

# Power Quality Disturbance Detection Based on Improved Robust Random Cut Forest

Ge Zhang<sup>1</sup>, Feifei Bai<sup>1,2</sup>, Yi Cui<sup>1</sup>

<sup>1</sup>School of Information Technology and Electrical Engineering, University of Queensland

<sup>2</sup>School of Engineering and Built Environment, Griffith University  
Brisbane, Australia

[ge.zhang@uq.edu.au](mailto:ge.zhang@uq.edu.au), [f.bai@uq.edu.au](mailto:f.bai@uq.edu.au), [y.cui3@uq.edu.au](mailto:y.cui3@uq.edu.au)

David Dart<sup>3</sup>, Jalil Yaghoobi<sup>3</sup>

<sup>3</sup>Department of R&D, NOJA Power

Brisbane, Australia

[davidd@nojapower.com.au](mailto:davidd@nojapower.com.au)

[JalilY@nojapower.com.au](mailto:JalilY@nojapower.com.au)

Matthew Zillmann<sup>4</sup>

<sup>4</sup>Department of Renewables & Distributed Energy, Energy Queensland

Brisbane, Australia

[matthew.zillmann@energyq.com.au](mailto:matthew.zillmann@energyq.com.au)

**Abstract**— Many distribution network disturbances exhibit unique electrical signatures which can be observed from voltage and current waveforms. With the continuous enhancement of Power Quality Data (PQD) acquisition capabilities, it is feasible to continuously monitor the operating status of power nodes from a more detailed perspective. In the past decade, an increasing amount of high-quality PQD has been collected accumulating a massive and unique high-resolution power grid data asset. However, how to obtain valuable information, such as Power Quality Disturbance Events (PQDEs) from massive PQD remains challenging in the research community. In this paper, a reliable PQDEs detection method is proposed based on the improved Robust Random Cut Forest (RRCF). This method achieves accurate detection of PQDEs through the adaptive improvement of pre-filtering based on Ensemble Empirical Mode Decomposition (EEMD) and redundant interpolation. Numerical test results on the synthetic PQD and the realistic pollution experiment of a silicone rubber insulator in a salt fog chamber demonstrate the reliability, efficiency, and scalability of the proposed approach in practical online detection.

**Keywords**—Power quality, Robust Random Cut Forest (RRCF), abnormality detection, big data, arcing

## I. INTRODUCTION

The current development of the power grid has entered a new era of digitization and intelligence. As more and more power electronic-based devices are integrated into the power grids, there is a sharply increased risk of PQDEs occurring. With the continuous improvement of condition monitoring methods and data acquisition infrastructures, researchers can interpret and analyze the operating status of the power grid from a more microscopic view.

High-precision power operation information can be collected through power quality meters installed at the key power nodes. This information usually includes phase voltage, phase current, power angle, etc. In terms of the nature of the problem, the occurrence of PQDEs can cause voltage and current to fluctuate to varying degrees. Thus, by analyzing the collected voltage and current data, the abstract power quality problem can be refined into an anomaly detection problem for time-series data.

Traditional time-series abnormality detection methods mostly rely on manual feature extraction. Most of them focus on data analytics of PQD in the time domain, frequency domain or both. For example, the Fast Fourier Transform (FFT) built on the time-series waveform can reveal the distribution in the frequency spectrum of the time series data, and these distributions can be used as the signatures for detecting PQDEs [1]-[2]. However, the detection methods based on FFT mainly focus on the frequency characteristics of

the time series, and they may lose valuable disturbance characteristics in the time domain. The Short-Time Fourier Transform (STFT) has been introduced to overcome the shortcoming of the FFT-based methods and it demonstrates to have a better detection accuracy [3]-[4]. STFT samples the time series by introducing a dynamic moving window with an adjustable scale so that it can obtain the time-frequency domain characteristics simultaneously. However, the selection of the window length will affect the performance of the STFT-based method [5]-[6]. To extract the signatures in the time-frequency domain more efficiently and avoid the limitations of STFT and FFT algorithms, the Wavelet Transform (WT) was introduced [7]-[8]. Although these traditional methods have been widely used in various fields, and some of them achieved relatively high detection accuracy, the truth is that they heavily rely on experts' knowledge for feature extraction and feature selection, and they mostly focus on events associated with specific datasets. The effort to find a generic detection method that can be applied to any kind of situation is very limited.

