A Bayesian Approach for Fault Location in Medium Voltage Grids With Underground Cables

YU XIANG[1] (Student Member, IEEE) AND JOSEPH F. G. COBBEN[1,2] (Member, IEEE)

[1]Electrical Energy Systems Group, Eindhoven University of Technology, Eindhoven 5612 AZ, The Netherlands
[2]Alliander, Arnhem 6812 AH, The Netherlands
CORRESPONDING AUTHOR: Y. XIANG (y.xiang@tue.nl)

ABSTRACT This paper proposes a statistical approach for section-based fault location in medium voltage (MV) grids with underground cables, using Bayesian inference. The proposed approach considers several important uncertainties in the MV grid, including measurement errors, fault breakdown resistance, and the inaccuracies of zero-sequence parameters. The approach first obtains the prior distribution of the fault position from the component failure database, the readings of the transmitted fault indicators, and the relevant digging activity record. With the estimated prefault grid status and the measured transient voltages/currents, the posterior distribution is then calculated based on Bayes' theorem. To solve the problem numerically, the Monte Carlo integration is applied and a two-step calculation procedure is proposed. Simulations are performed on a typical European MV feeder to demonstrate the feasibility of the approach. The distribution grid operators can use the calculated posterior distribution to rank the possible faulted sections and to facilitate the restoration process, which can reduce the interruption duration of power supply.

INDEX TERMS Bayesian inference, Monte Carlo integration, section-based fault location, transient measurements, underground cables.

I. INTRODUCTION

WHEN a fault occurs in a medium voltage (MV) distribution grid, the protection relay will clear the fault and some region will experience power outage. To restore the power supply, the grid operator needs to first locate and isolate the faulted section between two MV points of connection (PoCs). Because the cables in the MV grids are usually buried underground, traditionally, the grid operator relies on the readings of the fault indicators along the faulted feeder. Since the readings need to be retrieved locally, the fault location is the most time-consuming task during the restoration process [1].

Several studies aimed to locate the fault efficiently and accurately. The most common method is to calculate the fault loop impedance using the measured data of transient currents and voltages, and to compare it with the preknown grid model [2]–[4]. It was suggested that only the reactance part is considered due to the presence of the fault breakdown resistance. The limitation of this type of method

is the difficulty to cope with the uncertainties in the grids. Zhu *et al.* [5] analyzed the influence of the uncertainties and developed an approach to estimate the fault position range. Mora-Flòrez *et al.* [6] compared the performance index of different impedance-based fault location methods.

Other than the impedance-based methods, Choi *et al.* [7], [8] presented a method based on a direct circuit analysis, and enhanced its robustness to load impedance uncertainty using impedance compensation with voltage and current measurements. Moreover, Nouri *et al.* [9] and Thomas *et al.* [10] approached the problem by applying wavelet transform. It has been proposed to measure the time difference between the the traveling waves during the fault. To compare different methods, Lotfifard *et al.* [11] developed a systematical approach to rank the fault location methods qualitatively and quantitatively.

Besides the deterministic methods, statistical methods have also been proposed to properly consider the uncertainties. Cormane *et al.* [12] presented an approach

using data clustering. Mora-Flórez *et al.* [13] developed a statistical classification algorithm based on fuzzy probability functions to locate single-phase faults. Chien *et al.* [14] adopted a Bayesian network on the basis of expert knowledge and historical data for fault diagnosis. The Bayesian network imitates the causal relations between the fault equipment and the evidences of observations, and gives the possibilities that the failure is on each equipment.

Due to the high degree of uncertainties in the distribution grid, the statistical methods proposed by previous studies managed to solve the fault location problem effectively. However, the distribution grid operators desire an approach that considers the specific characteristics of the MV grids with underground cables. To bridge this gap, in [15], we have introduced an impedance-based statistical method for fault location in simple situations. This paper takes a step further, and presents a fault location approach with Bayesian inference based on the measured transient electrical quantities during the fault. It considers the typical uncertainties in the MV grids, including the measurement errors, the influence of loads, the dispersed generations (DGs), and the inaccuracies in the zero-sequence parameters of the cables. The proposed procedure calculates a ranking of possible faulted sections, which can be used by the grid operators to facilitate the power restoration process.

Section II defines the problem of section-based fault location and introduces the proposed procedure. In Section III, the prior distribution of the fault location is calculated considering the component failure, the transmitted fault indicators (TFIs), and the digging activities. Section IV defines the uncertainties as a grid condition vector, and calculates the posterior distribution. In Section V, the numerical issues are addressed to guarantee that the calculation can be finished within a reasonable time. To demonstrate the feasibility of the approach, a case study on a sample grid is given in Section VI.

II. PROBLEM DEFINITION
A. SECTION-BASED AND POSITION-BASED FAULT LOCATION

MV distribution grids are usually constructed as a ring or meshed structure, and operated in a radial structure [16]. In case of a fault, the grid operators will isolate the faulted section, and reconfigure the grid to restore the power supply [1]. A reparation activity will be planned to further examine and repair/replace the faulted cable section.

During the outage restoring procedure, the most time-consuming task is to find the faulted section along the feeder. Therefore, instead of locating the exact fault position, the speed and the accuracy of the section-based fault location are more of interest for the grid operators. Therefore, this paper focuses on section-based fault location, and develops a procedure to give a ranking of possible faulted sections, which can facilitate the restoring process and minimize the interruption duration. Due to the time constraint, the procedure has to rely on the online measured transient data, e.g., transient fault currents/voltages and/or fault indicator readings.

The position-based fault location is to locate the exact fault position during the reparation of the (already isolated and de-energized) faulted cable section. Without a strict time limit, more sophisticated offline diagnostic methods are usually used to locate the exact fault position. In this case, the (posterior) distribution of the fault position resulted from this paper's approach can give an extra indication besides the offline methods.

