A Method of Fault Line Selection in Small Current Grounding System Based on VMD Energy Entropy and Optimized K-means Clustering

H Li¹, Y Sun¹, K Li¹, J Liu¹, S Li¹ and M Xu²
¹School of Electrical Engineering, Shandong University, Jinan 250061, China
²Laiyang Power Supply Company, State Grid Shandong Electric Power Company, Yantai, China

Email: 969012980@qq.com; sy@sdu.edu.cn; lkjun@sdu.edu.cn; 249716289@qq.com; dereklsy@163.com; guoanqi630@163.com;

Abstract. When single-phase to ground fault occurs in small current grounding system, the zero-sequence transient current contains a lot of complex fault information. In order to extract characteristic from the zero-sequence current accurately as the criterion of fault line selection, this paper puts forward a method of fault line selection based on VMD energy entropy and optimized K-means clustering. Firstly, the zero-sequence current of each line is decomposed by VMD (variational mode decomposition, VMD), and each sub mode of the original current is obtained. Secondly, the energy entropy of each sub mode is calculated as the characteristic quantity of each line state. Finally, the optimized K-means clustering algorithm is used to cluster the fault feature to realize fault line selection. In this paper, Simulink is used to simulate the single-phase to ground fault in resonant grounded system and the results show that the method that has high accuracy can provide reference for practical engineering application.

1. Introduction
In recent years, there have been a lot of fault line selection methods, mainly divided into the line selection method applying the steady-state signal and the transient signal [1]. The transient current signal has large intensity, and is not affected by the arc suppression coil, so the method of line selection applying transient signal is more suitable for the identification of faults [2-3].

A method of Hilbert transform and probabilistic neural network routing is proposed in [4]. In the proposed method, it is shown that the zero-sequence transient current of each line is decomposed into a series of intrinsic mode functions by the empirical mode decomposition (EMD) and the intrinsic mode function energy is calculated as the input of PNN network to select the fault line. However, the empirical mode decomposition is prone to the phenomenon of modal aliasing, which cannot accurately extract the fault information[5]. In view of the problem of mode mixing in EMD, the Hilbert transform is improved by ensemble empirical mode decomposition (EEMD), and the fault feature components are extracted to identify the fault line[6]. However, EEMD is affected by the addition of white noise amplitude and integration times, and the computational complexity is large.

Variational modal decomposition [7] is a new signal processing method that decomposes the original signal into k finite bandwidths with different center frequencies. This method not only effectively avoids the modal aliasing problem of EMD, but also greatly reduces the computational complexity and improves the computational efficiency relative to EEMD. K-means algorithm is a widely used clustering method [8], usually in European distance as a similarity measure. The algorithm is simple and has a fast convergence rate.
Based on the above analysis, this paper presents a fault line selection method based on variational mode decomposition energy entropy and optimized K-means clustering. Firstly, the zero-sequence current signal of each line is decomposed into sub-modal with the main frequency component by using VMD. Secondly, calculate the energy entropy of each sub modal, which is used as the fault characteristic. Finally, the optimized K-means clustering algorithm is used to cluster the energy entropy, so as to select the fault line more effectively. The simulation results show that the method proposed in this paper can accurately identify the fault line, which has good effectiveness and feasibility.

2. VMD and energy entropy

2.1. Variational modal decomposition principle

The core of VMD is to decompose the input signal into k finite bandwidths with central frequencies, minimizing the sum of the bandwidth estimates for each modal, which is called the mode $u_k$. The signal decomposition process is the solution of the variational problem [9]. The variational problem model with constraints is shown in (1).

$$\min_{\{u_k\} \in \mathbb{C}^k} \left\{ \sum_k \mathcal{E}_k\left[ (\delta(t) + \frac{i}{\pi t}) \cdot u_k(t) \right] \right\}$$

subject to $\sum_k u_k = f$

Where $\{u_k\} = \{u_1, \ldots, u_k\}$ is each modal component. $\{\omega_k\} = \{\omega_1, \ldots, \omega_k\}$ is the center frequency of modal component. $\delta(t)$ is the impulse function. $\partial_t$ is the partial derivative of t. $f$ is the original signal.

In order to obtain the optimal solution of the constrained variational problem, Lagrange multiplication operator $\lambda(t)$ is introduced and the constrained variational problem is transformed into a non-constrained variational problem which is given by (2).

$$L(u_k, \{\omega_k\}, \lambda(t)) := \alpha \sum_k \mathcal{E}_k\left[ (\delta(t) + \frac{i}{\pi t}) \cdot u_k(t) \right] + \left\| f(t) - \sum_k u_k(t) \right\| + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle$$

Where $\alpha$ is the second penalty factor. The saddle point of the above Lagrange function is obtained by alternate direction method of multipliers (ADMM), which is optimal solution. The principle of variational modal decomposition is described in detailly in [10], so this article will not repeat.

2.2. Energy entropy

In order to further extract the fault information contained in the zero sequence current, on the basis of the variational modal decomposition, this paper calculates the sub-mode energy entropy that is used as an important criterion for fault line selection. The principle of energy entropy and the solving process are described as follows [11].

