





TRANSMISSION LINE FAULT LOCATION USING MFCC AND LS-SVR

Hermes M. G. Castelo Branco^a 
James B. O. Reis^a 
Luan M. M. Pereira^b
Lucas da C. Sá^a 
Ricardo de A. L. Rabelo^a 

^aCentro de Tecnologia
Universidade Federal do Piauí
Teresina, Brazil

^bCentro de Tecnologia e Urbanismo
Universidade Estadual do Piauí, Teresina, Brazil

emails: {hermesobl@ufpi.edu.br, jamesblayne@ufpi.edu.br, luanpereira@aluno.uespi.br, lucas.duh@hotmail.com
ricardoalr@ufpi.edu.br}

Abstract- The location of Transmission Line (TL) Faults is a major problem in Electrical Power Systems (EPSs), since precisely identifying the point of occurrence of a fault in a TL it is possible to perform a faster restoration of the operation to the desired normal conditions. In this work we used a Least-Squares Support Vector Regression (LS-SVR) to locate faults in a TL with inputs provided by MFCC (Mel-Frequency Cepstral Coefficients) obtained from voltage signals during the fault. A modelled line based on parameters of a real line was used, with a total of 4008 fault situations being simulated on this Transmission Line. It is important to point out that MFCC are not used in applications involving EPS's, and, according to the bibliographic research conducted by the team so far, no application of this feature extraction tool has been detected for the TL fault location problem. 3006 faults were used to train the model with cross-validation by the k-fold method, and 1002 faults were used for testing. The proposed methodology presented a good performance in the tests carried out, with a mean relative error of $0.000419 \pm 0.000640\%$ when models are trained and tested with noiseless voltage signals. For models trained with voltage signals that present SNR ranging from 100 dB to 25 dB, the relative mean error ranged from $0.00334 \pm 0.00459\%$, in the first case, to $0.030580 \pm 0.043160\%$ in the last.

Keywords- fault location; transmission line; LS-SVR; MFCC.

1 Introduction

Electric Power Systems (EPS's) are subject to several situations that may interrupt the power supply. The re-establishment of the energy system must be carried out as quickly as possible in order to reduce the impacts and damages caused to the consumers. Fault situations may occur in various Electric Power System components. Transmission Lines (TL's) stand out as the most susceptible elements, mainly because of their extent, as pointed out by [1]. The correct Transmission Line fault location is an important task to be performed, since the repair time, in the case of permanent faults, affects the reliability of the EPS and leads to increase losses caused by the stoppage of customers processes, especially industrial processes [2]. By identifying, as precisely as possible, the point of occurrence of a fault in TL, less time is required in maintenance and repair services, which allows a faster restoration of the operation to the normal conditions desired.

Several approaches has developed and applied to the TL fault location problem, both using traditional techniques and Machine Learning (ML) tools. These approaches can be divided into two main categories:

- 1) Fundamental frequency-based methods;
- 2) High frequency-based methods.

Methods that use fundamental components generally determine the fault location by estimating the line impedance between the measuring point and the fault point [3]. The estimation is made by calculating the phase angles, observing the absolute value

and phase variations of the voltage and/or current waves during the fault [4]. The accuracy of these methods is affected by several factors, such as: fault resistance, remote terminal current contribution, line loading, and source parameters [5]. Among the methods based on fundamental frequency, it is possible to mention, among others, approaches that employ: least squares theory [6]; Fourier Transform [7]; Wavelet Transform [8] and some other approaches with filters [9].

High frequency-based methods are grounded on the traveling waves theory [1]. Such methods determine the traveling time interval of the voltage or current wave during the short circuit from the fault point to the line terminal, as well as to determine the propagation speed of this wave in TL. By having the information of the travel time and the wave speed during the transients coming from the short circuit, it is possible to infer the fault distance, overcoming some of the limitations of the methods based on fundamental frequency components [10]. In these methods three aspects are important: to accurately detect the moment of incidence of traveling waves at the terminals of the line; to extract relevant characteristics of the transitory disturbances; and to develop mathematical expressions that relate the characteristics to the fault distance [11]. Some works, such as [12], seek to find solutions for the fast detection of transients. Other works use disturbance extraction characteristics through the spectral analysis of transients in the period of fault or right after their isolation. [13]. The great limitation related to the method based on traveling waves lies in the requirement to work with high sampling rates [3]; however, with the current technological development, this barrier has been overcome.

Concepts associated with ML applied to EPS's is attracting an increasing attention of researchers over the past years, several proposals in the literature seek to build an intelligent system in order to monitor and to diagnose faults, including, among other tasks, the transmission lines fault locating. Some approaches use Artificial Neural Networks [14], Genetic Algorithms [15] and Fuzzy Logic [16]. Recently, Support Vector Regression (SVR) based proposals have also been highlighted [17], as well as Least Squares Support Vector Regression (LS-SVR) based proposals [18]. The application of machine learning techniques together with signal processing tools has allowed a more accurate estimation of the fault distance in transmission lines.