B. PROCEDURE DESIGN

The fault location procedure designed in this paper is shown in Fig. 1. It involves: 1) the preprocessing to acquire the prior distribution of the fault position; 2) the application of Bayes' theorem to calculate the posterior distribution; and 3) the postprocessing to obtain a ranking of possible faulted sections.
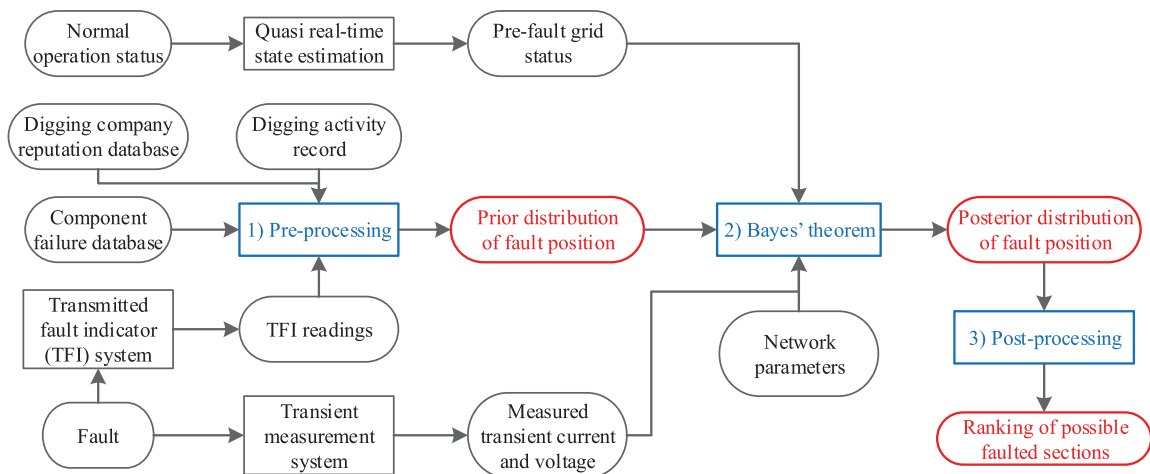

FIGURE 1. Designed fault location procedure.

In the preprocessing, the prior (marginal) distribution of the fault position is acquired from various input data. First, the component failure database determines the basic fault distribution due to normal component failure. In addition, the digging activities during the fault and the relevant digging company reputation should be considered. Moreover, if the TFIs are present at the faulted feeder, the range of the possible fault position can be shortened by the TFI readings.

At the moment before the fault occurs, the prefault grid status can be obtained from the quasi real-time state estimation, which has been presented in [17]–[20]. During the fault, the transient short circuit currents and voltages will be measured and recorded. Given the measured transients, the prefault grid status, and the network parameters, the posterior (conditional) distribution is calculated based on Bayes' theorem.

In the postprocessing, the probability that this fault is on each section is obtained, and a ranking of possible faulted sections is available for the grid operator to facilitate the outage restoring process.

III. PRIOR DISTRIBUTION OF FAULT POSITION

In this section, the calculation of the prior distribution is discussed. The basic distribution of the fault position is first calculated with the component failure rate, followed by the consideration of the digging activity and the TFI system.

A. BASIC DISTRIBUTION WITH MIXED RANDOM VARIABLE

From the component failure rate database, the basic distribution of a potential fault position can be calculated. Although the Weibull distribution is more suitable for the modeling of component aging, practical experience suggests that only minority of the faults are caused by aging [21]. Therefore, this paper adopts the exponential distribution (which is memoryless and has a constant failure rate) to model the failure of components.
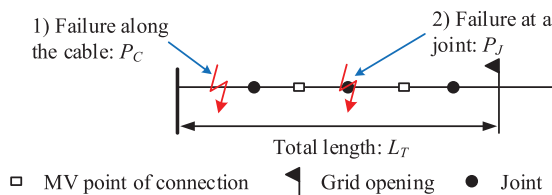

FIGURE 2. Fault: 1) along the cable or 2) at a joint.

If a fault occurs along an MV feeder due to component failure, as shown in Fig. 2, the breakdown can take place either: 1) along the cable or 2) at a joint position [15]. The probability of the fault being at a joint position P_J and along the cable P_C is calculated by

$$P_J = \frac{N_J \lambda_J}{L_T \lambda'_C + N_J \lambda_J}, \quad P_C = 1 - P_J \quad (1)$$

where

L_T total length of the cable;
λ'_C failure rate per length of the cable;
N_J total number of joints;
λ_J failure rate of each joint.

The potential fault position L is a mixed random variable, since it has a discrete part at the joint positions and a continuous part along the cable. Thus, this paper uses both the probability mass function (PMF) $f_{L,M}(l)$ and the probability density function (pdf) $f_{L,D}(l)$ to represent its distribution. It is reasonable to assume that the basic prior distribution is uniform, and is shown as

$$f_{L,M}(l) = \begin{cases} P_J/N_J, & l \in S_J \\ 0, & other \end{cases} \quad (2)$$

$$f_{L,D}(l) = \begin{cases} P_C/L_T, & l \in S_L \backslash S_J \\ 0, & other \end{cases} \quad (3)$$

where

S_J set of all joint positions;
$S_L = [0, L_T]$ interval of fault position along the cable.

B. DIGGING ACTIVITY

Besides the faults caused by the component failure, a large number of faults in the MV grids with the underground cables are caused by the incautious digging activities damaging the cable. For each country, a governmental organization usually manages a digging company reputation database, from which a (equivalent) failure rate of digging damage by a certain digging company can be obtained.

