed as clustering index.

TABLE 4: Clustering results of 4-group wind turbines under different indexes.

<i>(a) 13 state variables</i>	
Equivalent unit number	Clustering results
1	1, 5, 9, 13, 17, 21
2	2, 3, 6, 7, 10, 14, 15, 18, 22, 23
3	4, 8, 12, 16, 20, 24
4	11, 19
<i>(b) Dominant variables</i>	
1	1, 5, 9, 13, 17, 21
2	2, 3, 6, 7, 10, 14, 15, 18, 22, 23
3	4, 8, 12, 16, 20, 24
4	11, 19
<i>(c) Wind speed</i>	
1	1, 5, 9, 13, 17, 21
2	2, 6, 10, 14, 18, 22
3	3, 4, 7, 8, 11, 15, 19, 23
4	12, 16, 20, 24

It is worth mentioning that only a part of the curves with faults or wind speed fluctuations is captured to show the differences among three models, while the curves in other periods are not captured.

4.4. *Three-Phase Short-Circuit Fault on System Side.* The simulation curve from the beginning stage to the stable stage is not captured. When the simulation curve reaches a stable state and lasts for 14 seconds, a man-made three-phase short-circuit fault is imposed to the system side, which lasts for 0.2 seconds. After 14.2 seconds, the output curve is relatively stable and there is no abnormal change. Hence, we only capture the curves for a period of time around 14 seconds. Within this period of time, the active power and reactive power output curves of wind farm are shown in Figure 3.

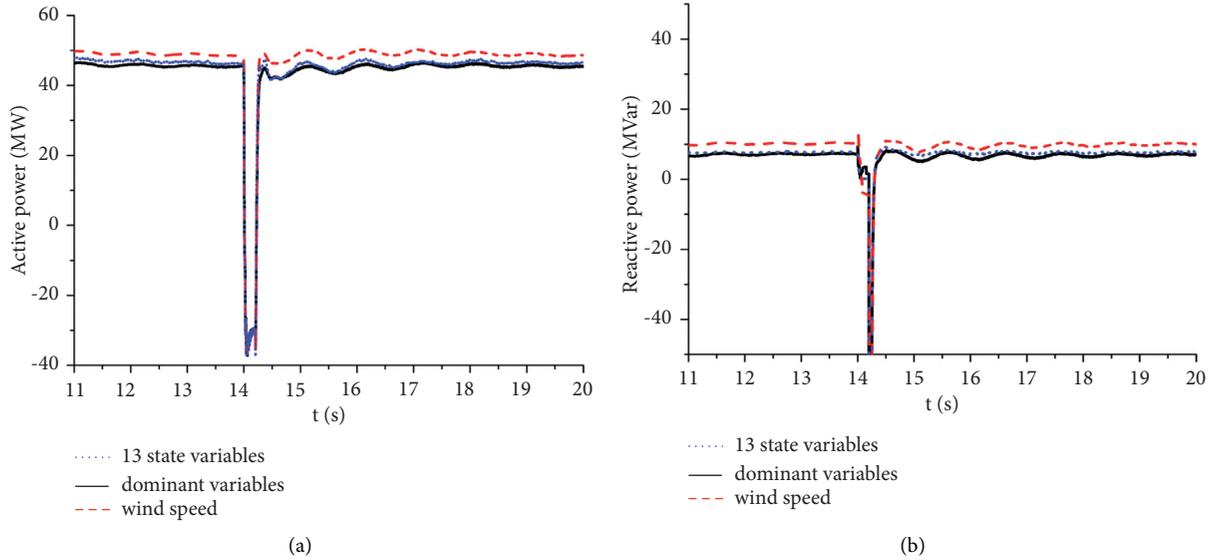


FIGURE 3: The output curve of wind farm in case of the three-phase short-circuit fault on system side. (a) Active power curve of three models. (b) Reactive power curve of three models. Blue line, black line, and red line represent 13 state variables, dominant variables, and wind speed, respectively.

