IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received September 2, 2020, accepted September 12, 2020, date of publication September 28, 2020,
date of current version October 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3027435

Risk-Oriented Renewable Energy Scenario Clustering for Power System Reliability Assessment and Tracing

WENHUA YANG[1], MAOSEN CAO[2], PENGJIANG GE[1], BO HU[2], (Member, IEEE), GAOQIANG QU[1], KAIGUI XIE[2], (Senior Member, IEEE), XIN CHENG[2], LVBIN PENG[2], JIAHAO YAN[2], (Student Member, IEEE), AND YUFEI LI[2]

[1]Economics and Technology Research Institute, State Grid Ningxia Electric Power Company, Yinchuan 750004, China
[2]School of Electrical Engineering, Chongqing University, Chongqing 400044, China

Corresponding authors: Maosen Cao (caomaosen@foxmail.com) and Bo Hu (hboy8361@163.com)

This work was supported by State Grid Ningxia Electric Power Company Ltd., through the Science and Technology Project, under Grant SGNXJY00GHJS1900041.

ABSTRACT The integration of large-scale renewable energy significantly increases the computation time of reliability assessment and tracing. To solve this problem, the traditional methods cluster the scenarios directly based on renewable energy data. However, it could lead to errors in reliability assessment due to scenarios with similar risks. In this paper, a multi-scenario risk-oriented clustering algorithm considering renewable energy is proposed. The enumeration method is used to calculate the risk for different scenarios. According to the risk of each scenario, the Fuzzy C-means clustering method is adopted to cluster the scenarios, which maximizes the similarity of scenarios in the same cluster. The high-risk scenarios that contribute more to the reliability index are retained. The system reliability assessment and tracing are conducted based on the clustered scenarios. Case studies on the IEEE RBTS-6 and RTS-79 systems verify the accuracy and efficiency of the proposed method.

INDEX TERMS Reliability assessment, Fuzzy C-means clustering, reliability tracing, risk-oriented.

NOMENCLATURE

c_d	load shedding price at load point d
PG_i	active power output of generator i
PD_d	load shedding at load bus d
G	number of generators
NC	total number of buses
NL	total number of lines
ND	the number of load buses
PL_i(s)	active power flow of line i in state s
H	line-bus incidence matrix
PG_i^{min}	lower limit of the generator output
PG_i^{max}	upper limit of the generator output
PL_i^{max}	active power flow capacity of line i

I. INTRODUCTION

Due to the global shortage of traditional fossil energy, renewable energy has been widely used in the world. However, the power output of renewable energy sources suffers from intermittency and fluctuation [1]. When connected to the power system, renewable energy sources will have an impact on the safe and stable operation of power system. There are two important means to ensure the reliable operation of the power system from the perspective of planning and design. The first is to accurately assess the reliability level of power systems. The second is to identify and eliminate the weak links after the integration of renewable energy sources.

Traditional reliability assessment methods are mainly divided into two categories: analytical method and simulation method [2]. When the system scale increases, the computation time of analytical method will increase exponentially. The Monte Carlo simulation method estimates the system index through random sampling of the system states and its computation time mainly depends on the reliability level of the system [3]. Existing studies have optimized and improved the analytical method and simulation method [4], [5].

Another important objective of power system reliability assessment is to identify the weak links of the system. Different components in the system have different influence on the system reliability index. It is necessary to quantify

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

the contribution of each individual equipment component to the overall reliability index. At present, this topic is mainly addressed by equivalent reliability sensitivity modeling [6], [7] and reliability tracing theory [8]–[10]. They can effectively identify the weak links of power system reliability and provide information for operators and planners in their decision-making. Reference [6] proposes a new sequential Monte Carlo simulation (sMCs) sensitivity model based on the derivative of state duration with respect to the reliability parameters. The model is able to solve the problem that the state probability cannot be explicitly expressed. Reference [7] applies reliability tracing theory in the reliability evaluation of High Voltage Direct Current systems(HVDC). Reference [8] establishes an unreliability tracing model based on the principle of minimum cut set and proportional contribution. Reference [9] proposes a transmission network blocking tracing technology. The proportional sharing principle (PSP) of failed components is adopted to identify the weak links of the power system. In [10], PSP is used in parallel with failure component sharing principle (FCSP) to allocate the unreliability index to each component of the bulk power system. Reliability tracing is used to identify the weak links.

With the increase in the penetration of renewable energy, the uncertainty of its output has seriously affected the reliability of the system. At the same time, ''curse of dimensionality'' could occur in the system states during the evaluation process. Some researches assess the reliability of power systems connected with renewable energy from the aspects of evaluation methods [11]. The probabilistic models of renewable energy output are proposed in [12], [13] and their characteristics are analyzed. In [11], the ARMA model is used to obtain the probability distribution of wind speed and a five-state wind energy conversion reliability model is established. References [12] and [13] build a multi-variate wind speed distribution with the copula function, which can represent the correlations of wind speeds. A reliability test system is used to analyze the correlations in various wind conditions. In [14], wake effect is explicitly incorporated with a wind farm in the reliability evaluation and its effectiveness is proved.