To solve the shortcomings of artificial decision-making in feature engineering, abnormality detection methods based on Deep Learning (DL) and Neural networks (NN) have begun to emerge. The methods based on DL and NN mainly solve detection problems from the perspective of classification and clustering. These methods first automatically learn the relationship between the features and different classes of PQDEs from historical datasets, and then make classification of the PQDE using the power quality data of interest. Among these methods, artificial neural networks and support vector machine [9, 10] are the two most used methods in abnormality detection. Probabilistic neural network based on Bayes minimum risk criteria can give the possibility of sample classification into different categories [11]-[12], which can also be used in the PQDEs detection. Moreover, multi-layer perceptron neural network [13] with more hidden layers is regarded as an improved algorithm of function neural network. The capabilities of the model have been improved to a certain extent. But at the same time, due to its complex structure, the substantial increase in the number of parameters also has a non-negligible impact on the model training and classification speed. Compared with the above methods, the convolutional layer in Convolutional Neural Network (CNN) [14]-[15] has stronger automatic feature learning capabilities. By setting the appropriate number of filters and the size of the convolution kernel, the CNN model can be well qualified for multi-classification based on one-dimensional time series PQD. However, it is worth noting that due to the polymorphism of PQDEs and the shortage of real-life datasets with confirmed events, the model training process is still a

This work was supported by the Queensland State Government, Australia under the Advance Queensland Industry Research Fellowship AQIRF0022018 and the Amplify Women's Academic Research Equity project (UQAWARE 617678).

very challenging task. The current mainstream solution is to use standard-recommended simulation models to establish synthetic datasets. Some researchers also use their own measured data combined with synthetic data for model training and verification. Differences in datasets and operating equipment will also affect the final performance of the NN model.

There are many practical problems in the power grid, such as complex and diverse operating states, relatively high noise levels and integration of PQDEs. Restricted by the limited authenticity and accuracy of PQDE datasets, most PQDEs detection algorithms based on traditional neural networks can only achieve better results on their specially designed datasets. When in the face of these polymorphism problems, the models trained based on the specific datasets are often unable to make correct judgments on the real situation due to the limited information learned. This problem can greatly affect the generalization capability of these models. Unsupervised learning (UL) methods do not require a specific training process and can effectively overcome the problem of low model availability due to limited dataset quality.

This paper proposed a UL-based method for PQDEs detection to fill this gap. Accurate detection of PQDEs in the real-world grid is achieved without requiring training with the specific dataset and manual feature selection as support. The rest of this paper is organized as follows: Section II elaborates on the proposed PQDE detection method which combines EEMD with improved RRCF. Then the performance of the proposed method is verified in Section III using both synthetic PQDEs data and the real-world data. The synthetic data is generated based on the IEEE 1159-2009 standard and the real-world data is collected from extensive insulator pollution experiments in a salt fog chamber. Finally, the conclusions are provided in Section IV.

## II. PROPOSED POWER QUALITY DISTURBANCE EVENT DETECTION METHOD

The workflow of the proposed method is shown in Fig 1. Firstly, through the EEMD based filter, the environmental noise in the original signal can be accurately removed, thereby it can improve the signal-to-noise ratio (SNR) and can facilitate in realising unique PQDEs features. Afterwards, the RRCF-based module is designed to determine the abnormality score of the specific signal point and achieve accurate detection of complex PQDEs.

