Probabilistic characterization of fault resistance on transmission networks

Patrícia D. F. Duarte
Instituto Superior Técnico
Universidade de Lisboa
Lisboa, Portugal

Abstract—Electric power systems are daily exposed to service interruptions caused by faults due to different causes. Any fault in the power system has a certain resistance value, referred to as fault resistance. Since it depends on several factors, it can be unpredictable and therefore it is relevant to study its value probabilistically in order to understand its behavior.

This paper proposes a probabilistic model that describes the fault resistance, through the study of fault resistance values obtained for 100 faults, recorded in the RNT, studied in this paper. Fault resistance values are computed from the implementation and development of a fault location algorithm that uses voltage and current measurements at both ends of the line where the fault occurred.

Keywords—fault resistance, fault

I. INTRODUCTION

Electric power systems are constantly exposed to service interruptions caused by faults due to different causes, such as lightning, insulators pollution, birds and fires. Each fault in the power system has a certain resistance value, referred to as fault resistance. Fault resistance is composed by electric arc resistance, effective grounding resistance of towers and resistance of strange objects placed between the conductor and the ground [8].

The contribution of fault resistance in the analysis of faults in transmission lines has been studied and worked by several authors for some years [7], [8], and resistance values were obtained through studies for lines of different voltage levels. Because it is uncertain and dependent on several factors, it can present unpredictable values and therefore it is important to study them probabilistically in order to understand their behavior. Previous studies, [9] and [10] present probabilistic models proposed by other authors to characterize the fault resistance in transmission lines. This paper proposes a probabilistic model that describes the fault resistance, using for this, values calculated from records collected in the RNT. This calculation is performed by applying a fault location algorithm, which uses voltage and current measurements, recorded at both ends of the line where the fault occurred.

The algorithm is applied to 100 faults recorded in the network considered in this paper, corresponding to the RNT network. Two probabilistic models are proposed, and based on the exploratory data analysis (EDA) is chosen the most appropriate model to describe fault resistance.

II. COMPONENTS OF FAULT RESISTANCE

The resistance which characterizes each fault depends on the type and cause of the fault. Only faults which involve ground will be considered and so the value of fault resistance, $R_F$, is a combination of three factors, being them: the electric arc resistance, $R_{arco}$, the effective grounding resistance of towers, $R_{ref}$ and the resistance of strange objects between the line and ground, $R_0$, and so it’s defined by:

$$R_F = R_{arco} + R_{ref} + R_0$$ (1)

A. Electric Arc Resistance

The electric arc is a non-linear phenomenon which depends on several factors, however is common to consider the electric arc as a resistance, this one being dependent of the short-circuit current and the electric arc length, which can vary with the type of fault, and so it can correspond to the length of the insulators chains if the fault is on the transmission towers or the distance between the lines and the ground, if the fault is on the lines.

Therefore, Warrington [1] defines the electric arc resistance as:

$$R_{arco} = \frac{28707,351}{I^2}$$ (2)

where $R_{arco}$ represents the electric arc resistance [Ω], $I$ represents the electric arc length [m] and $I$ the short-circuit current [A].

B. Resistance of objects between conductor and ground.

The value of fault resistance is unpredictable in the way that strange objects can be placed between the line and ground. The resistance of the objects is a fixed value, varying only with these ones and is considered resistive.

If the object is metallic then its resistance is very low and therefore despicable. If the object is, for example, a tree the value can be very high [2]. As will be seen, the obtained
resistance values are always very low so this event is very rare in transport networks.

C. Effective Grounding Resistance of towers

In order to determine the effective grounding resistance of towers, is assumed that the studied network has ground wires and so, on a phase to earth fault in a tower, part of the current flows through each tower and the other flows through the ground wire. This way the current flows to ground through several ground resistances of towers in parallel.

The resistance of towers and ground wires can be represented by several T linked in series. This means that the effective ground resistance of towers, $R_{T\text{-apoio}}$, is different from the tower resistance individually, $R_{T\text{-apoio}}$.

The ground resistance of each tower is given by:

$$ R_{T\text{-apoio}} = k \times \rho $$  \hspace{1cm} (3)

where $k$ represents a constant which varies with the electrode shape and $\rho$ represents the ground resistivity [$\Omega m$] which is the same for all the towers involved since they are located in the same region because ground resistivity is characteristic of each region.