In this research we use a machine learning tool that performs well for many problems and that is suitable for training with relatively small databases, the LS-SVR. The LS-SVR presents the good performance already demonstrated of the SVR and, due to its training method, is suitable for relatively small bases. The input characteristic vectors for estimating the distance of fault are Mel-Frequency Cepstral Coefficients (MFCC) obtained from the voltage signals of the first cycle after the fault. The main contribution of this work is the use of the MFCC as a feature extractor, which allows obtaining information from both the high-frequency components of the signal and the low-frequency ones. The unprecedented application of this tool to fault location problems can contribute to obtaining attributes that facilitate the process of estimating the fault distance by ML tools. Another important point is that a modeled line based on parameters of a real line was used, and the results so far have been promising. This article is organized in the following sections: II Used tools; III Electrical System Studied and Methodology proposed; IV Results; V Conclusions.

2 Theoretical Foundations

This section aims to present the basic concepts related to Mel-Frequency Cepstral Coefficients and least squares support vector regression.

2.1 Mel-Frequency Cepstral Coefficients

One of the first steps when performing the process of prediction or classification of a given signal is the extraction of characteristics, that is, identifying components that represent the characteristics of the signal, discarding information that is not useful as the noise present in most digital signals. For this purpose, several signal processing and feature extraction tools have been proposed over the years. In 1980, Davis and Mermelstein developed a new type of discrete-time Fourier transform representation (DTFT) using logarithm of the signal amplitude [19], known as Mel-Frequency Cepstral Coefficients (MFCC), which is widely used in voice signal processing [19]. According to [20], MFCC have been widely used in speech recognition because they can lead with their dynamic characteristics as they extract linear and non-linear properties from the signal. MFCC is a representation defined as the cepstrum of a signal defined by a window. The difference from the real spectrum is in the use of a non-linear frequency scaled by triangular filters, which approximates the behavior of the auditory system [20].

Thus, the basic idea of MFCC is to calculate frequency analysis based on filters bank, also called Mel filters, which have a nonlinear spacing. Mel scale is described by equation (1):

$$mel = 2597 \times \log_{10} \left(1 + \frac{F(Hz)}{700} \right), \quad (1)$$

mel is the Mel-Frequency whose unit is $mel(m)$, F is the frequency in Hertz (Hz). Analogously, substituting the logarithm in base 10 by natural logarithm in equation (1), we have:

$$mel = 1127 \times \ln \left(1 + \frac{F(Hz)}{700} \right). \quad (2)$$

The Mel scale filter bank can be obtained in different ways; the main variations are in the number of filters; in triangular filter, square or Schroeder format. The function of these filters is to calculate the average spectrum around each center frequency with increasing bandwidths as illustrated in Figure 1.

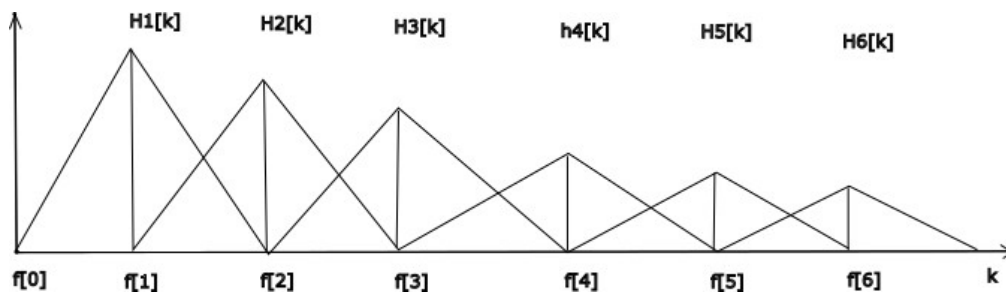


Figure 1: Triangular Mel filters.

Coefficients are basically the power spectrum transformed by the logarithm of the absolute value of the signal on the Mel scale.

The following steps guide the processing to extract the MFCC:

1. Split the signal into small windows or take the entire signal as a single window, in this work we take only one cycle window of voltage and current signals.
2. For each window, the Discrete Fourier Transform (DFT) is applied to calculate the periodogram of the power spectrum.

The Fourier transform $X(\omega)$ for discrete time signals $x[n]$ is defined by the equation 3:

$$X(\omega) = \sum_{-\infty}^{\infty} x[n] e^{-i\omega n} \quad (3)$$

3. A Mel scale filter bank is applied using equation 1.
4. The energy in each triangular filter is calculated, using the equation 4:

$$E_i = \sum_{\omega} (X(\omega))^2 \quad (4)$$

where i is the filter number, $i=1,2,\dots,J$.

5. The logarithm of all filter bank energies is calculated.
6. The Discrete Cosine Transform (DCT) is applied to the final energy signal.

The equation 5 is used to calculate the discrete cosine transform:

$$X(k) = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x[n] \cos \left(\frac{\pi(2n+1)k}{2N} \right) \quad (5)$$

$K = 1, 2, 3, \dots, N-1$.

Figure 2 illustrates the steps in order to obtain the Mel Coefficients:

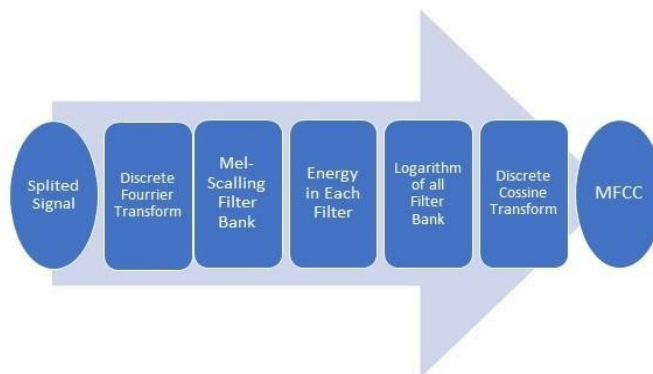


Figure 2: Flowchart to obtain the Mel-Frequency Coefficients.