1) Sub modal energy $E_k$. $E_k = \int_{-\infty}^{\infty} |u_k(t)|^2 dt \quad k = 1, 2, \ldots, K$ (3)

2) Energy normalization of K sub modal. $E = \sum_{k=1}^{K} E_k$ (4) $p(u_k) = E_k / E$ (5)

3) VMD energy entropy. $H(u) = -\sum_{k=1}^{K} p(u_k) \log p(u_k)$ (6)

3. The optimized K-means clustering algorithm

3.1. Linear decreasing weight PSO

PSO is an intelligent optimization algorithm, which is widely used in the optimization of dynamic and multi-objective problems. Compared with the traditional optimization algorithm, it has faster computing speed and better global search ability [12]. Therefore, PSO is applied to the optimization of K-means clustering algorithm to solve the problem that K-means algorithm is easy to fall into local
optimal problem improving the final clustering accuracy. Because PSO is easy to be premature and prone to oscillation in the vicinity of the global optimal solution, this paper improves the problem by the method of linear decreasing weight.

3.2. K-means clustering algorithm
The Euclidean distance is used as similarity measure in K-Means algorithm, and the squared error criterion function is used as the clustering criterion function [13]. For the problem of fault line selection, there is no need to consider the selection of the number of cluster centers and we only need to divide the number into two categories. However, the algorithm is sensitive to the initial cluster center selection, and it is prone to local optimum. So the algorithm needs to be optimized properly.

3.3. Linear decreasing weight PSO- K-means clustering algorithm
Based on the above analysis, this paper proposes a method of linear decreasing weight PSO-K-means clustering, which uses PSO algorithm to solve the problem of selecting the initial cluster centers of K-means. In this paper, we use the method of cluster center coding, and design the Euclidean distance function of the sample data points and the center of mass as the fitness function of linear decreasing weight PSO. The whole clustering process is described in Figure 1.

![Flow chart of linear decreasing weight PSO-K-means clustering](image1)

**Figure 1.** The algorithm flow chart of linear decreasing weight PSO-K-means clustering

![System simulation model](image2)

**Figure 2.** System simulation model

4. Simulation analysis
In this paper, a 10kV system is simulated by Simpower System module in Simulink, which is shown in Figure 2. Some of the parameters in the circuit are as follows: In the model there are 6 return lines in
total, and the cable length is shown in the figure. The positive sequence resistance of the line is 0.17Ω/km. The zero sequence resistance is 0.23Ω/km. The positive sequence inductance is 5.48mH/km. The positive sequence capacitance is 12.74×10⁻⁹F/km. The zero sequence inductance is 5.48mH/km. The positive sequence capacitance is 12.74×10⁻⁹F/km. The transformer rated voltage ratio is \( U_{N1}/U_{N2} = 35kV/10kV \).

Single phase to ground fault occurs in line 2 and the fault location is 4 km away from the bus-bar. The operating mode of the system is over compensation. The compensating inductance is 3.3H and the compensation degree is 10%. The sampling frequency is 6 kHz during the simulation.

VMD is used to decompose the zero sequence current signal of line 3 when the fault angle is 90°. The time frequency of zero sequence current for the line 3 is shown in Figure 3. The VMD results of original signal are shown in Figure 4, Figure 5 and Figure 6. The original signal is decomposed into three sub-modal.

The results show that VMD can effectively decompose the zero sequence current signal, and the frequency component of each IMF component is relatively concentrated.

When the fault angle is 30° and 60°, and the grounding resistance is 20Ω and 200Ω, the system is simulated. The VMD energy entropy of each line is calculated and then clustered by the linear decreasing weight PSO-K-means clustering algorithm. The clustering results are described in Table 1 and Table 2.

The clustering results of Table 1 and Table 2 show that whether it is a small fault angle or high resistance fault, the proposed method can accurately determine the fault line.
### Table 1. Clustering results of different fault angles

| Fault setting $R_f=20\Omega$ | Line Ener Entrop | Clustering results | Fault setting $\theta=0^\circ$ | Line Ener Entrop | Clustering results |
|------------------------------|------------------|--------------------|------------------------------|------------------|--------------------|
| $\theta=30^\circ$            |                  |                    | $\theta=0^\circ$              |                  |                    |
| $L_1$                        | 32.32 24.32      | Normal             | $L_1$                        | 29.22 17.34      | Normal             |
| $L_2$                        | 28.73 17.06      | Fault              | $L_2$                        | 22.25 16.94      | Normal             |
| $L_3$                        | 29.51 23.27      | Normal             | $L_3$                        | 24.06 20.71      | Normal             |
| $L_4$                        | 30.81 24.65      | Normal             | $L_4$                        | 25.69 19.14      | Normal             |
| $L_5$                        | 30.85 20.87      | Normal             | $L_5$                        | 24.42 20.49      | Normal             |
| $L_6$                        | 28.93 20.71      | Normal             | $L_6$                        | 25.55 21.68      |                    |

| $\theta=60^\circ$            |                  |                    | $\theta=200\Omega$           |                  |                    |
| $L_1$                        | 37.52 62.40 24.51| Normal             | $L_1$                        | 30.00 23.74      | Normal             |
| $L_2$                        | 36.82 60.51 72.14| Normal             | $L_2$                        | 26.52 20.99      | Normal             |
| $L_3$                        | 28.58 81.45 72.14| Normal             | $L_3$                        | 28.01 20.99      | Fault              |
| $L_4$                        | 25.06 29.49 12.11| Normal             | $L_4$                        | 21.50 20.99      | Normal             |
| $L_5$                        | 26.64 29.49 12.11| Normal             | $L_5$                        | 30.04 20.99      | Normal             |
| $L_6$                        | 27.74 26.64 12.11| Normal             | $L_6$                        | 27.01 20.99      | Normal             |