Since that ground wires interconnect towers, the effective ground resistance is dependent on the combination of several towers involved, namely the number of towers, and since that ground wires have low influence in ground effective resistance, this is also proportional to ground resistance of each tower.

$$ R_{T\text{-apoio}} = k \times \rho $$  \hspace{1cm} (2.4)

where $k$ is a constant, different from the previous one and $\rho$ represents the ground resistivity [$\Omega m$].

The fact that ground wires, by interconnecting all towers, placing ground resistance of each transmission tower individually in parallel with ground resistance of nearby transmission towers, creates a drastic reduction on effective grounding resistance [3].

III. FAULT RESISTANCE ESTIMATION

Computation of fault resistance is directly linked with the determination of fault location and therefore two algorithms are presented for locating the fault and posteriorly for compute fault resistance. Despite both algorithms being distinct, both are based on voltage and current measurements. It is also assumed for both algorithms a short line model (<100 km) and lumped parameters.

A. Algorithm based on measurements on a line end

The working principle of algorithm [4], is based on the analysis of the monophasic circuit of figure 1, the simplest for analyzing, which has a representation of the line where the fault occur, a representation of the rest of the network and the fault location.

This way, on the figure 1, F represents the location where the fault occurred, a pure resistive one, represented by a resistor $R_f$, representing the fault resistance which has to be computed. $Z$ represents the transmission line impedance, which is known, and each end of the line, $S$ and $R$, are represented by its Thévenin equivalent. Extremity $S$ is represented by voltage source $\bar{U}_S$ and by network impedance $\bar{Z}_S$ and extremity $R$ is represented by voltage source $\bar{U}_R$ and by network impedance $\bar{Z}_R$. $d$ represents distance to the fault from extremity $S$, to be computed. Since this algorithm uses current and voltage measures from one end of the line, only $\bar{V}_S$ and $\bar{I}_S$ are known values.

According to Kirchhoff laws:

$$ V_S = dZI_s + R_I f $$  \hspace{1cm} (4)

$$ I_f = I_s + I_e $$  \hspace{1cm} (5)

Once that $\bar{I}_e$ is unknown (since this algorithm only knows current and voltage on one end of the line) $\bar{I}_f$ is also unknown, since beyond that, $R_f$ is also unknown it’s not possible to get the value of $d$ from (4). For that, it is assumed that current $\bar{I}_s$ is proportional to the current variation in $S$, according to [4] so:

$$ I_s - I_{s0} = \Delta I_s = k_d I_f $$  \hspace{1cm} (6)

where $I_{s0}$ corresponds to the current before the fault and $k_d$ corresponds to the distribution factor which describes the distribution of the circuit currents, assumed as 1 according to [5].

Therefore is obtained:

$$ V_f = dZI_s + R_f(I_s - I_{s0}) $$  \hspace{1cm} (7)

where $d$ and $R_f$ are the only unknown values, which are the wanted values. Since equation (7) is a complex one, it can be separated into real and imaginary part,

$$ \begin{cases} \text{Real} \{V_f\} = d \times \text{Real} \{ZI_s\} + R_f \times \text{Real} \{(I_s - I_{s0})\} \\ \text{Imag} \{V_f\} = d \times \text{Imag} \{ZI_s\} + R_f \times \text{Imag} \{(I_s - I_{s0})\} \end{cases} $$  \hspace{1cm} (8)

and a system with two equations and two unknown variables is obtained, $d$ and $R_f$ whose solution is given by:

$$ x = A^{-1}b $$  \hspace{1cm} (9)

with

$$ x = \begin{bmatrix} d \\ R_f \end{bmatrix} $$  \hspace{1cm} (10)

being
\[
A = \begin{bmatrix}
Z_{s_{null}} I_{s_{null}} - Z_{i_{null}} I_{i_{null}} & I_{s_{null}} - I_{s_{null}} \\
Z_{i_{null}} I_{i_{null}} - Z_{i_{null}} I_{i_{null}} & I_{i_{null}} - I_{i_{null}}
\end{bmatrix}
\] (11)

and
\[
b = \begin{bmatrix}
V_{s} \\
V_{i}
\end{bmatrix}
\] (12)

This way, it is possible to compute the values of distance to fault, \(d\) and the fault resistance, \(R_f\).