The computation burden in the reliability assessment caused by the large number of renewable energy scenarios can be alleviated by selecting representative scenarios. The selection of representative scenarios is usually based on historical experience. Nonetheless, this approach will decrease the accuracy of weak link identification and the value of its results. To reduce the error of risk index in reliability assessment, the representative scenarios can be obtained by directly clustering wind, solar and load data, considering the uncertainty and correlation of renewable energy. On the other hand, the result of weak link identification is obtained by allocating the load loss scenarios generated during the reliability assessment process. The original power generation data and net load data are clustered in [15]–[20] for feature extraction. To achieve a probabilistic assessment on the total transmission capacity of power systems, [15] first proposes using clustering to assess the reliability of power system with renewable energy integrated which can improve the efficiency of assessment. Reference [16] uses K-means to cluster the wind farm output and load demand. Reference [17] uses a hybrid clustering algorithm combined with hierarchical clustering algorithm and K-means algorithm to quantitatively analyze the impact of large-scale grid-connected photovoltaics on grid performance. The time series representing the fluctuation of photovoltaic output is retained in the historical data. The principle of maximum membership of Fuzzy C-means clustering [18] is used to maximize the similarity between the samples that are classified into the same cluster. Reference [20] proposes an improved Fuzzy C-means clustering algorithm. The measured load response is used as a feature vector to cluster power loads, so that the parameters of the composite load model can be obtained. Reference [21] uses K-means and Fuzzy C-means clustering techniques to evaluate the reliability for a large number of failure states. The energy capacity and installation location are optimized. Reference [18] also clusters the wind and load data to reduce data scale before reliability assessment and compares the effect of different clustering approaches on the results of reliability index. However, the above clustering methods only work on the input data and the risk of different scenarios are not considered. They could leave out high-risk scenarios in the reliability assessment. Since high-risk scenarios have higher contribution to the reliability index, ignoring them will lead to a larger error of reliability index. In addition, K-means is only sensitive to data with hyperspherical distribution [22]. Some adaptive clustering algorithms such as DBSCAN [23] cannot cluster data sets with large density differences very well. Consequently, it is difficult to obtain the desired number of clusters according to the needs of the operators.

In order to decrease the computation cost of reliability assessment and tracing caused by large-scale renewable energy integration, this paper proposes a multi-scenarios risk-oriented clustering method. The traditional scenario clustering methods directly use the original renewable energy data. It ignores the differences between contributions of different scenarios to the reliability index and may lead to errors in reliability assessment results. In the proposed method, the risk label for each net load scenario is calculated by the state enumeration (SE) method. Then, the net load scenarios are clustered by the Fuzzy C-means clustering according to the risk label of each scenario. In this way, the high-risk scenarios that contribute more to the reliability index can be retained. Using the clustered scenarios, the reliability evaluation and tracing are performed to obtain the reliability index and weak links. Compared with the traditional methods, the proposed method can avoid choosing scenarios with similar risks. Therefore, more representative clustering results can be achieved, which improves the accuracy of reliability assessment. The results also provide effective information for power system reinforcement.

The structure of the remainder of this paper is as follows. Section 2 introduces reliability assessment and tracing methods. Section 3 proposes a risk-oriented renewable energy scenario clustering method for power system reliability assessment and tracing. Section 4 provides case studies on the IEEE RBTS-6 and RTS-79 systems. Section 5 gives the conclusion.

II. RELIABILITY ASSESSMENT AND TRACING METHODS

A. RELIABILITY ASSESSMENT METHOD

The probability and consequence of system failure events are considered simultaneously in the reliability assessment of power systems. This paper uses Monte Carlo simulation method for reliability assessment. The procedures are as follows:

① State sampling is carried out for each component. s_i and Q_i represent the state and failure probability of the component i, respectively. $s_i = 0$ represents the normal state and $s_i = 1$ represents the failure state. R_i is a random number generated for the component i that is uniformly distributed in the interval [0,1].

$$s_i = \begin{cases} 0 & R_i > Q_i \\ 1 & 0 \le R_i \le Q_i \end{cases} \tag{1}$$

② The system state with N components is denoted by a vector $s = (s_1, \ldots s_i, \ldots s_N)$. When the number of samples is large enough, the occurrence frequency of the system state s can be used as an unbiased estimate for its occurrence probability.

$$P(s) = \frac{m(s)}{M} \tag{2}$$

where M is the number of samples. $m(s)$ is the number of occurrences of s in sampling.

③ The DC load shedding model is used to calculate the optimal load shedding P_d for the system states obtained in step ②.

The optimal DC load shedding model is:

$$Min \sum_{d \in ND} c_d P_d \tag{3}$$

$$\text{subject to} \sum_{i \in G} PG_i + \sum_{d \in ND} P_d = \sum_{d \in ND} PD_d \tag{4}$$

$$PL_i(s) = \sum_{j=1}^{NC} H_{ij}(s)(PG_j + P_j - PD_j), \quad i = 1, 2, \cdots, NL \tag{5}$$

$$PG_i^{\min} \le PG_i \le PG_i^{\max} \tag{6}$$

$$0 \le P_d \le PD_d \tag{7}$$

$$|PL_i(s)| \le PL_i^{\max} \tag{8}$$

Formula (3) represents the objective function including the penalty cost of load shedding. Formula (4) ensures the system power balance. (5) represents the nodal power balance. (6) limits the generator output. (7) constrains the bound of load shedding. (8) denotes the line power flow constraint.

④ The convergence coefficient of variance is calculated according to (9). If the accuracy requirements are satisfied, go to step ⑤. otherwise, go to step ①.

$$\sigma = \frac{\sqrt{V[P_d]/M}}{E(P_d)} \tag{9}$$

⑤ Accumulate the reliability index $EENS$(Expected Energy not Supplied):

$$EENS = \sum_{i \in S} C_i P_i T \tag{10}$$

where P_i denotes the probability of the state S_i. C_i is the load shedding of the state S_i. T represents the operating period.

B. RELIABILITY TRACING METHOD

The reliable operation of the power system depends on the normal operation of its components. When one or more components are out of service, the system load loss may occur. It is necessary to quickly identify the contribution of each individual component to the overall reliability index. This paper adopts the reliability tracing method in [24]. It identifies the weak links that lead to the load curtailment risk and provides information for the decision-making of operators and planners.

There are two criteria in the reliability tracing [10]:

1. When a system failure event occurs, only the components that participate in the failure share the unreliability of the system.

2. The system unreliability should be allocated among the failed components according to their contributions.

Taking the allocation of the system $EENS$ as an example, the probability of failure event k is:

$$P(k) = \prod_{i \in A} q_i \times \prod_{j \in B} (1 - q_j) \tag{11}$$

q_j and q_i are the failure probability of components j and i respectively. A and B represent the set of failed components and normal components respectively.