According to [21], the successful use of MFCC for speech recognition is due to its ability to represent the amplitude spectrum of the signal in a compact way. This feature is important for fault location, since having few coefficients it is possible to use information both related to high frequency, as in approaches based on the traveling waves theory, obtained from higher order coefficients, as well as frequency information, lowest values obtained in the lowest order coefficients. It is important to note that the higher coefficients represent rapid changes in the filter bank energies, bringing information related to methods based on high frequency components for fault location. The lower coefficients bring information related to the fundamental frequency methods. One of the advantages of MFCC representation is that the filter energies are more robust to noise and spectral estimation errors [19], which may be useful for the intended application in this work.

2.2 LS-SVR

The Support Vector Machine (SVM), proposed by [23], besides pattern recognition, can also be successfully employed for regressions, called Support Vector Regression (SVR) [24]. Authors report that in several situations the results presented by SVM surpass other machine learning algorithms as can be seen in [17]. As pointed out in [23], most of the machine learning techniques present difficulties in the treatment of high dimensional data, among other reasons, due to the possibility of converging to local optimum. The application of SVM and SVR leads to the optimization of a quadratic function, which has only a minimum. This is an advantage over other approaches, such as Neural Networks, which may have several local minima.

The main characteristic of the SVM algorithm is learning by means of hyperplanes of separation, or hyperplanes of maximum margin [24]. The maximum margin hyperplane provides the maximum separation between classes, and is generated from the support vectors, which in turn are formed by the records that are closest to the separation limit. For a case where the data are linearly separable and considering a dimensionality of the n -degree data, the maximum margin hyperplane can be defined by equation (6):

$$y = w_0 + \sum_{i=1}^n w_i x_i \quad (6)$$

where y is the output, x is attribute value, and w are parameters to be calculated by the algorithm that defines the maximum margin hyperplane. The maximum separation hyperplane equation can still be defined as a function of the supporting vectors, as shown in equation (7):

$$y = b + \sum_{i=1}^n \alpha_i y_i x(i) \cdot x \quad (7)$$

Therefore, y_i is the value of the class used in training, $x(i) \cdot x$, The inter product between the support vectors and the training records, b e α are parameters that determine the hyperplane. When data is non-linearly separable, a kernel function is used to

modify the dimensionality of the data by enabling linear separation. The equation (8) presents a version of equation (7) for the non-linearly separable case, using a Kernel function (K) [23]:

$$y = b + \sum_{i=1}^n \alpha_i y_i K(x(i), x) \quad (8)$$

Different kernel functions can be used in SVMs, among the most commonly used are Linear, Polynomial and Gaussian. The proper choice of the core and its respective parameters has a strong influence on the results obtained by an SVM.

A SVM can have a hard border, when the training is running the model must completely separate two classes with a hyperplane. However, a hyperplane that separates all patterns belonging to two classes does not always exist in real problems, or when it does, it is defined by a rather complex function. Therefore, an approach in which it is possible to relax some restrictions of SVM, allowing it to misclassify some standards, resulting in what is conventionally called soft margin SVM. Soft margin SVM is achieved by including slack variables and constant regularization in the model.

In 1999 Suykens and Vandewalle proposed an extension of the SVM theory using the least squares method for multivariate calibration of a linear and nonlinear data set [25], which became known as LS-SVM, that formulates the classification problem as a minimization problem described by equation (9):

$$J_{LS}(w, b, e) = \frac{1}{2} w^T w + C \frac{1}{2} \sum_{i=1}^N e_i^2 \quad (9)$$

$$s.t. y_i = w \cdot K(x_i) + b + e_i$$

In which C is a regularization constant, $K(x)$ is a kernel function that converts the input data into high-dimensional data with linear relation, b is the threshold, w is the slack variable associated with the classification, and e_i is the error.

LS-SVM algorithm brings two important changes in the formulation of primary optimization problems related to the SVM algorithm [25]. The first change is in the restriction, which is presented to the LS-SVM classifier as equality, while in the SVM it is inequality. The second change refers to incorporation of the sum on float variables, ek to the square in the cost function, weighted by a smoothing constant, C . When applied to regression problems, the algorithm is called LS-SVR, however it still follows equation (9) as presented. In this case, w is a regression weights matrix and the optimization for this parameter is normally converted to a Lagrangian function, in equation (10):

$$\frac{1}{2} w^T w + C \frac{1}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \lambda_i (w \cdot K(x_i) + b + e_i - y_i) \quad (10)$$

where λ_i is the Lagrange multiplier, which depends on the input data x_i and the regularization parameter C ; it is expressed in equation (11):

$$\lambda_i = \left(x_i^T x_i + \frac{1}{2C} \right)^{-1} \quad (11)$$

Considering that minimum L is determined based on the principle that the first derivative equals zero, the final form of the LS-SVR is expressed in equation (12):

$$y = \sum_{i=1}^N \lambda_i w \cdot K(x_i) + b. \quad (12)$$

The simplifications of LS-SVR related to SVR algorithm have some limitations. One of them refers to the loss of the sparse nature in solution of the problem, thus, the Lagrange multipliers obtained in LS-SVR learning process are non-zero, therefore, after the learning process, it is necessary to store all training standards as well as the associated Lagrange multipliers for discriminating function composition purposes.

