Harmonic Source Region Location Based on Fast ICA and Scene Clustering

Haobo Xu¹², Zhenguo Shao¹² and Feixiong Chen¹²*

¹ College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, Fujian, 350108, China
² Fujian Smart Electrical Engineering Technology Research Center, Fuzhou, Fujian 350108, China
*Corresponding author’s e-mail: feixiongchen@fzu.edu.cn

Abstract. The widespread application of power electronic devices and renewable power generation has resulted in significant harmonic pollution. Harmonic source location is of great importance since it is a prerequisite to mitigate harmonic pollution. In this paper, a scene-based harmonic source region location method is proposed. The proposed method employs the fast independent component analysis (ICA) to locate the harmonic source region as the harmonic measurement matrix is unknown and the measuring equation is not observable. In this way, the range of suspect harmonic sources is narrowed and the cost of harmonic source location is reduced. Furthermore, the improved K-means algorithm is adopted to divide the measurement data into several scenes to improve location accuracy. Finally, a case study is conducted to verify the validity and accuracy of the proposed method.

1. Introduction

The growing adoption of power electronic devices and renewable power generation has exacerbated harmonic-related power quality issues in the power system[1-2]. Excessive harmonic distortion may result in the equipment failures, shorten the life expectancy of transformers, and increases in equipment power losses[3-5]. Therefore, the harmonic source location is essential for effective harmonic mitigation to ensure the safe and efficient operation of the power system.

In recent years, various harmonic source location methods have been proposed. For instance, the least-squares estimation (LSE) is adopted to estimate the harmonic state and locate the harmonic source node[6-7]. However, the above methods will lead to large error as the number of measurement points is limited, because LSE requires that the number of measurement points is greater than that of the network nodes. In this case, considering that the distribution of harmonic sources is relatively sparse, the compressive sensing (CS) and sparse Bayesian learning (SBL) are utilized to estimate the harmonic state and locate the harmonic source node as the measuring equation is observable[8-9]. Compared with LSE, the number of measurement points required in CS and SBL is smaller, but the limited measurement points in practice are still difficult to satisfy the requirements of CS and SBL. Therefore, it is necessary to locate the harmonic source regions as the measuring equation is not observable. In this way, the range of suspect harmonic sources can be narrowed and the cost of harmonic source location can be reduced.

Besides, the existing methods discussed above have a fundamental shortcoming which renders it of limited practical value, i.e., the above methods assume the knowledge of the harmonic measurement...
matrix. However, it is difficult to obtain an accurate harmonic measurement matrix in practice. To deal with these issues, Wei Zhou et al. adopt the supervised learning technique and independent component analysis (ICA) to predict the harmonic measurement matrix and locate the harmonic source [10-11]. However, the above methods assume that the measurement data contains only one scene (the number and location of the harmonic sources remain the same). If the measurement data contains several scenes (the number and location of the harmonic sources are variable), the above methods will result in large errors. Therefore, it is critical to divide the measurement data into several scenes to improve the accuracy of the harmonic source location.

Based on the discussion above, the scene-based harmonic source region location method is proposed in this paper. First, the fast ICA is employed to locate the harmonic source regions as the harmonic measurement matrix is unknown and the measuring equation is not observable. In this way, the range of suspect harmonic sources can be narrowed and the cost of harmonic source location can be reduced. Second, the improved K-means algorithm is adopted to divide the measurement data into several scenes. The optimal number of clusters and initial clustering centers are calculated to improve the clustering accuracy. Therefore, the accuracy of harmonic source region location can be improved.

The rest of the paper is organized as follows. Section 2 presents the scene-based harmonic source region location method. Section 3 summarizes the simulation results. Finally, Section 4 concludes the paper.

2. Scene-based harmonic source region location
In this section, the harmonic source region location method and scene clustering method for measurement data are proposed, respectively.

2.1. Harmonic source region location
Let the node $h$-th harmonic voltages and injection $h$-th harmonic currents be the measurement and state variables, respectively. It is assumed that the measurement data contains only one scene. The $h$-th harmonic measurement equation is shown as follows,

$$U_h = Z_h \cdot I_h$$

where $Z_h$ is the $m \times n$ matrix that denotes the $h$-th harmonic measurement matrix, $U_h$ is the $m \times T$ matrix that represents the node $h$-th harmonic voltages, $I_h$ is the $n \times T$ matrix that represents the injection $h$-th harmonic currents, $m$ denotes the number of measurement points, $n$ denotes the number of harmonic sources, and $T$ denotes the number of measurement data.