If (according to the record) a digging activity is present during the fault, as shown in Fig. 3, this information should be considered. In this case, the probability that the fault is caused by the digging damage P_D and by the component failure P_B can be calculated by

$$P_D = \frac{L_D \lambda'_D}{L_D \lambda'_D + L_T \lambda'_C + N_J \lambda_J}, \quad P_B = 1 - P_D \quad (4)$$

where

L_D length of the cable involved in digging;
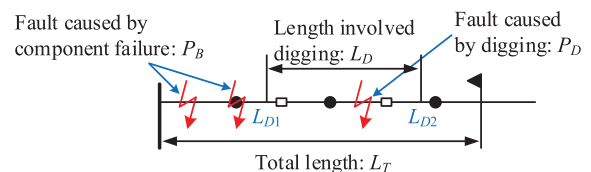λ'_D failure rate per length for the digging company.

FIGURE 3. Fault position involving digging activity.

Given that the fault is caused by the digging damage, the pdf of the fault position is calculated by (5). It should be noted that there is no discrete part (PMF) for the digging

fault position, since a continuous section of the cable is involved. Finally, combining the component failure and the digging damage, the resulted PMF and pdf can be calculated in (6) and (7)

$$f_{L,D}(l|\text{dig}) = \begin{cases} 1/L_D, & l \in [L_{D1}, L_{D2}] \\ 0, & \text{other} \end{cases} \quad (5)$$

where $[L_{D1}, L_{D2}]$ is the interval of the cable section involved in digging

$$f_{L,M,\text{dig}}(l) = P_B f_{L,M}(l) \quad (6)$$
$$f_{L,D,\text{dig}}(l) = P_B f_{L,D}(l) + P_D f_{L,D}(l|\text{dig}). \quad (7)$$

C. TRANSMITTED FAULT INDICATOR

Traditionally, grid operators rely on fault indicators to locate the faulted section. A fault indicator trips if a large (short circuit) current has flowed through it. Usually, this reading is only available locally, while a limited number of fault indicators are equipped with communication devices, so that the readings can be retrieved remotely from the control center.

If the TFIs are present along the faulted feeder, the TFI readings can limit the range of the fault location and, therefore, influence the prior distribution. As shown in Fig. 4, N_{TFI} TFIs divide the whole feeder into $N_{\text{TFI}} + 1$ subintervals: $S_{\text{FI},i}, i = 1 \sim N_{\text{TFI}} + 1$. If the TFI readings indicate that the fault is within the subinterval $S_{\text{FI},i}$, the range of the fault position is shortened, and the PMF and the pdf should be modified as in (8) and (9). In (8) and (9), $f_{L,M}$ and $f_{L,D}$ should be replaced by $f_{L,M,\text{dig}}$ and $f_{L,D,\text{dig}}$ if a digging activity is present during the fault

$$f_{L,M,\text{FI},i}(l) = \begin{cases} f_{L,M}(l)/P_{\text{FI},i}, & l \in S_{\text{FI},i} \\ 0, & \text{other} \end{cases} \quad (8)$$

$$f_{L,D,\text{FI},i}(l) = \begin{cases} f_{L,D}(l)/P_{\text{FI},i}, & l \in S_{\text{FI},i} \\ 0, & \text{other} \end{cases} \quad (9)$$

where $P_{\text{FI},i}$ is the marginal probability that the fault is within the subinterval $S_{\text{FI},i}$, and is calculated as follows:

$$P_{\text{FI},i} = \sum_{l \in S_{\text{FI},i} \bigcap S_J} f_{L,M}(l) + \int_{S_{\text{FI},i}} f_{L,D}(l)\mathrm{d}l. \quad (10)$$
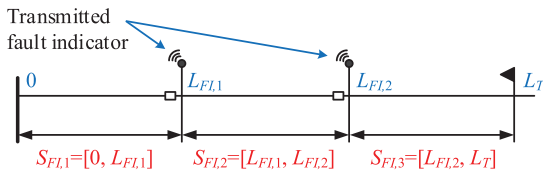

FIGURE 4. Subintervals divided by TFIs.

If during the fault there are DGs connected along the faulted feeder, the fault current contribution from the DGs can trigger the TFIs from the opposite direction, and the TFI readings can no longer indicate the faulted subinterval. Therefore, the TFI readings should be discarded in those cases.

IV. POSTERIOR DISTRIBUTION OF FAULT POSITION AND RANKING OF POSSIBLE FAULTED SECTIONS

Given the measured fault transient currents/voltages and the estimated prefault grid status, the posterior distribution of the fault position is calculated in this section. The measurement vector and the grid condition vector are first defined, followed by the application of Bayes' theorem.

A. MEASUREMENT VECTOR

This paper defines the vector M as the real values of the measured quantities applied in the analysis. The number of the dimensions of M is $9 \times N_{\text{mea}}$, where N_{mea} is the number of points along the faulted feeder with transient measurements. Usually, meters are installed at the beginning of the feeder, while sometimes, there are additional measured points along the feeder. The composition of M for single measured points is shown in (11), while that for multiple measured points is the duplication of the corresponding components

$$M = [U_{a,b,c} \quad I_{a,b,c} \quad \varphi_\alpha]^T \quad (11)$$

where

$U_{a,b,c}$	rms values of line-to-ground three-phase voltages;
$I_{a,b,c}$	rms values of three-phase currents;
φ_α	phase angle vector;
α	indication on the type of the fault: 0 for three-phase fault, 1 for single-phase fault, 2 for two-phase fault, and 3 for two-phase-to-ground fault.