B. Algorithm based on measurements on both line ends

The working principle of the algorithm is also based on the analysis of meshes of the circuit from figure 1, however since are known the measurements of both ends of the line, the right mesh is also going to be analyzed [6], obtaining for this mesh:
\[
\overline{V}_s = (1-d)\overline{Z}_l \overline{I}_s + \overline{V}_r
\] (13)

where
\[
\overline{V}_r = R_f \overline{I}_r
\] (14)

It is now possible to obtain two equations in order to \(\overline{V}_r\) through equations (4) and (13), respectively:
\[
\overline{V}_r = \overline{V}_s - d\overline{Z}_l \overline{I}_s
\] (15)

and
\[
\overline{V}_r = \overline{V}_s - (1-d)\overline{Z}_l \overline{I}_s
\] (16)

By matching equations (15) and (16) and solving in order to \(d\) it is obtained:
\[
d = \frac{\overline{V}_r - \overline{V}_s + \overline{Z}_l \overline{I}_s}{\overline{Z}_l (\overline{I}_s + \overline{I}_s)}
\] (17)

This way it can be computed the value of distance to fault seen from end S.

Having the distance to the fault computed, \(\overline{V}_r\) is computed from (16). \(\overline{I}_r\) can also be computed from (5) and once computed \(\overline{V}_r\) and \(\overline{I}_r\) it’s now possible to compute the fault resistance, \(R_f\), through
\[
R_f = \frac{\overline{V}_r}{\overline{I}_r}
\] (18)

C. Admeasurement of the application of the fault location algorithms to the computation of fault resistance

In order to test the correct working and viability of the previous algorithms some simulations were made in the EMTP-RV software, for a phase to ground short-circuit on the test network presented in figure 2, relative to “RNT Sul”.

For each algorithm fault resistance values were computed and compared with a defined set of fault resistance values, in EMTP-RV simulation, for six different cases, namely considering resistances of 0, 10, 20, 30, 40 and 50 \(\Omega\).

![Test Network – “RNT Sul”](image)

Tables I and II present the errors associated to the use of each algorithm, for the short-circuit near Sines, showed in figure 2 and simulated in EMTP-RV.

Through the analysis of table I it is possible to notice extremely high errors associated to the algorithm using only information from one end of the line for each one of the six cases studied, with errors between 270 and 290%. It is also noticeable that as the fault resistance values increase so does the absolute error. This doesn’t happen with the relative error since that all values are fairly close, however are quiet high which leads to conclude that this algorithm is inaccurate in its results and therefore not viable for being used on getting fault resistance values.

Table II allows to see much more acceptable values for both absolute and relative errors with the algorithm using information on both ends of the line, being verified errors between 5 and 23%. To note that as fault resistance values increase absolute error slightly increases and relative error decreases.

According to the previous analysis it can be concluded that the algorithm using information of both ends of the line is much more accurate for computing fault resistance values on phase to neutral faults in the studied network.

| TABLE I |
|-----------------|-----------------|-----------------|
| Fault Resistance [\(\Omega\)] | Fault Resistance value through 1st algorithm [\(\Omega\)] | Absolute error [\(\Omega\)] | Relative error [%] |
| 0                | 1,013            | 1,013            | -               |
| 10               | 38,851           | 28,851           | 288,512         |
| 20               | 76,197           | 56,197           | 280,986         |
| 30               | 113,476          | 83,476           | 278,252         |
| 40               | 150,749          | 110,749          | 276,8715        |
| 50               | 188,022          | 138,022          | 276,044         |
IV. FAULT RESISTANCE ESTIMATION ON RNT

In order to use the algorithm chosen in III to real cases, 200 records corresponding to 100 incidents occurred along three years (2013, 2014 and 2015) were used. Each incident has two records matching the same fault, and so to the same transmission line but seen from both ends of the line, however fault resistance value will be the same since both refers to the same incident. So, in order to avoid repetition of results these will be shown by incident. This way, each incident has information about the transmission line were the fault occurred, for a total of 44 different transmission lines for voltage levels of 150, 220 and 400 kV and yet substations connected to the transmission line, for a total of 44.

Table V presents fault resistance values for each incident occurred, from the algorithm using information on both ends of the line, taking into account the transmission line were the fault was verified, the matching substations on each end of the line and the voltage level in question.

Transmission lines are represented by the letter L followed by a number identifying the line in question, the same way substations are represented by the letter S followed by a number identifying the substation.