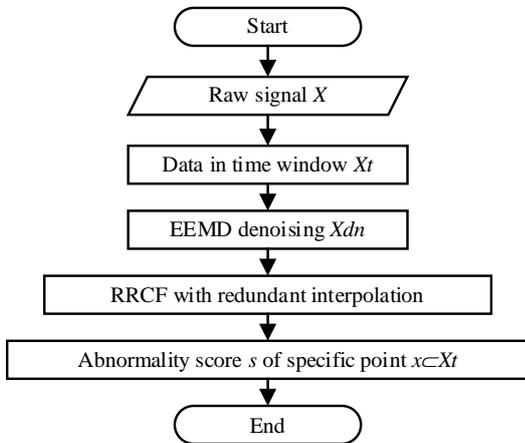


Fig 1. Flowchart of the proposed method

### A. Original Robust Random Cut Forest (RRCF)

A typical RRCF can be understood as consisting of several Robust Random Cut Trees (RRCTs) with adjustable sizes. As the basic parts of RRCF, RRCTs on point set  $S$  can be generated by choosing random dimensions and numerical operations based on normalized values [16]-[17]. To form the structure of an RRCT, the dynamic maintenance should be supported by the random sampling processes with point inserting and deleting operations, among which the recency-based weighted random sampling [18] and the reservoir sampling [19] are used.

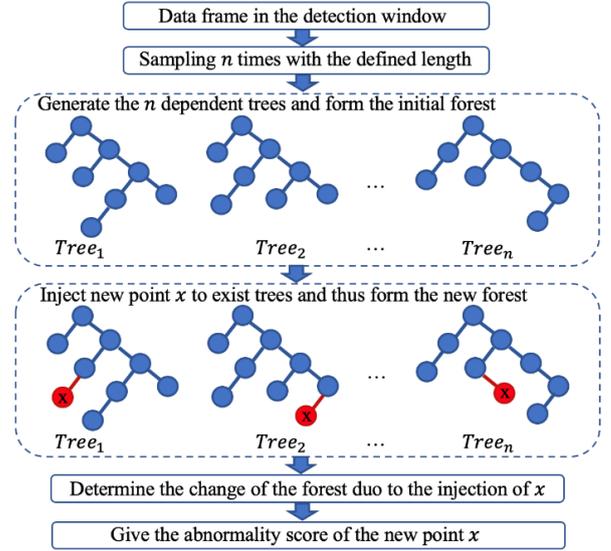


Fig 2. Flowchart of original RRCF

Considering the given sample  $S$ , the corresponding tree  $T(S)$ , and the point  $p$ , a random tree can be easily calculated as  $T(S \cup p)$  and the depth of  $p$  in tree  $T(S \cup p)$  is denoted as  $d(p, S \cup p, T)$ [16]. The anomaly will be identified when the difference between the joint distribution including the point  $p$  and the differential distribution becomes too significant. In other words, the anomaly can be determined by the impact of sampling point  $p$  on the original distribution.

To quantify the influence of the presence or absence of the new point  $p$  on the original distribution, literature [16] proposed the concept of displacement  $DISP(p, S)$  and the collusive displacement  $CODISP(p, S, |S|)$  to determine the degree of the abnormality of a single point. Assuming  $C$  is a point set that includes point  $p$ ,  $y$  is a point in  $S - p$  and  $y \in C \in S$ , random tree  $T' = T(S - x)$  and  $T'' = T(S - C)$ . The difference is that the  $CODISP(p, S, |S|)$  remove  $C$ , a point set that includes point  $p$ , while  $DISP(p, S, |S|)$  removes a single point  $p$  in each trail. Additionally, the  $DISP(p, S, |S|)$  mainly focuses on the difference between  $d(y, S, T)$  and  $d(y, S - p, T')$  while the  $CODISP(p, S, |S|)$  mainly takes  $d(y, S, T) - d(y, S - C, T'')$  into consideration. If the abnormality score of all data points in the signal segment is always lower than the predetermined threshold, the segment will be considered normal. If the abnormality score exceeds a certain threshold, the abnormality is detected. The process and parameter design of the original RRCF is detailed in [18]. This part summarizes the key steps and concepts of the original RRCF, the rest of the details will not be further repeated since they are not the key point of this paper.

### B. Redundant Interpolation

RRCF is a powerful tool to detect abnormality in the periodical data. However, the initial section failure at the beginning of the moving window greatly affects the accuracy of the original RRCF. In addition, the PQD collected from the power grid often contains noise which will also have a significant impact on the detection results. The pre-filter based on EEMD and the redundant interpolation are designed to solve this problem.