The existing harmonic source location methods generally utilize $U_h$ and $Z_h$ to estimate $I_h$ and locate the harmonic source. However, it is difficult to obtain the accurate harmonic measurement matrix $Z_h$ in practice. Therefore, it is necessary to locate the harmonic sources as $Z_h$ is unknown. Additionally, the existing harmonic source location methods generally locate the harmonic source nodes as the measuring equation is observable. The location error is large as the measuring equation is not observable. Therefore, if we can accurately locate the harmonic source regions as the measuring equation is not observable, it is helpful to narrow the range of suspect harmonic sources and reduce the cost of the harmonic source location.

Aiming at the above problems, the fast ICA is adopted to estimate $Z_h$ and locate the harmonic source regions even the harmonic measurement matrix is unknown and the measuring equation is not observable.

The fast ICA model is shown as follows,

$$X = A \cdot S$$

where $X$ is the observation signals, $A$ denotes the hybrid matrix, and $S$ represents the original signals.

The fast ICA model can estimate $A$ and $S$ by utilizing $X$ as the following three conditions are met: (a) The original signals $S$ are statistically independent for each other. (b) At most one original signal is
Gaussian distribution. (c) The number of observation signals is greater than or equal to the number of original signals.

Since the distribution of harmonic sources is sparse, i.e., the number of harmonic sources is much smaller than that of network nodes. Therefore, the condition (c) is usually met as the measuring equation is not observable. On the other hand, to satisfy conditions (a) and (b), each row of $X$ has been centered to have mean zero (that is, the row means of $X$ are zero) and $X$ has been whitened. Therefore, $U_h^*$ is obtained by centering each row of $U_h$. Assumed that $Q$ denotes the whiten matrix for $U_h$. It can be found that $X$ corresponds to $Q \cdot U_h^*$.

Additionally, since there is the uncertainty of estimated original signals in the fast ICA model\cite{12-14}. Therefore, $S$ corresponds to $D \cdot P \cdot I_h^*$, where $D$ is an $n \times n$ matrix whose each row and each column has one and only one element equal to 1, the remaining elements of $D$ are zero, $P$ is a diagonal matrix whose diagonal elements are not zero, and $I_h^*$ is the matrix obtained by centering each row of $I_h$.

In conclusion, Equation (2) can be rewritten as follows.

$$Q \cdot U_h^* = A \cdot (D \cdot P \cdot I_h^*)$$  \hspace{1cm} (3)

Since

$$U_h^* = Z_h \cdot I_h^*$$  \hspace{1cm} (4)

therefore

$$Z_h = Q^{-1} \cdot A \cdot D \cdot P$$  \hspace{1cm} (5)

Let $Z_h^E = Q^{-1} \cdot A$, since $Q$ and $A$ can be calculated by a fast ICA model, therefore, $Z_h^E$ can be obtained. Since $Z_h = Z_h^E \cdot D \cdot P$ and the column correlation between $Z_h$ and $Z_h^E$ is consistent. Therefore, instead of $Z_h$, $Z_h^E$ can be used to locate the harmonic source region.

It is not sufficient for $Z_h^E$ to locate the harmonic source region. We also need to calculate the column correlation between $Z_h^E$ and the location matrix $L$. The process of the harmonic source region location method in the $N$-node network is shown in Figure 1.

![Figure 1. The process of the harmonic source region location method.](image-url)
First, \( Z_h^E \) is calculated based on fast ICA. Second, assuming that all network impedances are 1, the admittance matrix \( Y \) is calculated. Furthermore, the impedance matrix \( Z \) is obtained by calculating the inverse of \( Y \), i.e., \( Z = Y^{-1} \). Then the \( m \) rows corresponding to the serial numbers of measurement points are extracted from \( Z \) to form \( L \). For instance, if the serial numbers of measurement points are 3, 7, 8, and 12, the rows 3, 7, 8, and 12 of \( Z \) will be extracted to form \( L \). Thereafter, the correlation coefficient \( C_{ij} \) is calculated according to the Equation (6). Let \( \lambda \) be the location threshold and \( 0 < \lambda < 1 \). If \( C_{ij} \geq \lambda \), the \( j \)-th node is considered as a suspect harmonic source node. Finally, the harmonic source regions are located according to the all suspect harmonic source nodes.

\[
C_{ij} = \frac{\text{cov}(Z_i^{E(i)}, L_j^{(j)})}{\sigma(Z_i^{E(i)}) \cdot \sigma(L_j^{(j)})}
\]

where \( Z_i^{E(i)} \) is \( i \)-th column of \( Z_h^E \), \( L_j^{(j)} \) is \( j \)-th column of \( L \), \( \text{cov}(\cdot) \) denotes the covariance, and \( \sigma(\cdot) \) denotes the standard deviation.