<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSOLUTE AND RELATIVE ERRORS OF ALGORITHM USING INFORMATION OF BOTH ENDS OF THE LINE FOR THE SINES-PORTIMÃO FAULT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1,943</td>
<td>1,943</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>12,298</td>
<td>2,299</td>
<td>22,997</td>
</tr>
<tr>
<td>20</td>
<td>22,491</td>
<td>2,491</td>
<td>12,454</td>
</tr>
<tr>
<td>30</td>
<td>32,653</td>
<td>2,653</td>
<td>8,843</td>
</tr>
<tr>
<td>40</td>
<td>42,808</td>
<td>2,808</td>
<td>7,020</td>
</tr>
<tr>
<td>50</td>
<td>52,948</td>
<td>2,948</td>
<td>5,896</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE V</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAULT RESISTANCE VALUES OBTAINED THROUGH THE USAGE OF ALGORITHM USING INFORMATION OF BOTH ENDS OF THE LINE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incident</th>
<th>Voltage level [Kv]</th>
<th>Transmission Line</th>
<th>Substations</th>
<th>Fault Resistance [Ω]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>L1</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>L1</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>L2</td>
<td>S3</td>
<td>S4</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>L3</td>
<td>S5</td>
<td>S6</td>
</tr>
<tr>
<td>5</td>
<td>150</td>
<td>L4</td>
<td>S7</td>
<td>S8</td>
</tr>
<tr>
<td>6</td>
<td>400</td>
<td>L5</td>
<td>S9</td>
<td>S10</td>
</tr>
<tr>
<td>7</td>
<td>220</td>
<td>L6</td>
<td>S11</td>
<td>S12</td>
</tr>
<tr>
<td>8</td>
<td>400</td>
<td>L7</td>
<td>S13</td>
<td>S1</td>
</tr>
<tr>
<td>9</td>
<td>400</td>
<td>L7</td>
<td>S13</td>
<td>S1</td>
</tr>
<tr>
<td>10</td>
<td>400</td>
<td>L1</td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>11</td>
<td>150</td>
<td>L8</td>
<td>S14</td>
<td>S15</td>
</tr>
<tr>
<td>12</td>
<td>400</td>
<td>L2</td>
<td>S3</td>
<td>S4</td>
</tr>
<tr>
<td>13</td>
<td>400</td>
<td>L2</td>
<td>S3</td>
<td>S4</td>
</tr>
<tr>
<td>14</td>
<td>400</td>
<td>L2</td>
<td>S3</td>
<td>S4</td>
</tr>
<tr>
<td>15</td>
<td>150</td>
<td>L9</td>
<td>S16</td>
<td>S17</td>
</tr>
<tr>
<td>16</td>
<td>150</td>
<td>L10</td>
<td>S17</td>
<td>S18</td>
</tr>
</tbody>
</table>
According to the fault resistance obtained values, an analysis is done by voltage level, namely, the percentage of cases verified for different intervals of fault resistance values and a comparison is made, in figure 3, 4 and 5, with values published in literature for the same range of values [7], [8].

<table>
<thead>
<tr>
<th>Voltage Level</th>
<th>Fault Resistance (Ω)</th>
<th>Number of Incidents</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>220 kV</td>
<td>0-1</td>
<td>28</td>
<td>5.76</td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>29</td>
<td>5.64</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>37</td>
<td>7.14</td>
</tr>
<tr>
<td></td>
<td>&gt;10</td>
<td>43</td>
<td>8.56</td>
</tr>
<tr>
<td>230 kV</td>
<td>0-1</td>
<td>28</td>
<td>5.28</td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>28</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>34</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>&gt;10</td>
<td>42</td>
<td>8.32</td>
</tr>
<tr>
<td>240 kV</td>
<td>0-1</td>
<td>28</td>
<td>5.16</td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>29</td>
<td>5.71</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>36</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>&gt;10</td>
<td>40</td>
<td>7.93</td>
</tr>
</tbody>
</table>

In figure 3 some differences are noted in the obtained results since that in this paper no values were obtained below 3 \( \Omega \) whereas according to [8] about 52\% of the noticed cases were in this range, however it is noteworthy the big difference in the dimension of the studied population for both sources. For fault resistance values above 10 \( \Omega \) the results are similar.

For figure 4 there are some noticeable differences between the obtained results, especially in the maximum percentages obtained. In this paper the maximum percentage of cases observed, about 67\%, was verified for resistance values between 1 and 3 \( \Omega \). For [8] the maximum percentage, 70\%, was verified for values between 3 and 10 \( \Omega \) and for [7] the maximum percentage, 75\%, was verified for values between 0 and 1 \( \Omega \). Only for values greater than 10 \( \Omega \) are seen similarities between the obtained results. The dimension of the population is similar between this paper and [8] with a slight increase for [7].