Considering the original sampling window as  $W(t, t+\Delta t)$  where  $t$  is the initial moment of sampling,  $\Delta t$  is the length of the moving window which is usually set as an integer multiple of the wave cycles. Redundant interpolation is achieved by connecting a short normal signal without event before the target waveform in the moving window. Let  $\Delta t'$  be the length of the interpolation segment. Then, the interpolation segment can be denoted as  $W'(t-\Delta t', t)$  where  $\Delta t'$  is adjustable and usually can be set as half period. Then, the new segment  $W'$  will replace the original  $W$  as the input of the RRCF. Such redundant interpolation can help to overcome the initial segment failure of the original RRCF and greatly improve the accuracy of the proposed method for PQDEs detection.

### C. EEMD-based Filter

EEMD refers to the principle of noise-assisted signal processing [20], which can effectively fix the modal aliasing problem existing in traditional EMD-based methods. As a significant improvement compared with the traditional EMD-based method [21]-[22], EEMD uses the Gaussian white noise whose zero-mean property can help to characterise the original signal. Specifically, EEMD can decompose a signal into several component waves in the time-frequency domain and these component waves are called Intrinsic Mode Functions (IMFs). The main steps of the EEMD method can be summarised as follows: In the first step, the white noise signal  $N_n(t)$  is added to the original signal  $X(t)$  in each trial to form a new composite signal  $Y_n(t)$ .

$$Y_n(t) = N_n(t) + X(t) \quad (1)$$

Then,  $Y_n(t)$  is decomposed using the traditional EMD algorithm to obtain a series of IMFs and a residual signal.

$$Y_n(t) = \sum_{m=1}^{M-1} IMF_m^n(t) + r_M^n(t) \quad (2)$$

where  $IMF_m^n(t)$  is the  $m$ -th IMF of  $Y_n(t)$  in the  $n$ -th trial and  $M-1$  is the total number of IMFs based on the EMD decomposing. And  $r_M^n(t)$  is the residual signal in the  $n$ -th trail. Then, the previous two steps will be repeated  $N$  times since the ensemble number was set to  $N$ . At the final step, the results output from the EEMD ( $IMF_M^{avg}$ ) are the average values of the corresponding order IMFs obtained from  $N$  trials:

$$IMF_M^{avg}(t) = \frac{1}{N} \sum_{n=1}^N IMF_m^n(t) \quad (3)$$

Based on the above discussion, it is easy to find that the number of trials  $N$  and the amplitude  $A$  of the white noise added to each trial will have an impact on the decomposition results. The final standard deviation  $\varepsilon$  was defined in [20] to evaluate the performance of the EEMD decomposition:

$$\varepsilon = \frac{A}{\sqrt{N}} \quad (4)$$

where  $N$  is the total number of trials and  $A$  is the amplitude of the white noise signal  $N_n(t)$  added in each trial.

The noise signal and the pure signal can be accurately synthesized by the IMFs decomposed by the EEMD. In order to preserve the information of the original signal as much as possible on the premise of accurately extracting the noise signal, the correlation between the different order IMFs and the original signal needs to be clarified. Specifically, those IMF components that are closely related to noise have strong random normal distribution characteristics, so they will have a low correlation with the original signal. For those components that are closely related to the original signal, they will have a similar distribution to the original signal. Thus, these IMFs will have higher correlated coefficients to the original signal. In this paper, Pearson correlation is used to evaluate the correlation between different order IMFs and the original signal, so that the IMFs representing the noise and the pure signal can be determined.

The correlation characterisation  $r$  and  $p$  of each IMF can be obtained from the Pearson analysis. The parameter  $r$  is the Pearson correlation coefficient, which represents the strength of the correlation.  $R$  can represent the degree of covariance of the two time-series, and the range of  $r$  is  $(-1, 1)$ .  $P$  is a parameter that characterizes the significance of the IMF. When  $p < 0.05$ , the correlation is significant, which means the correlation between the two time-series can be observed under the current sample. The larger  $r$  and smaller  $p$  indicate the higher the correlation between the two waveforms.

This section mainly presents the adaptive improvements to the original RRCF, including redundant imputation and the EEMD-based filter. The principle and processes of the proposed method have been detailedly discussed. Furthermore, the effectiveness of the proposed method will be verified in the following section.

