

Fault Location in Power Distribution Networks using Arbitrary Similarity Criteria in the Principal Component Subspace

L. Souto, J. Meléndez, S. Herraiz
Intelligent Systems and Control Engineering group
Universitat de Girona
Girona, Girona, Spain
laiz.souto, joaquim.melendez, sergio.herraiz @udg.edu

Abstract—This paper presents a new strategy to support fault location in power distribution networks based on arbitrary similarity criteria in the principal component subspace. Input data consist of impedance measurements at the secondary transformer of one or more substation buses, which are used to build data-driven models of the grid operating conditions. The statistical models of actual operating conditions are further compared with a few reference scenarios to determine the network configuration and the type and location of the fault based on arbitrary features which minimize the variability of the data. Furthermore, this paper includes a case study with a real-based low voltage power distribution network to test the method under different faults.

Index Terms—fault location, machine learning algorithms, power distribution faults, smart grids, statistical learning

I. INTRODUCTION

Accurate fault location in power distribution networks still faces challenges at present, particularly at low voltage (LV) level [1], [2]. Stumbling blocks include the large number of lateral branches, different conductor sizes, unbalanced operation, increasing penetration level of distributed generation (DG), implementation of flexible network configurations (e.g. switches), limited measurements along the feeders (usually available only at the distribution substation), and inaccurate representation of components (e.g. DG, loads, lines) [3], [4].

In this scenario, there are great opportunities for computational intelligence techniques due to the increasing amount of data gathered by phasor measurement units (PMU) and smart meters (SM) combined to the availability of computers with high processing power [5]. However, state-of-the-art fault location methods based on computational intelligence are complex to implement and computationally expensive to train [2], [6], [7] and have to be simplified.

Thereby, the application of dimensionality reduction techniques for fault location is appealing, as they are easily implementable and reduce the amount of information required for the monitoring of the grid. This article fits into this

This research was supported by the European Union's Horizon 2020 research and innovation programme, call LCE-01-2016-2017, under the auspices of the project *Renewable penetration levered by Efficient Low Voltage Distribution grids*, grant agreement number 773715, and University of Girona scholarship.

context and extends the methodology presented in [8] with an alternative calculation of similarity criteria to support location of permanent faults in power distribution networks with DG and variable loads connected to the feeders. Input data consist of impedance measurements gathered at the secondary of substation transformers, which are then used build a multivariate statistical model of the network operating conditions. Dimensionality reduction is achieved with an arbitrary selection of principal components that capture relevant trends of power system operating conditions. Thus, the algorithm relies on historic data of a few reference scenarios representing standard operation and short-circuits at different buses to distinguish between different fault types and perform fault location, supported by information about the network topology. Moreover, the methodology is general enough to consider changes in the grid configuration and identify them correctly. As a result, this approach makes an attempt to capture the main trends of power system operation that are relevant to identify, classify, and locate faults. Furthermore, testing is conducted in a real-based LV power distribution network under different fault types, fault impedance values, and grid settings.

II. METHODOLOGY

The fault location method presented in this article is based on principal component analysis (PCA). Some theoretical background is summarized in the following paragraphs.

First, consider a generic power distribution network with J substation buses equipped with PMUs and K reference scenarios of operating conditions including standard operation and short-circuits at different buses. For a generic reference scenario k , $k = 1, \dots, K$, let $\mathbf{X}(k)$ be an $n \times m$ centered and scaled observation matrix with n valid samples over time and m variables referred to distinct phasor quantities measured at substation buses $j = 1, \dots, J$. Then, compute the covariance matrix of $\mathbf{X}(k)$ and apply eigenvalue decomposition to the resulting matrix such that the $m \times m$ matrices $\mathbf{V}(k)$ and $\mathbf{\Lambda}(k)$ are obtained. Columns in $\mathbf{V}(k)$ are the eigenvectors or principal components $\mathbf{v}_i(k)$, $i = 1, \dots, m$, which consist of orthonormal vectors whose directions represent the major variability of the data and the relative loadings of the original

variables. In turn, $\Lambda(k)$ is a diagonal matrix containing the eigenvalues $\lambda_i(k)$, $i = 1, \dots, m$, which express variability in the direction of each principal component or column of $\mathbf{V}(k)$.