According to the reliability tracing principle, for failure event k, the allocation of failure probability to any component i in the system can be expressed as:

$$\begin{cases} P(k \to 1) = P(k) \times \dfrac{q_i}{\sum_{j \in A} q_j}, & i \in A \\ P(k \to i) = 0, & i \in B \end{cases} \tag{12}$$

The $EENS$ index allocated for any failure component i in the outage event k can be obtained by:

$$EENS(k \to i) = C_k P(k \to i) \times 8760 \tag{13}$$

where C_k is the power shedding in event k.

The total amount of $EENS$ allocated for component i, $EENS_i$ is:

$$EENS_i = \sum_{k \in \varphi} EENS(k \to i) \tag{14}$$

Then the contribution ratio of component i to system *EENS* is obtained by:

$$R_{EENS_i} = \frac{EENS_i}{\sum_{i=1}^{N} EENS_i} \tag{15}$$

III. RISK-ORIENTED RENEWABLE ENERGY SCENARIO CLUSTERING ALGORITHM

Due to the integration of renewable energy, the number of scenarios for power output and power load increases sharply. A large number of random events should be generated and analyzed to calculate the reliability index for each renewable energy scenario, thus adding the computation burden to the reliability assessment.

The traditional method of scenario clustering [25] is directly performed based on the historical data of net load. The net load is calculated by load minus wind and solar output.

The traditional method has the following disadvantages:

1. Highly complex and nonlinear relationships exist between the input data and the decision variables of model (3)-(8). If the input space directly serves as the clustering domain, the clustering results could fail to represent the optimal load shedding, which is directly used in the calculation of reliability index.

2. The effect of net load scenarios towards the system reliability is not considered because only the features of net load are extracted. In fact, clustered scenarios with significant difference in input domain may yield close reliability results. This is commonly seen in scenarios with low net load. Conversely, the scenarios with similar inputs may lead to disparate reliability results, which often occurs on high net load scenarios. It is necessary to cluster the scenarios that have the similar effects on the system reliability into the same cluster.

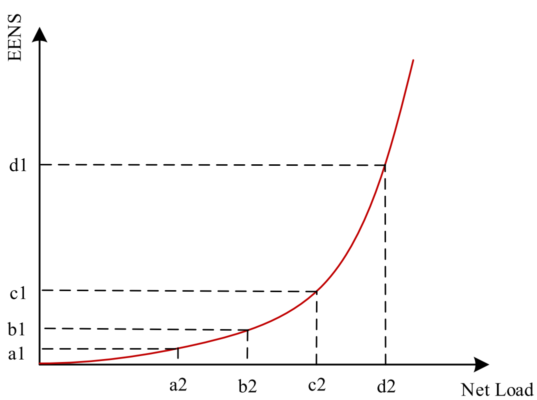

FIGURE 1. ***EENS* at different net load levels.**

3. High load scenarios contribute more to the risk. The traditional methods tend to ignore the high load scenarios. Therefore, the reliability results obtained by these methods could be inaccurate. As shown in Fig 1, the *EENS* increases nonlinearly as the net load increases. Note that *EENS* increases slowly in the low net load interval [a2, b2] but more rapidly in the high net load interval [c2, d2]. In the traditional methods, the number of clusters in the segment [a2, b2] may be similar with that in the [c2, d2]. However, the *EENS* of the high load level [c2, d2] varies in a wider range and there should more clusters in [c2, d2].

In this paper, a risk-oriented clustering method is proposed. The risk label for each net load scenario is calculated. Then, the net load scenarios are clustered by the Fuzzy C-means clustering. The clustered scenarios are applied to the reliability assessment and tracing.

A. AN INTRODUCTION AND APPLICATION OF FUZZY C-MEANS CLUSTERING ALGORITHM

The algorithm classifies the scenarios into different clusters according to their degree of similarity [26], [27]. Similar scenarios should be gathered into the same cluster but the scenarios with large differences should be categorized into different clusters. Fuzzy control is a classic method in the field of automatic control. The fuzzy set introduces the concept of membership degree. It does not follow the convention "either 0 or 1" in the classic mathematics and provides more flexible clustering results. The membership degree is represented by a weight assigned to each scenario and cluster.

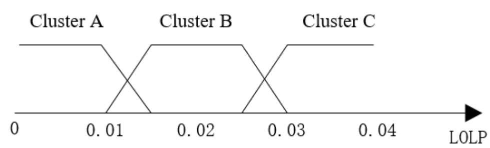

FIGURE 2. **Application of Fuzzy clustering in reliability assessment.**

Assume that we have obtained the reliability index for each scenario based on the procedures in Section II. If the failure probability of a certain scenario is 0.015, the scenario belongs to cluster A and cluster B at the same time as shown in Fig.2. Obviously, it could lead to errors if the scenario is classified forcibly into A or B using the traditional clustering method. The membership degree introduced by Fuzzy C-means clustering can obtain the weight of this scenario in cluster A and cluster B, respectively. The weights will determine to what degree the scenario belongs to these two clusters. Thus, the accuracy of clustering will be improved greatly.