## III. EXPERIMENT AND VALIDATION

### A. Validation on Synthetic PQDE Data

In this section, seven basic PQDEs, namely swell, sag, oscillation, interruption, notch, flicker and arcing are used to construct the simulation dataset (as shown in Fig. 3). The PQDEs are generated by IEEE 1159-2009 standard [23]. In this paper, the synthetic dataset contains 200 samples for each type of PQDE.

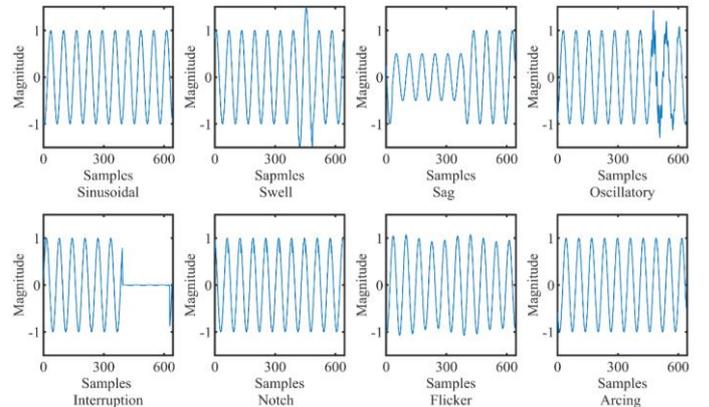


Fig 3. Examples of PQDEs including swell, sag, oscillatory, interruption, notch, flicker and arcing

Based on this synthetic dataset, the comparison between the proposed method and the original RRCF and the widely adopted Kullback-Leibler divergence (KLD) method [24] is performed to verify the performance of these methods. As shown in Table 1, the proposed method exhibits comprehensive advantages in the detection of each type of PQDEs. Overall, the proposed method achieves a comprehensive detection accuracy of 98.71% on the synthetic dataset of multiple types PQDEs, which is 12.57% higher than the original RRCF method and 26.43% higher than the traditional KLD-based method. Among seven types of PQDEs, the detection accuracy of arcing has been significantly improved by 24%. This is because the energy fluctuations caused by arcing are relatively small and are easily mixed with ambient noise making it difficult to separate them. The pre-filter design for eliminating environmental noise in the proposed method can better suppress the environmental noise and make the arcing feature more obvious. The above simulation results also demonstrate the advantages of the pre-filter design.

TABLE I. COMPARISON OF DETECTION ACCURACY AMONG DIFFERENT METHODS BASED ON SYNTHETIC PQDES DATASET

PQDEs	Original RRCF	KLD	Proposed method	Improvement (RRCF/KLD)
Arcing	76%	50%	100%	+24%/+50%
Voltage Swell	89%	74%	100%	+11%/+26%
Voltage Sag	91%	74%	100%	+9%/+26%
Flicker	88%	76%	98%	+10%/+22%
Oscillation	86%	68%	95%	+9%/+27%
Interruption	90%	92%	100%	+10%/+8%
Notch	83%	72%	98%	+15%/26%
Overall	86.14%	72.28%	<b>98.71%</b>	<b>+12.57%</b> <b>+26.43%</b>

### B. Validation on Real-Life Arcing Data

In this section, the effectiveness of the proposed method is verified by using real-life arcing data collected from the insulator pollution experiment in a salt fog chamber. Fig 4 shows the schematic diagram of the insulator pollution experiment. An electrical switchgear is installed in a salt fog chamber. The fusible salt can significantly reduce the electrical breakdown strength of the silicone rubber [25] so that the discharge phenomenon will appear more frequently on the surface of the insulator string. When the discharge occurs, there will be a short circuit at S. A high-performance power quality meter is used to record the voltage and current on the low voltage side. Since the duration of arcing is relatively short (mainly within 100ms to 800ms). Thus, the sampling rate is set to  $1024 \times 50\text{Hz}$  to obtain a more detailed waveform of the discharge.