Then, consider a generic testing scenario denoted by k' with a few observations of an unknown operating condition and let $\mathbf{V}(k')$ and $\Lambda(k')$ be its $m \times m$ eigenvector matrix and eigenvalue matrix, respectively. Under the assumption that the operating conditions of some scenarios \hat{k} of all reference scenarios K are nearly the same as k' , then $\mathbf{V}(\hat{k})$ and $\mathbf{V}(k')$ are expected to be very similar. However, in the event of a fault whose impedance is different from the reference scenarios, the principal components whose relative weights or loadings cannot be neglected in the variables involved in the fault should not be selected to perform fault location, as they are expected to deviate more than the others.

As a matter of fact, the equivalent impedance \bar{Z}_{eq} seen at the substation bus in the event of a fault is given by (1), where \bar{Z}_{us} is the upstream equivalent impedance, \bar{Z}_F is the fault impedance, and \bar{Z}_{ds} is the downstream equivalent impedance.

$$\bar{Z}_{eq} = \bar{Z}_{us} + \frac{\bar{Z}_F \times \bar{Z}_{ds}}{\bar{Z}_F + \bar{Z}_{ds}} \quad (1)$$

In the event of a fault with $\bar{Z}'_F \neq \bar{Z}_F$, where \bar{Z}_F is the fault impedance of a reference scenario and \bar{Z}'_F is the fault impedance of a testing scenario k' , the difference between the equivalent impedance calculated with (1) is given by (2)

$$\bar{Z}_{eq} - \bar{Z}'_{eq} = \bar{Z}_{ds} \frac{\bar{Z}_F (\bar{Z}'_F + \bar{Z}_{ds}) - \bar{Z}'_F (\bar{Z}_F + \bar{Z}_{ds})}{(\bar{Z}_F + \bar{Z}_{ds}) (\bar{Z}'_F + \bar{Z}_{ds})} \quad (2)$$

Assuming that the fault impedance is very small in comparison with the downstream equivalent impedance, $\bar{Z}_F + \bar{Z}_{ds} \approx \bar{Z}_{ds}$ and $\bar{Z}'_F + \bar{Z}_{ds} \approx \bar{Z}_{ds}$. Then (2) can be simplified as (3)

$$\bar{Z}_{eq} - \bar{Z}'_{eq} = \bar{Z}_F - \bar{Z}'_F \quad (3)$$

Therefore, dimensionality reduction is achieved by retaining $r < m$ principal components or columns of $\mathbf{V}(k)$ and $\mathbf{V}(k')$ which represent the major trends of the data set. As a result, $\mathbf{V}(k)$ is reduced with some loss of information to an $m \times r$ matrix $\mathbf{P}(k)$ given by (4). It is noteworthy that the assembling of $\mathbf{P}(k)$ is arbitrary and $\mathbf{p}_i(k) \neq \mathbf{v}_i(k)$ may hold, $i = 1, \dots, r$.

$$\mathbf{P}(k) = [\mathbf{p}_1(k) \quad \dots \quad \mathbf{p}_r(k)] \quad (4)$$

with $\mathbf{p}_i = [p_{1,i} \quad \dots \quad p_{m,i}]^T$, $i = \{1, \dots, m\}$

In the event of short-circuit operation in k' , further investigation is required to locate the fault over all candidate scenarios \hat{k} of K representing a given fault type and grid setting. The selection of r principal components is arbitrary and depends on the following aspects:

- 1) Network configuration: The selection of principal components may consider a single PMU or more PMUs, depending on the status of the switches in the network. The status of the switch is determined by calculating the

dot product between $\mathbf{p}_i(\hat{k})$ and $\mathbf{p}_i(k')$, which should be very close to the unity for all $i = 1, \dots, m$ if the switch connecting the PMU j of J is off. If the switch mode is off, then investigation proceeds with PMU data recorded at the substation where the fault is detected only; otherwise, if the switch mode is on, then investigation proceeds with PMU data recorded at all substations involved in the fault.

- 2) Fault type: the selection of r principal components is made separately for each reference scenario \hat{k} with short-circuit operation and is based on the loadings of the principal components $\mathbf{p}_i(\hat{k})$ of $\mathbf{P}(\hat{k})$, $i = 1, \dots, m$. The principal components whose relative weights or loadings are minimum in the variables involved in the fault are selected to perform fault location, as they are expected to be very similar to the testing scenario k' due to the same operating conditions.