From Figure 3, one can see that the active curve and reactive curve of dominant variables almost coincide with that of 13 state variables, which demonstrates that it is reasonable to extract the dominant variables using dominant variable hierarchical clustering algorithm proposed in this paper. The dominant variable can represent actual operating state of wind turbines in wind farm in the case of the three-phase short-circuit fault on system side. On the contrary, there is a large error between wind speed curve and 13-state-variable curve, which demonstrates that it is inappropriate to only take the wind speed as the clustering index.

**4.5. Wind Speed Fluctuation.** The simulation curve from the beginning stage to the stable stage is not captured. When the simulation curve reaches a stable state and lasts for 14 seconds, a man-made gust fluctuation is imposed, which lasts for 2 seconds and the maximum wind speed is 6 m/s. Because of the short lag and persistence effect of wind speed fluctuation, the curve is relatively stable after 20 seconds. Hence, we only capture the curves from 11 seconds to 20 seconds. Within this period of time, the active power and reactive power output curves of wind farm are shown in Figure 4.

From Figure 4, one can see that the active curve and reactive curve of dominant variables almost coincide with that of 13 state variables, which demonstrates that it is reasonable to extract the dominant variables using dominant variable hierarchical clustering algorithm proposed in this paper. The dominant variable can represent actual operating state of wind turbines in wind farm even in the case of wind speed fluctuation. On the contrary, there is a large error between wind speed curve and 13-state-variable curve, which demonstrates that it is inappropriate to only take the wind speed as the clustering index.

**4.6. Results and Discussion.** From Figures 3 and 4, the following conclusions can be drawn:

- (1) In the case of the three-phase short-circuit fault on system side or wind speed fluctuation, the maximal relative errors of output power curve of wind farm between the wind speed model and 13-state-variable model and between the dominant variable equivalent model and 13-state-variable model are 50.14% and 9.9%, respectively.
- (2) Reference [19] indicates that the maximal relative error of output curve of wind farm between the neural network model and 13-state-variable model is 12.1% and between the support vector machine model and 13-state-variable model is 11.4%, respectively. Reference [33] indicates that the maximal relative error of output power curve of wind farm between the K-means clustering model and 13-state-variable model is 18.06%.
- (3) Compared with the wind speed model, K-means clustering model, neural network model, and support vector machine model, the maximal relative error of the dominant variable equivalent model is the smallest, which demonstrates that this model is more accurate and much closer to the actual operating state of wind farm.
- (4) The dominant variable can also represent actual operating state of wind turbines in wind farm, even in the case of the three-phase short-circuit fault on system side or wind speed fluctuation. Therefore, it is reasonable to extract slip rate ( $S$ ) and wind turbine torque ( $T_m$ ) as dominant variables from 13 variables, and it is also reasonable to classify 24 wind turbines using hierarchical clustering algorithm.