For figure 5, the shown results are pretty similar, in the way that the obtained percentages by increasing order occur for the same intervals of resistance. Therefore, the maximum percentages for the two sources of results occur for the interval between 1 and 3 \( \Omega \), followed by the interval between 3 to 10 \( \Omega \) and in the end for values between 0 and 1 \( \Omega \). For [7] 3.6\% of incidents occur for values above 10 \( \Omega \), not being verified any case in this paper. The difference in obtained percentages between the two sources of results is also similar for each interval, with a slight increase for the second one. The dimension of the population is similar, with just one incident of difference.

V. FAULT RESISTANCE PROBABLISTIC ANALYSIS

A. Adopted probabilistic distributions

1) Observed values histograms

Table VI shows a pre-analysis of values get in IV were a substantial difference between samples size is noticeable for the several voltage levels.
Figures 5 to 8 show histograms of the distribution of fault resistance values by voltage level separately and together. The red circle is the data mean value. Each bar represents observation density given by:

\[
\text{density} = \frac{\text{number of observations}}{\text{total number of observations} \times \text{class width}}
\]  

(19)

Table VI

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>62</td>
<td>37.5100</td>
<td>0.7700</td>
<td>4.3253</td>
<td>2.9550</td>
</tr>
<tr>
<td>220</td>
<td>9</td>
<td>61.0000</td>
<td>0.0100</td>
<td>2.5844</td>
<td>2.3800</td>
</tr>
<tr>
<td>150</td>
<td>29</td>
<td>4.3600</td>
<td>0.1000</td>
<td>2.1431</td>
<td>1.9700</td>
</tr>
<tr>
<td>All</td>
<td>100</td>
<td>37.5100</td>
<td>0.0100</td>
<td>3.5357</td>
<td>2.7250</td>
</tr>
</tbody>
</table>

Through the analysis of histograms, it’s possible to check, for voltage level 400 kV and for all the set of voltage levels a great density of observations with resistance values below 10 Ω, with an accentuated decrease, tending to zero, as fault resistance values increase.

For voltage level 150 kV this characteristic is not as clear, however it’s noticeable the short range of fault resistance values for this voltage level when comparing with the value range of the others voltage levels which is much more embracing.

For the voltage level 220 kV it’s important to emphasize the dimension of the population in study, which, because it’s rather small, does not allow to conclude the correct, or not, adequacy of a possible theoretical distribution to the data distribution and so, this voltage level will not be studied individually, being only use in the set of all voltage levels.

From the probabilistic distributions who present the characteristic identified initially, two were identified as the most potentially adequate, the Weibull distribution and the lognormal distribution.

1) Weibull distribution

Weibull distribution is adopted to describe fault resistance according to previous work developed by several authors [9], [10] where it was concluded that the Weibull probability distribution was the one who suited better to the obtained fault resistance distribution.

Weibull distribution is characterized by a probability density function given by:

\[
f(R_f) = \frac{\beta}{\delta} \left( \frac{R_f}{\delta} \right)^{\beta-1} \exp \left[ -\left( \frac{R_f}{\delta} \right)^\beta \right]
\]  

(20)

and a cumulative distribution function given by

\[
F(R_f) = 1 - \exp \left[ -\left( \frac{R_f}{\delta} \right)^\beta \right]
\]  

(21)

were \( R_f \) represents the fault resistance and \( \beta \) and \( \delta \) represents the shape parameter and scale parameter.

2) Normal distribution in a logarithmic scale

The adequacy of the Normal distribution in a logarithmic scale is substantiated in two studies [11] and [12] where it is concluded that soil resistivity, of uncertain nature, and dependent of several factors, characteristics of each soil properties [13], can be described as a Normal distribution in a logarithmic scale.

Considering that soil has a linear behavior, soil resistivity is directly related, in a logarithmic scale, towers resistance for certain electrodes geometry. Since soil resistivity can be statistically described by a normal distribution in a logarithmic scale, so the towers resistance is too.

With towers resistance being a factor with a considerable weight in fault resistance, it is also assumed that fault resistance can also be described by a normal probability density function in a logarithmic scale.