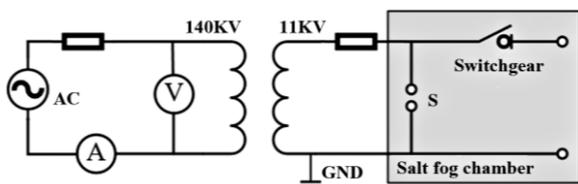


Fig 4. Schematic diagram of insulator pollution experiment

A moving window with 10 cycles window length is used to scan the original waveform for event detection. To determine the noise and the true signal, EEMD decomposition

and Pearson analysis are conducted, and the results are shown in Fig 5 and Fig 6.

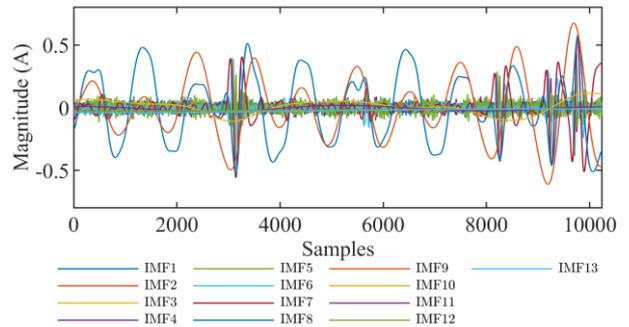


Fig 5. IMFs of the waveform decomposition based on the EEMD method

As shown in Fig 6,  $p$  decreases sharply from the IMF\_2 and reaches a value close to 0 after the IMF\_3, which means that from the perspective of the  $p$  value, IMFs are more related to the original signal from the third IMF. The values of  $r$  for IMF\_6 to IMF\_10 increase significantly meanwhile the peak value is close to 1, which means that these five IMFs have stronger covariance with the original waveform.

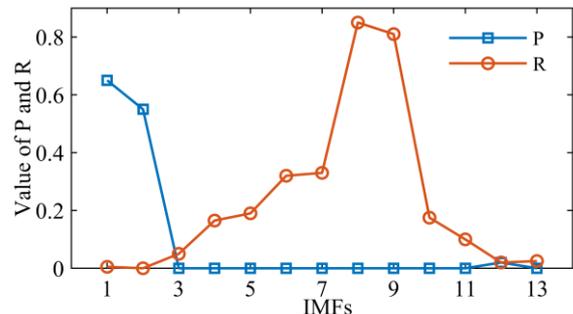


Fig 6. Results of the Pearson analysis for each IMF after EEMD-based filter

In most cases, the correlation between the 6<sup>th</sup> to 10<sup>th</sup> IMF and the original signal is more prominent than the other IMFs, in other words, the composite wave of the 6th to 10th IMF should be used as the main component of the waveform after noise reduction. In addition, from the results of the waveform decomposition in Fig 5, the IMFs after the 10<sup>th</sup> are all low-frequency and low-amplitude signals. These signals may contain long-term trend characteristics or base value fluctuations caused by equipment ageing or environmental factors. Consequently, they should also be taken into consideration when constructing the final denoised signal. According to Fig 6 and the previous discussion, the first five IMFs are used to construct the environmental noise.

As shown in Fig 7, the noise level in the noise-reduced waveform is well suppressed, while the amplitude and shape characteristics of the main event are also well preserved. This result is also consistent with the Pearson analysis results discussed before. Fig 8 shows the distribution of the reconstructed noise, which shows that the noise signal follows the Gaussian distribution and can be treated as white noise. The proposed noise filter based on EEMD performs well on the real-world arcing data.

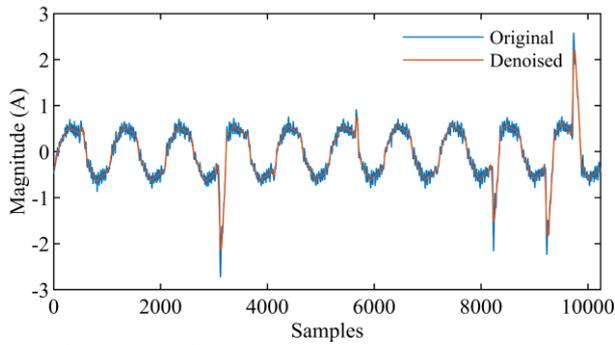


Fig 7. Comparison of the original signal and the denoised signal based on the EEMD denoising filter