The phase angle vector φ_α varies for different types of the faults, depending on how the fault loop impedance is calculated [4]. Its composition for three-phase, single-phase (phase A), and two-phase (phases B and C) faults are shown in (12)–(14), respectively

$$\varphi_0 = [\varphi(\underline{I}_a/\underline{U}_a) \quad \varphi(\underline{I}_b/\underline{U}_b) \quad \varphi(\underline{I}_c/\underline{U}_c)] \quad (12)$$
$$\varphi_1 = [\varphi(\underline{I}_a/\underline{U}_a) \quad \varphi(\underline{I}_b/\underline{U}_a) \quad \varphi(\underline{I}_c/\underline{U}_a)] \quad (13)$$
$$\varphi_{2,3} = [\varphi(\underline{U}_c/\underline{U}_b) \quad \varphi(\underline{I}_b/\underline{U}_b) \quad \varphi(\underline{I}_c/\underline{U}_b)] \quad (14)$$

where $\varphi(\underline{X}/\underline{Y})$ represents the angle difference between the phasors $\underline{X}$ and $\underline{Y}$.

Instead of the real values of the measured quantities, the meter readings only give the measured values (which include errors). Thus, the vector $\widetilde{M}$ is defined as the corresponding measured values of M. It should be noted that the measurement errors of the transient currents and voltages are usually larger than those of the steady states, because: 1) the transient processes are involved; 2) the higher current measuring ranges are needed for large fault current; and 3) the current transformers are inclined to be saturated. Nevertheless, to mitigate the influence of the transient processes, it has been suggested to pick up the (current/voltage) values at the moment before the fault is cleared, when the waveforms are the steadiest [1].

B. GRID CONDITION

Besides the fault position, a number of other factors have influence on the measurement vector $\boldsymbol{M}$. To consider these uncertainties, the random variable grid condition vector $\boldsymbol{G}_c$ is defined. It consists of five independent components shown in (15). The sample space of $\boldsymbol{G}_c$ is defined as $\Omega_{\boldsymbol{G}_c}$, and its joint pdf is designated as $f_{\boldsymbol{G}_c}(\boldsymbol{g}_c)$

$$\boldsymbol{G}_c = [R_f \quad U_g \quad \boldsymbol{S}_{\text{Load}} \quad \boldsymbol{E}'_{\text{DG}} \quad \boldsymbol{Z}'_0]^T. \tag{15}$$

R_f represents the fault breakdown impedance. It only has a resistive part, because the reactance of the fault is usually zero [3]. It can be assumed that R_f follows a uniform distribution between 0 and $R_{f,\max}$. $R_{f,\max}$ is the upper bound of the breakdown resistance.

U_g, $\boldsymbol{S}_{\text{Load}}$, and $\boldsymbol{E}'_{\text{DG}}$ represent the prefault grid status, which is obtained through the state estimation. In particular, U_g is the voltage level of the equivalent external grid, $\boldsymbol{S}_{\text{Load}}$ is the load vector in the MV grid, and $\boldsymbol{E}'_{\text{DG}}$ is the vector of transient excitation voltage of the DGs with fault ride through capability. To ensure the numerical stability, with a given known estimated prefault grid status, the errors of the state estimation (and the real grid status) can be assumed to follow the truncated normal distribution.

$\boldsymbol{Z}'_0$ represents the inaccuracies in the zero-sequence parameters of the cables. Similarly, given a measured zero-sequence impedance per length of the cable, the real $\boldsymbol{Z}'_0$ follows the truncated normal distribution. Note that this component can be omitted for nonground faults.

C. CONDITIONAL DISTRIBUTION OF MEASURED VALUE

With L and $\boldsymbol{G}_c$, the real values of the measured quantity $\boldsymbol{M}$ can be explicitly calculated (using common short circuit calculation methods), designated with $h_\alpha(\cdot, \cdot)$, and is shown as

$$\boldsymbol{M} = h_\alpha(L, \boldsymbol{G}_c). \tag{16}$$

Given a particular known fault position l and the grid condition $\boldsymbol{g}_c$, as well as the calculated real-value vector of the measured quantities $\boldsymbol{m}$, it is common practice to assume that the measured value vector $\widetilde{\boldsymbol{M}}$ conditionally follows the normal distribution. The conditional pdf of $\widetilde{\boldsymbol{M}}$ is calculated in (17), where the covariance matrix represents the measurement error

$$f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}|l, \boldsymbol{g}_c) = \mathfrak{N}(\widetilde{\boldsymbol{m}}|\boldsymbol{m}, \boldsymbol{\Sigma}_e) \tag{17}$$

where

$\mathfrak{N}(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pdf of the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$;
$\boldsymbol{\Sigma}_e$ covariance matrix of the measurement error.

D. POSTERIOR DISTRIBUTION OF FAULT POSITION

After measuring the transient voltages and currents during a fault, a particular instance of the measured value vector is obtained as $\widetilde{\boldsymbol{m}}^*$. To consider the influence of different grid conditions, the probability density of $\widetilde{\boldsymbol{m}}^*$ under the condition of given l can be calculated through the integral with the sample space of the grid conditions, and is shown as

$$f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l) = \int_{\Omega_{\boldsymbol{G}_c}} f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l, \boldsymbol{g}_c) f_{\boldsymbol{G}_c}(\boldsymbol{g}_c) \mathrm{d}\boldsymbol{g}_c. \tag{18}$$

Since L is a mixed random variable, the calculation of the marginal probability density of $\widetilde{\boldsymbol{m}}^*$ includes a summation and an integral, using the law of total probability, and is shown as

$$f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*) = \sum_{l \in S_J} f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l) f_{L,M}(l) + \int_{S_L} f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l) f_{L,D}(l) \mathrm{d}l. \tag{19}$$

Finally, based on Bayes' theorem, the posterior conditional PMF and pdf of the fault position L given the measured value vector $\widetilde{\boldsymbol{m}}^*$ can be calculated by (20) and (21), respectively

$$f_{L,M}(l|\widetilde{\boldsymbol{m}}^*) = \frac{f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l) f_{L,M}(l)}{f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*)} \tag{20}$$

$$f_{L,D}(l|\widetilde{\boldsymbol{m}}^*) = \frac{f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l) f_{L,D}(l)}{f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*)}. \tag{21}$$

In (19)–(21), $f_{L,M}$ and $f_{L,D}$ should be replaced by $f_{L,M,\text{dig}}$ and $f_{L,D,\text{dig}}$ if a digging activity is present during the fault. Similarly, they should be replaced by $f_{L,M,\text{FI},i}$ and $f_{L,D,\text{FI},i}$ when the TFI readings are available and indicate that the fault is in the subinterval i.