The goal of the algorithm is to minimize the Euclidean distance between each candidate clustering scenario and its cluster center. The optimization objective function is:

$$J_{FCM}^{m}(U, A, X) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} \left\| x_j - a_i \right\| \tag{16}$$

where U is a $c \times n$ membership matrix and u_{ij} represents the membership of x_i to cluster j. c denotes the number of clusters. n denotes the number of candidate scenarios. A is a clustering center matrix and a_i represents the clustering center of each cluster. X is a one-dimensional matrix of candidate clustering scenarios, and x_j indicates the j_{th} scenario to be clustered.

d_{ij} is the Euclidean distance between x_j and a_i. m refers to the fuzzy coefficient. The value of m is generally determined by experience. If m is too small or large, the effect of the clustering algorithm will deteriorate [21].

$$a_i = \frac{\sum_{i=1}^{c} u_{ij}^m \cdot x_i}{\sum_{i=1}^{c} u_{ij}^m} \tag{17}$$

$$u_{ij} = 1/\sum_{k=1}^{c}\left(\frac{\|x_i - a_j\|}{\|x_i - a_k\|}\right)^{\frac{2}{m-1}} \tag{18}$$

$$\sum_{i=1}^{c} u_{ij} = 1 \quad (1 \le j \le n) \tag{19}$$

$$u_{ij} \ge 0 \quad (1 \le i \le c, 1 \le j \le n) \tag{20}$$

Equations (17)-(18) represent the iterations of membership degree and clustering center. Equations (19)-(20) indicate that the membership degree for each cluster is greater than 0. The sum of membership degree of each cluster is 1.

The flow chart of the algorithm is shown in Fig.3:

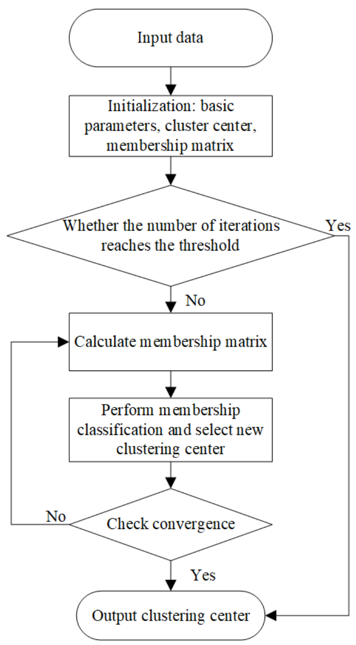

FIGURE 3. **Flow chart of Fuzzy clustering algorithm.**

B. RISK-ORIENTED RENEWABLE ENERGY SCENARIO CLUSTERING ALGORITHM

The situation that scenarios with similar risk are classified into different clusters can be avoided by using Fuzzy C-means clustering. High risks scenarios that contribute more to the reliability index can be preserved in reliability assessment and tracing. Overall, the proposed method ensures the validity and representativeness of clustered scenarios.

The procedures of the risk-oriented clustering algorithm for reliability assessment and tracing in this paper is as follows:

In the first stage, the wind and solar data (P_i^{wind}, P_i^{solar}, $i = 1, \ldots..M$), power load data (PD_i , $i = 1, \ldots..M$) and net load $P_i^{net}(i = 1, \ldots.M)$ with annual operating hours M are defined. The k-order events are enumerated by analytical method for $k = 1, \ldots.N$. The k-order system event denotes the event where k components are in the failure state simultaneously. N is the maximum order considered in the analytical method. The reliability index $EENS_i(i = 1, \ldots, M)$ under each net load scenario i is calculated as its risk label.

In the second stage, the risk labels of M scenarios are obtained. The Fuzzy C-means clustering is used to cluster the M scenarios based on the risk label. Fuzzy clustering is conducted according to formulas (17)-(20) to obtain L representative net load scenarios ($w_1, w_2, \ldots\ldots w_L$).

In the third stage, the reliability assessment method described previously is used to evaluate the reliability index given the L representative net load scenarios. The *EENS* of the system is obtained. According to the criteria of reliability tracing, the *EENS* allocated for each failed component is calculated so as to identify the weak links of the system.

The specific flow chart is shown in Fig.4.

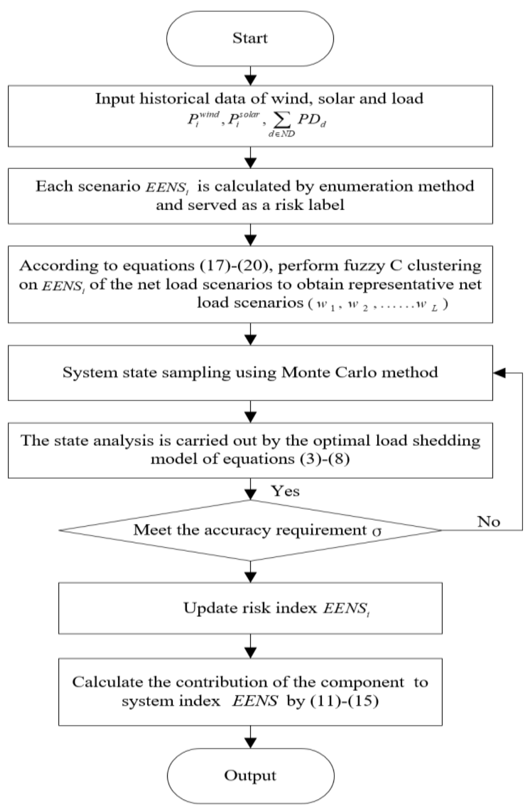

FIGURE 4. **Flow chart of risk-oriented renewable energy scenario clustering algorithm.**

IV. CASE STUDIES

The proposed method is tested on the IEEE RBTS-6 node system and the RTS-79 system. The data of these international standard systems can be obtained easily by [28], [29]. Wind turbines and photovoltaic units are connected to nodes 3 and 4 in RBTS and nodes 18 and 21 in RTS-79, respectively. Annual wind power and photovoltaic output data of a provincial power grid are used as the initial input data. The failures of the power lines and generators are considered this paper. The experiments are performed on a computer with 3.2 GHz Core-i5 6500 CPU and 8G RAM in MATLAB 2018b environment. The linear optimization for each iteration is solved by Gurobi solver. This paper uses Monte Carlo simulation method for reliability assessment. The variance convergence coefficient in the Monte Carlo method is 0.02. The DC optimal load shedding model is used to calculate the optimal load shedding.

In order to verify the effectiveness of the multi-scenarios risk-oriented clustering algorithm proposed in this paper. Three methods are selected for comparison.

