Deployability of a probabilistic earth fault location method for MV distribution feeders

Alexandre Bach, Trung Dung Le, Marc Petit

TPWRD-00656-2023

Deployability of a probabilistic earth fault location method for MV distribution feeders

Alexandre Bach, *Student Member, IEEE,* Trung Dung Le, Marc Petit, *Member, IEEE*

***Abstract*— This paper presents a probabilistic impedance-based fault location method (FLM) designed for medium voltage (MV) distribution feeders using zero-sequence (ZS) components with few additional voltage measurements on secondary substations. It is based on a previous work in which a novel FLM has been developed and showed promising locating performance on the most challenging ramified rural feeders. However, this method has been found to be too sensitive to measurement errors to be deployable with the current generation of sensors. Thus, this paper presents a probabilistic formulation of the fault location problem with Monte-Carlo simulations which fits two four-dimensional Gaussian probability density functions (PDFs) of the ZS voltage and current on the considered paths (between the busbar and each measurement node) on a MV feeder. This way, our fault location algorithm is robust to both measurement and line impedance errors. A criterion for the optimal number of Monte-Carlo simulations as well as an efficient way to compute the similarity between two different PDFs in several dimensions are proposed. Finally, it is shown that the performances of this algorithm are highly correlated to the value of the threshold probability.**

***Index Terms*— Probability, Fault location, Power distribution feeders, Model and measurement errors, Monte-Carlo methods, Similarity measurement**

I. Introduction

DISTRIBUTION system operators (DSOs) are looking for faster and more accurate fault location methods deployable on real MV grids. Indeed, the growing part of electricity in the energy demand paired with its increase in price lead to an increase in the value of lost load. To compensate for this fact, an increase in reliability of the distribution grids is needed, especially on the MV feeders to which the distributed producers (like the renewable energy sources) are connected. That motivates the researches carried out on FLMs at the distribution level [1], [2]. [3] quantified the gains in reliability from having accurate FLMs for distribution grid, measured by the SAIDI index. An improvement of around 5% is shown when having accurate FLMs with respect to visual inspection of the feeder by maintenance teams. With the economic impact of power outage, this level of improvement is far from negligible. Indeed, in France, the SAIDI index in MV grids is around 40 min with an incentive given by the regulator of 6M€ per minutes below this value (or with a malus when above). This mean that a reduction of SAIDI of 5% would lead to an economic gain of at least 12M€ by the regulator incentive only. The distribution feeders are traditionally operated with a radial configuration. Besides, the only usually instrumented node is the busbar of the primary HV/MV substation where the currents in the neutral grounding as well as at the head of all feeders are measured, and voltages at busbar are also measured. Fault location in radial feeder is more challenging than in transmission lines where the full observability of the network is often reached. That explains the difficult transposition of FLMs tailored for transmission grids at the distribution level.

Three main categories of FLMs can be found in the literature. First, the algorithms which are using the phasors at the fundamental frequency are called impedance-based FLMs [4]. They represent a well-known category of location methods and have proven to be cost effective at the transmission level since they need only low sampling frequency measurement devices. However, when deployed at the distribution level with only the busbar instrumented, they faced the multiple estimation problem. Indeed, there is no way to differentiate between two nodes which are located at the same apparent impedance from the busbar while being potentially located onto different branches which can be far from each other. This is one of the main limitations of these methods at the distribution level. This issue can be solved with increase in instrumentation of the distribution grids, as considered by some DSOs with the rise of *smart grids*. For instance, [5] presents a French demonstrator of smarter MV feeders where 130 secondary substations have been instrumented across six feeders of a primary substation. The choice of the number of additional measurements and their placement is then of utmost importance since DSOs are facing stricter economical constraints than TSOs. Coming from control theory, one of the most common methods to place measurement on transmission grids is called the 1-bus spaced placement method [6]. While enabling the exact state estimation of the feeders, this method is not economically viable for MV feeders since it would need around 30% of the nodes to be instrumented [7]. That is why there is a need for impedance-based FLMs leveraging a limited number of additional measurements. [8] proposed a FLM leveraging both synchronized and unsynchronized measurements placed at each extremity of a feeder. This method is based on the fact that two voltage estimate sets from both ends of a line are equal on the node on which the fault is the closest, meaning that the fault is occurring on a lateral branch of the exhibited node. The method proposed in this paper, and from our previous work [9], is based on this property. However, when deployed on long and highly ramified feeders, [8] chooses not to instrument every extremity node without quantifying its impact on the locating performance of the method. Moreover, since this method is based on variational phasors (the difference between pre-fault and during-fault phasors), it supposes that there is no load variation when the fault is occurring, which seems to be a strong hypothesis. Besides, [10] and [11] proposed a FLM needing only one additional instrumented node. These methods also use variational phasors and by such imply that the load is of constant impedance in order to be able to take into account its variation when the fault occurs. In [12] and

[13], the variation of the load current is taken into account in the case of static loads only. This requires the voltage at each substation to be estimated, which would need more measurements in the case of complex feeders. Another way to solve the load estimation problem is to use the ZS components. Indeed, in case that the MV windings of all MV/LV transformers are mainly delta-coupled (or at least with an isolated neutral grounding), there is no ZS current that can go from a voltage level to another. This is the case in France and the approach was used in our previous work [9]. With ZS component, the main difficulties are the low voltage drop across the feeder (especially with compensated neutral grounding) which can be inferior to measurement errors coupled with the difficult estimation of the ZS line impedances which depend on the earth resistivity. There are very few methods proposed in the literature that deals with the combination of these two sources of error. [14] proposes a method to correct the wrong line impedance values from the DSO database and proposes to filter out the bad data coming from misfunctioning phasor measurement units (PMUs). In [15], a probabilistic FLM is formulated which enables the estimation of a PDF of the fault location which takes into account the potential errors coming from voltage transformers (VTs).

Secondly, transient-based methods try to estimate the propagation time of the electromagnetic waves induced by the fault [16], [17]. These methods are said to be able to achieve more accurate location of faults than the impedance-based ones. However, since the propagation time is very small, they need the deployment of high sampling frequency measurement devices (around or more than 1MHz), which are not found yet at the distribution level.

Finally, we chose to gather in a third group 1) the hybrid methods, which try to merge data coming from different sources (such as fault indicators, calls from clients etc.) [18], sometimes even with uncertainties taken into account [19], [20] and 2) the machine-learning based methods [21], [22] which are learning their statistical properties from large dataset of faults. Given that fault location on MV feeders might become of strategical importance, and given that building a unbiased database from simulations is difficult to achieve, the lack of interpretability of such methods is their main weakness [23].

Seeing the advantages and drawbacks of each method family, this paper proposes a ZS impedance-based method which leverages sparse additional voltage measurement on some extremity secondary substations. A deterministic formulation with optimal measurement placement has been proposed in [9] and further studied in [24]. While being robust to fault impedance value and to position of the faults even on complex feeders, this method proves to be sensitive to line impedance and measurement uncertainties. That is why this paper proposes a reformulation of the fault location method in terms of probability density functions of both ZS voltage and ZS current along the feeder to perform a probabilistic fault location which is robust to both impedance and measurement uncertainties by design.

This paper is structured as follows. Section II presents the probabilistic formulation of the FLM dealing with both measurement and line impedance value errors. Then, Section III demonstrates the convergence of the Monte-Carlo estimation and proposes a criterion to choose the optimal number of simulations N_{MC} and presents different ways to compute intersection of the two PDFs. The following Section IV shows some simulation results obtained on a reconstructed realistic feeder and compares the sensitivity analysis of the new probabilistic method to what was obtained with the previous one. Finally, Section V concludes the paper and presents some perspectives.

II. Probabilistic fault location formulation

The presented probabilistic FLM is an extension of a previous work [9], [24] which is based on the comparison of two sets of ZS voltage estimates. This method is able to locate any type of earth faults (single phase to ground or multiple phases to ground, except for balanced three phase faults which present no ZS current component).

A. Overview of the deterministic fault location formulation

Under the hypothesis that all loads and producers are connected to the MV grid with a transformer having either a delta-coupled winding or with an isolated neutral grounding (which is true in all cases in France), it is possible to compute the ZS voltage on all nodes knowing only the ZS voltage and current at the busbar (at the secondary winding of the HV/MV transformer in the primary substation) and all the line impedance values when the grid is sound, meaning that fault location method is not impacted by presence of distributed generators.