E. RANKING OF POSSIBLE FAULTED SECTIONS

With the posterior distribution of the fault position, the probability that the fault is in a particular section can be calculated by (22). By calculating the probabilities for each cable section, a ranking of possible faulted sections can be obtained, which can be used by the grid operator to facilitate the restoring process and to reduce the outage time

$$P_i = \sum_{l \in S_i \bigcap S_J} f_{L,M}(l|\widetilde{\boldsymbol{m}}^*) + \int_{S_i} f_{L,D}(l|\widetilde{\boldsymbol{m}}^*) \mathrm{d}l \tag{22}$$

where

P_i probability that the fault is in section i;
S_i position range of section i.

V. NUMERICAL CONCERNS

The analytic mathematical principle of the approach has been introduced in Sections III and IV. However, in practice, the grid operator needs to calculate the aforementioned formulas numerically. To reduce the outage duration, the calculation should be finished within a strict time constraint, which is defined as less than 5 min in this paper. The numerical complexity is, in general, one of the main difficulties when applying Bayesian inference [22]. To ensure that the proposed fault location procedure is practically feasible, this section discusses the related numerical concerns.

A. MONTE CARLO INTEGRATION WITH INDEPENDENT IMPORTANCE SAMPLING

Given the measured value vector $\widetilde{\boldsymbol{m}}^*$, for each particular sampling l^* of the fault position L, the integral in (18) has to be calculated to obtain the corresponding conditional pdf $f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*)$. The number of the dimensions of the integral is the same as the grid condition vector $\boldsymbol{G}_c$, and is shown as

$$\dim(\boldsymbol{G}_c) = 2 + 2N_{\text{Load}} + 2N_{\text{DG}} + 2N_{\text{CableType}} \quad (23)$$

where

N_{Load}	number of MV loads;
N_{DG}	number of MV DGs;
$N_{\text{CableType}}$	number of cable types in the faulted feeder (this item can be omitted for nonground faults).

It is usually not viable to evaluate this high-dimensional integral using the deterministic methods. Therefore, this paper applies the Monte Carlo integration with independent importance sampling [22]. Considering the composition of the integrand, the pdf $p(\boldsymbol{g}_c)$ of the sampling distribution is chosen to be the same as the joint distribution of the grid condition vector, i.e., $p(\boldsymbol{g}_c) = f_{\boldsymbol{G}_c}(\boldsymbol{g}_c)$. Because the components in $\boldsymbol{G}_c$ are independent to each other, it would be feasible to sample the components individually and to combine them together. Two types of sampling are involved: 1) uniform and 2) truncated normal distribution, where the latter needs special sampling techniques [23], [24].

With N_{MCI} random samples generated, shown in (24), the numerical approximation of the integral in (18) can be calculated by (25). With the principles of the Monte Carlo integration, the error of the approximation decreases as $1/\sqrt{N_{\text{MCI}}}$, and does not depend on the number of the dimensions

$$p(\boldsymbol{g}_c) \xrightarrow{\text{sampling}} \boldsymbol{g}_c^{(1)}, \dots, \boldsymbol{g}_c^{(N_{\text{MCI}})} \in \Omega_{\boldsymbol{G}_c} \quad (24)$$

$$f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*) \approx \frac{1}{N_{\text{MCI}}} \sum_{k=1}^{N_{\text{MCI}}} \frac{f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*, \boldsymbol{g}_c^{(k)}) f_{\boldsymbol{G}_c}(\boldsymbol{g}_c^{(k)})}{p(\boldsymbol{g}_c^{(k)})}$$

$$= \frac{1}{N_{\text{MCI}}} \sum_{k=1}^{N_{\text{MCI}}} f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*, \boldsymbol{g}_c^{(k)}). \quad (25)$$

B. SAMPLING OF FAULT POSITION AND TWO-STEP CALCULATION PROCEDURE

For the accuracy and the smoothness of the posterior distribution, it is preferred to calculate its values at the densely sampled points of the fault position L. On the other hand, for each sampled point l^*, the time consuming integral in (18) has to be evaluated. Thus, a proper sampling strategy for L is essential for the numerical approach, where two important aspects need to be considered.

1) The joint positions and the edges of the cable section involved in the digging activity (if present) will lead to discontinuous or nondifferentiable points in the cumulative distribution function. Thus, these positions are defined as critical points in this paper, and should be included in the sampling of L.

2) The posterior PMF and pdf only have significant values within a limited interval near the real fault position, while their values at other positions are (near to) zero. Therefore, it would be inefficient to sample L homogeneously and to calculate the corresponding Monte Carlo integration with the same number of points.

This paper proposes a two-step calculation procedure. First, the rough posterior distribution is obtained using lower calculation intensity, followed by a finer calculation within the 95% confidence interval of L, where the PMF and the pdf have significant values. The detailed procedure is shown below:

1) sample L homogeneously with step 0.1 km, and include all the critical points, with N_{L1} points in total;
2) given each sampled l^*, calculate $f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*)$, using $N_{\text{MCI},1}$ points in the integral;
3) calculate the rough posterior PMF and pdf of the fault position, and identify the 95% confidence interval;
4) resample L (more densely) in the 95% confidence interval (also including all the critical points inside the interval) with N_{L2} points;
5) given each resampled l^*, calculate $f_{\widetilde{\boldsymbol{M}}}(\widetilde{\boldsymbol{m}}^*|l^*)$, using $N_{\text{MCI},2}$ (higher than $N_{\text{MCI},1}$) points in the integral;
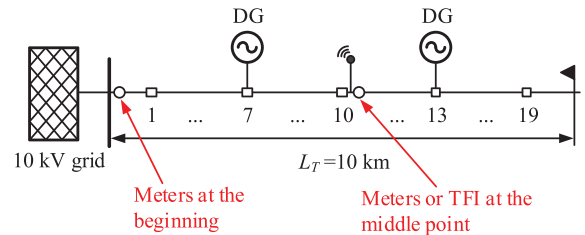6) combine the results from Steps 2 and 5, and calculate the final posterior PMF and pdf.