3. Electrical System Studied and proposed Methodology

For the development of this work, a real high-voltage (500kV) transmission system of 330km located in the Northeast of Brazil was modeled.

The transmission line modeled consists of guyed VX type towers, as illustrated in Figure 3. It has three phases, with four conductors each and two guard cables, a horizontal phase arrangement of 11 meters and 4 sub-conductors 954 MCMRAIL. The external radius of the conductors is 14.795 mm, and the internal radius 3.7 mm and DC resistance is 0.05995 ohm/km at a height of 41m and an average arrow at half span of 13.43 m.

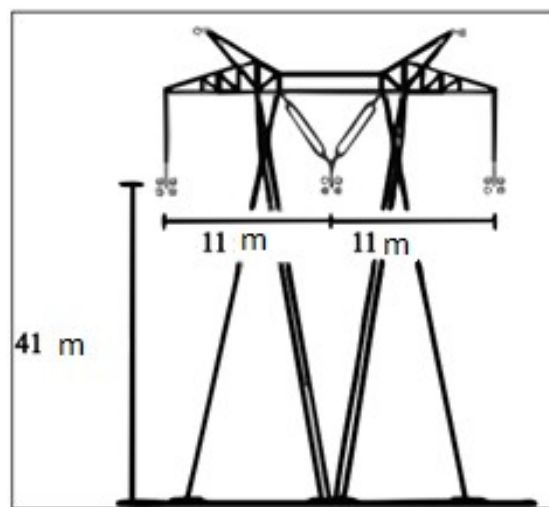


Figure 3: Line model with 500 kV Symmetric Expanded Beam (VX - Symmetric) tower.

The system was modeled using the ATP (Electromagnetic Transients Program) software, through the ATPDraw interface. Figure 4 represents the system modeled in the ATPDraw. Fault simulations were performed for different distances varying resistance and fault angle for faults: single-phase to ground (LG), two-phase to ground (LLG) and three-phase to ground (LLLG). The sampling rate used for the simulations was 400 KHz.

In the simulations performed, the presence of noise in the voltage or current signals was not considered, that way the acquisition was performed with a sampling rate of 400 kHz, a low-pass filter was applied with a cut-off frequency of 200 kHz, to satisfy the Nyquist sampling theorem, thus avoiding the Aliasing effect. One cycle window of voltage and current signals was used, signal data has been collected on both terminals, and these must be synchronized.

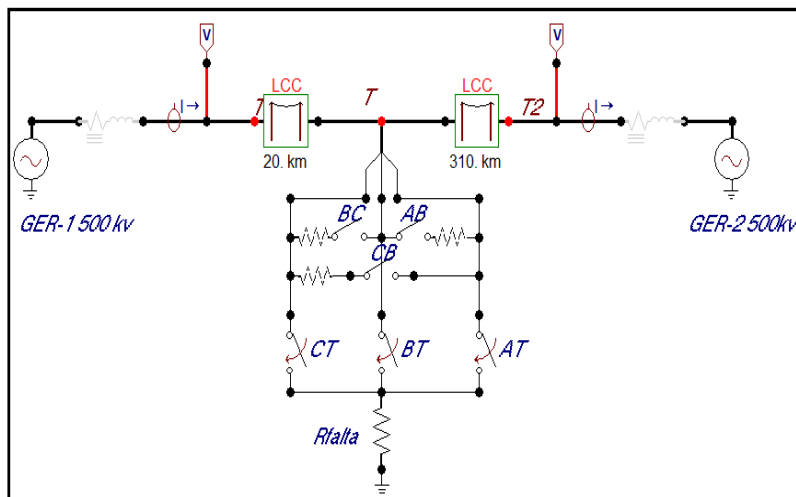


Figure 4: Transmission System modeled on ATPDraw software.

In total, 4008 fault situations were simulated and Table I below summarizes the characteristics of the simulated faults. From the total of simulated faults, 3006 (75% of all faults) were randomly selected and assigned to the LS-SVR training and validation. The remaining 1002 cases were reserved for testing of the proposed methodology. The training was carried out using the cross-validation technique, for which the set was divided into 6 equal parts. In the process of adjusting the weights 5 parts are used for training and one part for validation until all six parts have been used as test data. A Gaussian kernel function has been adopted.

After the training stage, using the best model obtained in the cross-validation, the test was carried out with the 1002 cases separated initially.

| Table I: Quantities of simulated faults per type of fault. | | | | |
|---|-------------------------|------------------------|----------------------|-----------------|
| Type of Fault | Resistance (Ohm) | Angle (Degrees) | Distance (KM) | Quantity |
| LG | 0.1 to 500 | 0; 45; 90 | 0.9 to 329.1 | 1269 |
| LLG | 0.1 to 500 | 0; 45; 90 | 0.9 to 329.1 | 1269 |
| LLLG | 0.1 to 500 | 0; 45; 90 | 0.9 to 329.1 | 1269 |
| LL | ---- | 0; 45; 90 | 0.9 to 329.1 | 201 |
| Total Simulated Faults | | | | 4008 |

The inputs for the LS-SVR were obtained from the voltage signals of the 3 phases, on each of two terminals, after the fault, totaling 6 voltage signals. From each voltage signal the MFCC were calculated. For the MFCC calculation on each phase, only one window with 1 voltage signal cycle was used just after the fault input. We used 40 triangular filters, and as input the first 20 MFCC of each phase, obtained from the first 20 triangular filters, totaling 120 inputs. It should be noted that the choice of using 40 triangular filters to calculate the MFCC and the first 20 MFCC as LS-SVR inputs was made after some tests considering different amounts of triangular functions. After the training stage, the LS-SVR was used to estimate the fault distance of the validation set. Fig. 5 illustrates the steps of the proposed methodology.