With M optimally placed additional voltage measurements (on some secondary MV/LV substations where there is no ZS current), it is possible to perform two ZS voltage and current estimates when a fault occurs using Fortescue decomposition of the fundamental phasors extracted from the measurements with discrete Fourier transform. From the busbar to each additional measurement node (noted $m_i, i \in [\![1, M]\!]$), a Top-Down estimation is performed, while a Bottom-Up estimation is performed on the reverse paths. For every path, the node on which the distance (error) between the two estimate sets is the lowest is called the projection node and is the projection of the faulty node onto the considered path (busbar $\leftrightarrow$ measurement node m_i). It is then possible to define a fault location area $\Omega_{m_i}^{sol}$ as the set of nodes being lateral to the projection node with respect to the considered path. In the case of M additional voltage measurements optimally placed on the feeder [24], each node n is associated with a number ($N_\Omega(n) \in [\![0, M]\!]$) which counts the number of times it is located in a solution area. The algorithm returns areas of increasing likelihood for location of the fault, defined in (1):

$$\Omega_{sol}^{k} = \{n \in nodes : N_\Omega(n) \in [\![k, M]\!]\} \quad (1)$$

This method presents a degree of freedom being the number k that can be chosen less or equal to the number of additional measurements M, which enables the algorithm to deal with wrong or missing data from some measurement nodes. A sensitivity analysis [24] has been carried out and has shown that this method is not sensible to the fault resistance value (faults with resistances up to $500\ \Omega$ have been tested) and to the complexity of the feeder. It means that this method is able to perform as efficiently on a simple feeder (typically a

straight feeder with very few ramification) as on a long and highly ramified rural feeders. This fact shows that this algorithm is particularly well tailored for the distribution level.

B. Limitations and motivations for a probabilistic method

However, it appears that the method is sensitive to line impedance uncertainties (mainly series line impedance), as most impedance-based methods. Even though the use of synchronized measurements helps to reduce this sensitivity, the reliable estimation of ZS line impedance values for distribution grids is known to be difficult [25]. So, the real values of line impedances might be quite different from the ones stored in the DSO database so that the FLM is not able to perform as expected.

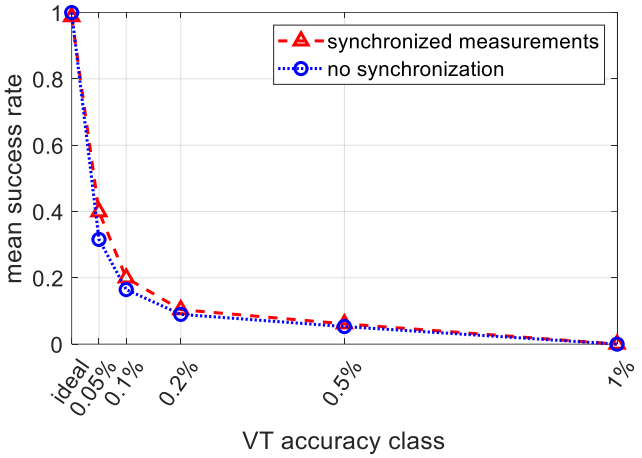

Fig. 1: Impact of VT accuracy class on the mean success rate of the non-probabilistic method

Besides, it appears that the deterministic FLM is also very sensitive to measurement errors, to a level that compromises its deployability with the current generation of sensors. A sensitivity analysis of our method has been carried out considering 20 reconstructed realistic feeders [26] on which one single-phase-to-the-ground fault has been randomly placed with different measurement errors following the standard for voltage transformers (VT) and current transformers (CT). A total of 87 measurement locations have been tested across the twenty faults considering exact knowledge of the ZS line impedances. The standard [27] defines accuracy classes which bound the magnitude and phase angle errors. Some classes have been built by interpolation (the 0.05% class) in order to study the behavior of the algorithm when dealing with non-ideal measurements. We defined a success indicator being set to 1 if a measurement is able to accurately find the projection node of the fault and set to 0 otherwise. On Fig. 1, the success rate (mean of success indicator) is displayed with respect to measurement errors considering uniformly distributed and independent magnitude and phase angle errors. It appears that while the method is able to always locate the projection node with ideal measurements (the success rate is 1), it is only able to behave as expected in 20% of the cases considering the 0.1% accuracy class measurement, which is the best available today. We can even see that it is almost always wrong when using 0.5% accuracy class measurements. The exponential sensitivity of the success rate to measurement errors calls for a change in formalism and shows a need to take into account the potential measurement errors in the fault location algorithm if we want to design an algorithm deployable in real field. While being wrong, it appears from the results that in most cases the estimated projection node is a neighbor of the real one. Thus, a method that could be able to select more than one projection node when the two voltage estimates are close enough to each other on more than one node might be robust to measurement errors. Given the usual topology of MV feeders, where there are rarely two long lateral branches on two consecutive nodes of the main feeder (being the set of nodes on the path between the primary substation and the farthest away node), this method could highly enhance the robustness while increasing the fault location area size by a small factor. That is why this paper proposes to replace the deterministic estimations of phasors by estimations of probability density functions along the paths.

C. Dealing with measurement errors: Estimation of a 4-dimension Gaussian PDF

To take into account potential measurement errors, a Monte-Carlo process is used. The measurement values are treated as a random variable vector R, with $R \in \mathbb{R}^{12\times1}$ at the busbar (three complex voltage measurements – with real and imaginary parts - and three complex current measurements) and $R \in \mathbb{R}^{6\times1}$ for any additional voltage at a remote node. R is described by its joint PDF f_R which is supposed known, and a Monte-Carlo process is used to estimate the PDF of a 4-dimension random variable being the ZS components $X = [\Re(V_0) \quad \Im(V_0) \quad \Re(I_0) \quad \Im(I_0)]^T$, where $\Re(V_0)$ is the real part of the ZS voltage. As verified later on Fig. 4, it seems reasonable to suppose that the PDF of X - being an average of random variables - can be well described by a normal law with mean $\mu \in \mathbb{R}^4$ and covariance matrix $\Sigma \in \mathbb{R}^{4\times4}$, noted $X \sim \mathcal{N}(\mu, \Sigma)$, in concordance with the central limit theorem. The Monte-Carlo process involves N_{MC} random samplings of the variable R according to its PDF f_R and $X(i)$ is the sample of the random variable X drawn during the ith simulation. Then, the two parameters of the Gaussian PDF - the mean vector and the covariance matrix - are estimated with unbiased estimators defined in (2) and (3):

$$\hat{\mu} = \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} X(i) \tag{2}$$

$$\hat{\Sigma} = \frac{1}{N_{MC}-1} \sum_{i=1}^{N_{MC}} (X(i)-\mu)(X(i)-\mu)^T \tag{3}$$

The chosen value for N_{MC} has a high impact on the accuracy of the proposed process to estimate the PDF of X. Indeed, if we consider each $X(i), i \in [\![1, N_{MC}]\!]$ to be a realization of the same random variable with mean μ and variance Σ, then we can say that the estimator $\hat{\mu}$ is not biased since $\mathbb{E}[\hat{\mu}] = \mu$ and its variance is $\mathbb{V}(\hat{\mu}) = \frac{1}{N_{MC}}\Sigma$. So, the choice of N_{MC} is directly linked to the variance of $\hat{\mu}$, which means to the probability of having an estimator $\hat{\mu}$ distant from μ - this is discussed in section III.B.