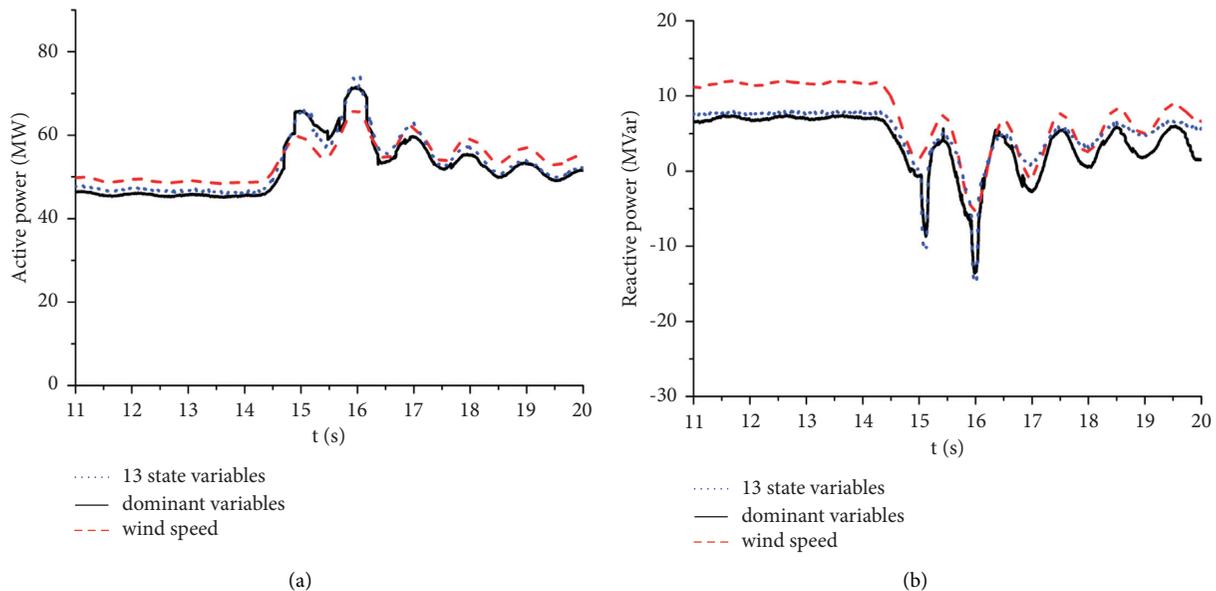


FIGURE 4: The output curve of wind farm in case of wind speed fluctuation. (a) Active power curve of three models. (b) Reactive power curve of three models. Blue line, black line, and red line represent 13 state variables, dominant variables, and wind speed, respectively.

## 5. Conclusions

In this paper, the mathematical model of the doubly fed induction generator is established under the consideration of the actual industrial applications of large-scale wind farm, and 13 state variables which can represent the dynamic operation process of wind turbines are calculated. The principal component analysis is proposed to extract dominant variables from 13 state variables, which can be used as clustering index. Then, three equivalent models are established, which use 13 state variables, dominant variables, and wind speed as clustering indexes, respectively. In order to compare the advantages and disadvantages of the above three models, the active power and reactive power output curves of wind farm are simulated in the case of the three-phase short-circuit fault on the system side or wind speed fluctuation, respectively. The simulation results demonstrate the following conclusions. (1) It is reasonable and effective to extract dominant variables by principal component analysis. (2) The dominant variable equivalent model has higher accuracy than the wind speed model, K-means clustering model, neural network model, and support vector machine model. (3) When the scale of wind farm increases, the advantage of this model will be more obvious. (4) The research result is especially suitable for characteristics analysis of large-scale wind farm connected to power grid. The question of how to conduct the systematic experimental research on large-scale wind farm, establish a better model to be as close to the actual operation state of large-scale wind farm as possible, and quantify and evaluate the errors between various equivalent models, complete models, and actual large-scale wind farm will be focus of future research work.

## Data Availability

The data supporting the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was supported by Sichuan Provincial Academic and Technical Leader Training Plan and Xihua Scholars Training Plan of Xihua University. Wenbo Jiang would like to acknowledge the Overseas Training Plan of Xihua University (09/2014–09/2015, University of Michigan, Ann Arbor, US).

## References

- [1] V. Nelson, *Introduction to Renewable Energy*, CRC Press, Boca Raton, FL, USA, 2010.
- [2] M. A. Chowdhury, W. X. Shen, N. Hosseinzadeh, and H. R. Pota, "A novel aggregated DFIG wind farm model using mechanical torque compensating factor," *Energy Conversion and Management*, vol. 67, no. 3, pp. 265–274, 2013.
- [3] J. Zou, C. Peng, Y. Yan, H. Zheng, and Y. Li, "A survey of dynamic equivalent modeling for wind farm," *Renewable and Sustainable Energy Reviews*, vol. 40, no. 12, pp. 956–963, 2014.
- [4] M. A. Chowdhury, W. Shen, N. Hosseinzadeh, and H. R. Pota, "Transient stability of power system integrated with doubly fed induction generator wind farms," *IET Renewable Power Generation*, vol. 9, no. 2, pp. 184–194, 2015.
- [5] W. Du, W. Dong, and H. Wang, "A method of reduced-order modal computation for planning grid connection of a large-scale wind farm," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1185–1198, 2019.
- [6] X. Yang, C. Milliren, M. Kistner et al., "High-fidelity simulations and field measurements for characterizing wind fields in a utility-scale wind farm," *Applied Energy*, vol. 281, no. 1–17, Article ID 116115, 2021.