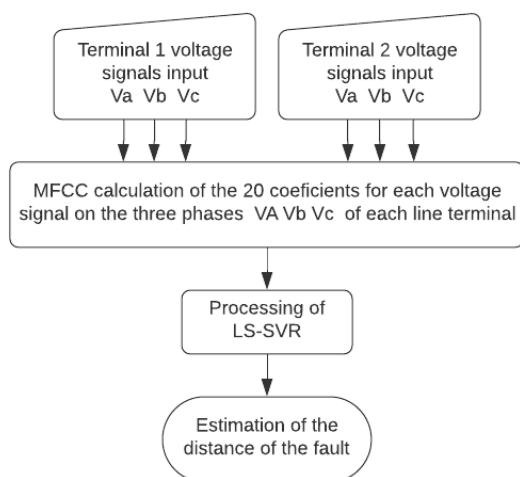


Fig. 5: Flowchart of the proposed methodology for Transmission Lines Fault Location.

In order to evaluate the performance of the methodology with signals that present noise, white noise was inserted in the initially generated signals, at the following levels of signal/noise ratio: 100 dB, 75 dB, 50 dB and 25 dB. Then a new model was trained for the signals with each noise level, and finally, these models were evaluated considering voltage test signals with the same noise levels applied to the model training signals.

4. Results

The proposed methodology obtained its best results for fault location with a Mean Absolute Error (MAE) of 0.001350 ± 0.002118 km, for fault signals without the presence of noise. The Mean Relative Error (MRE) was 0.000419% and standard deviation 0.000640% for the same signals. The MAE and MRE errors were calculated according to equations (13) and (14):

$$MAE = \frac{1}{N} \sum_{i=1}^N E_i \quad (13)$$

$$MRE = \frac{1}{N} \sum_{i=1}^N \frac{E_i}{L} \times 100 \quad (14)$$

Where E_i is the fault location error in km, given by the difference module, between the actual and estimated distance, N is the number of faults, and L the line length in km. We use these methods to calculate the error with the purpose of better comparison with the methods of other authors, but the error in different EPSs, may be influenced by the characteristics of the study system and the quantity and variability of samples used in training and tests.

It is important to highlight that even for very noisy signals, with an SNR of 25 dB, the proposed methodology obtained excellent results, with Mean Absolute Error (MAE) of 0.100910 ± 0.142430 km. Table II illustrates the influence of fault type and SNR on the location error. In this table it is possible to notice that LG faults had higher MRE from fault signals with different SNRs. According to Table II, it is also possible to notice that when noisier the line, error increases in fault distance estimating. Analyzing the results, we noticed that the biggest errors were from single-phase faults with SNR of 25 dB, which presented errors of approximately 200 meters.

Table II: Influence of the type of fault and SNR on the proposed location methodology.

| Type of Fault | SNR dB | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|---------------|-----------|----------|------------------------|----------|------------------------|
| LG | noiseless | 0.002580 | 0.002786 | 0.000782 | 0.000844 |
| LLG | noiseless | 0.001440 | 0.001574 | 0.000436 | 0.000477 |
| LL | noiseless | 0.000031 | 0.000084 | 0.000009 | 0.000018 |
| LLLG | noiseless | 0.000030 | 0.000056 | 0.000009 | 0.000017 |

Table II: Influence of the type of fault and SNR on the proposed location methodology.