D. Dealing with line parameters uncertainties: Progression of the estimation along a path and definition of fault location

area

Once the probability of the ZS voltage and current is known at both extremities of a path, the proposed FLM needs a Top-Down and a Bottom-up estimation of the ZS voltage along the considered path of nodes between the measurements. We note V_{0k} the ZS voltage at node k, $X_k = [\Re(V_{0k}) \quad \Im(V_{0k}) \quad \Re(I_{0k}) \quad \Im(I_{0k})]^T$ is the random vector at node k and the next node on the path is named m. As presented in [9] - using a PI lumped parameter model of the distribution lines - we can compute the ZS voltage and current and node m, as shown in (4):

$$\begin{pmatrix} \Re(V_{0m}) \\ \Im(V_{0m}) \\ \Re(I_{0m}) \\ \Im(I_{0m}) \end{pmatrix} = T. \begin{pmatrix} \Re(V_{0k}) \\ \Im(V_{0k}) \\ -\Re(I_{0k}) \\ -\Im(I_{0k}) \end{pmatrix} \quad (4)$$

As presented in the previous paragraph, we supposed that $X_k \sim \mathcal{N}(\mu_k, \Sigma_k)$. Besides, $T \in \mathbb{R}^{4\times4}$ is a matrix composed of the ZS line impedance parameters of the line between node k and m and of the lateral lines of node k (merged by their equivalent at node k). We suppose the joint PDF of all line impedances in T to be known. Given that, a Monte-Carlo process similar to what has been presented in II.C is used with N_{MC} random samplings of the matrix T and vector X_k according to their respective PDFs. The computed N_{MC} samples of the ZS voltage and current at node m, noted X_m, enable the estimation of the mean vector and covariance matrix at node m following (2) and (3). X_m is then supposed to follow a normal law of parameter (μ_m, Σ_m) for the next estimation along the path. An example of implementation is presented in III.C and the impact of the choice of N_{MC} value is shown in Fig. 5.

After all the ZS voltages and currents estimations for both the Top-Down and Bottom-up processes, the projection node needs to be exhibited to define the fault location area for the considered path. In the non-probabilistic formulation, the projection node is the node p on the path on which the difference between the Top-Down estimation and the Bottom-up estimation is minimal. With the probabilistic formulation, we can say that for each node p along the path, the intersection of the two PDFs is equal to the probability of the fault being lateral of p, which is the same as the probability of p being the projection node. There are two cases to consider (cf. Fig. 2): First, when dealing with unsynchronized measurements, the phase angle difference between the two sets of estimates cannot be computed. Then, a Monte-Carlo process is used to infer a 1-dimension Gaussian PDF for the ZS voltage magnitude from the marginal 2-dimension PDF of the complex ZS voltage. For each node, we get two normal PDFs (the Top-Down and Bottom-up ones) that we chose to note with indices 1 and 2 (f_1^{1D} is the PDF of $|V_{0,1}| \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and f_2^{1D} describes $|V_{0,2}| \sim \mathcal{N}(\mu_2, \sigma_2^2)$). The indices 1 and 2 are chosen so that $\mu_1 \leq \mu_2$. To compute the intersection of these two PDFs, we need to compute the area located below both curves. We search for the intersection point by solving the quadratic equation $f_1^{1D}(c) = f_2^{1D}(c)$ that leads to $c = \frac{\mu_1+\mu_2}{2}$ if $\sigma_1 = \sigma_2$ (which is very unlikely when using floating number operations) or leads to (5) when $\sigma_1 \neq \sigma_2$:

$$c = \frac{\mu_2\sigma_1^2 - \sigma_2\left\{\mu_1\sigma_2 + \sigma_1\left[(\Delta\mu)^2 + 2\Delta\sigma^2.\log\left(\frac{\sigma_1}{\sigma_2}\right)\right]^{\frac{1}{2}}\right\}}{\Delta\sigma^2} \quad (5)$$

With $\Delta\mu = \mu_1 - \mu_2$ and $\Delta\sigma^2 = \sigma_1^2 - \sigma_2^2$. Then, the wanted intersection area $\mathcal{A}(f_1, f_2)$ can be written as function of the cumulative distribution functions (CDFs) F_1 and F_2: $\mathcal{A}(f_1^{1D}, f_2^{1D}) = \mathbb{P}(|V_{0,1}| > c) + \mathbb{P}(|V_{0,2}| < c) = 1 - F_1(c) + F_2(c)$. We use the error function $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$ to compute $\mathcal{A}$ (6):

$$\mathcal{A}(f_1^{1D}, f_2^{1D}) = 1 - \frac{1}{2}\text{erf}\left(\frac{c-\mu_1}{\sqrt{2}\sigma_2}\right) + \frac{1}{2}\text{erf}\left(\frac{c-\mu_2}{\sqrt{2}\sigma_2}\right) \quad (6)$$

In the second case, synchronized measurements are used. The complex values of the ZS voltages can be directly compared in the two set of estimates. The marginal voltage PDFs are 2-dimensional Gaussians (real and imaginary part), and the intersection is the volume $\mathcal{V}$ under both surfaces (cf. Fig. 7). There is no more closed-form expression to compute said volume. A widely used method to compute such volume is to use the rectangle integration method (RIM), which divides the space into N_{rec} rectangles of surface $\mathcal{A}_{N_{rec}}$ and of barycenter coordinates noted (x_k, y_k) to approximate the volume by (7):

$$\mathcal{V}(f_1^{2D}, f_2^{2D}, N_{rect}) = \sum_{i=1}^{N_{rect}}\sum_{j=1}^{N_{rect}}\left[\min\left(f_1^{2D}(x_i, y_j), f_2^{2D}(x_i, y_j)\right).\mathcal{A}_{N_{rect}}\right] \quad (7)$$

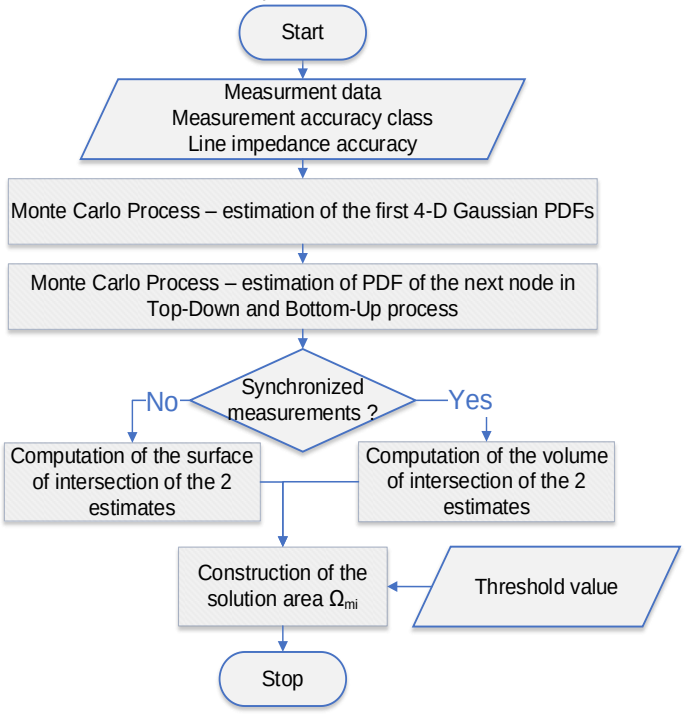

Fig. 2: Flowchart of the construction of one the M fault location areas Ω_{m_i}

While being easy to implement, this method presents a quadratic complexity and can lead to high computation time when dealing with long feeders having paths with a high number of nodes. That is why another criterion is proposed: the Bhattacharyya coefficient. It is a well-known measure for similarity between two different probability density functions [28]. It differs from the intersection measure since it is defined

as $BC(f_1^{2D}, f_2^{2D}) = \iint_{\mathbb{R}^2} \sqrt{f_1^{2D}} . \sqrt{f_2^{2D}}$ but can be seen as the projection of one PDF onto the other in a hypersphere (of infinite dimensions) whose cosine directors are the square root of the probabilities. This measure of similarity presents the advantage of having a closed-form expression when considering two multivariate gaussian PDFs f_1^{2D} and f_2^{2D} (8):

$$BC(f_1^{2D}, f_2^{2D}) = \left(\frac{\sqrt{|\Sigma_1| . |\Sigma_2|}}{|\Sigma_{12}|} \right)^{\frac{1}{2}} . \exp\left(-\frac{\Delta\mu^T . \Sigma_{12}^{-1} . \Delta\mu}{8} \right) \quad (8)$$

With $\Sigma_{12} = \frac{\Sigma_1 + \Sigma_2}{2}$ and $|\Sigma|$ being the determinant of Σ.

Using this similarity measure instead of computing the intersection volume using the rectangle method for instance leads to major computation time savings. Only two matrix exponentials are needed to compute this coefficient instead of two by point computed for the rectangle method. A detailed comparison of the two metrics is discussed in III.D.