### 2.2. Scene clustering for measurement data

The above harmonic source region location method assumes that the measurement data contains only one scene. However, the number and location of the harmonic sources are generally variable. Therefore, the measurement data generally contains several scenes. Note that the scene refers to the harmonic source distributions, i.e., the number and location of the harmonic sources. The above harmonic source region location method may result in a large error if the measurement data contains several scenes. Therefore, it is critical to divide the measurement data into several scenes.

In this paper, the K-means algorithm based on particle swarm optimization (PSO) is adopted to divide the measurement data into several scenes. We adopt the SSE and PSO, which is different from the conventional K-means algorithm, to calculate the optimal number of clusters and initial clustering centers to improve clustering accuracy.

Considering that the number of clusters of the K-means algorithm needs to be set beforehand, we introduce SSE formulated as follows to determine the optimal number of clusters,

\[
SSE = \sum_{i=1}^{k} \sum_{q \in R_i} (q - v_i)^2
\]

where \( k \) is the number of clusters, \( R_i \) is the \( i \)-th cluster, \( q \) denotes a data point belonging to \( Y \), and \( v_i \) is the center of \( R_i \).

With the increasing number of clusters, SSE decreases, and SSE will decrease markedly if \( k \) approximates the optimal number of clusters. Therefore, we calculate SSEs under different conditions \((k = 1, 2, \ldots, 10)\) and take \( k \) at the marked decline of SSEs as the optimal number of clusters.

Additionally, PSO is employed to obtain the initial clustering center of the K-means algorithm rather than randomly initializing the K-means algorithm. Compared with the random initialization K-means algorithm, it is more efficient for the PSO-based K-means algorithm to search for the near-global solution or global optimal solution and enhance the clustering accuracy and computational efficiency\(^{[15-16]}\).

Based on the discussion above, the PSO-based K-means algorithm can divide the measurement data into several scenes accurately. Therefore, the accuracy of harmonic source region location can be improved significantly even the measurement data contains several scenes.

### 3. Case study

#### 3.1. Test system

Since field data and real system models are scarce and not readily available, we evaluate the proposed method on an IEEE 33-bus test system in MATLAB. The topological graph of the IEEE 33-bus test system is shown in Figure 2, node 0 is the balancing node, and nodes 1-32 are load nodes. Taking the
5-th harmonic source as an example \((h = 5)\), the interval constant current source model is adopted to simulate the harmonic source. Assumed that there are three scenes in the test system, and the harmonic source distributions of three scenes are shown in Table 1. The nodes 1, 7, 16, 19, 22, and 27 are selected as the measurement points. The harmonic voltage interval of measurement points of each scene is obtained by harmonic power-flow calculation, and the harmonic voltage interval is randomly sampled 100 times to obtain measurement data of each scene.

![Topological graph of the IEEE 33-bus test system](image_url)

**Figure 2.** The topological graph of the IEEE 33-bus test system.

### Table 1. The harmonic source distributions of three scenes.

| Scene | Harmonic Source Node | Harmonic Current Interval / A |
|-------|----------------------|-----------------------------|
| Scene 1 | 3                    | [10, 12]                    |
|        | 17                   | [12, 14]                    |
|        | 24                   | [8, 10]                     |
| Scene 2 | 3                    | [4, 5]                      |
|        | 21                   | [12, 13]                    |
|        | 32                   | [9, 10]                     |
| Scene 3 | 5                    | [10, 12]                    |
|        | 21                   | [4, 6]                      |

#### 3.2. Implementation of scene clustering

In this subsection, to simulate the measurement data that contains several scenes, the measurement data of three scenes is combined into a dataset. Thereafter, the PSO-based K-means algorithm is employed to divide the dataset into several scenes. The clustering results are shown in Table 2. It can be found that the 100 data points of each scene are accurately aggregated into a cluster. Therefore, the clustering accuracy \(CA\) is 100% according to equation (8) and the PSO-based K-means algorithm can divide the measurement data into several scenes accurately.

\[
CA = \frac{\text{sum(diag}(CM))}{\text{sum}(CM)}
\]  

(8)

where \(CM\) denotes the confusion matrix of clustering results, \(\text{sum}[x]\) represents the sum of all the elements of the square matrix \(x\), and \(\text{diag}(x)\) is a diagonal matrix whose diagonal elements are equal to those of the square matrix \(x\).