For instance, consider $m = 3$ and different faults Φ given by a unitary direction vector ϕ multiplied by a fault magnitude f . For three-phase (ABC) faults, make $\Phi = f[1, 1, 1]^T$; for phase-phase (BC) and phase-phase-ground (BCG) faults, make $\Phi = f[0, 1, 1]^T$; for phase-ground (AG) faults, make $\Phi = f[1, 0, 0]^T$. The principal components whose relative weights or loadings are minimum in the variables involved in the fault (i.e. ones in ϕ) are selected to perform fault location. Therefore, for three-phase (ABC) faults, the sum of $p_{1,i}(k)$, $p_{2,i}(k)$, and $p_{3,i}(k)$ should be close to zero; for phase-phase (BC) and phase-phase-ground (BCG) faults, the sum of $p_{2,i}(k)$ and $p_{3,i}(k)$ should be close to zero; for phase-ground (AG) faults, $v_{1,i}(k)$ should be close to zero, $i = \{1, 2, 3\}$.

The similarity between \hat{k} and k' (assembled with the principal components used in \hat{k}) can be computed through the dot product between the principal components $\mathbf{p}_i(\cdot)$ of those matrices, $i = 1, \dots, m$, which is equal to the unity if $\mathbf{p}_i(\hat{k})$ and $\mathbf{p}_i(k')$ are identical. This procedure allows to identify the operating condition of k' , including the presence of faults and phases involved, as well as the grid setting of the network in the presence of switches.

Finally, the overall similarity criteria $\varphi_{\hat{k}, k'}$ is calculated in (5) as the sum of the dot product of all $\mathbf{p}(\cdot)$ in the resulting matrices $\mathbf{P}(\cdot)$. Since r may be different for distinct reference scenarios, the sum should be averaged over r to ensure consistency of results. The results calculated with (5) are ranked in descending order such that the reference scenario \hat{k} with the highest φ is the most probable location of the fault in k' . If there is a tie between two or more buses, the similarity criteria is also computed in the nodes connecting lateral branches.

$$\varphi_{\hat{k}, k'} = \frac{\sum_{i=1}^r |\mathbf{p}_i(\hat{k}) \cdot \mathbf{p}_i(k')|}{r} \quad (5)$$

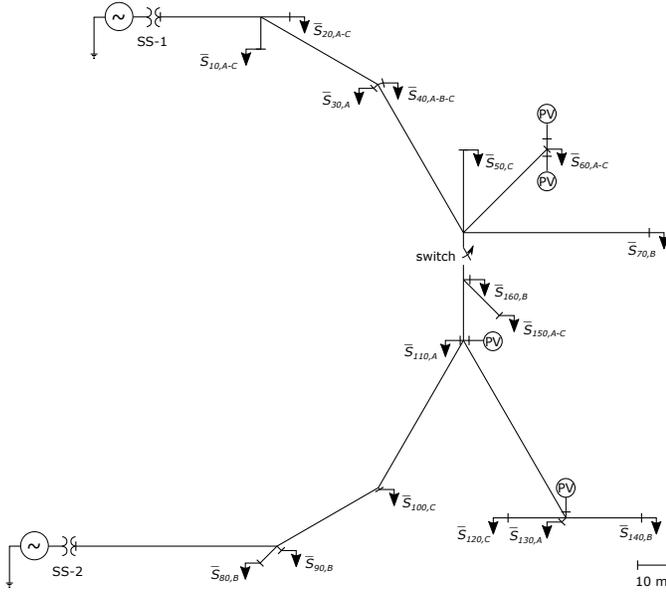


Fig. 1. Single-line diagram of the LV distribution network, with 2 distribution substations (SS-1 and SS-2) equipped with PMUs, 28 feeders drawn to scale, 1 switch, 4 PV panels, 1 industrial three-phase customer ($\bar{S}_{40,A-B-C}$), and 19 residential single-phase customers ($\bar{S}_{10,A}$, $\bar{S}_{10,C}$, $\bar{S}_{20,A}$, $\bar{S}_{20,C}$, $\bar{S}_{30,A}$, $\bar{S}_{50,C}$, $\bar{S}_{60,A}$, $\bar{S}_{60,C}$, $\bar{S}_{70,B}$, $\bar{S}_{80,B}$, $\bar{S}_{90,B}$, $\bar{S}_{100,C}$, $\bar{S}_{110,A}$, $\bar{S}_{120,C}$, $\bar{S}_{130,A}$, $\bar{S}_{140,B}$, $\bar{S}_{150,A}$, $\bar{S}_{150,C}$, $\bar{S}_{160,B}$).