Once the intersection probability $\mathbb{P}_\cap$ is computed for each node on the path, the algorithm needs a threshold value τ to select as the set of projection nodes all the nodes having a higher-than-τ intersection probability. For example, 5 nodes (from 952 to 974) would be selected in Fig. 8 with $\tau = 50\,\%$ and using the rectangle method (red curve). Then, the fault location area Ω_{m_i} for this path is the set of nodes being lateral to one of the selected projection nodes with respect to the considered path. The probabilistic algorithm that builds the *M* solution areas is summarized in Fig. 2. The impact of the threshold value is discussed in IV.B.

III. Accuracy of the probabilistic estimation

A. Use case

In this paper, we chose to assess the performance of the presented method with the worst-case scenario, being uniformly distributed and independent random errors on magnitude and phase displacement of measured quantities within the boundaries given by the accuracy class of the devices for measurement errors. Given that we need three measurements to compute the ZS component coming from three different VTs or CTs, we made the hypothesis that the errors on the different phases are also independent from each other. In this section, the figures are obtained using measurement errors up to 0.1%, as being the best yet deployed accuracy class. Concerning impedance value uncertainties, uniformly distributed errors are also implemented within an extent given by a confidence interval. In this section, an uncertainty of $\pm 20\%$ has been chosen since we believe it is a plausible margin of error on the ZS impedances.

Given that this method is specially designed for long and ramified feeders, the FLM has been tested in simulation on reconstructed realistic feeders from the open data of the French DSO Enedis [26]. The longest and most ramified out of the 20 reconstructed feeders has been chosen to be exhibited in this paper as it is the worst-case scenario. It is a 211-node feeder with a total length $L_{tot} = 75.05\,km$. Its topology is further described in [24] and shown on Fig. 3 where overhead lines (OHLs) are represented in dark red, underground cables (UC) in bright blue and large section lines ($r_d \leq 0.4\,\Omega.km^{-1}$) are thick while small section lines are thin.

The ten most priority measurement locations have been instrumented. Ten fault locations have been chosen (shown in red stars in Fig. 3). They are the barycenter of a clustering of the grid in ten connected areas of equivalent fault probability (considering ten times more fault probability on OHLs than UCs) with a K-means algorithm. This ensures that all parts of this complex grid are considered. The simulated faults in this paper are single-line-to-ground faults, which represent more than 75% of the detected faults [29], with a fault resistance of $500\,\Omega$, which is higher than 90% of the cases [29]. The FLM has been simulated using MATLAB/Simulink running on a PC equipped with an Intel(R) Core(TM) i7-10700 CPU and 32 GB of RAM with compensated neutral grounding, which leads to low values of fault current and can then be seen as the most challenging case for an impedance-based FLM. The DFT is used to estimate the phasors at each measurement node in steady-state and Fortescue transform is applied to get the ZS component so that the presence of harmonics does not impact the method.

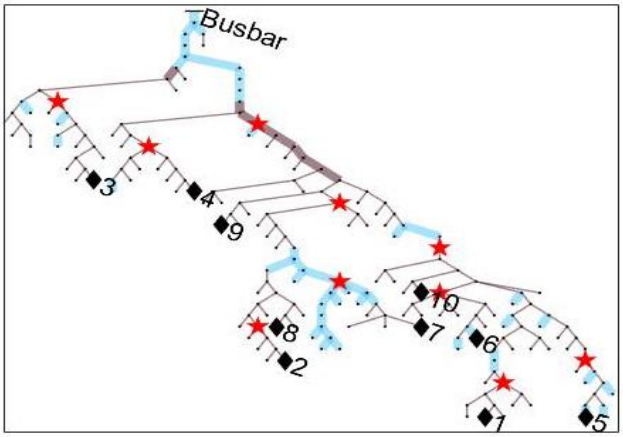

Fig. 3: Topology of the simulated grid with additional measurements (black diamond nodes labeled by placement priority) and fault positions (red pentagrams)

B. Convergence of Monte-Carlo process and uncertainties: optimal N_{MC} value

The complexity – and by such the computation time – of the Monte-Carlo estimation process grows linearly with the value of N_{MC}. That is why this number should be kept as low as possible. However, a dependency of the variance of the estimated average value with N_{MC}^{-1} has been theoretically stated in II.C and observed with our simulations. That is why there is a need to find an optimal value for N_{MC} which ensures a tradeoff between potential errors and computation time. We want to ensure that the first estimation process (the estimation of the ZS component *X* from the phase measurements *R* as in II.C.) does not induce a biased error that would lead to a non-working algorithm. As we can see on Fig. 1, if the error in the estimated mean of the ZS voltage (and current at the busbar) is superior to 0.1%, then the deterministic algorithm would be accurately exhibiting the projection node in less than 20% of the cases.

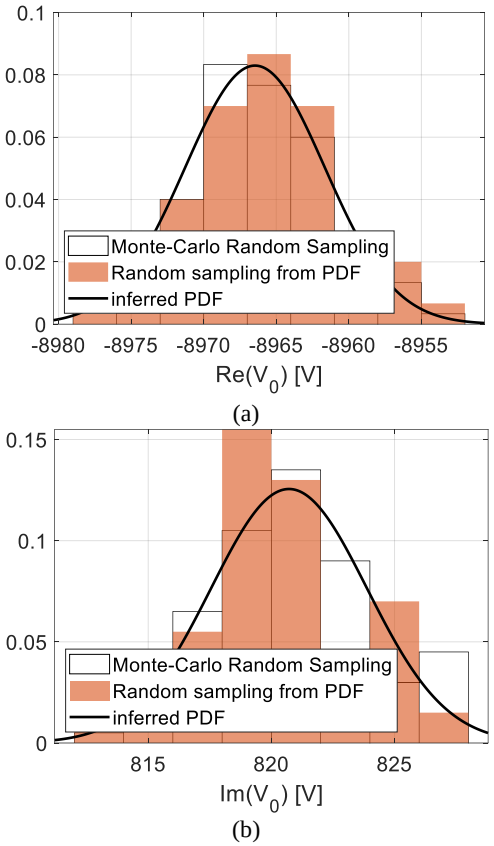

Fig. 4: Random sampling (RS) from Monte-Carlo simulations (white histogram), inferred PDF (line) and RS from inferred pdf (colored histogram) for real (a) and imaginary (b) parts of the ZS voltage

To find an optimal value for N_{MC}, we propose to compute the total vector error (TVE) between the estimated ZS voltage $V_{estimated}$ (the mean of the samples obtained from the Monte-Carlo process) and the real ZS voltage V_{real} as defined in (9) while considering voltage measurement PDFs centered around the real measurement for each phase. In this case, the TVE quantifies the bias introduced by the estimation process only, since the measurements are not biased.

$$TVE = \frac{|V_{estimated} - V_{real}|}{|V_{real}|} \quad (9)$$

Fig. 4 presents one of the 250 obtained histograms with a chosen value of $N_{MC} = 100$. We can see that the mean computation has a low pass filtering effect which enables us to get the accurate estimates of both real and imaginary parts of the ZS voltage even though the histogram is not close to following a Gaussian curve. This means that using mean and covariance as a description of the PDF (black line in the figure) is better than reusing the previous histogram, which could lead to error accumulations (especially when the value of N_{MC} is low) when propagating along a path of nodes.

To compute this function, the estimation processes have been performed 250 times for different values of N_{MC} with a considered theoretical ZS voltage measured of $V_0^{mes} = -8{,}967 + 821\mathrm{j}\,\mathrm{V}$. The CDF is built by finding how many times over the 250 different estimation processes the TVE is inferior to a given value. The results are presented in Fig. 5 when considering 0.1% accuracy class measurement devices.

We can observe that with such an accuracy class, drawing as less as 5 samples is enough to ensure that in all cases, the TVE remains inferior to 0.1%. However, Fig. 1 shows that this level of accuracy is not enough to ensure a good behavior of the FLM. An accuracy level of 0.01% has been chosen as target for the choice of N_{MC}. That is why the results shown in this paper are obtained considering $N_{MC} = 100$, which ensures that the TVE remains inferior to 0.02% in all cases and even inferior to 0.01% in 90% of the cases. This led us to believe that this value presents the best tradeoff between computing time and estimation variance for the targeted application.

