

Article

# Electrical Insulator Fault Forecasting Based on a Wavelet Neuro-Fuzzy System

Stéfano Frizzo Stefanon <sup>1,\*</sup> , Roberto Zanetti Freire <sup>2,\*</sup> , Leandro dos Santos Coelho <sup>2,3</sup>, Luiz Henrique Meyer <sup>4</sup>, Rafael Bartnik Grebogi <sup>5</sup>, William Gouvêa Buratto <sup>1</sup> and Ademir Nied <sup>1</sup>

<sup>1</sup> Electrical Engineering Graduate Program, Department of Electrical Engineering, Santa Catarina State University (UDESC), Joinville 89219-710, Brazil; williamburatto@gmail.com (W.G.B.); ademir.nied@udesc.br (A.N.)

<sup>2</sup> Industrial and Systems Engineering Graduate Program (PPGEPS), Pontifical Catholic University of Parana (PUCPR), Curitiba 80215-901, Brazil; leandro.coelho@pucpr.br

<sup>3</sup> Department of Electrical Engineering, Federal University of Parana (UFPR), Curitiba 81530-000, Brazil

<sup>4</sup> Electrical Engineering Graduate Program, Regional University of Blumenau (FURB), Electrical Engineering, Blumenau 89030-000, Brazil; meyer@furb.br

<sup>5</sup> Department of Computer Science, Federal Institute of Education Science and Technology of Santa Catarina (IFSC), Lages 88506-400, Brazil; rafagrebogi@gmail.com

\* Correspondence: stefanostefanon@gmail.com (S.F.S.); roberto.freire@pucpr.br (R.Z.F.)

Received: 9 December 2019; Accepted: 16 January 2020; Published: 19 January 2020



**Abstract:** The surface contamination of electrical insulators can increase the electrical conductivity of these components, which may lead to faults in the electrical power system. During inspections, ultrasound equipment is employed to detect defective insulators or those that may cause failures within a certain period. Assuming that the signal collected by the ultrasound device can be processed and used for both the detection of defective insulators and prediction of failures, this study starts by presenting an experimental procedure considering a contaminated insulator removed from the distribution line for data acquisition. Based on the obtained data set, an offline time series forecasting approach with an Adaptive Neuro-Fuzzy Inference System (ANFIS) was conducted. To improve the time series forecasting performance and to reduce the noise, Wavelet Packets Transform (WPT) was associated to the ANFIS model. Once the ANFIS model associated with WPT has distinct parameters to be adjusted, a complete evaluation concerning different model configurations was conducted. In this case, three inference system structures were evaluated: grid partition, fuzzy c-means clustering, and subtractive clustering. A performance analysis focusing on computational effort and the coefficient of determination provided additional parameter configurations for the model. Taking into account both parametrical and statistical analysis, the Wavelet Neuro-Fuzzy System with fuzzy c-means showed that it is possible to achieve impressive accuracy, even when compared to classical approaches, in the prediction of electrical insulators conditions.

**Keywords:** Adaptive Neuro-Fuzzy Inference System; insulator fault forecast; wavelet packets; time series forecasting

## 1. Introduction

Power grid insulators are responsible for supporting cables and keeping the system isolated from the ground and the other voltage phases. As these insulators are exposed to the environment, they may get contaminated by small particle deposits on their surface. The contamination does not necessarily mean that the insulator needs to be replaced, but if this contamination remains or increases, it may lead to a system failure [1]. In practice, the protection switchgear (recloser) would disconnect the line. If the insulator was seriously damaged and the defect was permanent, field personnel would have to

be sent to replace the insulator; otherwise, the recloser will put the line back in service and it will work as mentioned before.

As presented in [2], fault location and identification associated with the electrical system is considered an important issue in order to ensure the efficiency of the services associated with energy distribution. For the inspection of the electrical system and faulty insulators location, ultrasound detectors are used, which capture the ultrasonic noise of the network components. The signal generated by this equipment is an audio signal, which is electronically sampled in a time series form [3]. In order to predict the continuity of the signal generated by the ultrasound detector, an evaluation based on a modified version of the Wavelet Neuro-Fuzzy is presented in this article.