III. CASE STUDY

The methodology was tested in a real-based LV distribution network simulated in Simulink, whose topology is illustrated in Fig. 1. It represents a LV distribution network located in Catalonia, Spain, which consist of primary distribution feeders with branches connecting the substation node to the customers (i.e. local energy producers or consumers). In total, the network has 2 distribution substations (one with a 250-kVA and the other with a 630-kVA transformer Dyn11, 400 V secondary), 28 feeders modeled as short R-L lines (with $\frac{R}{X} = 5.4$ for overhead lines and $\frac{R}{X} = 2.7$ for underground cables), 1 switch, 20 different energy consumers (among them, 1 industrial, three-phase with 70 kW of contracted power and 19 residential, single-phase with less than 10 kW of contracted power), and DG from 4 PV modules (10 kWp each). In total, the length of the primary distribution feeder connecting SS-1 to SS-2 is 325 m and the length of the longest lateral branch is 95 m.

PMUs are installed at both substation nodes and sample phase voltage and line current phasor quantities from which the equivalent impedance is calculated. A statistical model is built for each PMU in separate, as they are supposed to represent systems that are possibly independent from each other. This choice provides 2 distinct statistical models per scenario with $m = 3$ variables each. It is noteworthy that the statistical models are built with impedance magnitudes only to suit the linear algorithm in use. This choice is not expected to make a negative impact on the accuracy of fault location, despite the nonlinear behavior of the faults, since the lines and loads are mainly resistive and the faults are purely resistive.

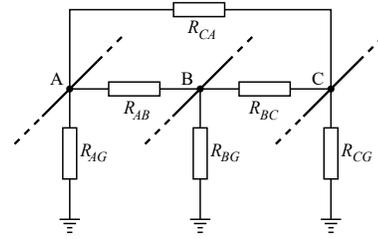


Fig. 2. Fault impedance

Historical data are obtained from simulations of standard operation and permanent short-circuits with typical hourly values of PV generation and load consumption profiles over a year. This choice provides $n = 365$ observations per reference scenario and is suitable to a specific hour of the day. Although this time intervals are chosen due to the real PV generation and load consumption profiles available for the simulations, it is noteworthy that a shorter or longer time interval can be used to build the statistical models without loss of generality. In addition, the reference fault scenarios include different types of faults applied at the substation and load buses with fault resistance $R_F = 0.001\Omega$ between phases and/or phases and ground. Both switch statuses on and off are considered in all training and testing scenarios.

In turn, testing scenarios consist of permanent three-phase (ABC), phase-ground (AG), phase-phase (BC), and phase-phase-ground (BCG) faults simulated at the load buses, considering the same PV generation and load consumption profiles as the reference scenarios. The fault impedance model is illustrated in Fig. 2. Different values of fault resistance are considered as a function of the fault type: $R_F = 0.1\Omega$ and $R_F = 1.0\Omega$ between phases (i.e. R_{AB} , R_{BC} , R_{CA} in Fig. 2) and/or phases and ground (i.e. R_{AG} , R_{BG} , R_{CG} in Fig. 2), whenever applicable. In total, 256 fault scenarios are tested per hour (i.e. 16 faulty buses times 4 fault types times 2 fault resistance values times 2 switch modes).

IV. RESULTS

The algorithm is capable of identifying the status of the switch (i.e. on, off) and the type of fault (i.e. three-phase, phase-ground, phase-phase, phase-phase-ground) correctly for all simulated faults (k') using the similarity criteria presented in Section II. Calculation times are between 0.025 s and 0.050 s for all testing scenarios, whereas the file size of all reference scenarios together for each fault type is 20 kB.

The fault location results computed under different fault operating conditions at midday are displayed in Tables I to VIII, obtained for three-phase, phase-ground, phase-phase, and double-phase-ground faults with $R_F = 0.1\Omega$ and $R_F = 1.0\Omega$ at different buses (i.e. fault scenarios k') with the switch off and on. The faulty bus k' is displayed in the first column and the results obtained with the similarity criteria in the principal component subspace are shown in the following columns, including the calculated faulty bus \bar{k} determined by considering the sum of dot products calculated with (5) for all candidate buses, given information about the right network