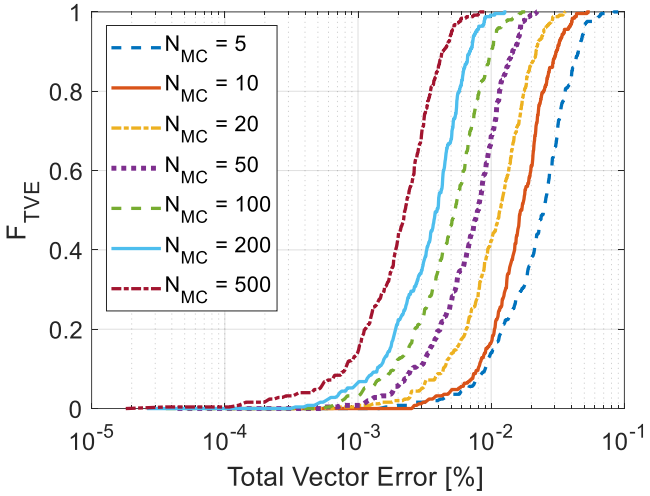

Fig. 5: CDF of the TVE depending on the value of N_{MC} considering 0.1% accuracy class measurement devices

C. Propagation along paths

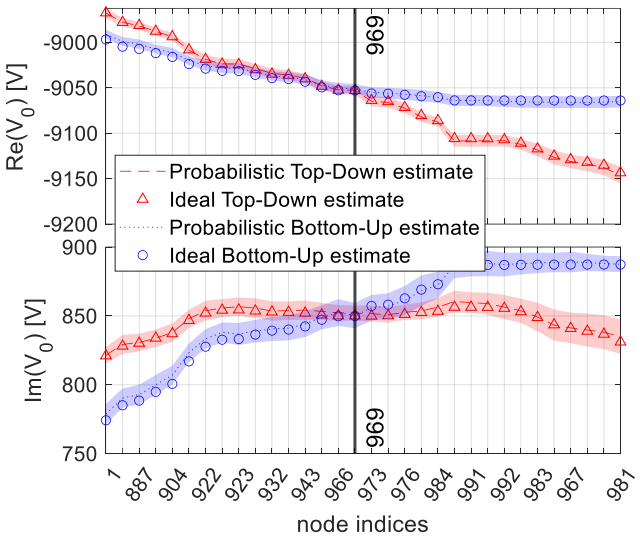

Fig. 6: Comparison of probabilistic (with $\pm\sigma$ intervals in shaded areas) and ideal estimation of the ZS voltage on a path (from busbar 1 to measurement node 981, only one node out of two is labeled on the figure)

The estimation process has been implemented on the presented feeder for all presented faults. The results of the ZS voltage estimation along one of the paths is shown in Fig. 6. We observe no divergence between the Monte-Carlo-based estimation and the ideal one – being the ZS voltage estimation with exact knowledge of all impedances and ideal

measurements. This means that there is no accumulation of errors with the 32 iterative processes (corresponding to the 32 nodes on the path between the busbar 1 and the considered measurement node 981). Node 981 is the second most priority node from the optimal placement algorithm and is labelled 2 in Fig. 3. In other words, even with N_{MC} being as low as 100, the number of random samplings is sufficient to ensure that the estimated mean is the theoretical one. As expected, the node on which the two estimate sets are the closest to each other is the projection of the considered fault on this path of nodes: node 969 (the real projection node). However, it is clear that node 966 is also a node on which the difference between the two sets is close to zero. Thus, without a probabilistic formulation, a FLM could easily exhibit node 966 as the projection node. Besides, it appears that the standard deviation (shown by the shaded area in Fig. 6) does not increase substantially in this example. A visible increase of the standard deviation during the progression along the path is only observed on the Top-Down estimation in this case.

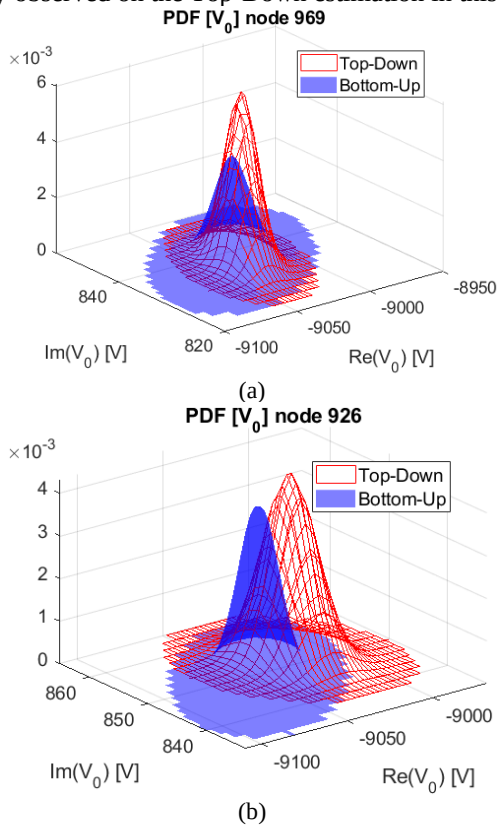

Fig. 7: The two marginal PDFs (Top-Down is colored and Bottom-Up in black) on two different nodes: the projection node (a) and another node far away (b)

Once the mean and covariance matrix of the marginal 2-dimensional ZS voltage have been estimated (the covariance is not shown in Fig. 6), the two PDFs can be built on each node. Fig. 7 shows the two gaussian laws on two different nodes, with node 969 (Fig. 7 a) being the real projection node. As expected, the two marginal PDF functions have the same mean on the projection node, meaning that they are centered around the same location. However, their covariance matrices are not identical, meaning that the intersection probability - i.e. the volume under both surfaces – is not equal to 1. On node 926 that is far away from the projection node (Fig. 7 b), the two PDFs are centered around locations that are far enough from each other with respect to their variances so that their intersection volume is low. The results presented in this subsection lead us to have confidence in the values of μ, Σ obtained with the probabilistic estimation process.

D. Computation of the intersection area or volume

The critical next step in the algorithm (Fig. 2) is the computation of the intersection of the two ZS voltage estimates on all nodes on each path - being the probability of the fault being lateral to a node. When using traditional unsynchronized measurements, the intersection area $\mathcal{A}$ between the two 1-dimension ZS voltage magnitude PDFs is computable with a closed-form equation (6). This is a low-complexity computation which is compliant with the targeted computation time of the whole algorithm of a few minutes at maximum. Besides, when dealing with synchronized measurements, there is no closed-form expression of the volume $\mathcal{V}$ under the intersection of the two PDFs. A possible way to compute this volume is to use the rectangle integration method (7), which depends on the choice of the value N_{rect} for its accuracy, and the computing complexity is quadratically linked to this value since there is a need to compute $2.N_{rect}^2$ PDF values. To ensure that the computed volume is a good approximation of the real one, the time needed might be too important with respect to the time constraints of fault location methods, even at the distribution level. For instance, with $N_{rect} = 1000$ as chosen value, with our up-to-date PC configuration, it took 260s to compute the intersection probabilities on a path of 32 nodes on the presented feeder. This time does not account for the time needed to do the 32×2 Monte-Carlo estimations of the ZS voltage. That is why there is a need to either find a reduced value for N_{rect} whilst keeping a good estimation of $\mathcal{V}$ or find another expression for the intersection which is lighter in computation time.

On one hand, we searched for an optimal value for N_{rect}. On the considered feeder instrumented with the four most needed measurement nodes (Fig. 3), the probabilistic estimation process has been simulated, leading to the estimation of two sets (μ, Σ) on a total of 4 paths composed of 102×2 Monte-Carlo estimations. From all these estimations, the intersection volume has been computed with the rectangle method for 7 different values of N_{rect} ranging from 10 to 1,000. We computed the relative difference of intersection volume for each case with the value obtained for $N_{rect} = 1000$ being chosen as reference. The histograms of these relative errors show that around 99% of the relative errors between the intersections computed with $N_{rect} = 500$ and the ones computed with $N_{rect} = 1000$ are inferior to 0.05%. This proves that with $N_{rect} = 1000$, the intersection computed with rectangle method converges almost to the exact intersection value. Choosing $N_{rect} = 50$ leads to 99% of the errors being in the $\pm 10\%$ interval. This range is too important when knowing that for some nodes along a path, the estimated mean will be close to each other, resulting in close values for the intersection. So, choosing this value might change the ranking order of the node in function of intersection probability. In this

paper, the results presented when using the rectangle method are shown with $N_{rect} = 100$, which ensures that the errors remain in a $\pm 2\%$ interval while being 100 times faster to compute than with $N_{rect} = 1000$.