| Type of Fault | SNR dB | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|----------------|----------------------------------|-----------------|------------------------|-----------------|------------------------|
| AVERAGE | noiseless | 0.001350 | 0.002118 | 0.000419 | 0.000640 |
| LG | 100 | 0.020704 | 0.018991 | 0.006274 | 0.005755 |
| LLG | 100 | 0.010657 | 0.012056 | 0.003229 | 0.003653 |
| LL | 100 | 0.001727 | 0.002152 | 0.000522 | 0.000610 |
| LLLG | 100 | 0.001720 | 0.002002 | 0.000521 | 0.000606 |
| AVERAGE | 100 | 0.01103 | 0.01516 | 0.00334 | 0.00459 |
| LG | 75 | 0.054340 | 0.048230 | 0.016467 | 0.014635 |
| LLG | 75 | 0.025041 | 0.022793 | 0.007588 | 0.006907 |
| LL | 75 | 0.008248 | 0.010234 | 0.002503 | 0.003049 |
| LLLG | 75 | 0.008256 | 0.010035 | 0.002502 | 0.003041 |
| AVERAGE | 75 | 0.02921 | 0.03667 | 0.00885 | 0.01111 |
| LG | 50 | 0.049764 | 0.041164 | 0.015080 | 0.012474 |
| LLG | 50 | 0.017145 | 0.018676 | 0.005196 | 0.005660 |
| LL | 50 | 0.022600 | 0.020984 | 0.007662 | 0.007273 |
| LLLG | 50 | 0.027594 | 0.030662 | 0.008362 | 0.009292 |
| AVERAGE | 50 | 0.03150 | 0.03431 | 0.00955 | 0.01040 |
| LG | 25 | 0.192782 | 0.183327 | 0.058419 | 0.055554 |
| LLG | 25 | 0.091004 | 0.106860 | 0.027577 | 0.032382 |
| LL | 25 | 0.027872 | 0.034759 | 0.010896 | 0.010957 |
| LLLG | 25 | 0.018949 | 0.025903 | 0.005742 | 0.007849 |
| AVERAGE | 25 | 0.100910 | 0.142430 | 0.030580 | 0.043160 |
| LG | average of all SNR tested | 0.064034 | 0.0588996 | 0.0194044 | 0.0178524 |
| LLG | average of all SNR tested | 0.0290574 | 0.0323918 | 0.0088052 | 0.0098158 |
| LL | average of all SNR tested | 0.0120956 | 0.0136426 | 0.0043184 | 0.0043814 |
| LLLG | average of all SNR tested | 0.0113098 | 0.0137316 | 0.0034272 | 0.004161 |
| AVERAGE | average of all SNR tested | 0.034800 | 0.046138 | 0.010548 | 0.013980 |

Still analyzing Table II, for the most common SNR observed in LT, between 100 dB and 50 dB [1], MAE in the estimation of the fault distance, are of the same order of magnitude, with values between 11 m and 32 m, indicating good tolerance to the expected noise levels in LT. Even for voltage signals collected on TL with SNR above those expected, the proposed method proved to be efficient, with MAE of approximately 100m. Considering the results obtained for the location of faults with the different noise levels tested, the general average of the error obtained with the models was approximately 35 meters. Table II also shows the average performance of all models trained for different noise levels, considering each type of fault.

The authors of [1] proved that by applying independent component analysis (ICA) to fault signals it is possible to separate fault signal from noise signal. Thus, since we demonstrated that the proposed methodology presents good results even for very noisy voltage signals, which an adequate separation of noise was not possible, we will present the stratified results considering the variation of the fault angle, fault resistance and fault distance, only for the noiseless fault signals.

The angle of fault insertion is also a factor that influences the accuracy of the fault location methodologies. Table III presents data that show how the precision of the response varies according to the variation of the Angle of fault, considering noiseless signals of faults. It is possible to notice that the difference between the estimation errors of simulated faults with different angles is small for simulated faults without the noise insertion in the line. Therefore, considering the simulated dataset and the modeled system, the proposed method is not influenced by the fault insertion angle.

Table III: Influence of the fault insertion angle in noiseless signals of faults.

| Fault Angle | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|-------------|----------|------------------------|---------|------------------------|
| 0° | 0.00162 | 0.00244 | 0.00049 | 0.00074 |

Table III: Influence of the fault insertion angle in noiseless signals of faults.

| Fault Angle | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|-------------|----------|------------------------|---------|------------------------|
| 45° | 0.00120 | 0.00212 | 0.00036 | 0.00064 |
| 90° | 0.00124 | 0.00174 | 0.00038 | 0.00053 |

Table IV presents the fault location statistics for different distances. It was possible to observe that the proposed methodology presented a slightly larger error for faults located closer to the terminal 1, while the smallest errors are mid-line. However, once again, we detected that in the worst situations the errors are about 2.25 meters \pm 3.43 meters, demonstrating the good performance of the proposed approach.

Table IV: Influence of fault distance on proposed location methodology in noiseless signals of faults.

| Fault Distance (KM) | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|---------------------|----------|------------------------|---------|------------------------|
| 0.9 to 30 | 0.00225 | 0.00343 | 0.00068 | 0.00104 |
| 31 to 60 | 0.00200 | 0.00264 | 0.00060 | 0.00080 |
| 61 to 90 | 0.00158 | 0.00213 | 0.00048 | 0.00064 |
| 91 to 120 | 0.00095 | 0.00141 | 0.00029 | 0.00043 |
| 121 to 150 | 0.00118 | 0.00186 | 0.00036 | 0.00056 |
| 151 to 180 | 0.00085 | 0.00131 | 0.00026 | 0.00040 |
| 181 to 210 | 0.00075 | 0.00132 | 0.00023 | 0.00040 |
| 211 to 240 | 0.00124 | 0.00161 | 0.00037 | 0.00049 |
| 241 to 270 | 0.00142 | 0.00188 | 0.00043 | 0.00057 |
| 271 to 300 | 0.00129 | 0.00199 | 0.00039 | 0.00060 |
| 301 to 329.1 | 0.00102 | 0.00198 | 0.00031 | 0.00060 |

It is also important to analyze the influence of fault impedance on the performance of the proposed fault location methodology. For this purpose, the data were divided into 10 groups of 50 Ω to 50 Ω . It was possible to observe in Table V that the proposed methodology presented a slightly larger error for faults considered of low impedance.