TABLE I

RESULTS: THREE-PHASE FAULT, SWITCH STATUS OFF

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	10	0.0	40	58.5
20	10	20.6	40	57.5
30	40	3.8	10	58.5
40	40	0.0	40	0.0
50	40	75.4	40	75.4
60	50	68.0	40	87.5
70	40	110.5	40	110.5
80	80	0.0	80	0.0
90	80	10.2	80	10.2
100	100	0.0	100	0.0
110	160	24.7	120	90.9
120	120	0.0	120	0.0
130	120	21.5	120	21.5
140	120	45.1	120	45.1
150	160	19.3	160	19.3
160	160	0.0	160	0.0

TABLE II

RESULTS: THREE-PHASE FAULT, SWITCH STATUS ON

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	20	20.6	10	0.0
20	20	0.0	20	0.0
30	30	0.0	20	57.5
40	40	0.0	20	57.5
50	50	0.0	50	0.0
60	50	68.0	40	87.5
70	50	90.9	40	110.5
80	80	0.0	80	0.0
90	90	0.0	80	10.2
100	110	60.2	90	42.5
110	110	0.0	130	72.2
120	140	45.1	130	21.5
130	130	0.0	130	0.0
140	140	0.0	130	27.5
150	160	19.3	150	0.0
160	160	0.0	150	19.3

setting, and the cable length between the actual fault bus k' and the calculated bus \bar{k} . As the results vary from simulation to simulation, the cable length between k' and \bar{k} is a good indicator of how far the calculated point of fault \bar{k} is from the actual point of fault k' .

V. DISCUSSION

Overall, the methodology is effective to identify the area where permanent faults occurred, as well as the grid configuration and the fault type. On the top of this, the calculation times and storage space requirements show the feasibility of an online implementation of the method for detection and location of different faults in a small LV network.

In the event of a fault, the methodology identifies the grid configuration given by the status of switches and the part of the network where the fault occurred correctly in all cases, regardless of the fault type, location, and impedance. Moreover, the calculated location of the fault is accurate in most cases, as the results displayed in Tables I to VIII represent the exact point of fault or another bus in the same branch of the network in most cases.

TABLE III

RESULTS: PHASE-GROUND FAULT, SWITCH STATUS OFF

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	10	0.0	20	20.6
20	10	20.6	20	0.0
30	40	3.8	40	3.8
40	40	0.0	50	75.4
50	50	0.0	50	0.0
60	40	87.5	50	68.0
70	50	90.9	50	90.9
80	80	0.0	80	0.0
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	160	24.7	100	60.2
120	140	45.1	100	153.3
130	120	21.5	100	135.7
140	140	0.0	100	159.2
150	160	19.3	100	96.0
160	160	0.0	100	81.1

TABLE IV

RESULTS: PHASE-GROUND FAULT, SWITCH STATUS ON

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	10	0.0	20	20.6
20	20	0.0	20	0.0
30	40	3.8	30	0.0
40	40	0.0	50	75.4
50	50	0.0	40	75.4
60	50	68.0	60	0.0
70	70	0.0	50	90.9
80	90	10.2	90	10.2
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	140	96.9	110	0.0
120	160	111.8	140	45.1
130	160	94.2	140	27.5
140	160	117.8	140	0.0
150	160	19.3	110	39.7
160	110	24.7	140	117.8

TABLE V

RESULTS: PHASE-PHASE-GROUND FAULT, SWITCH STATUS OFF

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	20	20.6	20	20.6
20	20	0.0	10	20.6
30	60	87.5	10	58.5
40	50	75.4	10	58.5
50	50	0.0	50	0.0
60	50	68.0	60	0.0
70	50	90.9	50	90.9
80	80	0.0	80	0.0
90	80	10.2	80	10.2
100	100	0.0	100	0.0
110	120	96.9	100	60.2
120	140	45.1	120	0.0
130	140	27.5	140	27.5
140	140	0.0	140	0.0
150	140	132.7	120	126.7
160	140	117.8	120	111.8

TABLE VI
RESULTS: PHASE-PHASE-GROUND FAULT, SWITCH STATUS ON

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	20	20.6	20	20.6
20	20	0.0	20	0.0
30	30	0.0	50	75.4
40	30	3.8	50	75.4
50	40	75.4	60	68.0
60	60	0.0	20	141.2
70	40	110.5	50	90.9
80	90	10.2	90	10.2
90	90	0.0	90	0.0
100	100	0.0	100	0.0
110	100	60.2	100	60.2
120	100	153.3	100	153.3
130	100	135.7	100	135.7
140	100	159.2	140	0.0
150	100	96.0	150	0.0
160	100	81.1	100	81.1