An Adaptive Neuro-Fuzzy Inference System (ANFIS) is a particular type of Artificial Neural Network (ANN) based on the Takagi–Sugeno–Kang inference model. The ANFIS method couples the benefits of both feedforward ANNs and fuzzy system techniques in the same framework [4]. Considering the best characteristics of each technique, the neuro-fuzzy network can be used to handle systems that involve inaccurate, complex and nonlinear data [5].

Neuro-fuzzy systems inherit learning and classification capacity, robustness, adaptation, nonlinear mapping, and clustering characteristics from ANNs. The behavior of these models can be understood through the observation of variables associated with the membership functions, the relationship between inputs and outputs, and from fuzzy rules due to similarities to human languages. From these aspects, the ANFIS method could be adopted for chaotic time series forecasting [6–9].

The idea of using ANFIS in this study was based on the success of applications of hybrid models. Actually, many techniques are available for the purpose of prediction, but hybrid techniques present consistent results when applied to both classification and time series forecasting applications [10,11]. In [10], assuming public datasets with concept drift, the authors proposed an ensemble technique based on the Random Forest algorithm. The algorithm exploits ensemble pruning as a forgetting strategy, and the results performed better in classification when compared to other state-of-the-art concept drift classifiers. Additionally, in [11], both wind speed and power were assumed as case studies to propose a hybrid strategy, named the ultra-short-time forecasting method, based on the Takagi–Sugeno fuzzy model. The antecedent and the consequent parts of the inference system were identified by the fuzzy *c*-means clustering algorithm, which was associated with the recursive least squares method. Considering wind farms from both China and Ireland, the proposed approach was compared with Support Vector Machines (SVM), empirical mode decomposition, and a classical back-propagation neural network, where the proposed method was shown to better predict short-term wind power.

In this article, the ANFIS model was employed for time series forecasting with the objective of evaluating its performance in predicting electrical insulator conditions, those available in the distribution network and which are susceptible to different climate and environment conditions. In this study, the signal adopted as an input for the model came from ultrasonic equipment used for electrical network inspection. Considering a normalized time series, feature extraction was performed by Wavelet Packets Transform (WPT) [9], which allows signal simplification in both time and frequency domains considering its entropy, energy, and variation.

By associating wavelets and ANNs based on a fuzzy system, recent research has shown promising results in distinct applications. In the work presented in [12], a novel fuzzy neural network structure assuming a cerebellar model neural network (CMNN) was proposed. Combining the advantages of wavelets associated with CMNN and the Takagi–Sugeno–Kang inference model, the authors compared the proposed method with traditional ANN structures, showing promising results for uncertain nonlinear systems identification.

In [13], a hybrid fuzzy wavelet neural network (HFWNN) was proposed, and the algorithm parameters were initialized considering the fuzzy *c*-means clustering method (FCM). The proposed approach considered the first layer of the network to reflect data uncertainties, while a flexible second layer performed linear combinations of the wavelet function. In this case, the HFWNN parameters were adjusted assuming a genetic algorithm optimization procedure.

Another application involving both fuzzy and wavelet methods was presented in [14], where a polynomial neural network, also assuming FCM, was applied in the premise operator to overcome dimensionality problems, while the consequence part was determined by means of wavelet functions whose parameters were estimated with the aid of the least squares method. The proposed algorithm showed an impressive ability to describe nonlinear relations between input and output variables, especially in regression and system identification problems.

Based on features extraction, an approach considering the ANFIS method associated with both wavelet and Fourier transforms was presented in [15] to solve a classification task, with the main purpose of identifying the electrical energy quality provided to an electrical system. Similar works assuming ANFIS to deal with identification or classification of electrical systems failures were presented in [16,17].

A comparison between the fuzzy learning vector quantization used in clustering, Levenberg–Marquardt, and ANFIS based on input signals provided from the wavelet transform was presented in [18]. Considering a classification case study, the objective was to evaluate fundus eye images in order to identify retinal abdominal eye disease. In this case, all methods presented 100% of success in solving this task.

An application concerning electrical energy price prediction based on both wavelets and ANFIS was presented in [19]. Following the same line as previous works mentioned in this article, the technique provided consistent results in terms of prediction even considering the nonlinear characteristic of the data set. A study assuming three performance indices to compare ANFIS with both classical ANN structure and Multivariate Linear Regression (MLR) models was presented in [20]. The main idea was to solve the prediction problem associated with the wastewater quality of the Las Vegas Wash, which is a 12-mile-long channel that feeds most of the Las Vegas Valley. The authors showed that ANFIS provided better results in terms of prediction when compared to classical ANN and MLR techniques.