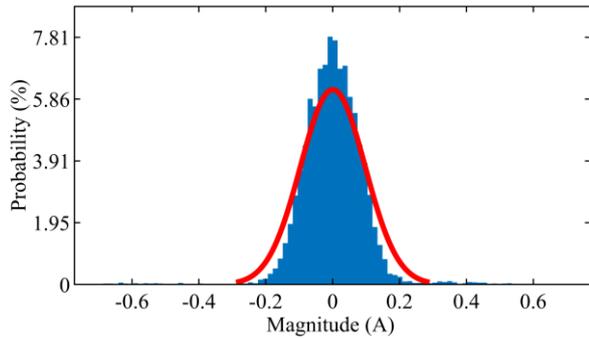


Fig 8. The distribution of the noise signal

Fig 9 shows the abnormality scores determined by the proposed method where an obvious turning point is observed. The existence of the turning point means that the abnormality scores for PQDEs in the detection result are much greater than the scores for normal conditions. Additionally, based on this, the threshold can be selected as 54.31 as shown in Fig 9. Through the previous discussion, the effectiveness of the proposed method is well validated in both synthetic and real datasets. It has been proved that the RRCF improved by pre-denoising and redundant interpolation can be well qualified for the adaptive PQDEs detection.

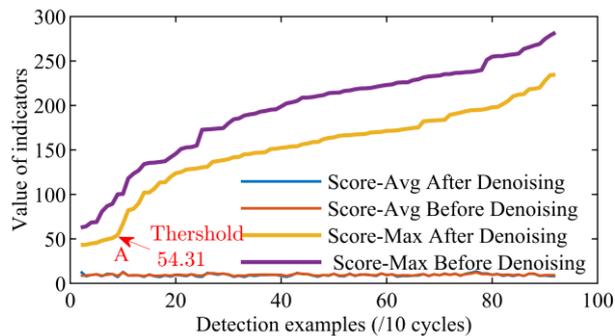


Fig 9. Comparison of RRCF detection results before and after improvement

#### IV. CONCLUSION

This paper proposed an unsupervised PQDEs detection method based on the improved RRCF and EEMD. Due to the improvements in the redundant interpolation, the problem of initial section failure of the original RRCF is solved and the PQDEs at the beginning of the moving window can also be detected. At the same time, the noise filter based on the EEMD segregates the noise from the desired true signal which further improves the difference between the abnormality score obtained by the normal conditions and PQDEs. The proposed

method is validated on both synthetic and real-world PQDEs datasets. It is expected that the proposed method would lay a foundation for fully exploring the value of PQD and realizing the application of real-time monitoring and automatic classification of complicated PQDEs.