(a) Risk-oriented Fuzzy Clustering (ROFC): The proposed risk-oriented clustering method using Fuzzy C-means algorithm.

(b) KM (K-means): K-means algorithm that directly cluster the original data.

(c) FCM (Fuzzy C-means): The Fuzzy C-means algorithm that directly cluster the original data.

The effectiveness of the proposed method is verified by comparing the results of reliability assessment and weak link identification based on the three methods.

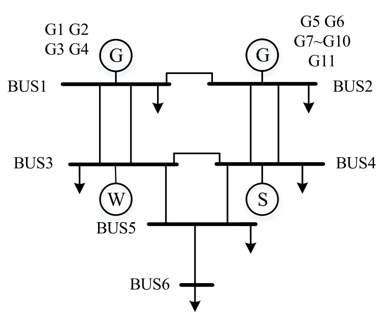

FIGURE 5. Diagram of RBTS-6 system.

A. A RBTS-6 SYSTEM

Fig.5 shows the diagram of RBTS-6 system where node 3 and node 4 are connected to wind turbines and photovoltaic units, respectively.

1) ANALYSIS OF RELIABILITY ASSESSMENT

Representative net load scenarios are obtained by the above three methods. The corresponding reliability index are calculated. The number of clusters varies from 4 to 120. The relative error is obtained by comparing with the standard results based on the annual hour load curve. The three methods are used to evaluate the reliability of the RBTS system. Under different number of clusters, the relative error calculated by each method is shown in Fig.6 and Fig.7. Overall, in Fig. 8, for ROFC, the median and the average value of the relative error

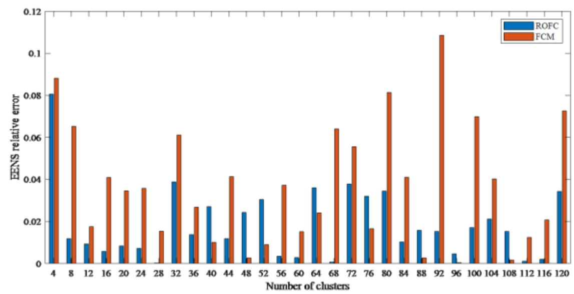

FIGURE 6. ***EENS*** **relative error of different number of clusters compared ROFC and FCM.**

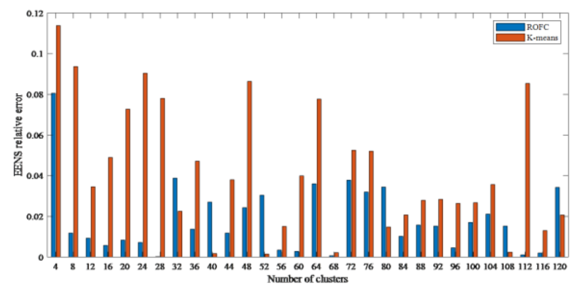

FIGURE 7. ***EENS*** **relative error of different number of clusters compared ROFC and K-means.**

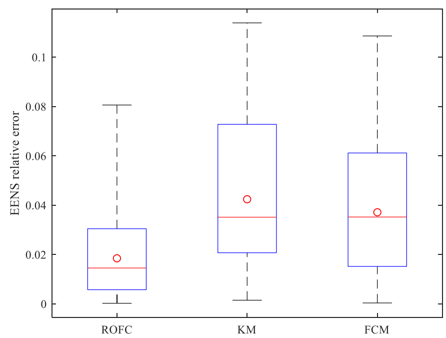

FIGURE 8. Comparison of relative errors of different clustering methods.

are 0.0145 and 0.0184 respectively. The median of relative error for KM is 0.0351 and the average value is 0.0424. The median of relative error for FCM is 0.0352 and the average value is 0.0371. The average relative error for ROFC is 56.6% and 50.4% lower than that of KM and FCM, respectively. It can be seen from Fig.6 to Fig.8 that the representative net load scenarios derived from ROFC greatly reduce the relative error of the reliability index. In addition, the variation range of relative error corresponding to the reliability index obtained by ROFC is the smallest among the three methods, indicating the stability of ROFC in RBTS system.

2) ANALYSIS OF RELIABILITY TRACING

In this section, the three methods are used to evaluate the reliability tracing of the RBTS-6 system. The contribution of each component to the *EENS* index is obtained based on the reliability tracing. Table 1 lists the relative error

TABLE 1. **Error percentage of *EENS* reliability index when the number of cluster is 8.**

	ROFC	KM	FCM
EENS(MWh)	2047.6	1667.8	1843.7
Error Percentage(%)	0.6	18.1	9.4

TABLE 2. **Reliability index allocation percentage of each component(CP) in RBTS-6 system.**

CP	ROFC	KM	FCM	CP	ROFC	KM	FCM
G1	12.6131	11.9538	9.9841	G11	6.7623	6.5376	6.1498
G2	8.6789	8.0890	9.0592	L1	0.0979	0.1675	0.0987
G3	13.9287	11.6000	13.3893	L2	0.2526	0.3017	0.1700
G4	11.7832	11.5064	10.3261	L3	0.1355	0.1350	0.1348
G5	0.6618	0.6284	0.4802	L4	0.0312	0.0796	0.1436
G6	0.5019	0.5661	0.5448	L5	0.0337	0.0530	0.0940
G7	8.5903	8.3369	9.3525	L6	0.0045	0.0028	0.0078
G8	8.8170	8.7532	8.8193	L7	0.0029	0.0000	0.0032
G9	8.2780	8.7852	9.2519	L8	0.0050	0.0079	0.0054
G10	10.3486	13.1617	13.2655	L9	8.4730	9.3342	8.7198

between the calculated result and the standard value. Table 2 lists the percentage of the contribution of each component to the *EENS*.