On the other hand, a second way to compute the wanted intersection is the Bhattacharyya coefficient (BC), which has a closed-form expression (8) and depends only on the values of the couples (μ, Σ). This coefficient is not equal to the intersection of the two probabilities but measures their similarity with a single computation. In Fig. 8, the intersection volume computed by rectangle method (red) is compared to the BC (blue). It has been observed that the BC always overestimates the value of the intersection volume - meaning that using the similarity measure would be more conservative than the intersection volume. This could be a computationally efficient way to compute the intersection probability $\mathbb{P}_\cap$ but comes at the cost of a lower locating potential of the algorithm since the number of projection nodes will be higher, enlarging the size of the fault location area.

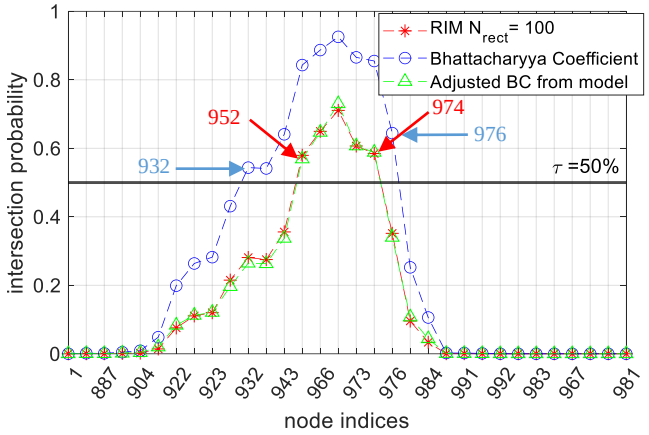

Fig. 8: Intersection probability indices on a path between node 1 and node 981 (one node out of two is labeled)

Moreover, we observed that the BC follows the same variations as the intersection volume. This means that a bijection should exist between the BC values and the intersection volume. Indeed, Fig. 9 shows the compiled values of BC for all tested nodes on all tested paths on the considered reconstructed feeder with respect to the value of the intersection volume obtained with $N_{rect} = 1000$ as reference. Different possible models have been tested. In this paper, we chose to present a model with trigonometric functions, as the BC is a projection of one PDF to another in an infinite dimension space. The fitted model describes the values of BC in function of intersection volume $\mathcal{V}$, as $BC = 0.81 \times \tan^{-1}(3\mathcal{V})$. The model for BC values is presented in red in Fig. 9. The total root mean square error (RMSE) between the two curves is $RMSE = 0.131$. This value is small enough to be confident in the fact that this model is a good representation of the link between BC values and intersection volume values. Indeed, we can observe in Fig. 8 the adjusted BC from the inversion of the empirical model, which projects the values of BC obtained from the model to the intersection values obtained from rectangle integration, in black on the figure. These values are very close to the volume computed with the rectangle method while needing only two computations for each node on the path.

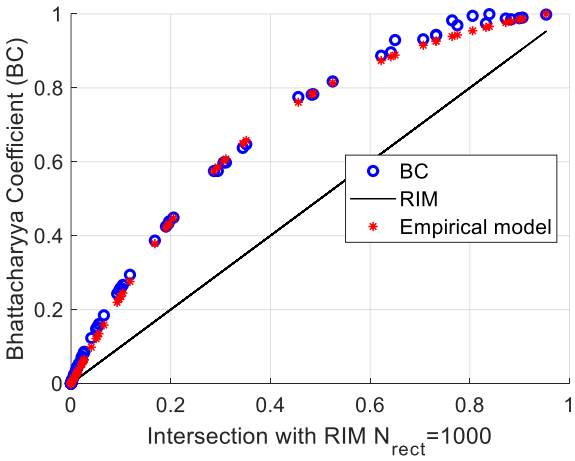

Fig. 9: BC values (blue) and model (red) as functions of intersection volume with rectangle method

IV. Locating Potential and Sensibility Analysis

A. Use case

In order to assess the sensitivity of the method with respect to the measurement errors, impedance uncertainties, and the probability threshold value, we clustered the feeder into 10 connected areas with equivalent fault probability and selected their barycenter as fault locations (taking into account the fact that underground cables are ten times less likely to be faulty than overhead lines). This way, the 10 considered fault location covers all parts of the feeder (near and far away from the busbar) with equal fault probability. On each fault position, a single phase to the ground fault with resistance of $500\,\Omega$ has been simulated. Moreover, for these 10 fault positions, a design of experiment has been built with 12 sets of possible impedance and measurement errors such that $\Delta Z_0 \in \pm 20\%$, $\Delta C_0 \in \pm 15\%$ and considering measurement errors up to the 1% accuracy class. In order to have a margin between the real variance of the random variable and the variance considered in the estimation process, the impedance and measurement errors are supposed normally distributed with a standard deviation being the third of the limit given by the norm. In the estimation process, the variable is still supposed uniformly distributed, this way the probability for the real value of the voltage to be in the lower probability parts of the estimated PDF is kept low. For each experiment, the FLM is performed 100 times with $N_{MC} = 100$ as in the previous part. Counting the fact that the 10 most priority nodes are simulated with measurements (cf. Fig. 3), a total of 327,600 different probability estimations are performed. For each case, 4 ways of computing $\mathbb{P}_\cap$ are tested: the intersection area with magnitudes (no synchronization), the volume intersection with rectangle method ($N_{rect} = 100$) with phasors (with synchronization), the Bhattacharyya coefficient and its model of the rectangle integration approximation (projected BC).

For each of the ten measurement nodes considered (the top ten measurement placements according to the optimal placement algorithm), the success rate is stored. The definition of the success rate is extended to the probabilistic method by counting the relative number of times (rate) in which the

accurate projection node is located inside the set of projection nodes exhibited.

Regarding the computation time, the first estimation of the ZS Gaussian parameters on a measurement node needs 135 ms when run on our computer equipped with an intel i7-10700 and 32GB of RAM. Then, the computation of the two estimate sets on the longest path (37 nodes) needs around 6.5 s to finish. Finally, the intersection volume computation needs around 6.9 s with the RIM while only needing 8 ms using the projected BC.

B. Success rate considering measurement errors

As expected, the probability threshold value has a paramount influence on the success rate of the method, as we can see on Fig. 10. Theoretically, the success rate (SR) should tend to 1 when the threshold is tending to 0. This is what can be seen on the figure. However, we observe that with the intersection volume (or with the projected BC) when using phasor, choosing a threshold of 50% does not lead to an acceptable value for the success rate when using measurements of 0.1% accuracy class (SR is around 43%). This can be explained by the fact that sometimes the erroneous measurement or line impedance value is drawn with high error, so that the real value is not in the central part of the distribution anymore, leading to a low intersection probability between the two estimates at the projection node. Besides, we observe that, at constant value for the threshold and the method to compute $\mathbb{P}_\cap$, there is no variation of the success rate in function of measurement errors with all intersection metrics except for the intersection volume using phasors and with higher (greater than 50%) values for the threshold.