#### REFERENCES

- [1] J. Johnson *et al.*, "Differentiating series and parallel photovoltaic arc-faults," *38th IEEE Photovoltaic Specialists Conference*, 2012: IEEE, pp. 000720-000726.
- [2] J.-C. Gu, D.-S. Lai, J.-M. Wang, J.-J. Huang, and M.-T. Yang, "Design of a DC series arc fault detector for photovoltaic system protection," *IEEE Transactions on Industry Applications*, vol. 55, no. 3, pp. 2464-2471, 2019.
- [3] H. Cheng, X. Chen, F. Liu, and C. Wang, "Series arc fault detection and implementation based on the short-time fourier transform," *Asia-Pacific Power and Energy Engineering Conference*, 2010: IEEE, pp. 1-4.
- [4] S. Chen, X. Li, and J. Xiong, "Series arc fault identification for photovoltaic system based on time-domain and time-frequency-domain analysis," *IEEE Journal of Photovoltaics*, vol. 7, no. 4, pp. 1105-1114, 2017.
- [5] R. Balamurugan, F. Al-Janahi, O. Bouhali, S. Shukri, K. Abdulmawjood, and R. S. Balog, "Fourier Transform and Short-Time Fourier Transform Decomposition for Photovoltaic Arc Fault Detection," *47th IEEE Photovoltaic Specialists Conference (PVSC)*, 2020: IEEE, pp. 2737-2742.
- [6] S. Lu, B. Phung, and D. Zhang, "A comprehensive review on DC arc faults and their diagnosis methods in photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 89, pp. 88-98, 2018.
- [7] C.-H. Kim, H. Kim, Y.-H. Ko, S.-H. Byun, R. K. Aggarwal, and A. T. Johns, "A novel fault-detection technique of high-impedance arcing faults in transmission lines using the wavelet transform," *IEEE Transactions on Power Delivery*, vol. 17, no. 4, pp. 921-929, 2002.
- [8] S. Jamali and N. Ghaffarzadeh, "A new method for arcing fault location using discrete wavelet transform and wavelet networks," *European Transactions on Electrical Power*, vol. 22, no. 5, pp. 601-615, 2012.
- [9] W. Miao, Q. Xu, K. Lam, P. W. Pong, and H. V. Poor, "DC arc-fault detection based on empirical mode decomposition of arc signatures and support vector machine," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 7024-7033, 2020.
- [10] J. Morales, E. Muñoz, E. Orduña, and G. Idarraga-Ospina, "A novel approach to arcing faults characterization using multivariable analysis and support vector machine," *Energies*, vol. 12, no. 11, p. 2126, 2019.
- [11] Z. Moravej, S. H. Mortazavi, and S. M. Shahrtash, "DT - CWT based event feature extraction for high impedance faults detection in distribution system," *International Transactions on Electrical Energy Systems*, vol. 25, no. 12, pp. 3288-3303, 2015.
- [12] Z. Moravej, M. Pazoki, M. Niasati, and A. A. Abdoos, "A hybrid intelligence approach for power quality disturbances detection and classification," *International Transactions on Electrical Energy Systems*, vol. 23, no. 7, pp. 914-929, 2013.
- [13] M. Sarlak and S. Shahrtash, "High impedance fault detection using combination of multi-layer perceptron neural networks based on multi-resolution morphological gradient features of current waveform," *IET Generation, Transmission & Distribution*, vol. 5, no. 5, pp. 588-595, 2011.
- [14] K. Cai, W. Cao, L. Aarniovuori, H. Pang, Y. Lin, and G. Li, "Classification of power quality disturbances using wigner-ville distribution and deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 119099-119109, 2019.
- [15] K. Cai, T. Hu, W. Cao, and G. Li, "Classifying power quality disturbances based on phase space reconstruction and a convolutional neural network," *Applied Sciences*, vol. 9, no. 18, p. 3681, 2019.
- [16] S. Guha, N. Mishra, G. Roy, and O. Schrijvers, "Robust random cut forest based anomaly detection on streams," *International Conference on Machine Learning*, 2016: PMLR, pp. 2712-2721.
- [17] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1-39, 2012.
- [18] P. S. Efraimidis and P. G. Spirakis, "Weighted random sampling with a reservoir," *Information Processing Letters*, vol. 97, no. 5, pp. 181-185, 2006.

- [19] M. Al-Kateb, B. S. Lee, and X. S. Wang, "Adaptive-size reservoir sampling over data streams," *19th International Conference on Scientific and Statistical Database Management (SSDBM 2007)*, 2007: IEEE, pp. 22-22.
- [20] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 01, pp. 1-41, 2009.
- [21] A.-O. Boudraa and J.-C. Cexus, "EMD-based signal filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 6, pp. 2196-2202, 2007.
- [22] T. Wang, M. Zhang, Q. Yu, and H. Zhang, "Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal," *Journal of Applied Geophysics*, vol. 83, pp. 29-34, 2012.
- [23] R. Igual, C. Medrano, F. J. Arcega, and G. Mantescu, "Integral mathematical model of power quality disturbances," *18th International Conference on Harmonics and Quality of Power (ICHQP)*, 2018: IEEE, pp. 1-6.
- [24] B. Li, Y. Jing, and W. Xu, "A generic waveform abnormality detection method for utility equipment condition monitoring," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 162-171, 2016.
- [25] M. Yaacob, M. A. Alsaedi, J. Rashed, A. Dakhil, and S. Atyah, "Review on partial discharge detection techniques related to high voltage power equipment using different sensors," *Photonic Sensors*, vol. 4, no. 4, pp. 325-337, 2014.