Normal distribution is characterized by a probability density function given by:

\[
f(\log_{10}(R_f)) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{(\log_{10}(R_f) - \mu)^2}{2\sigma^2} \right]
\]  

(22)

and a cumulative distribution function given by:

\[
F(\log_{10}(R_f)) = \Phi \left( \frac{\log_{10}(R_f) - \mu}{\sigma} \right)
\]  

(23)

where \( R_f \) represents fault resistance and \( \mu \) and \( \sigma \) represents the location parameter, which corresponds to the mean value and
scale parameter which corresponds to standard deviation, respectively.

\( \Phi(\log_{e}(R)) \) corresponds to the standard normal deviation, this is, the normal distribution with mean equal to 0 \((\mu=0)\) and variance equal to 1 \((\sigma^2 = 1)\).

**B. Adopted procedure for fitment analysis of the chosen distributions - EDA**

Exploratory data analysis is an approach that focuses on identifying general patterns in data and mark atypical values that show deviations in relation to others. Contrary to classical methods, who usually start to assume a model data, EDA, through data analysis suggests models that can be appropriate.

Most EDA techniques are of graphical nature, allowing a study more qualitative study, meeting this way EDA main objectives.

The theoretical quantile-quantile plot is a graphical tool, based on EDA, which allows evaluating if a given data set follows a theoretical distribution since it provides a match between the empirical quantiles, coming from the data set and the related quantiles from the theoretical distribution assumed.

For the plot construction \([14]\) it is assumed \(y_i\) for \(i=1\ldots n\), the set of fault resistance values and \(y_{0i}\) the set of the same values sorted by ascending order. Being \(p\) any fraction between 0 and 1, it’s defined the empirical quantile \(Q_i(p)\) corresponding to \(p\) as being \(y_{0i}\) always that \(p \in [0,1]\). It is also assumed \(F(y)\) as being the cumulative distribution function, in this way, the plot points will be close of the theoretical distribution parameter.

The theoretical quantile \(p\) of \(F\), \(Q_i(p)\), satisfies

\[ F(Q_i(p)) = p \leftrightarrow Q_i(p) = F^{-1}(p). \]

According to Chambers \([12]\), for a normal distribution, empirical quantiles, represented in the ordinate axis, are given by \(y_{0i}\), and the theoretical quantiles represented in the abscissa axis are given by \(\Phi^{-1}(p)\), where \(\Phi\) represents the standard distribution function of the normal distribution, for \(p_i = (i-0.5)/n\).

In another way, for a Weibull distribution, the empirical quantiles, represented in the ordinate axis, are given by \(y_{0i}\), and the theoretical quantiles, represented in the abscissa axis, corresponding to the inverse function of the cumulative distribution function, are given by \(\log_{i}[-\log_{i}(1-p_i)]\), for \(p_i = (i-0.5)/n\).

The theoretical quantile-quantile plot is only a visual tool, which somehow makes it subjective, however it allows to evaluate if a theoretical distribution fits to a set of data, once, if it suits, the points on the plot follow a well-defined behavior, this, about a straight line.

In a situation for which the theoretical distribution is a good approximation of the empirical one, is when the data quantiles will match, approximately, to the theoretical quantiles and in this way, the plot points will be close of the straight \(y = x\). Sometimes the behavior of the plot points doesn’t match exactly to the straight \(y = x\), which may be displaced (the distributions present different locations) or compressed/expanded (the distributions presents different scales). Sometimes, in addition to the distributions presents different scales and locations, individual points emerge, or a set of points, diverted from the normal behavior of the remaining ones, which can compromise the match between two distributions, being so necessary having them into consideration and check if its influence is relevant to the point of the assumed theoretical distribution is no longer valid.

An outlier is an observation which presents an abnormal distance in relation to the others values observed in a data set, being so inconsistent with them.

The study of outliers of a given data set starts with identifying possible outliers, through a visual analysis of the theoretical quantile-quantile plots, signaling all the points who present an abnormal distance to the estimated straight through the theoretical distribution parameters.

The question that arises is whether those possible points have spread really bigger than expected. This way, in order to eliminate subjectivity it’s necessary to check them through methods already existent in literature, namely the Tukey method \([15]\).

A second cause, also common, for the existence of irregularities in the theoretical quantile-quantile plot, corresponding to spreads in relation to the estimate straight is the existence of tails in both ends of the plot. This case is frequently observed in theoretical quantile-quantile plots corresponding to normal distributions, where is noticeable that the straight is adjusted to the central points and that comparatively to the points mean value in the plot, points in the tail have lots of dispersion and are spread away from the straight.