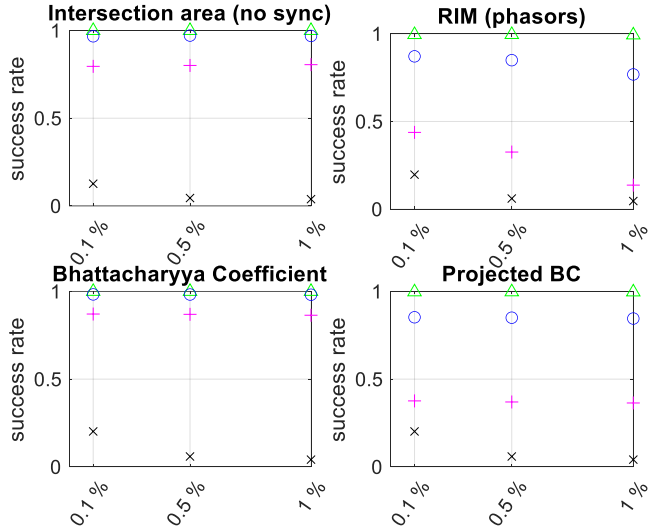

Fig. 10: Success rate with the 4 different intersection measures in function of the measurement accuracy classes. In each subfigure, 3 values of threshold are shown: $\tau = 5\%$ (green triangle), $\tau = 25\%$ (blue circle), $\tau = 50\%$ (red cross) compared with the non-robust method (black cross)

This is one of the biggest improvement of this formulation with respect to [24]. Indeed, the non-robust deterministic method would need measurement accuracy of 0.01% (which does not exist today) in order to reach 80% SR. With a threshold of 50%, the probabilistic formulation is able to maintain a SR higher than 80% on all tested measurement accuracy classes up to 1% (knowing that 0.5% is the most commonly found and 0.1% is commercially available). Considering 0.5% measurement accuracy, the SR have been increased by a factor of 8 from previous formulation. There is no more exponentially decreasing success rate and the method is behaving the same way considering 0.1%-accuracy class or 1%-accuracy class measurement devices. This means that this formulation is compatible with the use of real field measurement devices used by DSOs.

Moreover, we observe that the SR is higher when considering the BC or the intersection area without synchronization. This is the result of the fact that both BC and intersection area are higher or equal than the real volume under both surfaces, as seen on Fig. 8. Indeed, the intersection area is computed from projections of 2-D phasors onto the 1-D magnitude axis. This projection results in the loss of some information so that two PDFs that were disjointed could be seen as close to each other if they were to have close mean magnitudes. By comparison to the same threshold, using the BC (or intersection area) would lead to more nodes being exhibited as the projection ones (8 nodes instead of 5 on Fig. 8), which naturally increases the chance of the actual projection node not being inside the set of exhibited ones. However, one should keep in mind that this comes at the expense of an increase in the size of the solution area, meaning that while being more robust, the fault location will be less precise. This aspect will be studied in future work.

C. Success rate considering impedance values uncertainties

In this study, the performance is shown with respect to the presence and magnitude of uncertainties in impedance values. 4 cases are considered in which the uncertainty is quantified by a range $\Delta(R_0, X_0, C_0)$ so that the random variables (R_0, X_0, C_0) are drawn following a normal law of standard deviation the third of the Δ value while we consider this variable uniformly distributed in the estimation part (from the point of view of the algorithm):

- Case 1: low uncertainty $\Delta(R_0, X_0) \in [-5\%, +5\%]^2$, $\Delta C_0 \in [-5\%, +5\%]$
- Case 2: medium uncertainty $\Delta(R_0, X_0) \in [-10\%, +10\%]^2$, $\Delta C_0 \in [-5\%, +5\%]$
- Case 3: high uncertainty $\Delta(R_0, X_0) \in [-20\%, +20\%]^2$, $\Delta C_0 \in [-10\%, +10\%]$
- Case 4: very high uncertainty $\Delta(R_0, X_0) \in [-20\%, +20\%]^2$, $\Delta C_0 \in [-15\%, +15\%]$

Fig. 11 shows that the impedance uncertainties does not affect the ability of the algorithm to select a projection node anymore with the proposed method. Indeed, we observe that, at fixed value for the threshold value and the intersection estimation method, there is no visible difference in SR when knowing all the impedances parameters with a $\pm 5\%$ uncertainty or with $\pm 20\%$ for series impedances (R_0, X_0) and $\pm 15\%$ for C_0. This verifies the fact that the probabilistic FLM is robust to impedance uncertainties by design and can be applied to feeders with imperfect knowledge of the impedances. This is an important increase for the deployability of the method when we know that ZS impedances, especially those of distribution grids, are not well known.

Besides, we observe that the SR is decreasing with the value of the threshold. From this, a DSO could fit a model in order

to choose the optimal value for the threshold. Indeed, there is a tradeoff between SR and size of the solution area since lowering the value of the threshold leads to exhibiting more projection nodes, which increases the probability of the right projection node being in the selected ones while increasing the size of the solution area. A DSO could perform an interpolation of the SR value with the data corresponding to the accuracy class of the deployed measurement devices and then could choose the threshold value according to the targeted success rate value and given a tradeoff between probability of the fault to be outside of the solution area Ω_{sol}^{m} (for m additional ZS voltage measurements) and the size of the solution area. The deterministic algorithm presents a linearly decreasing SR when considering uncertainties on ZS series line impedance (without measurement errors). We observe that the decrease coefficient is twice as low as when using synchronized data than with unsynchronized ones. For instance, the SR of the deterministic method is below 80% when considering 20% uncertainties on (R_0, X_0). This is explained by the fact that ZS voltage magnitudes are very close to each other along the feeder in this case with compensated neutral grounding, meaning that a large set of nodes are selected as projection nodes with this data. This means that the greater the uncertainties, the larger the added value of the probabilistic formulation.

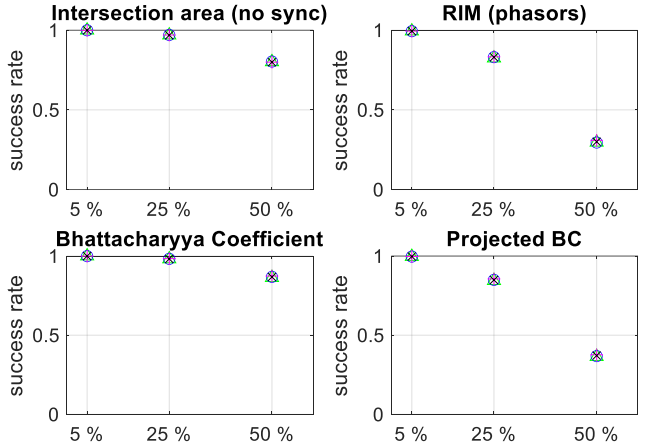

Fig. 11: Success rate with the 4 intersection measures in function of the threshold value. In each subfigure, 4 cases of line impedance uncertainties are shown: case 1 (green triangle), case 2 (blue circle), case 3 (red cross) and case 4 (black cross)

V. Conclusion

This paper has presented a probabilistic formulation which achieves more robust fault location on MV feeders with respect to both measurement and impedance uncertainties while leveraging few optimally placed additional measurement nodes. First, the ability of the Monte-Carlo process to accurately infer the PDFs parameters of a measurement has been demonstrated when considering its accuracy class. Besides, no accumulation of errors along the paths have been seen when considering possible errors on the ZS impedance values, even with very low number of random samplings. This explains why this method has been shown to be robust to those two potential error sources, which is a great improvement with respect to the deterministic formulation. However, the performances of this probabilistic algorithm are very sensitive to the choice of the value of the probability threshold τ. Indeed, a small threshold increases the probability of the fault being located laterally to one of the projection nodes, whilst having a large value of this threshold limits the number of projection nodes to the ones which are the most likely, meaning that the solution area would be of a size comparable to the best theoretical one (without any uncertainties). A detailed study of the locating performances, especially considering the size of the solution areas, still need to be carried out, as well as a more detailed study for the optimal choice of τ. A method without any probability threshold will be investigated in future works. Moreover, a more detailed study of the impact of the number of samplings N_{MC} would enable us to design a FLM with variable N_{MC} across the path, so that when the PDF variance is very low, a low number of samplings is used to estimate its parameters (ensuring a faster estimation process) while a higher number could be used when large variance variables are encountered (ensuring the accuracy of the estimation process along all nodes of all paths). Furthermore, in the case where the ratio μ/σ is large enough, a linearization and Taylor expansion could be used to obtain the PDFs parameters along the paths without need for Monte-Carlo processes. If the hypothesis were to be met, this would enable the design of a faster FLM. Finally, we could say that this zero-sequence fault location method is not impacted by presence of distributed generators if they are connected to the MV grid in a way that ensures no zero-sequence current injection (when the neutral is not distributed). However, if this hypothesis were not met, the DSO should place a zero-sequence current and voltage measurement in order to apply the method. This case should be investigated in future works.

References

[1] M. Shafiullah and M. A. Abido, "A Review on Distribution Grid Fault Location Techniques," *Electr. Power Compon. Syst.*, vol. 45, no. 8, pp. 807–824, May 2017, doi: 10.1080/15325008.2017.1310772.