Taking into account the necessity of predictive maintenance to avoid electrical system failures, those associated with electrical insulator conditions, and the consistent results provided by the ANFIS method in time series forecasting applications presented above, this research proposes the use of Wavelet Packets Transform for both signal preprocessing and feature extraction based on a data set obtained from ultrasonic equipment considering a laboratory experiment in which a contaminated electrical insulator removed from an actual transmission line was assumed for data acquisition.

As mentioned before, contaminated insulators could be the reason for electrical system failures. To avoid this situation, the prediction of the insulator condition assuming a modified ANFIS method was performed in this study considering three approaches: (i) grid partition [21]; (ii) subtractive clustering [22]; and (iii) fuzzy c-means clustering [23]. This paper presents a complete statistical evaluation of the capabilities of the ANFIS algorithm combined with WPT to predict the development of a fault in insulators of the electrical distribution system based on time series forecasting procedures.

The next section of this paper describes the problem related to the contamination of electrical insulators and their proper classification. Section 3 presents experimental procedures for data acquisition, and Section 4 addresses the proposed method assumed for time series forecasting. Section 5 shows the results and discusses the method performance. Finally, Section 6 reports the conclusions and future works associated with this research.

## 2. Description of the Electrical Insulator Problem

For more than a century, porcelain insulators have been used to support and insulate aerial conductors on transmission and distribution systems. Despite recent polymeric insulators being lighter, ceramic insulators are still being used, and some utilities still prefer them over the polymeric ones [24]. Since transmission and distribution systems run over wide and open areas, the insulators used in these systems are subjected to environmental stresses, such as pollution and contamination, along with the normally applied voltage and mechanical loads. Transient voltage due to lightning or transient mechanical stress due to strong winds are examples of stresses imposed on the insulation system [25].

The stresses which these insulators must withstand during an operational lifetime may weaken their electrical and mechanical characteristics, leading to failure. A failure would be when the voltage applied finds a way through the insulator's surface to the ground, leading to a short circuit, taking the transmission line or distribution feeder out of operation. A failure could also be mechanical, when the insulator breaks and the line or feeder may get to the ground, in this case leading to a short circuit [26].

The contamination of the insulator's surface is a great concern [27], as it may lead to other possible failure mechanisms. As contamination deposits on the insulator surface, it may increase the leakage current that flows from the live side to the ground and/or to the other phases of a polyphasic system. The increased leakage current increases the level of electrical losses, intensifies electromagnetic interference, and increases the flashover probability. Proximity to unpaved roads, coastal areas, and polluted environments—especially due to the proximity of industry, mining and agricultural activities—may increase the level of contamination and threaten the insulators' surfaces of transmission and distribution to electrical systems.

To avoid or mitigate the possibility of an insulator failure, it is important to monitor its condition. Among the various techniques available, ultrasound is one of the most employed by utilities in order to find defective insulators [28]. This method is based on the capture (and processing) of the ultrasound emitted by partial discharges that would happen in an insulator that is not working correctly.

Inspectors should be able to identify a defective insulator based on an audio signal provided by the ultrasound equipment. To identify a defective insulator, inspectors must be trained and able to detect differences in the audio signal provided by the ultrasound equipment, which is not a simple task [1]. Additionally, contaminated insulators do not represent a failure in the system, and do not need to be replaced. However, this situation may lead to failures [29]. In this way, through time series forecasting methods based on ultrasound signals of contaminated insulators, techniques can be assumed to predict failures in the system.

### 3. Data Acquisition Experiment

This section describes the data acquisition method that was performed in order to detect contaminated insulators according to a common procedure adopted by utilities.

#### 3.1. Contaminated Insulator

An actual 25 kV class insulator was taken from the local utility (CELESC—*Centrais Elétricas de Santa Catarina*, Brazil) distribution feeder, in a rural area. A controlled environment for data acquisition was prepared in a laboratory environment simulating the inspection routine. The ultrasound data were captured using ultrasound equipment. Figure 1 shows the contaminated insulator adopted in this research.