### Table 2. Confusion matrix of clustering results.

| True value | Scene 1 | Scene 2 | Scene 3 |
|------------|---------|---------|---------|
| Scene 1    |         |         |         |
| Scene 2    |         |         |         |
| Scene 3    |         |         |         |
3.3. Influence of location threshold on harmonic source region location

In this subsection, the influence of location threshold on harmonic source region location is discussed. The measurement data of Scenes 1 obtained by the PSO-based K-means algorithm is employed to locate the harmonic source regions as the location threshold $\lambda$ is equal to 0.99, 0.95, 0.90, and 0.85, respectively. The suspect harmonic source nodes are obtained by the proposed method, and the results are shown in Table 3. Thereafter, the harmonic source regions are located according to the suspect harmonic source nodes and the results are shown in Figure 3. In Figure 3, the regions in the yellow polygons are the harmonic source regions found by the proposed method. It can be observed that as location threshold $\lambda$ increases, so does the range of harmonic source regions. When the value of $\lambda$ is large, some harmonic source regions may be omitted (see $\lambda = 0.99$). Alternatively, when the value of $\lambda$ is small, although the regions containing the harmonic sources are located, the range of harmonic source regions is large (see $\lambda = 0.85$). Therefore, it is relatively suitable to let the location threshold $\lambda$ be 0.95 for the proposed method in the IEEE 33-bus test system.

| Location threshold | Suspect harmonic source nodes |
|--------------------|-------------------------------|
| 0.99               | 3, 4, 15, 16, 17              |
| 0.95               | 2, 3, 4, 12, 13, 14, 15, 16, 17, 22, 23, 24 |
| 0.90               | 1, 2, 3, 4, 10, 11, 12, 13, 14, 15, 16, 17, 22, 23, 24 |
| 0.85               | 1, 2, 3, 4, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 22, 23, 24 |

Figure 3. Harmonic source region location results: (a) $\lambda = 0.99$, (b) $\lambda = 0.95$, (c) $\lambda = 0.90$, (d) $\lambda = 0.85$.

3.4. Implementation of harmonic source region location

In this subsection, the measurement data of Scenes 1, 2, and 3 obtained by the PSO-based K-means algorithm is employed to locate the harmonic source regions, respectively. First, let the location threshold $\lambda$ be 0.95. Then the suspect harmonic source nodes of three scenes are obtained by the proposed method, respectively, and the results are shown in Table 4. Thereafter, the harmonic source
regions are located according to the suspect harmonic source nodes and the results are shown in Figure 4(a), (b), and (c). In Figure 4(a), (b), and (c), the regions in the red polygons are the harmonic source regions found by the proposed method. It can be seen that the regions containing the harmonic sources can be located by the proposed method, and the regions that do not contain the harmonic sources are not mistakenly identified. Besides, the proposed method is implemented as the measuring equation is not observable, since the measuring equation of the IEEE 33-bus test system is observable only as the number of measurement points is greater than 19. Therefore, the proposed method can accurately locate the harmonic source regions as the measuring equation is not observable.

Additionally, the measurement data without scene clustering is adopted to locate the harmonic source regions by the proposed method. The results are shown in Figure 4(d), the regions in the blue polygons are the harmonic source regions. It can be found that the regions containing the harmonic sources are not all located, and the nodes 25 and 26 are incorrectly identified as the harmonic source region. Therefore, compared with the results of Figure 4(a), (b), and (c), it can be seen that the scene clustering can improve the accuracy of the harmonic source region location.

Table 4. The suspect harmonic source nodes of three scenes.

| Scene          | Suspect harmonic source nodes |
|---------------|-------------------------------|
| Scene 1       | 2, 3, 4, 12, 13, 14, 15, 16, 17, 22, 23, 24 |
| Scene 2       | 3, 4, 5, 18, 19, 20, 21, 28, 29, 30, 31, 32 |
| Scene 3       | 4, 5, 18, 19, 20, 21             |
| Without scene clustering | 2, 3, 4, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 25, 26 |

Figure 4. Harmonic source region location results: (a) Scene 1, (b) Scene 2, (c) Scene 3, (d) Without scene clustering.