TABLE VII

RESULTS: PHASE-PHASE FAULT, SWITCH STATUS OFF

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	10	0.0	10	0.0
20	10	20.6	10	20.6
30	40	3.8	70	110.5
40	40	0.0	70	110.5
50	60	68.0	50	0.0
60	40	87.5	10	142.2
70	50	90.9	10	155.2
80	80	0.0	120	194.3
90	80	10.2	120	188.3
100	110	60.2	120	153.3
110	110	0.0	120	90.9
120	120	0.0	120	0.0
130	120	21.5	120	21.5
140	110	96.9	140	0.0
150	160	19.3	160	19.3
160	150	19.3	160	0.0

TABLE VIII

RESULTS: PHASE-PHASE FAULT, SWITCH STATUS ON

Faulty bus (k')	Results $R_F = 0.1\Omega$		Results $R_F = 1.0\Omega$	
	Bus (\bar{k})	Cable length [m]	Bus (\bar{k})	Cable length [m]
10	20	20.6	10	0.0
20	20	0.0	10	20.6
30	60	87.6	10	58.5
40	60	87.6	10	58.5
50	60	68.0	50	0.0
60	70	103.0	60	0.0
70	60	103.0	50	90.9
80	80	0.0	80	0.0
90	80	0.0	80	10.2
100	100	0.0	150	96.0
110	120	96.9	160	24.7
120	120	0.0	120	0.0
130	120	21.5	130	0.0
140	120	45.1	120	45.1
150	150	0.0	150	0.0
160	150	19.3	150	19.3

However, the results indicate diverse effects on the fault location errors with an increase in the fault resistance. In many cases, the errors remain roughly the same; in others, however, they increase or decrease considerably. This is attributed to the small impedance of the line sections (few $m\Omega$), to the statistical models of the network (main features) under different operating conditions, and also to the loadings of the variables involved in the fault in the principal components subspace (not negligible in a few reference scenarios).

Nonetheless, it is noteworthy that the results calculated with (5) may be inaccurate over a range of scenarios due to the differences between the training and testing scenarios in use, such as variations in the standard and short-circuit operating conditions, different faults, and inaccurate representations of the network provided by the statistical models in use. Therefore, the method could be further improved by including more reference scenarios in the training data sets, with different faults, timescales, operating conditions, more measurement devices at different locations, other pieces of information about the operating conditions, etc.

VI. CONCLUSIONS

The methodology is capable of locating and classifying different types of faults with good accuracy and also identifies the correct grid configuration in the presence of switches. The identification of the correct grid setting and fault type as a previous step to fault location improves the accuracy of the method, as it limits the search to the right part of the network in a reduced set of reference scenarios. As a result, the errors lie in the same part of the network in all tested scenarios. Nonetheless, it may be inaccurate over a range of scenarios due to the differences between training and testing scenarios.

REFERENCES

- [1] A. Estebansari, E. Pons, E. Bompard, A. Bahmanyar, and S. Jamali, "An improved fault location method for distribution networks exploiting emerging lv smart meters," in *2016 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*, June 2016, pp. 1–6.
- [2] Y. Jiang, "Data driven fault location of electric power distribution systems with distributed generation," *IEEE Transactions on Smart Grid*, pp. 1–1, 2019.
- [3] M. Kezunovic, "Smart fault location for smart grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 1, pp. 11–22, March 2011.
- [4] J. Ren, S. S. Venkata, and E. Sortomme, "An accurate synchrophasor based fault location method for emerging distribution systems," *IEEE Transactions on Power Delivery*, vol. 29, no. 1, pp. 297–298, Feb 2014.
- [5] S. E. McHann, "Grid analytics: How much data do you really need?" in *2013 IEEE Rural Electric Power Conference (REPC)*, April 2013, pp. C3–1–C3–4.
- [6] Z. Yin, G. Sun, Z. Wei, H. Zang, J. Ding, F. Lu, and H. Yuan, "The research of fault section location for distribution network based on transient feature extraction and fuzzy clustering," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, May 2018, pp. 1279–1283.
- [7] E. A. Reche, J. V. d. Sousa, D. V. Coury, and R. A. S. Fernandes, "Data mining-based method to reduce multiple estimation for fault location in radial distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3612–3619, July 2019.
- [8] L. Souto, J. Meléndez, and S. Herraiz, "Fault location in power distribution networks based on similarity criteria in the principal component subspace," in *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2020, pp. 1–5.