**Figure 1.** Contaminated insulator removed from a 25 kV rural area distribution system.

### 3.2. Laboratory Setup and Data Acquisition

The sample removed from the electrical system was fixed in a crossarm of a pole inside the laboratory as it would be in the field, according to the local energy utility company. The ultrasound equipment was positioned 2.2 m away from the sample. A nominal voltage of 13.8 kV (RMS, root mean square) and 60 Hz—the same as that provided by electrical feeders in both urban and rural areas in the south of Brazil—was assumed.

The ultrasound detector model 250 from Radar Engineers® was used during the experiment. The sensitivity of the equipment (gain) can be adjusted according to the intensity of the ultrasonic noise. The gain of the equipment varies from 0 to 10; in this work, the adequate gain for signal recording was 0.5, considering that, from 1.0, the signal was saturated in some measurements and thus was not considered. The signal that can be captured ranges from 1 kHz to 1 GHz [30,31].

The detector's audio output was connected to a computer through a sound card controlled by a LabVIEW® interface for data acquisition. The sampling rate assumed for data acquisition was 48 kHz, which is sufficient to process signals with a frequency lower than 20 kHz. The signal was recorded in a time series of 6.25 s, totaling a signal of 300,000 samples. For data recording and all software analysis, an Intel Core i7-3520M, with 8 GB of Random-Access Memory (RAM), with MATLAB® software was used. The signal was recorded and analyzed offline with the same computer. After the data acquisition procedures, the time series was divided into distinct data sets to perform the statistical analysis associated with the time series forecasting method presented in this work. More information about this division was presented in Section 5 in the sequence of this article.

## 4. Time Series Forecasting

The present section describes the technique employed for time series forecasting based on the data collected in the experiment described in the previous section. At first, a brief introduction about time series forecasting concepts is presented, followed by the feature extraction method assumed in this study. The ANFIS approach is presented in the sequence. Finally, an overview of the time series strategy proposed in this study is addressed.

A time series can be defined as a data set obtained considering a sampling rate in time [7]. The data set can be presumed to build a prediction model considering previous values of the time series to perform both one-step or n-steps ahead forecasting. Primarily, models were built based on the probability distribution of the data set.

According to [32], assuming the time  $t$  of available observations from a time series to forecast their value at some future time  $t + D$ , the time series can be considered stationary if no significant variations are found in the variance analysis over time. In this case, the time series is stable and shows regular behavior. If a short time series is considered, it is not usually possible to evaluate tendencies, seasonality, and irregularity in the data set [9].

Supposing that observations are available at discrete samples, at equally spaced intervals of time, a sample at instant  $t$  might be described as  $x_t$ , and previous observations that can be used to forecast the time series considering a prediction horizon  $D$  are  $\varphi(t) = [x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-\tau}]$ , where  $\tau$  represents the number of regressors assumed in the model.

A parametric autoregressive model for nonlinear time series forecasting can be defined as [33]

$$\hat{x}_{t+D}(t|\theta) = y[\varphi(t), \theta] \quad (1)$$

where  $\varphi(t)$  represents the regression vector while  $\theta$  is the vector containing the adjustable parameters of the model. Additionally,  $y$  is the function realized by the selected model. In this research,  $y$  represents the function provided by the ANFIS technique that will be addressed in the sequence of this section.

#### 4.1. Features Extraction

The present research adopted WPT for feature extraction, which represents the generalization of the wavelet transform. At each iteration, WPT performs a new decomposition based on coefficients of previous iterations. Consequently, it indicates that the final number of coefficients depends on the number of iterations (decompositions) [34].

By considering an orthogonal wavelet decomposition ( $W$ ) in the wavelet packet node level (WP), the division of approximation coefficients creates a tree structure of two vectors: the first one is the approximation coefficient vector, and the second one can be defined as a detailed vector [35]. The information lost during the approximation procedure is captured in the previously mentioned coefficients and a new vector is created. In this case, successive details are not reanalyzed [18].