4. Conclusion
The scene-based harmonic source region location method is proposed in this paper based on the fast ICA and PSO-based K-means algorithm. The proposed method accurately locates the harmonic source regions as the harmonic measurement matrix is unknown, the measuring equation is not observable, and the measurement data contains several scenes. The case study performed on the IEEE 33-bus test system corroborates the validity and accuracy of the proposed method. Finally, the proposed method can help electric utilities to narrow the range of suspect harmonic sources and reduce the cost of the harmonic source location.
In future work, we intend to research the optimal distribution of the measurement points to further reduce the cost of the harmonic source location.

Acknowledgments
This work is supported by the National Natural Science Foundation of China (51777035).

References
[1] Molina-Moreno, I., Medina, A., Cisneros-Magaña, R., Anaya-Lara, O. (2017) Time domain harmonic state estimation in unbalanced power networks based on optimal number of meters and the principle of half-wave symmetry. IET Generation, Transmission & Distribution, 11(15): 3871-3880.
[2] Sharma, H., Rylander, M., Dorr, D. (2016) Grid Impacts Due to Increased Penetration of Newer Harmonic Sources. IEEE Transactions on Industry Applications, 52(1): 99-104.
[3] Xu, W., Liu, X., Liu, Y. (2003) An investigation on the validity of power-direction method for harmonic source determination. IEEE Transactions on Power Delivery, 18(1): 214-219.
[4] Sinvula, R., Abo-Al-Ez, K. M., Kahn, M. T. (2019) Harmonic Source Detection Methods: A Systematic Literature Review. IEEE Access, 7: 74283-74299.
[5] Ko, W., Gu, J. (2016) Impact of Shunt Active Harmonic Filter on Harmonic Current Distortion of Voltage Source Inverter-Fed Drives. IEEE Transactions on Industry Applications, 52(4): 2816-2825.
[6] Najjar, M. Y., Heydt, G. T. (1991) A hybrid nonlinear-least squares estimation of harmonic signal levels in power systems. IEEE Transactions on Power Delivery, 6(1): 282-288.
[7] Kanao, N., Yamashita, M., Yanagida, H., Mizukami, M., Hayashi, Y., Matsuki, J. (2005) Power system harmonic analysis using state-estimation method for Japanese field data. IEEE Transactions on Power Delivery, 20(2): 970-977.
[8] Carta, D., Muscas, C., Pegoraro, P. A., Sulis, S. (2019) Identification and Estimation of Harmonic Sources Based on Compressive Sensing. IEEE Transactions on Instrumentation and Measurement, 68(1): 95-104.
[9] Yuan, Y., Zhou, W., Zhang, H., Ping, Z., Ardakanian O. (2018) Sparse Bayesian Harmonic State Estimation. In: IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids. Aalborg. pp. 1-6.
[10] Zhou, W., Ardakanian, O., Zhang, H., Yuan, Y. (2020) Bayesian Learning-Based Harmonic State Estimation in Distribution Systems With Smart Meter and DPMU Data. IEEE Transactions on Smart Grid, 11(1): 832-845.
[11] Gursoy, E., Niebur, D. (2009) Harmonic Load Identification Using Complex Independent Component Analysis. IEEE Transactions on Power Delivery, 24(1): 285-292.
[12] Shi, Y., Zeng, W., Wang, N., Zhao, L. (2018) A New Constrained Spatiotemporal ICA Method Based on Multi-Objective Optimization for fMRI Data Analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(9): 1690-1699.
[13] Rodriguez, P. A., Anderson, M., Calhoun, V. D., Adali, T. (2015) General Nonunitary Constrained ICA and its Application to Complex-Valued fMRI Data. IEEE Transactions on Biomedical Engineering, 62(3): 922-929.
[14] Fontgalland, G., Ferreira, P. I. L. (2017) Combining Antenna Array Elements by Using ICA Method for Remote Sensing of Sources. IEEE Antennas and Wireless Propagation Letters, 16: 234-237.
[15] Peng, K., Leung, V.C.M., Huang, Q. (2018) Clustering approach based on mini batch kmeans for intrusion detection system over big data. IEEE Access, 6: 11897–11906.
[16] Zhang, Y., Chiang, H. (2017) A novel consensus-based particle swarm optimization-assisted trust-tech methodology for large-scale global optimization. IEEE Transactions on Cybernetics, 47(9): 2717–2729.