[2] P. Stefanidou-Voziki, N. Sapountzoglou, B. Raison, and J. L. Dominguez-Garcia, "A review of fault location and classification methods in distribution grids," *Electr. Power Syst. Res.*, vol. 209, p. 108031, Aug. 2022, doi: 10.1016/j.epsr.2022.108031.

[3] M. Z. Habib, M. T. Hoq, S. Duvnjak Žarković, and N. Taylor, "Impact of the fault location methods on SAIDI of a resonant-earthed distribution system," in *2020 IEEE International Conference on Power Systems Technology (POWERCON)*, Sep. 2020, pp. 1–6. doi: 10.1109/POWERCON48463.2020.9230614.

[4] T. D. Le and M. Petit, "Earth fault location based on a Modified Takagi Method for MV distribution networks," in *2016 IEEE International Energy Conference (ENERGYCON)*, Apr. 2016, pp. 1–6. doi: 10.1109/ENERGYCON.2016.7513910.

[5] "VENTEEA," Smart Grids - Le site édité par la CRE. Accessed: Jun. 02, 2022. [Online]. Available: https://www.smartgrids-cre.fr/projets/venteea

[6] Kai-Ping Lien, Chih-Wen Liu, Chi-Shan Yu, and J.- Jiang, "Transmission network fault location observability with minimal PMU placement," *IEEE Trans. Power Deliv.*, vol. 21, no. 3, pp. 1128–1136, Jul. 2006, doi: 10.1109/TPWRD.2005.858806.

[7] Y. Liao, "Fault location observability analysis and optimal meter placement based on voltage measurements," *Electr. Power Syst. Res.*, vol. 79, no. 7, pp. 1062–1068, Jul. 2009, doi: 10.1016/j.epsr.2009.01.005.

[8] R. F. Buzo, H. M. Barradas, and F. B. Leão, "A New Method for Fault Location in Distribution Networks Based on Voltage Sag Measurements," *IEEE Trans. Power Deliv.*, vol. 36, no. 2, pp. 651–662, Apr. 2021, doi: 10.1109/TPWRD.2020.2987892.
[9] A. Bach, T.-D. Le, and M. Petit, "A zero-sequence impedance-based fault location method for MV distribution feeders with sparse measurements," presented at the 16th International Conference on Developments in Power System Protection (DPSP 2022), Newcastle, UK: IET Digital Library, Mar. 2022, pp. 7–12. doi: 10.1049/icp.2022.0903.
[10] C. A. Apostolopoulos, C. G. Arsoniadis, P. S. Georgilakis, and V. C. Nikolaidis, "Fault location algorithms for active distribution systems utilizing two-point synchronized or unsynchronized measurements," *Sustain. Energy Grids Netw.*, vol. 32, p. 100798, Dec. 2022, doi: 10.1016/j.segan.2022.100798.
[11] C. G. Arsoniadis, C. A. Apostolopoulos, P. S. Georgilakis, and V. C. Nikolaidis, "A voltage-based fault location algorithm for medium voltage active distribution systems," *Electr. Power Syst. Res.*, vol. 196, p. 107236, Jul. 2021, doi: 10.1016/j.epsr.2021.107236.
[12] 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 Deliv.*, vol. 22, no. 4, pp. 2541–2547, Oct. 2007, doi: 10.1109/TPWRD.2007.905535.
[13] S.-J. Lee *et al.*, "An intelligent and efficient fault location and diagnosis scheme for radial distribution systems," *IEEE Trans. Power Deliv.*, vol. 19, no. 2, pp. 524–532, Apr. 2004, doi: 10.1109/TPWRD.2003.820431.
[14] S. H. Mortazavi, M. H. Javidi, and E. Kamyab, "Robust Wide Area Fault Location Considering Network Parameters Error," *IEEE Trans. Power Deliv.*, vol. 34, no. 3, pp. 786–794, Jun. 2019, doi: 10.1109/TPWRD.2019.2897402.
[15] N. C. Woolley, M. Avendaño-Mora, J. V. Milanovic, and A. P. Woolley, "Probabilistic fault location using erroneous measurement devices," in *2011 IEEE International Conference on Smart Measurements of Future Grids (SMFG) Proceedings*, Nov. 2011, pp. 101–106. doi: 10.1109/SMFG.2011.6125781.
[16] E. O. Schweitzer, A. Guzmán, M. V. Mynam, V. Skendzic, B. Kasztenny, and S. Marx, "Locating faults by the traveling waves they launch," in *2014 67th Annual Conference for Protective Relay Engineers*, Mar. 2014, pp. 95–110. doi: 10.1109/CPRE.2014.6798997.
[17] P. Li *et al.*, "Precise Fault Location Method of Traveling Wave in Distribution Grid Based on Multiple Measuring Point," in *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, Oct. 2020, pp. 1867–1872. doi: 10.1109/EI250167.2020.9346873.
[18] I. Džafić, R. A. Jabr, S. Henselmeyer, and T. Đonlagić, "Fault Location in Distribution Networks Through Graph Marking," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1345–1353, Mar. 2018, doi: 10.1109/TSG.2016.2587583.
[19] L. Al-Kanj, B. Bouzaiene-Ayari, and W. B. Powell, "A Probability Model for Grid Faults Using Incomplete Information," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 956–968, Mar. 2017, doi: 10.1109/TSG.2015.2447275.
[20] Y. Jiang, "Data-Driven Probabilistic Fault Location of Electric Power Distribution Systems Incorporating Data Uncertainties," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4522–4534, Sep. 2021, doi: 10.1109/TSG.2021.3070550.
[21] J.-M. Li, "A single-phase-to-ground fault location method based on convolutional deep belief network," *Electr. Power Syst. Res.*, p. 11, 2022.
[22] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault Location in Power Distribution Systems via Deep Graph Convolutional Networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 1, pp. 119–131, Jan. 2020, doi: 10.1109/JSAC.2019.2951964.
[23] R. Vaish, U. D. Dwivedi, S. Tewari, and S. M. Tripathi, "Machine learning applications in power system fault diagnosis: Research advancements and perspectives," *Eng. Appl. Artif. Intell.*, vol. 106, p. 104504, Nov. 2021, doi: 10.1016/j.engappai.2021.104504.
[24] A. Bach, T. D. Le, and M. Petit, "Sensitivity assessment of a novel earth fault location method with optimally placed distributed measurements for MV networks," *IET Gener. Transm. Distrib.*, vol. 17, no. 6, pp. 1358–1367, 2023, doi: 10.1049/gtd2.12740.
[25] N. George and O. D. Naidu, "Estimation of Zero-Sequence Impedance Parameters in Double-Circuit Lines Using Disturbance Recorder Data," in *2020 IEEE International Conference on Power Systems Technology (POWERCON)*, Sep. 2020, pp. 1–6. doi: 10.1109/POWERCON48463.2020.9230574.
[26] F. G. Venegas, "Electric vehicle integration into distribution systems : Considerations of user behavior and frameworks for flexibility implementation," phdthesis, Université Paris-Saclay, 2021.
[27] "IEC 61869-3 - Instrument transformers – Part 3: Additional requirements for inductive voltage transformers | Engineering360."
[28] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," 1943.
[29] STE EDF GDF services, "PLAN DE PROTECTION DES RESEAUX HTA. B.61-21. PRINCIPES." Feb. 01, 1994.

Alexandre BACH graduated in electrical engineering from the École Normale Supérieure Paris-Saclay, Gif-sur-Yvette in 2020. He received a Ph.D. degree from University Paris-Saclay, Gif-sur-Yvette in 2023. He is focusing on fault location and real-time simulation.

Trung-Dung LE received a Ph.D. degree in electrical engineering from CentraleSupélec, Gif-sur-Yvette, in 2014. He is currently an Assistant Professor with CentraleSupélec and a researcher with the GeePs Lab, Gif-sur-Yvette. His research interests include power system protection, distribution, automation, and power quality.

Marc Petit Graduated from the Ecole Normale Supérieure de Cachan, Cachan and received the Ph.D. degree in 2002 from the University Paris-Sud, Orsay. He is currently a Professor with CentraleSupélec, Gif-sur-Yvette, and GeePs Lab where he leads the Electrical Network Research Group. He is the Co-head of the Armand Peugeot chair about the economy of electromobility. His research interests include smart grids, which include EV and RES integration, power quality, and power system protection.