The correct adequation of the data theoretical distribution is evaluated by the proximity of the theoretical quantile-quantile plot points to the straight that defines the majority ones, being this straight the linear regression one.

For defining the straight of a simple linear regression, it’s computed, using the theoretical quantile-quantile plot points, two estimators \(\hat{\beta}_0\) and \(\hat{\beta}_1\), through the least squares method \([16]\), corresponding, respectively, to the y-intercept and to the slope of the straight.

Therefore, the estimated linear regression straight is given by:

\[ y = \hat{\beta}_0 + \hat{\beta}_1x \quad (24) \]

Having determined the equation of the linear regression straight, it’s possible to obtain the parameters for the theoretical distribution in question from its slope and y-intercept, given by the estimators.

According to Chambers \([14]\), to a normal distribution, the slope and y-intercept, of the linear regression straight estimated from the theoretical quantile-quantile plot, correspond exactly to the \(\sigma\) and \(\mu\) parameters of that distribution, respectively.

Regarding the Weibull distribution, the relation it’s not exactly direct, since the slope of the linear regression straight, of the theoretical quantile-quantile plot matches to the inverse of the \(\beta\) parameter and the y-intercept matches to the
natural logarithm of the $\delta$ parameter. This way, these distribution parameters are computed by:

$$\beta = \left(\hat{\beta}\right)^{-1}$$

(25)

and

$$\delta = e^{\hat{\delta}}.$$  

(26)

Knowing the parameters of the distribution in question, it is possible to characterize and define the probability density function.

C. Adjustment to theoretical distributions

For each voltage level, 400 kV and 150 kV, separately and together (with the contribution of the 220 kV level), are presented the Weibull distribution (figure 9 to 17) and normal distribution in a logarithmic scale (figure 17 to 26), respectively: a quantile-quantile plot, a histogram of the data distribution with the respective probability density function adjusted to the data and obtained through the calculated parameters and a cumulative distribution histogram of data with the respective cumulative distribution function obtained too through the computed parameters.

With Weibull distribution, for 400 kV voltage level it’s verified for theoretical quantile-quantile plot that the linear regression straight is well fitted to the plot points, with an area where is seen a big point density over the straight, however in the extremities some tails are noticed (distant values from the data mean value), with greater incidence on the left where some points more scattered from the other, which indicates an asymmetry to the mean value, but yet close enough to the straight. It is also noticeable a good adequation of Weibulls probability density and cumulative distribution function to the data distribution in the way that both are well adjusted to the defined classes. For 150 kV there is a clear discrepancy between Weibull probability density function and cumulative distribution function and data distribution since that probability density doesn’t fit to each bar density. When comparing with theoretical quantile-quantile plot the straight is relatively well adjusted to the majority of points, however as the points get closer to the edged it’s noticeable some data scattering to the linear regression straight. For the set of all voltage levels data points are well adjusted to the theoretical quantile-quantile plot straight. It can be observed two points very far away from the linear regression straight, but the re-
main points are well defined over this one. It’s also noticeable that probability density function and cumulative distribution function are correctly adjusted to data distribution. Since this model presents very unreliable results for voltage level 150 kV this one is assumed as inappropriate for describe fault resistance.

Figure 18 – Theoretical quantile-quantile plot for the normal distribution in a logarithmic scale for 400 kV.

Figure 19 – Normal probability density function in a logarithmic scale fitted to fault resistance values for 400 kV.

Figure 20 – Normal cumulative distribution function in a logarithmic scale fitted to fault resistance values for 400 kV.

Figure 21 – Theoretical quantile-quantile plot for the normal distribution in a logarithmic scale for 150 kV.

Figure 22 – Normal probability density function in a logarithmic scale fitted to fault resistance values for 150 kV.

Figure 23 – Normal cumulative distribution function in a logarithmic scale fitted to fault resistance values for 150 kV.

Figure 24 – Theoretical quantile-quantile plot for the normal distribution in a logarithmic scale for all voltage levels.

Figure 25 – Normal probability density function in a logarithmic scale fitted to fault resistance values for all voltage levels.

Figure 26 – Normal cumulative distribution function in a logarithmic scale fitted to fault resistance values for all voltage levels.