Table V: Influence of fault resistance on proposed approach in noiseless signals of faults.

| Fault Resistance(Ω) | MAE (KM) | Standard Deviation (%) | MRE (%) | Standard Deviation (%) |
|------------------------------|----------|------------------------|---------|------------------------|
| 0.1 to 50 | 0.00060 | 0.00120 | 0.00018 | 0.00036 |
| 55 to 100 | 0.00102 | 0.00191 | 0.00031 | 0.00058 |
| 105 to 150 | 0.00121 | 0.00180 | 0.00037 | 0.00055 |
| 155 to 200 | 0.00173 | 0.00242 | 0.00053 | 0.00073 |
| 205 to 250 | 0.00186 | 0.00228 | 0.00056 | 0.00069 |
| 255 to 300 | 0.00144 | 0.00206 | 0.00044 | 0.00062 |
| 305 to 350 | 0.00136 | 0.00231 | 0.00041 | 0.00070 |
| 355 to 400 | 0.00143 | 0.00212 | 0.00043 | 0.00064 |
| 405 to 450 | 0.00109 | 0.00157 | 0.00033 | 0.00048 |
| 455 to 500 | 0.00164 | 0.00272 | 0.00050 | 0.00082 |

After analyzing the results, it is possible to conclude that, although in some specific situations, the proposed methodology presents a slightly better performance than other methods; in general, it was possible to estimate with high precision the faults in the simulated transmission line. Table VI presents a comparison with results obtained by other methodologies presented in the literature detailing the tools used for each proposal and the number of cycles of the input signals used by each proposal.

It is possible to observe that the proposed methodology, using voltage signals without noise during the fault as input, has a better performance than other approaches presented in Table VI. Considering the proposed models trained and tested with voltage signals presenting SNR between 100db and 25dB, the proposed approach presents better results than references [1], [4], [5] and [26]. It is noteworthy that all approaches presented use noiseless signals in their training and testing processes. The results presented suggest that MFCC was able to extract features that allowed a good location of faults by LS-SVR, however new investigations should be carried out, such as analyzing the use of voltage and current data, aiming to obtain an even better MRE.

Table VI: Comparison of the results with techniques that use machine learning

| Methodology | Features Extraction | ML Tool | (%) | Cycles |
|---|---------------------|-------------|--------|--------|
| [1] | ICA/TWP | MLP | 0,1568 | |
| [4] | DWT | RBF | 0.0500 | 1 |
| [5] | HS-Transform | RBF | 0.8900 | 2 |
| [14] | WPT | RNA | 0.0010 | 1 |
| [17] | - | SVR | 0.0009 | 0.25 |
| [18] | S-Transform | LS-SVR | 0.0017 | 1 |
| [26] | DWT | RBF and SVR | 0.5100 | 2 |
| [27] | DWT | Neuro-Fuzzy | 0.0010 | 2 |
| [28] | WPT | SVR | 0.0021 | 0.5 |
| [29] | SWT | SVR | 0.0021 | 0.25 |
| Proposal (noiseless voltage signals) | MFCC | LS-SVR | 0.0004 | 1 |
| Proposal (voltage signals with 100 dB SNR) | MFCC | LS-SVR | 0.0033 | 1 |
| Proposal (voltage signals with 75 dB SNR) | MFCC | LS-SVR | 0.0088 | 1 |
| Proposal (voltage signals with 50 dB SNR) | MFCC | LS-SVR | 0.0096 | 1 |
| Proposal (voltage signals with 25 dB SNR) | MFCC | LS-SVR | 0.0306 | 1 |

5. Conclusions

The fault location in transmission lines is a problem of great relevance for EPSs, since pointing out the correct place of occurrence of short circuits allows the adoption of measures more efficiently to restore these problems. In the current context of Smarts Grids, several approaches using modern signal processing techniques and ML tools have been proposed in related literature. The approach proposed in this work uses an LS-SVR to estimate the distance of fault, receiving as inputs a vector of features composed of 20 coefficients obtained by the application of MFCC for each voltage signal of the 3 phases for each terminal of a line of two terminals. The proposed approach proved to be promising, presenting, when noiseless voltage signals are provided as input to the approach, a mean error of 0.000419% to the line length and a standard deviation of 0.000640% to the line length. The MFCC was able to provide adequate information for the LS-SVR, which in turn was quite accurate in locating the vast majority of simulated fault situations, including when noisy voltage signals are provided as input to the approach. As future work it is intended to investigate the use of the data obtained in a single terminal for the location of fault, following the proposed approach as well as the use of smaller windows. It is also intended to investigate whether the use of current and voltage data together helps to improve the performance of the algorithm that uses the MFCC for feature extraction. Finally, it is intended to propose fault location with a trained LS-SVR specifically for each type of fault.

6. Acknowledgements

The authors would like to thank CNPQ and CAPES for the financial support to conduct this research and UFPI for providing the entire necessary infrastructure.

7. References

- [1] Almeida, A. R., Almeida, O. M., Junior, B. F. S., Barreto, L. H. S. C., & Barros, A. K. (2017). ICA feature extraction for the location and classification of faults in high-voltage transmission lines. *Electric Power Systems Research*, 148, 254-263.
- [2] Carneiro, S.M., Rabelo, R. A. L. & Branco, H. M. G. C. (2018) A Multi-objective Approach for Optimized Monitoring of Voltage Sags in Distribution Systems. *J Control Autom Electr Syst* 29, 371–380.