FIGURE 5. Sample 10-kV grid with single feeder.

VI. CASE STUDY ON SAMPLE GRID

A. SIMULATION PARAMETERS

A case study is performed on a 10-kV sample grid with one MV feeder to demonstrate the feasibility of the fault location method. As shown in Fig. 5, the feeder has a 10-km cable and 38 joints. There are 19 MV PoCs, among which PoCs 7 and 13 are (optionally) connected with DGs. Other important parameters are shown below.

1) Short circuit capacity of a 10-kV grid: 325 MVA.
2) Cable impedance per length:
   a) $\underline{Z}'_1 = \underline{Z}'_2 = (0.126 + j0.109) \ \Omega \cdot \text{km}^{-1}$;
   b) $\underline{Z}'_0 = (0.83 + j0.05) \pm 10\% \ \Omega \cdot \text{km}^{-1}$.
3) Failure rate:
   a) cable: $\lambda'_C = 0.025 \ \text{year}^{-1} \cdot \text{km}^{-1}$;
   b) joint: $\lambda_J = 0.007 \ \text{year}^{-1}$;
   c) digging: $\lambda'_D = 0.25 \ \text{year}^{-1} \cdot \text{km}^{-1}$.
4) Fault breakdown resistance: $R_{f,\max} = 0.1 \ \Omega$.

5) Voltage meter, measuring range: 8.66 kV, error: 3%.
6) Current meter (two level):
   a) measuring range 1: 1 kA, error 1: 3%;
   b) measuring range 2: 15 kA, error 2: 5%.
7) Phase angle meter, error: 10°.
8) Two-step sampling and Monte Carlo integral:
   a) $N_{L1} \approx 135$, $N_{MCI,1} = 1 \times 10^4$;
   b) $N_{L2} = 120$, $N_{MCI,2} = 1 \times 10^6$.

Three measurement scenarios are considered: 1) Scenario 1 is the basic scenario where meters are only installed at the beginning of the feeder; 2) in Scenario 2, there is a second measured point at PoC 10, which is 5 km from the MV busbar; and 3) in Scenario 3, a TFI (instead of meters) is available at PoC 10. As described in Section III, Scenario 3 is only applicable for the situations without DGs.

In total, 100 random (fault) cases are simulated for three-phase fault and single-phase fault, wherein half of the cases, the digging activities are involved. As a reference, the fault position is first calculated with the X method (the method purely using fault-loop reactance). And the ranking of the possible faulted sections are calculated with Bayesian inference. The results from both the methods are compared with the real faulted section.

B. THREE-PHASE FAULT
First, 100 simulation cases with three-phase fault are executed, with and without DGs. Fig. 6 compares the count of cases that the calculated faulted section matches the real faulted section, between the X and Bayesian methods, with different measurement scenarios.
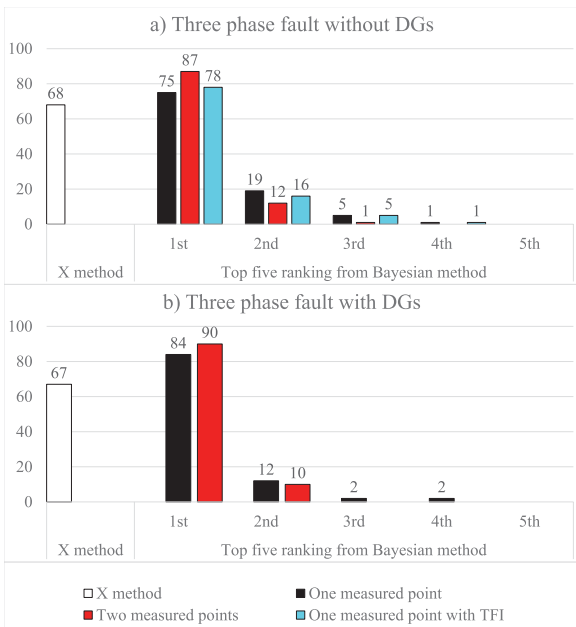

FIGURE 6. Count of cases that the calculated (ranking of) faulted section matches the real faulted section, for three-phase fault (a) without and (b) with DGs.

Regardless of the measurement scenario or the presence of DGs, the first rank of the calculated faulted sections from

the Bayesian method matches the real faulted section for at least 75% of all the cases, which is higher than the possibility of match for the X method. Moreover, the real faulted section always falls into the top four calculated sections from the Bayesian method. Therefore, the Bayesian method has a moderate improvement over the X method.

As for the impact of different measurement scenarios, one measured point can already achieve reasonably high accuracy. The additional measured point can improve the results slightly, while the TFI in the middle point provides little benefit. Furthermore, the results in the situations with DGs is better than those without DGs, because the prefault status (transient excitation voltage) of the DGs are known to the proposed method. This additional information (of the extra DG voltage sources apart from the equivalent external grid) can improve the accuracy of the algorithm in the Bayesian approach.