Regarding normal distribution in a logarithmic scale, for 400 kV voltage level and for all voltage levels set it’s seen for the theoretical quantile-quantile plot that, in relation to data central point, observed values in tails are more scattered meaning that they are more far away from the median than they should in a normal distribution, however this characteristic is pretty common for theoretical quantile-quantile plots for normal distributions. Yet, tail values are still close to the linear regression straight. It’s yet seen a good fitment of probability density function and cumulative distribution function to the data distribution. For 150 kV, to the theoretical quantile-quantile plot, despite a small data population there is a good fitment of linear regression straight to the plot points despite some noticeable tails on the edges, with just one value on the right edge more spread from the straight. It’s seen a good fitment from probability density function and cumulative distribution function to the data distribution with a slight deviation of probability density function maximum value in relation to data maximum density. This way it can be concluded that the probabilistic model based on lognormal distribution is well fitted to characterize fault resistance values for all voltage levels.

In a comparison with [9] and [10] is verified that there are incompatibilities on the proposed probabilistic methods. While in [9] and [10] is proposed a probabilistic method based on Weibull distribution, in this paper it was concluded that the Weibull distribution is inappropriate to describe the values of fault resistance to 150 kV, and so the normal distribution in a logarithmic scale is the chosen one to describe fault resistance of the voltage levels studied.
Tables VII and VIII presents the values of the parameters for both distributions, according to the previous procedure.

### TABLE VII
PARAMETERS OF WEIBULL DISTRIBUTION

<table>
<thead>
<tr>
<th>Voltage level [kV]</th>
<th>( \delta )</th>
<th>( \beta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>3.2086</td>
<td>3.0811</td>
</tr>
<tr>
<td>220</td>
<td>2.4610</td>
<td>1.7474</td>
</tr>
<tr>
<td>150</td>
<td>3.0215</td>
<td>1.8111</td>
</tr>
</tbody>
</table>

### TABLE VIII
PARAMETERS OF NORMAL DISTRIBUTION IN A LOGARITHMIC SCALE

<table>
<thead>
<tr>
<th>Voltage level [kV]</th>
<th>( \mu )</th>
<th>( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>0.4437</td>
<td>0.1736</td>
</tr>
<tr>
<td>220</td>
<td>0.2947</td>
<td>0.2197</td>
</tr>
<tr>
<td>150</td>
<td>0.3993</td>
<td>0.1944</td>
</tr>
</tbody>
</table>

Through analysis of table VII it is verified that the shape parameter is always greater than 1 for any of the voltage levels and that for the voltage level of 400 kV there is more symmetry since this parameter is higher for this level in comparison with voltage level of 150 kV.

Through the analysis of table VIII it is verified that the mean value increases as the voltage level increases, this means that the probability density function for voltage level of 400 kV is located more to the right than for the voltage level of 150 kV. The opposite is true for the standard deviation, since the standard deviation decreases as the stress level increases. This means that for the voltage level of 400 kV the values are closer to the mean value than to the voltage level of 150 kV. The opposite is true for the standard deviation, since the standard deviation decreases as the stress level increases. This means that for the voltage level of 400 kV the values are closer to the mean value than to the voltage level of 150 kV.

### VI. CONCLUSIONS

This paper presents a probabilistic approach of fault resistance in transmission lines. In this study was considered 100 phase-to-ground faults verified in 44 different transmission lines of 400, 220 or 150 kV, on the RNT network along 3 years.

To compute values of fault resistance were created two different algorithms, from fault location algorithms, based on current and voltage measurements in one or both line ends. To test and verify both algorithms, it was simulated one fault occurred in a line of the test network, assuming values of fault resistance and it was verified extremely high errors associated to the algorithm using only information from one end of the line. So it was concluded that the most viable algorithm to compute fault resistance is the algorithm that uses measurements of both line ends. After that, the value of fault resistance corresponding to each fault was computed and it was verified that the great majority of faults occurred to values less than 5 \( \Omega \), with some differences when compared to the values published from another authors.

To probabilistic characterization of fault resistance values were proposed two probabilistic models based on Weibull distribution and normal distribution in a logarithmic scale based on relevant bibliography. To evaluate the adequacy of both probabilistic methods it was adopted a procedure based on EDA with graphical tools, namely theoretical quantil-quantil plots and density and distribution function, to observe standard features and abnormal points that were analyzed with Tukey’s method. In relation to the Weibull distribution it was verified that for 150 kV this distribution doesn’t fit to data, so this distribution is incorrect to describe fault resistance. For normal distribution in a logarithmic scale, to all of voltage levels, the distribution fits to data, so this distribution is the right one to describe fault resistance.

### REFERENCES