- [3] Chen, K., Huang, C., & He, J. (2016). Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. *High voltage*, 1(1), 25-33.
- [4] Liang, F., & Jeyasurya, B. (2004). Transmission line distance protection using wavelet transform algorithm. *IEEE Transactions on Power Delivery*, 19(2), 545-553.
- [5] Samantaray, S. R., Dash, P. K., & Panda, G. (2006). Fault classification and location using HS-transform and radial basis function neural network. *Electric Power Systems Research*, 76(9-10), 897-905.
- [6] Ajaei, F. B., & Sanaye-Pasand, M. (2008, October). Minimizing the impact of transients of capacitive voltage transformers on distance relay. In *2008 Joint International Conference on Power System Technology and IEEE Power India Conference* (pp. 1-6). IEEE.
- [7] Phadke, A. G., & Thorp, J. S. (2009). *Computer relaying for power systems*. John Wiley & Sons.
- [8] Da Silva, M., Coury, D. V., Oleskovicz, M., & Segatto, E. C. (2008). An alternative fault location algorithm based on wavelet transforms for three-terminal lines. In *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century* (pp. 1-7). IEEE.
- [9] Lee, D. G., Kang, S. H., & Nam, S. R. (2008). New modified fourier algorithm to eliminate the effect of the DC offset on phasor estimation using DFT. In *2008 IEEE/PES Transmission and Distribution Conference and Exposition* (pp. 1-6). IEEE.
- [10] Bo, Z. Q., Jiang, F., Chen, Z., Dong, X. Z., Weller, G., & Redfern, M. A. (2000). Transient based protection for power transmission systems. In *2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No. 00CH37077)* (Vol. 3, pp. 1832-1837). IEEE.
- [11] Costa, F. B. (2013). Fault-induced transient detection based on real-time analysis of the wavelet coefficient energy. *IEEE transactions on power delivery*, 29(1), 140-153.
- [12] Lopes, F. V., Fernandes, D., & Neves, W. L. A. (2013). A traveling-wave detection method based on park's transformation for fault locators. *IEEE Transactions on Power Delivery*, 28(3), 1626-1634.
- [13] Mamiş, M. S., Arkan, M., & Keleş, C. (2013). Transmission line's fault location using transient signal spectrum. *International Journal of Electrical Power & Energy Systems*, 53, 714-718.
- [14] Ekici, S., Yildirim, S., & Poyraz, M. (2008). Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition. *Expert Systems with Applications*, 34(4), 2937-2944.
- [15] Davoudi, M. G., Sadeh, J., & Kamyab, K. (2012). Time domain fault location on transmission lines using genetic algorithm. In *2012 11th International Conference on Environment and Electrical Engineering* (pp. 1087-1092). IEEE.
- [16] Reddy, M. J. B., & Mohanta, D. K. (2008). Performance evaluation of an adaptive-network-based fuzzy inference system approach for location of faults on transmission lines using Monte Carlo simulation. *IEEE Transactions on Fuzzy Systems*, 16(4), 909-919.
- [17] Fei, C., Qi, G., & Li, C. (2018). Fault location on high voltage transmission line by applying support vector regression with fault signal amplitudes. *Electric Power Systems Research*, 160, 173-179.
- [18] Uyar, M. (2012). ST and LSSVR-based the fault location algorithm for the series compensated power transmission lines. *Energy Educ. Sci. Technol. Part A: Energy Sci. Res.*, 30(1), pages 75-88.
- [19] Furui, S. (2018). *Digital speech processing: synthesis, and recognition*. CRC Press.
- [20] Nelwamondo, F. V., & Marwala, T. (2006, October). Fault's detection using gaussian mixture models, mel-frequency cepstral coefficients and kurtosis. In *2006 IEEE International Conference on Systems, Man and Cybernetics* (Vol. 1, pp. 290 - 295). IEEE.
- [21] Logan, Beth. (2000). *Mel Frequency Cepstral Coefficients for Music Modeling*. Proc. 1st Int. Symposium Music Information Retrieval.
- [22] Brent, W. (2010). *Physical and Perceptual Aspects of Percussive Timbre*, doctoral thesis, University of California.
- [23] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- [24] Haykin, S. S. (2009). *Neural networks and learning machines*/Simon Haykin.
- [25] Suykens, J. A., Vandewalle, J., & De Moor, B. (2001). Optimal control by least squares support vector machines. *Neural*

networks, 14(1), 23-35.

[26] S.R. Samantaray, P.K. Dash, G. Panda. (2007) Distance relaying for transmission line using support vector machine and radial basis function neural network, International Journal of Electrical Power & Energy Systems 29. 551–556.

[27] F. Chunju, K.K. Li, W.L. Chan, Y. Weiyong, Z. Zhaoning (2007) Application of wavelet fuzzy neural network in locating single line to ground fault (slg) in distribution lines, International Journal of Electrical Power & Energy Systems 29, 497 –503.

[28] A.A. Yusuff, C. Fei, A.A. Jimoh, J.L. Munda, Fault location in a series compensated transmission line based on wavelet packet decomposition and support vector regression, Electric Power Systems Research, Volume 81, Issue 7, Pages 1258- 1265.

[29] A.A. Yusuff, A.A. Jimoh, J.L. Munda. (2014) “Fault location in transmission lines based on stationary wavelet transform, determinant function feature and support vector regression”.Electric Power Systems Research, Volume 110, Pages 73-83.