The WP function can be described in the following form:

$$W_{j,k}^n(t) = 2^{j/2} W^n(2^j t - k) \quad (2)$$

where  $j$  is a scalable parameter,  $k$  represents the translation operator, and  $n$  is the oscillation parameter. The two first WP functions for  $n = 0$  and  $n = 1$  are, respectively,

$$\begin{aligned} W_{0,0}^0(t) &= \phi(t), \\ W_{0,0}^1(t) &= \psi(t). \end{aligned} \quad (3)$$

The first function of Equation (3) represents the scale function, and the second one the main function [31]. The next functions, for  $n = 2, 3, \dots, N$ , can be defined according to the following relations:

$$\begin{aligned} W_{0,0}^{2n}(t) &= \sqrt{2} \sum_k \delta(k) W_{1,k}^n(2t - k), \\ W_{0,0}^{2n+1}(t) &= \sqrt{2} \sum_k \zeta(k) W_{1,k}^n(2t - k) \end{aligned} \quad (4)$$

where  $\delta(k)$  is a low-pass filter and  $\zeta(k)$  is a high-pass filter; these are associated with the predefined scaling function and the mother wavelet function. The coefficients  $\Omega_j^n(k)$  could be obtained assuming the product of functions  $x(t)$  and  $W_{j,k}^n$ :

$$\Omega_j^n(k) = \int_{-\infty}^{\infty} x(t) W_{j,k}^n dt. \quad (5)$$

Each coefficient WP can be defined according to a specific frequency level. The wavelet transform decomposes low-frequency elements, while WPT decomposes all the elements. In this way, the use of WPT results in components of both low and high frequencies; these are called low and high approximations.

In order to use WPT, entropy, energy and variation should be considered in the WP calculation procedure. Energy is assumed to define distinct classes, and in the proposed approach, it contains failure information associated with the insulator condition. The energy fluctuation corresponds to specific types of failures, similar to the approach presented in [36]. The signal is decomposed in  $J$  levels, resulting in orthogonal subspaces, where the frequency component can be obtained using

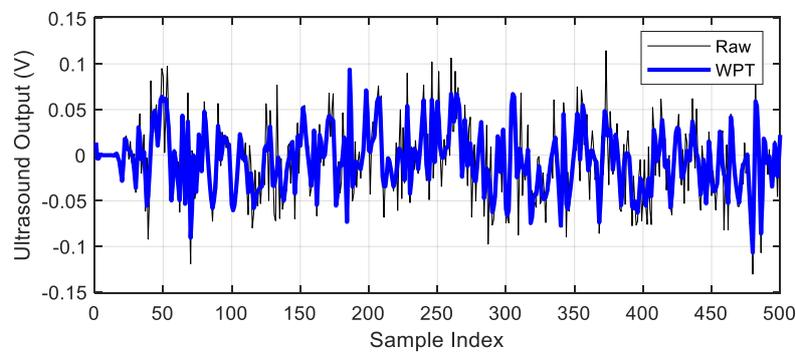
$$E_j^n = \sum_k [\Omega_j^n(k)]^2. \quad (6)$$

For energy normalization in each frequency bandwidth, the distribution percentage associated with the energy component is

$$e_j^n = \frac{E_j^n}{E_{total}} = \frac{E_j^n}{\sum_{n=1}^{2^J} E_j^n}. \quad (7)$$

The vector's relative energy describes the development in time considering subspaces of low and high frequencies. Changes in the distribution pattern describe the energy flow, which reveals the pattern to be identified. Assuming the tree structure that was previously mentioned, which was created from the division of the approximation coefficients, a binary optimal value is defined. In this way, it is possible to create new subdivisions (sub-trees) from the previous one considering the entropy criterion. Depending on the application, the resulting sub-tree can be much smaller than the original one. This technique considers that the objective is to find a minimum criterion in order to obtain an efficient algorithm [37].

The coefficients are allocated according to their Shannon entropy and are rebuilt to generate a filtered signal. Based on a data set obtained from experimental procedures described in Section 3, Figure 2 describes an example of the previously mentioned procedure considering 500 recorded points, representing 10.42 ms of data acquisition with a sampling frequency of 48 kHz. In this case, coefficients can be assumed quantitatively to represent signal distributions combining their characteristics; these could be used in an efficient way for training when associated with a time series forecasting problem.



**Figure 2.** Comparison between original and rebuild signal using Wavelet Packets Transform (WPT).

The Shannon entropy describes the energy content in a signal through the distribution of amplitude levels. The uncertainty definition is adopted in this case for probabilistic treatment purposes and can be defined as a logarithmic function  $H(\cdot)$ , given by

$$H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log(p_i) \quad (8)$$

where  $p_i$  is the occurrence probability associated with an event  $i$ . Thus, the entropy indicates the probabilistic uncertainty of a probability distribution [38]. After normalizing the input variables of the time series, the pertinence degree is calculated in the fuzzy layer. It corresponds to how the inputs satisfy the fuzzy sets associated with each input. In the rule layers, the firing level is calculated according to each rule.