The efficiency of reliability tracing can be tested by the effect of component reinforcement. The failure probability of those components selected for reinforcement will decrease. The reinforcement resources are assumed to be limited and only three components can be selected. According to the results of reliability tracing, the three components with the largest contribution should be selected for reinforcement, which are marked by red in Table 2. It can be seen that the weak links identified by three methods are not exactly the same. The weak links identified by ROFC are the units G1, G3 and G4. The corresponding contribution to the system *EENS* are 12.6131%, 13.9287% and 11.7832%. The weak links identified by KM are the units G1, G3 and G10. The corresponding proportions are 11.9538%, 11.6000% and 13.1617%. The weak links identified by the FCM are the units G3, G4 and G10. The corresponding proportions are 13.3893%, 10.3261% and 13.2655%.

TABLE 3. ***EENS* reliability index after reinforcement of weak link in RBTS-6 system.**

EENS(MWh)	ROFC	KM	FCM
After reinforcement	1547.4	1744.6	1750.6
Percentage(%)	24.0	14.3	14.0

Based on the annual hourly load curve, the standard *EENS* value of the RBTS system before enforcement is 2035.8MWh. Table 3 shows the system *EENS* index after reinforcing the components selected by the three methods. According to Table 3, the system reliability levels corresponding to the three methods are increased by 24.0%, 14.3%, and 14.0%, respectively. This shows that the system reliability level has the largest improvement when the weak links obtained by ROFC are reinforced. Therefore, ROFC has the best accuracy in weak link identification.

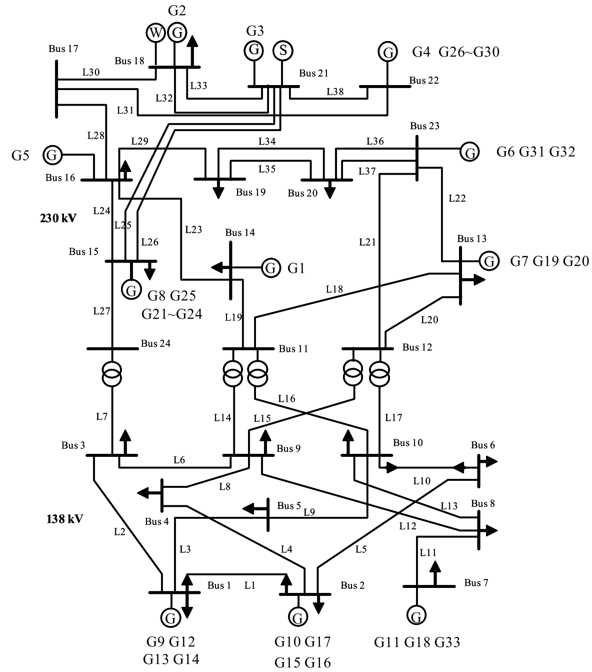

FIGURE 9. **Diagram of RTS-79 system.**

B. RTS-79 SYSTEM

1) ANALYSIS OF RELIABILITY ASSESSMENT

The three methods are used to evaluate the reliability of the RTS-79 system. Under different cluster numbers, the relative error calculated by each method is shown in the Fig. 10 and Fig.11. The median and average value of the relative error between the evaluation result and the standard value are shown in Table 4:

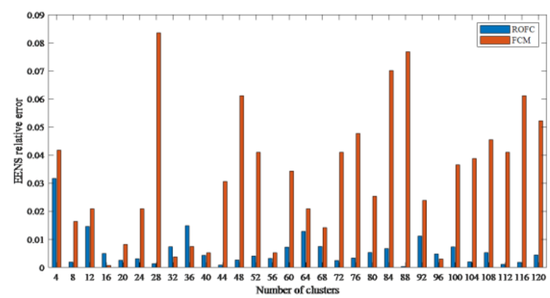

FIGURE 10. ***EENS* relative error of different number of clusters compared ROFC and FCM.**

TABLE 4. **Comparison of accuracy of different clustering methods in RTS-79 system.**

	ROFC	KM	FCM
Mean	0.0060	0.0299	0.0327
median	0.0044	0.0232	0.0323

Fig.10 to Fig.12 and Table 4 suggest that ROFC greatly reduces the relative error of the reliability index. Compared with KM and FCM, the evaluation accuracy of ROFC is

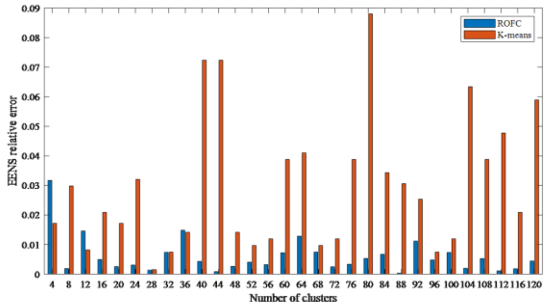

FIGURE 11. ***EENS*** **relative error of different number of clusters compared ROFC and K-means.**

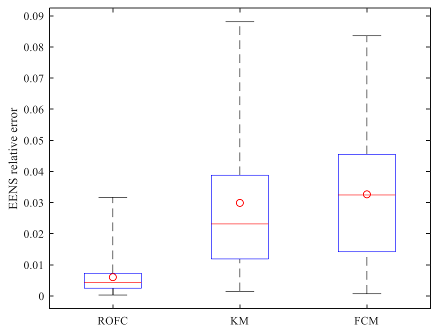

FIGURE 12. **Comparison of accuracy of different clustering methods in RTS-79 system.**

increased by 79.9% and 81.6%, respectively. It proves the accuracy of ROFC in RTS-79 system.

2) ANALYSIS OF RELIABILITY TRACING

The three methods are used to conduct the reliability tracing. The number of clustered representative net load scenarios is set to 44. Since the *EENS* value allocated for the transmission lines in RTS-79 system is relatively small, Table 5 only lists the contribution of each generating unit.