- [7] M. Zhang and R. J. A. M. Stevens, "Characterizing the coherent structures within and above large wind farms," *Boundary-Layer Meteorology*, vol. 174, no. 1, pp. 61–80, 2020.
- [8] T.-D. Pham, M.-C. Dinh, H.-M. Kim, and T.-T. Nguyen, "Simplified floating wind turbine for real-time simulation of large-scale floating offshore wind farms," *Energies*, vol. 14, no. 15, pp. 1–18, Article ID 4571, 2021.
- [9] S. K. Salman and A. L. J. Teo, "Windmill modeling consideration and factors influencing the stability of a grid-connected wind power-based embedded generator," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 763–802, 2003.
- [10] Z. Chen, J. M. Guerrero, and F. Blaabjerg, "A review of the state of the art of power electronics for wind turbines," *IEEE Transactions on Power Electronics*, vol. 24, no. 8, pp. 1859–1875, 2009.
- [11] A. S. Dobakhshari and M. Fotuhi-Firuzabad, "A reliability model of large wind farms for power system Adequacy studies," *IEEE Transactions on Energy Conversion*, vol. 24, no. 3, pp. 792–801, 2009.
- [12] J. Brochu, C. Larose, and R. Gagnon, "Validation of single- and multiple-machine equivalents for modeling wind power plants," *IEEE Transactions on Energy Conversion*, vol. 26, no. 2, pp. 532–541, 2011.
- [13] M. J. Mercado-Vargas, D. Gómez-Lorente, O. Rabaza, and E. Alameda-Hernandez, "Aggregated models of permanent magnet synchronous generators wind farms," *Renewable Energy*, vol. 83, no. 11, pp. 1287–1298, 2015.
- [14] W. Li, P. Chao, X. Liang, D. Xu, and X. Jin, "An improved single-machine equivalent method of wind power plants by calibrating power recovery behaviors," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4371–4381, 2018.
- [15] N. Shabanikia, A. A. Nia, A. Tabesh, and S. A. Khajehoddin, "Weighted dynamic aggregation modeling of induction machine-based wind farms," *IEEE Transactions on Sustainable Energy*, vol. 12, no. 3, pp. 1604–1614, 2021.
- [16] V. Akhmatov and H. Knudsen, "An aggregate model of a grid-connected, large-scale, offshore wind farm for power stability investigations-importance of windmill mechanical system," *International Journal of Electrical Power & Energy Systems*, vol. 24, no. 9, pp. 709–717, 2002.
- [17] D. J. Trudnowski, A. Gentile, J. M. Khan, and E. M. Petritz, "Fixed-speed wind-generator and wind-park modeling for transient stability studies," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1911–1917, 2004.
- [18] L. M. Fernández, J. R. Saenz, and F. Jurado, "Dynamic models of wind farms with fixed speed wind turbines," *Renewable Energy*, vol. 31, no. 8, pp. 1203–1230, 2006.
- [19] J. Yan, Y. Liu, S. Han, and M. Qiu, "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 613–621, 2013.
- [20] J. Liu, M. Zhu, Z. Lin, and S. Chen, "A dynamic clustering model of wind farm based on the operation data," *Journal of Renewable and Sustainable Energy*, vol. 7, no. 4, Article ID 043111, 2015.
- [21] G. Xia, L. Zhou, J. M. Freedman, S. Baidya Roy, R. A. Harris, and M. Charles Cervarich, "A case study of effects of atmospheric boundary layer turbulence, wind speed, and stability on wind farm induced temperature changes using observations from a field campaign," *Climate Dynamics*, vol. 46, no. 7–8, pp. 2179–2196, 2016.
- [22] D. C. Fabrizio, V. Alfredo, and V. Domenico, "Adaptive wind generation modeling by fuzzy clustering of experimental data," *Electronics*, vol. 7, no. 4, pp. 1–16, 2018.