To solve the forecasting problem, a data set is selected, and the mean, variance, and covariance values were used in the statistical analysis. The variance  $V_i$  of each variable can be defined as

$$V_i = \frac{1}{n-1} \sum_{m=1}^n (\hat{x}_{i,m} - \bar{x}_i)^2 \quad (9)$$

where  $\hat{x}_{i,m}$  is the value of the predicted output variable  $i$  in object  $m$ , and  $\bar{x}_i$  is the mean value.  $V_i$  indicates how far the predicted values are from expected values. The covariance  $C_{i,j}$  is the linear correlation between two random variables according to the following equation:

$$C_{i,j} = \frac{1}{n-1} \sum_{m=1}^n (\hat{x}_{i,m} - \bar{x}_i)(\hat{x}_{j,m} - \bar{x}_j). \quad (10)$$

where  $\hat{x}_{j,m}$  also represents the value of the predicted output—now for variable  $j$  in object  $m$ —and  $\bar{x}_j$  is the mean value. Here, the eigenvalues and eigenvectors are calculated and associated with the cumulative variability percentage in order to determine the main components (factors). Factors with the highest eigenvalues are selected, and indicators of each factor are then calculated. The influential characteristics are chosen based on the evaluation of indicators considering the most significant factors.

#### 4.2. Adaptive Neuro-Fuzzy Inference System

After the filtering procedures described in the previous section, the ANFIS method was applied for mapping input characteristics with the objective of creating input rules. These rules generate a set of characteristics associated with the desired output [39]. Considering an arbitrary selection of functions, the structures are predefined based on characteristics of the model variables [20]. The structure of ANFIS is a combination of a fuzzy inference system and a neural network; the summary of this architecture is presented in Figure 3.

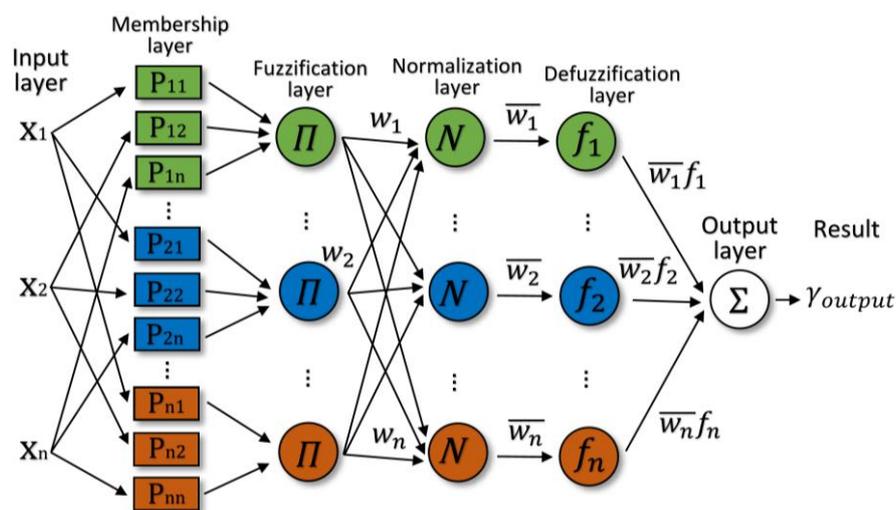


Figure 3. Adaptive Neuro-Fuzzy Inference System (ANFIS) structure for time series evaluation.

The fuzzy inference structure considering grid partitioning creates a single-output Sugeno fuzzy system, which is used as an initial condition for ANFIS training (see Figure 3). The grid partition method improves parallel processing performance, ensuring equality in the distribution of tasks to each core of the processor. For this type of cluster, a distinct rule is defined for each combination between the participation function and the correspondent output function [40]. Taking into account a subtractive cluster structure, which requires a separate data set and distinct arguments, it is possible to extract the rules sets that can identify the behavior of the time series. In this type of cluster exists a specific rule for each fuzzy cluster [41].