TABLE 5. **Reliability index allocation percentage of each component(CP) in RTS-79 system.**

CP	ROFC	KM	FCM	CP	ROFC	KM	FCM
G1	0	0	0	G18	2.3846	3.4027	2.6583
G2	**5.8740**	**5.8563**	**5.9522**	G19	3.6533	3.5789	3.9100
G3	**5.8570**	4.9071	**5.8905**	G20	3.8400	4.5708	3.8235
G4	0.1327	0.1105	0.1509	G21	0.3215	0.2869	0.3429
G5	5.3892	**5.9253**	**5.7949**	G22	0.3137	0.3158	0.3070
G6	**5.6579**	**6.0738**	5.5258	G23	0.3690	0.6066	0.4340
G7	**6.0758**	5.3283	5.6003	G24	0.3430	0.2917	0.3322
G8	0.3300	0.5145	0.3141	G25	3.7328	4.1184	3.3835
G9	5.1951	3.9911	4.5493	G26	0.1101	0.1564	0.1031
G10	0.6235	0.5624	0.6018	G27	0.1577	0.1571	0.1740
G11	2.3706	2.1826	2.3578	G28	0.1376	0.1560	0.1519
G12	5.5408	**5.6965**	**5.8182**	G29	0.1559	0.1461	0.1415
G13	0.6681	0.8587	0.7116	G30	0.1558	0.1894	0.1306
G14	0.6559	0.4383	0.5930	G31	3.6337	3.7942	3.6977
G15	5.3339	4.5166	5.2119	G32	**23.0369**	**23.5397**	**23.0679**
G16	4.7557	4.9356	5.0489	G33	2.4934	2.1372	2.5230
G17	0.6744	0.6281	0.6736	-	-	-	-

Assume that only 5 components can be selected for reinforcement. According to the results of reliability tracing, the five components with the largest contribution are selected for reinforcement. As can be seen from Table 5, the weak links determined by the three methods are different. The weak links identified by ROFC are unit G2, G3, G6, G7 and G32. The weak links identified by KM are unit G2, G5, G6, G12 and G32. The weak link identified by FCM are unit G2, G3, G5, G12 and G32.

Based on the original annual hourly load curve, the standard *EENS* of RTS-79 before reinforcement is 370318.4MW. Table 6 shows the system *EENS* is reduced by 32.2%, 30.4% and 28.5% respectively after reinforcing the components selected by the three methods. The system reliability level sees the largest improvement after reinforcing the weak links obtained by ROFC, which agrees with the findings in Section IV-A.

TABLE 6. ***EENS*** **reliability index after reinforcement of weak link in RTS-79 system.**

EENS(MWh)	ROFC	KM	FCM
After reinforcement	251025.1	257905.7	264789.8
Percentage(%)	32.2	30.4	28.5

V. CONCLUSION

This paper proposes risk-oriented renewable energy scenario clustering for power system reliability assessment and tracing. The clustered scenarios obtained by Fuzzy C-means algorithm are used to conduct reliability assessment and tracing.

Case studies show that the ROFC method effectively improves the representativeness of the clustered scenarios. It increases the accuracy of reliability assessment by up to 81.6% compared with KM and FCM. At the same time, in the system reinforcement, the greatest increase of reliability levels (32.2%) is achieved when using ROFC method for weak link identification. Therefore, the method can provide a reference for the reliability assessment and weak link identification of power systems with large-scale renewable energy.

In scenario clustering, the number of clustering categories, clustering methods, component reliability parameters, and system scale all have influence on the clustering result. We will consider more factors that may influence actual output of renewable units, such as wake effect described in [13], [18]. The wake effect can cause power loss which have an impact on reliability assessment indices. A comprehensive model and analysis is required to investigate these factors in future studies.

REFERENCES

[1] C. Shao, Y. Ding, J. Wang, and Y. Song, "Modeling and integration of flexible demand in heat and electricity integrated energy system," *IEEE Trans. Sustain. Energy*, vol. 9, no. 1, pp. 361–370, Jan. 2018.

[2] W. Li, J. Zhou, K. Xie, and X. Xiong, "Power system risk assessment using a hybrid method of fuzzy set and Monte Carlo simulation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 336–343, May 2008.

[3] A. M. L. D. Silva, R. A. G. Fernandez, and C. Singh, "Generating capacity reliability evaluation based on Monte Carlo simulation and cross-entropy methods," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 129–137, Feb. 2010.

[4] R. Sun, C. Singh, L. Cheng, and Y. Sun, "Short-term reliability evaluation using control variable based dagger sampling method," *Electr. Power Syst. Res.*, vol. 80, no. 6, pp. 682–689, Jun. 2010.

[5] Y. Wang, C. Guo, and Q. H. Wu, "A cross-entropy-based three-stage sequential importance sampling for composite power system short-term reliability evaluation," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4254–4263, Nov. 2013.

[6] Z. Wu and S. Li, "Reliability evaluation and sensitivity analysis to AC/UHVDC systems based on sequential Monte Carlo simulation," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3156–3167, Jul. 2019.

[7] B. Hu, K. Xie, and H.-M. Tai, "Reliability evaluation and weak component identification of ±500-kV HVDC transmission systems with double-circuit lines on the same tower," *IEEE Trans. Power Del.*, vol. 33, no. 4, pp. 1716–1726, Aug. 2018.

[8] K. Xie and R. Billinton, "Tracing the unreliability and recognizing the major unreliability contribution of network components," *Rel. Eng. Syst. Saf.*, vol. 94, no. 5, pp. 927–931, May 2009.

[9] M. Gan, K. Xie, and C. Li, "Transmission congestion tracing technique and its application to recognize weak parts of bulk power systems," *J. Mod. Power Syst. Clean Energy*, vol. 5, no. 5, pp. 1–10, 2016.

[10] K. Xie, B. Hu, and R. Karki, "Tracing the component unreliability contributions and recognizing the weak parts of a bulk power system," *Eur. Trans. Electr. Power*, vol. 21, no. 1, pp. 254–262, Jan. 2011.

[11] R. Billinton and Y. Gao, "Multistate wind energy conversion system models for adequacy assessment of generating systems incorporating wind energy," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 163–170, Mar. 2008.