### Table 2 Clustering results of different fault resistances

| Fault setting $R_f=20\Omega$ | Line Ener Entrop | Clustering results | Fault setting $R_f=200\Omega$ | Line Ener Entrop | Clustering results |
|------------------------------|------------------|--------------------|------------------------------|------------------|--------------------|
| $\theta=0^\circ$             |                  |                    | $\theta=60^\circ$            |                  |                    |
| $L_1$                        | 29.22 17.34      | Normal             | $L_1$                        | 26.52 20.99      | Normal             |
| $L_2$                        | 22.25 16.94      | Normal             | $L_2$                        | 28.01 20.99      | Fault              |
| $L_3$                        | 24.06 20.71      | Normal             | $L_3$                        | 21.50 20.99      | Normal             |
| $L_4$                        | 25.69 19.14      | Normal             | $L_4$                        | 30.04 20.99      | Normal             |
| $L_5$                        | 24.42 20.49      | Normal             | $L_5$                        | 27.01 20.99      | Normal             |
| $L_6$                        | 25.55 21.68      | Normal             | $L_6$                        | 27.51 22.00      |                    |

### 5. Conclusion

This paper presents a method of fault line selection based on VMD energy entropy and optimized K-means clustering. This method not only avoids the modal aliasing problem of EMD, but greatly reduces the computational complexity relative to EEMD. K-means algorithm is optimized by using linear decreasing weight PSO so that the problem has been solved that K-means algorithm is easy to fall into local optimal problem. It is shown that the proposed method in this paper can effectively identify the fault line, which provides a new idea for fault line detection in small current grounding.
The method of feature extracting based on VMD has its own drawbacks. For example, the value of sub modal $K$ should be given in advance, so the method needs to be further improved.

**Acknowledgments**
The authors gratefully acknowledge the assistance of Professor Ying Sun and Ke-Jun Li in research. And at the same time the authors gratefully acknowledge to Shandong University.

**References**

[1] Jiang B, Dong X and Shi S 2014 A method of single phase to ground fault feeder selection based on single phase current traveling wave for distribution lines *Proceedings of the Chinese Society of Electrical Engineering* **34** 6216-6227

[2] Zhang S, Yue M and Pan L 2015 Application of Improved Generalized Harmonic Wavelet Packet Decomposition and Chaos Oscillator to Fault Line Detection in Small Current Grounding System Transactions of China Electro-technical Society 30 13-20

[3] Lai P, Zhou X and Qiu D 2015 Research on transient-current frequency analysis and faulty line detecting method in indirectly grounding power system *Power System Protection & Control* **43** 51-57

[4] Li Z, Wang S and Wu L 2014 A new method and simulation for arcing high-impedance-grounding fault line selection in resonant grounded system *Power System Protection & Control* **42** 44-49

[5] Yang Y, Cheng J and Zhang K 2012 An ensemble local means decomposition method and its application to local rub-impact fault diagnosis of the rotor systems *Measurement* **561**-570

[6] Ren Z and Zhang Y 2015 Fault line selection of distribution network based on improved Hilbert-Huang Transform and identification confidence degree *Power System Protection & Control* **43** 8-13

[7] Konstantin D and Dominique Z 2014 Variational mode decomposition *IEEE Transactions on Signal Processing* **62** 531-544

[8] Renato C and Mirkin B 2012 Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering *Pattern Recognition* **45** 1061-1075

[9] Liu C, Wu Y and Zhen C 2015 Rolling bearing fault diagnosis based on variational mode decomposition and fuzzy C means clustering *Proceedings of the Chinese Society of Electrical Engineering* **35** 3358-3365

[10] Zhang N, Zhu Y and Gao Y 2016 An on-line detection method of transformer winding deformation based on variational mode decomposition and probability density estimation *Power System Technology* **40** 297-302

[11] Xie B, Chen W and Zhou Q 2016 Partition of the development stage of air-gap discharge in oil-paper insulation based on wavelet packet energy entropy *IEEE Transactions on Dielectrics & Electrical Insulation* **23** 866-872

[12] Benidris M, Elsaia S and Mitra J 2015 Power system reliability evaluation using a state space classification technique and particle swarm optimisation search method *IET Generation, Transmission & Distribution* **9** 1865-1873

[13] Boutsidis C and Magdon-Ismail M 2013 Deterministic Feature Selection for K-Means Clustering *IEEE Transactions on Information Theory* **59** 6099-6110