- [23] Z. H. Zhang, J. H. Li, and P. Zhuge, "Failure analysis of large-scale wind power structure under simulated typhoon," *Mathematical Problems in Engineering*, vol. 2014, no. 7, pp. 1–10, Article ID 486524, 2014.
- [24] X. Cheng, W. J. Lee, M. Sahni, Y. Cheng, and L. K. Lee, "Dynamic equivalent model development to improve the operation efficiency of wind farm," *IEEE Transactions on Industry Applications*, vol. 52, no. 4, pp. 1–9, 2016.
- [25] P. Wang, Z. Zhang, Q. Huang, N. Wang, X. Zhang, and W.-J. Lee, "Improved wind farm aggregated modeling method for large-scale power system stability studies," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6332–6342, 2018.
- [26] S. Y. Chen, C. Wang, and H. Shen, "Dynamic equivalence for wind farms based on clustering algorithm," *Proceedings of the Csee*, vol. 32, no. 4, pp. 11–19, 2012.
- [27] L. Lin and X. Pan, "Wind turbine grouping based on semi-supervised split-hierarchical spectral clustering algorithm for wind farm," *Electric Power Automation Equipment*, vol. 35, no. 2, pp. 8–14, 2015.
- [28] H. R. Ali, L. P. Kunjumammed, B. C. Pal, K. Vershinin, and A. G. Adamczyk, "A trajectory piecewise-linear approach to nonlinear model order reduction of wind farms," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 894–905, 2019.
- [29] S. Muller, M. Deicke, and R. W. De Doncker, "Doubly fed induction generator systems for wind turbines," *IEEE Industry Applications Magazine*, vol. 8, no. 3, pp. 26–33, 2002.
- [30] J. B. Ekanayake, L. Holdsworth, X. G. XueGuang Wu, and N. Jenkins, "Dynamic modeling of doubly fed induction generator wind turbines," *IEEE Transactions on Power Systems*, vol. 18, no. 2, pp. 803–809, 2003.
- [31] L. L. Fan and Z. X. Miao, *Modeling and Analysis of Doubly Fed Induction Generator Wind Energy Systems*, Academic Press, Cambridge, MA, USA, 2015.
- [32] M. Ahmad, *Operation and Control of Renewable Energy Systems*, Wiley Press, Hoboken, NJ, USA, 2017.
- [33] J. Zou, C. Peng, H. Xu, and Y. Yan, "A fuzzy clustering algorithm-based dynamic equivalent modeling method for wind farm with DFIG," *IEEE Transactions on Energy Conversion*, vol. 30, no. 4, pp. 1329–1337, 2015.
- [34] J. B. Daniel and J. O. Mark, "A study of power principal component analysis applied spatially distributed," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2084–2092, 2011.
- [35] L. Li, Z. Yan, D. H. Feng, and D. Han, "Discussion on intelligent technology evaluation of electrical power grid based on principal component analysis and production function," *Automation of Electric Power Systems*, vol. 38, no. 11, pp. 56–61 +73, 2014.
- [36] G. Fu, P. Hall, and N. J. Miles, "Performance evaluation for composites based on recycled polypropylene using principal component analysis and cluster analysis," *Journal of Cleaner Production*, vol. 115, no. 3, pp. 343–353, 2016.
- [37] Y. Wang, D. Xie, X. Wang, and Y. Zhang, "Prediction of wind turbine-grid interaction based on a principal component analysis-long short term memory model," *Energies*, vol. 11, no. 11, pp. 1–19, Article ID 3221, 2018.
- [38] G. Melendez-Melendez, D. Cruz-Paz, D. Cruz-Paz, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "An improved algorithm for partial clustering," *Expert Systems with Applications*, vol. 121, no. 5, pp. 282–291, 2019.