The fuzzy inference system based on c-means (FCM) automatically selects the number of clusters and randomly distributes the coefficients to each sample of the data set. The algorithm repeats this procedure until it reaches convergence, which means that each cluster centroid  $c_j$  should be calculated considering its membership level for  $n$  data points [42].

$$c_j = \frac{\sum_{k=1}^n w_{i,j}^m x_i}{\sum_{k=1}^n w_{i,j}^m}. \quad (11)$$

Any point  $x_i$  has a set of coefficients according to the cluster  $k$ -th degree, where  $w_{i,j}$  represents the clustering degree, and  $m$  the fuzzy partition matrix exponent. The FCM method tries to separate elements of the data set in a finite collection assuming a predefined criterion [43]. Thus, the objective function to be minimized, with  $\eta$  clusters, can be expressed by

$$\arg \min_{\eta} \sum_{i=1}^n \sum_{j=1}^{\eta} w_{i,j}^m \|x_i - c_j\|^2 \quad (12)$$

considering

$$w_{i,j} = \frac{1}{\sum_{k=1}^{\eta} \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}. \quad (13)$$

### 4.3. Algorithm Setup

Summarizing the technique procedures until this step, at first, a scalable filter was applied in the time series. In the sequence, a decomposition procedure was performed assuming Wavelet Packets Transform (WPT) from three to five levels. Previous tests showed that more levels did not improve the results obtained in this work [44]. We also considered two and three nodes during decomposition, and again, previous tests reported that, when more nodes were assumed, a loss of characteristics of the original signal was reported. The decomposition was performed to obtain a wavelet package tree; after that, WPT was applied.

For the fuzzy inference structure based on grid partition, two functions were associated with each input; in this case, Gaussian functions were utilized. The Gaussian function adopted here is given by

$$\gamma_{input}(x_i) = e^{-\frac{(x_i - u)^2}{2\sigma^2}} \quad (14)$$

where  $u$  is the center and  $\sigma$  represents the spreading parameter of the Gaussian function. For the output, a linear function was used.

In the FCM structure, 5 to 30 subtractive clusters were considered in the analysis. The influence range of each center was specified in each dimension to 0.5; i.e., for each cluster center, a spherical neighborhood with a radius equal to the previously mentioned value was assumed [14]. In order to apply standardized training procedures, the maximum number of iterations was set to 1000. Additionally, an adaptive algorithm was assumed with an initial step of 0.01, a decreasing rate equal to 0.9, and an increasing rate equal to 1.1. The hybrid neural network optimization method uses the combination of least-squares estimation and error back-propagation for training [13].

The error signal is calculated by the difference in net target  $\gamma_i$  to the net output  $\hat{\gamma}_i$  for both training and testing procedures. Finally, a metric of global error evaluation based on the root mean square error (RMSE) was assumed as a stopping criterion during training and also for testing, where

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2}. \quad (15)$$

This article presents other metrics for validation of the proposed method, such as mean absolute error (MAE) and mean absolute percentage error (MAPE). MAE denotes the mean of absolute difference between the observed value to the predicted one, given by:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\gamma_i - \hat{\gamma}_i|. \quad (16)$$

MAPE calculates the average error ratio to the correct values, where

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\gamma_i - \hat{\gamma}_i}{\gamma_i} \right|. \quad (17)$$

Based on recent studies focusing on time series forecasting [45–48], the coefficient of determination  $R^2$  was assumed as a performance criterion for model evaluation; see Equation (18). Thus,  $\bar{\gamma}_i$  is the mean of the targets ( $\gamma_i$ ), and these values represent the observed data—those acquired using the ultrasound equipment.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2}{\sum_{i=1}^n (\gamma_i - \bar{\gamma}_i)^2}. \quad (18)$$

With the objective of illustrating the procedures and methods described in this research, Figure 4 presents a flowchart of this research. The flowchart shows the analysis from the insulator which will probably develop the failure to predictability analysis.

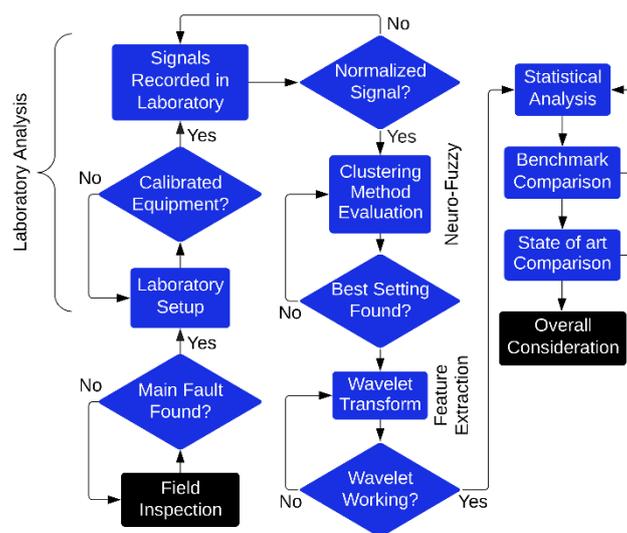


Figure 4. Method flowchart for insulator withdrawal and model evaluation.