C. SINGLE-PHASE FAULT
Fig. 7 shows the results for single-phase fault. In the MV grid with earthed neutral, the accuracy of the Bayesian method is
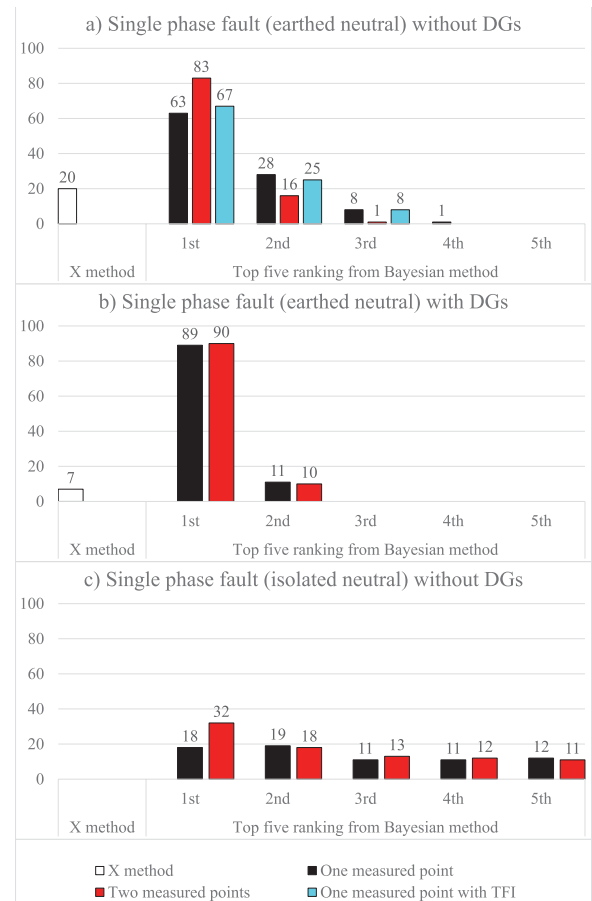

FIGURE 7. Count of cases that the calculated (ranking of) faulted section matches the real faulted section, for single-phase fault, in the MV grid (a) with earthed neutral and without DGs, (b) with earthed neutral and with DGs, and (c) with isolated neutral and without DGs.

significantly higher than the X method, especially when the DGs are present. In addition, in the cases without DGs, the scenario with two measured points has a considerable benefit for locating single-phase fault. Besides, the simulations are performed in the MV grid with isolated neutral, where the X method is not applicable anymore. The Bayesian method can still locate the faulted section within the top five ranking for at least 70% of the cases. In addition, in this situation, the second measured point has reasonable benefit.

The Bayesian method can be especially beneficial for single-phase fault (which is the most common fault in the MV grids), as the inaccuracies of the zero-sequence parameters and the errors of the phase angle measurement deteriorate the results of the X method. Using the ranking of the possible faulted sections, the grid operator can facilitate the fault location and reduce the outage time.

D. CALCULATION TIME

The simulation has been executed on a workstation with a dual-CPU (Intel Xeon E5-2680 v3 [25]) configuration and 128-GB RAM. The algorithm is programmed in Mathworks MATLAB with 24-thread parallel computing. The average calculation time of the Bayesian method is shown in Table 1. Obviously, the calculation time of both three-phase fault and single-phase fault satisfied the 5-min constraint proposed in this paper.

TABLE 1. Average calculation time of the Bayesian method.

Three phase fault	Single phase fault
136 sec	195 sec

It should be noted that MATLAB is an interpreted language and has generally lower efficiency. When programming the approach in a fully compiled environment (e.g., C++), even a shorter calculation time can be expected. Therefore, it can be concluded that the grid operators can realize the proposed fault location method within a reasonable calculation time and at an acceptable cost on hardware/software investment.

VII. CONCLUSION

This paper proposes a section-based fault location procedure for MV grids with underground cables, with Bayesian inference. The result of the approach is a ranking of possible faulted sections. The grid operators can use this ranking to facilitate the outage restoration process and to reduce the interruption duration.

When calculating the prior distribution of the fault location, the approach considers two types of major root causes of the faults in the underground cables: 1) component failure and 2) digging damage. In addition, the TFI readings can limit the interval of the fault positions. These readings can be especially useful when multiple branches exist in one MV feeder.

Considering the measurement errors, the posterior distribution of the fault position is calculated with Bayes' theorem. Besides, other uncertainties in the calculation are characterized as a grid condition vector that consists of fault

breakdown resistance, prefault grid status, and the inaccuracies of zero-sequence parameters.

To ensure the approach is computationally feasible, two major numerical issues are addressed. The Monte Carlo method is applied to calculate the integral on the grid condition vector. Moreover, a two-step calculation procedure is proposed to ensure the accuracy and the speed of the calculation.

The results from the case study show that the proposed approach has a significant advantage over the conventional impedance-based method for single-phase fault, which are the most common fault in the MV grids. It also merits moderate improvement for three-phase fault. Furthermore, the average calculation time is well below the proposed time limit.

The major advantage of the proposed Bayesian approach is the ability to integrate various measured data. Therefore, the accuracy can be improved by adding measured locations and/or electrical quantities. For future research, the paper suggests to integrate non-rms measurement quantities, such as the timing of traveling waves and/or the frequency of transient components.

REFERENCES

[1] F. Provoost, "Intelligent distribution network design," Ph.D. dissertation, Elect. Energy Syst. Group, Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2009.

[2] A. A. Girgis, C. M. Fallon, and D. L. Lubkeman, "A fault location technique for rural distribution feeders," IEEE Trans. Ind. Appl., vol. 29, no. 6, pp. 1170–1175, Nov./Dec. 1993.

[3] M. Lehtonen et al., "Automatic fault management in distribution networks," in Proc. 16th Int. Conf. Exhibit. Electr. Distrib., Contrib. (CIRED), vol. 3. Jun. 2001, pp. 1–10.