[12] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, "ARIMA-based time series model of stochastic wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–676, May 2010.

[13] K. Xie, Y. Li, and W. Li, "Modelling wind speed dependence in system reliability assessment using copulas," *IET Renew. Power Gener.*, vol. 6, no. 6, pp. 392–399, Nov. 2012.

[14] H. Kim, C. Singh, and A. Sprintson, "Simulation and estimation of reliability in a wind farm considering the wake effect," *IEEE Trans. Sustain. Energy*, vol. 3, no. 2, pp. 274–282, Apr. 2012.

[15] C. Singh and Y. Kim, "An efficient technique for reliability analysis of power systems including time dependent sources," *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 1090–1096, Aug. 1988.

[16] M. Ramezani, C. Singh, and M.-R. Haghifam, "Role of clustering in the probabilistic evaluation of TTC in power systems including wind power generation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 849–858, May 2009.

[17] W. A. Omran, M. Kazerani, and M. M. A. Salama, "A clustering-based method for quantifying the effects of large on-grid PV systems," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2617–2625, Oct. 2010.

[18] Z. Wang, "Comparison of four kinds of fuzzy C-means clustering methods," in *Proc. 3rd Int. Symp. Inf. Process.* Washington, DC, USA: IEEE Computer Society, Oct. 2010.

[19] H. Kim and C. Singh, "Comparison of clustering approaches for reliability simulation of a wind farm," in *Proc. IEEE Int. Conf. Power Syst. Technol. (POWERCON)*, Oct. 2012, pp. 1–6.

[20] G. Shi, J. Liang, and X. Liu, "Load clustering and synthetic modeling based on an improved fuzzy C means clustering algorithm," in *Proc. 4th Int. Conf. Electr. Utility Deregulation Restructuring Power Technol. (DRPT)*, Jul. 2011, pp. 859–865.

[21] M. M. Mahmoodi, E. Afjei, F. Nadimi, and A. Siadatan, "A new fuzzy-Monte Carlo simulation approach and its application in reliability evaluation of power systems protection," in *Proc. 22nd Iranian Conf. Electr. Eng. (ICEE)*, May 2014, pp. 831–835.

[22] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3029–3039, May 2018.

[23] S.-S. Li, "An improved DBSCAN algorithm based on the neighbor similarity and fast nearest neighbor query," *IEEE Access*, vol. 8, pp. 47468–47476, 2020.

[24] G. Li, G. Huang, Z. Bie, Y. Lin, and Y. Huang, "Component importance assessment of power systems for improving resilience under wind storms," *J. Mod. Power Syst. Clean Energy*, vol. 7, no. 4, pp. 676–687, Jul. 2019.

[25] Y. Lai, G. Li, Y. Wang, H. Li, and Z. Wang, "Reliability assessment for power systems with large-scale renewable energy sources," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Dec. 2013, pp. 1–5.

[26] B. Lami and K. Bhattacharya, "Clustering technique applied to nodal reliability indices for optimal planning of energy resources," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4679–4690, Nov. 2016.

[27] J. Dong and M. Qi, "K-means optimization algorithm for solving problem," *Knowl. Discovery Data Mining*, pp. 52–55, 2009.

[28] P. Subcommittee, "IEEE reliability test system," *IEEE Trans. Power App. Syst.*, vol. PAS-98, no. 6, pp. 2047–2054, Nov. 1979.

[29] R. Billinton, S. Kumar, N. Chowdhury, K. Chu, K. Debnath, L. Goel, E. Khan, P. Kos, G. Nourbakhsh, and J. Oteng-Adjei, "A reliability test system for educational purposes-basic data," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 1238–1244, Aug. 1989.

WENHUA YANG received the B.S. degree in electrical engineering from Xi'an Jiaotong University in 2005. He is currently a Senior Engineer with the Economics and Technology Research Institute, State Grid Ningxia Electric Power Company. His research interests include power grid and energy planning research.

MAOSEN CAO received the B.S. degree in electrical engineering from Southwest Jiaotong University, China, in 2018. He is currently pursuing the Ph.D. degree with Chongqing University, China. His research interests include reliability assessment and the optimization of the integrated energy systems.

PENGJIANG GE received the B.S. degree from the Lanzhou University of Technology, China, in 2012. He is currently a Senior Engineer with the Economics and Technology Research Institute, State Grid Ningxia Electric Power Company. His research interests include grid and energy planning research.

BO HU (Member, IEEE) received the Ph.D. degree in electrical engineering from Chongqing University, China, in 2010. He is currently working as a Full Professor with the School of Electrical Engineering, Chongqing University. His research interests include power system reliability and parallel computing techniques in power systems.

GAOQIANG QU received the B.S. degree from Ningxia University, China, in 2010. He is currently a Senior Engineer with the Economics and Technology Research Institute, State Grid Ningxia Electric Power Company. His research interests include grid and energy planning research.

KAIGUI XIE (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Chongqing University, China, in 2001. He is currently working as a Full Professor with the School of Electrical Engineering, Chongqing University. His main research interests include power system reliability, planning, and analysis.

XIN CHENG received the B.S. degree from the North University of China in 2019. She is currently pursuing the M.S. degree with Chongqing University China. Her research interest includes the optimization of integrated energy systems.

LVBIN PENG received the B.S. degree from the Hefei University of Technology, China, in 2015. He is currently pursuing the Ph.D. degree with Chongqing University, China. His research interests include power system reliability and planning.

JIAHAO YAN (Student Member, IEEE) received the B.S. degree from the Hefei University of Technology, China, in 2016. He is currently pursuing the Ph.D. degree with Chongqing University, China. His research interests include power system risk assessment and optimization.

YUFEI LI received the B.S. degree from Hehai University, China, in 2019. She is currently pursuing the M.S. degree with Chongqing University, China. Her research interest includes wind power forecasting.