[4] P. M. van Oirsouw and F. Provoost, "Fault localization in an MV distribution network," in Proc. 17th Int. Conf. Electr. Distrib. (CIRED), May 2003, pp. 12–15.

[5] J. Zhu, D. L. Lubkeman, and A. A. Girgis, "Automated fault location and diagnosis on electric power distribution feeders," IEEE Trans. Power Del., vol. 12, no. 2, pp. 801–809, Apr. 1997.

[6] J. Mora-Flòrez, J. Meléndez, and G. Carrillo-Caicedo, "Comparison of impedance based fault location methods for power distribution systems," Electr. Power Syst. Res., vol. 78, no. 4, pp. 657–666, Apr. 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S037877960700123X

[7] M.-S. Choi, S.-J. Lee, D.-S. Lee, and B.-G. Jin, "A new fault location algorithm using direct circuit analysis for distribution systems," IEEE Trans. Power Del., vol. 19, no. 1, pp. 35–41, Jan. 2004.

[8] M.-S. Choi, S.-J. Lee, S.-I. Lim, D.-S. Lee, and X. Yang, "A direct three-phase circuit analysis-based fault location for line-to-line fault," IEEE Trans. Power Del., vol. 22, no. 4, pp. 2541–2547, Oct. 2007.

[9] H. Nouri, C. Wang, and T. Davies, "An accurate fault location technique for distribution lines with tapped loads using wavelet transform," in Proc. IEEE Porto Power Tech, vol. 3. Sep. 2001, pp. 1–4.

[10] D. W. P. Thomas, R. J. O. Carvalho, and E. T. Pereira, "Fault location in distribution systems based on traveling waves," in Proc. IEEE Bologna Power Tech Conf., vol. 2. Jun. 2003, pp. 1–5.

[11] S. Lotfifard, M. Kezunovic, and M. J. Mousavi, "A systematic approach for ranking distribution systems fault location algorithms and eliminating false estimates," IEEE Trans. Power Del., vol. 28, no. 1, pp. 285–293, Jan. 2013.

[12] J. A. Cormane, H. R. Vargas, G. Ordonez, and G. Carrillo, "Fault location in distribution systems by means of a statistical model," in Proc. IEEE/PES Transmiss. Distrib. Conf. Expo., Latin Amer. (TDC), Aug. 2006, pp. 1–7.

[13] J. Mora-Flórez, N. Estrada-Cardona, and S. Perez-Londoo, "Fault location in radial power systems based on statistical analysis," in Proc. 10th Int. Conf. Probab. Methods Appl. Power Syst. (PMAPS), May 2008, pp. 1–6.

[14] C.-F. Chien, S.-L. Chen, and Y.-S. Lin, "Using Bayesian network for fault location on distribution feeder," *IEEE Trans. Power Del.*, vol. 17, no. 3, pp. 785–793, Jul. 2002.

[15] Y. Xiang and J. F. G. Cobben, "Fault location for medium voltage underground cables using Bayesian inference," in *Proc. IEEE Eindhoven PowerTech*, Jun./Jul. 2015, pp. 1–5.

[16] E. Lakervi and E. J. Holmes, *Electricity Distribution Network Design*. London, U.K.: IET, 1995.

[17] A. Shafiu, N. Jenkins, and G. Strbac, "Measurement location for state estimation of distribution networks with generation," *IEE Proc.-Generat., Transmiss. Distrib.*, vol. 152, no. 2, pp. 240–246, Mar. 2005.

[18] R. Singh, B. C. Pal, and R. B. Vinter, "Measurement placement in distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 668–675, May 2009.

[19] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement," *IET Generat., Transmiss. Distrib.*, vol. 4, no. 1, pp. 50–59, Jan. 2010.

[20] Y. Xiang, P. F. Ribeiro, and J. F. G. Cobben, "Optimization of state-estimator-based operation framework including measurement placement for medium voltage distribution grid," *IEEE Trans. Smart Grid*, vol. 5, no. 6, pp. 2929–2937, Nov. 2014.

[21] Q. Zhuang, N. Steentjes, R. Mehairjan, and J. Smit, "Expectation-maximization method for analyzing incomplete failure data from 10 kV cables," in *Proc. Int. Conf. Condition Monitor. Diagnosis*, 2014, pp. 734–737.

[22] M. Evans and T. Swartz, *Approximating Integrals Via Monte Carlo and Deterministic Methods*. London, U.K.: Oxford Univ. Press, 2000.

[23] C. P. Robert, "Simulation of truncated normal variables," *Statist. Comput.*, vol. 5, no. 2, pp. 121–125, Jun. 1995.

[24] J. A. Breslaw, "Random sampling from a truncated multivariate normal distribution," *Appl. Math. Lett.*, vol. 7, no. 1, pp. 1–6, Jan. 1994. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0893965994900426

[25] Intel Corp., Santa Clara, CA, USA. (2014). *Intel Xeon Processor E5-2680 v3 Specification*. [Online]. Available: http://ark.intel.com/products/81908/Intel-Xeon-Processor-E5-2680-v3-30M-Cache-2_50-GHz

YU XIANG (S'13) received the B.Sc. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2008, and the M.Sc. degree in electrical engineering from the Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, in 2011, and is currently working towards the Ph.D. degree at TU/e, in cooperation with Alliander, Arnheim, The Netherlands.

His current research interests include distribution grid state estimation, measurement placement, and data analysis.

JOSEPH F. G. COBBEN (M'11) received the Ph.D. degree in electrical engineering from the Eindhoven University of Technology (TU/e), Eindhoven, The Netherlands, in 2007.

He is currently a Research Scientist with Alliander, Arnheim, The Netherlands. He is also a part-time Professor with TU/e. He has authored several technical books.

Dr. Cobben is a member of several national and international standardization committees about requirements for low and high voltage installations and characteristics of the supply voltage.