## 5. Results and Discussion

Taking into account the parameters described in the previous section to configure both feature extraction and neuro-fuzzy methods, this section presents and discusses the results of the proposed model. This section was divided into four subsections: (i) analysis of the inference system; (ii) analysis of the fuzzy  $c$ -means clustering method; (iii) comparison of the proposed method with classical approaches; and (iv) a brief review about the state-of-the-art approaches that follow the same line of this research.

For the statistical analysis, the time series obtained in the experimental procedure presented in Section 3, which was based on a contaminated insulator, was divided into five data sets of 50,000 samples each. The percentages of each data set assumed for training, validation, and testing were 75%, 15%, and 10%, respectively. The amount of data assumed for the three phases previously mentioned was obtained based on prior evaluations of the model performance in order to avoid overfitting during both validation and testing phases. The mean results provided by the algorithms among all data sets were assumed and are presented in the next subsection. Data analysis was conducted assuming the signal obtained from the wavelet energy coefficient.

### 5.1. Analysis of the Inference System Structure

Three fuzzy inference structures were examined in this study: the first one from data using grid partition, the second one from data assuming subtractive clustering (FCM), and the third one from data using FCM clustering. Table 1 shows mean values considering the decomposed signal in wavelet packets until the third level, where one node was considered. In all tables, underlined results indicate the best result for each column.

**Table 1.** Analysis with different fuzzy inference structures. FCM: fuzzy c-means clustering method; RMSE: root mean square error.

Method	Time (s)	$R^2$			RMSE	MAE	MAPE	Standard Deviation
		Training	Validation	Testing				
Grid Partition	115.03	0.9588	0.9602	0.9592	0.0155	$8.5 \times 10^{-3}$	0.9309	$2.5 \times 10^{-3}$
FCM Clustering	158.52	0.9635	0.9643	0.9637	0.0142	$6.3 \times 10^{-3}$	0.8580	$2.6 \times 10^{-3}$
Subtractive Clustering	956.59	<u>0.9678</u>	<u>0.9685</u>	<u>0.9686</u>	<u>0.0141</u>	<u><math>9.6 \times 10^{-4}</math></u>	<u>0.1737</u>	<u><math>2.2 \times 10^{-3}</math></u>

As presented in Table 1, the grid partition structure provided the fastest results for training. However, the faster the method, the lower the performance in terms of the coefficient of determination. The subtractive clustering structure provided the best results. However, it was 87.97% more time-consuming when compared to the grid partition strategy.

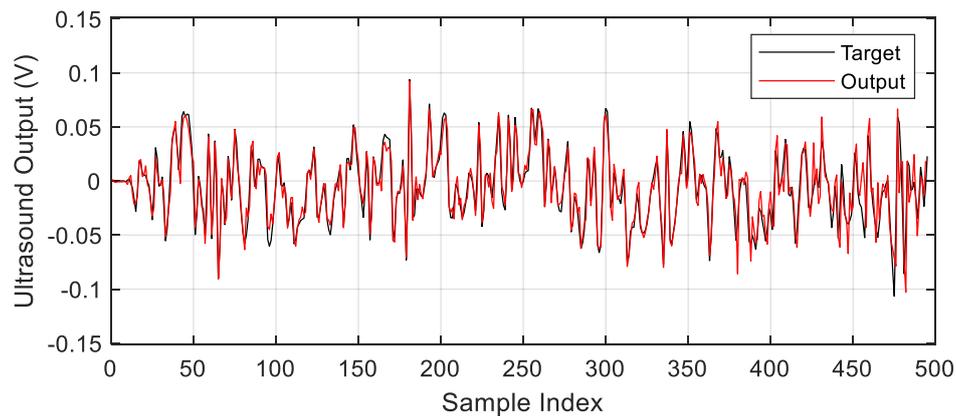
In all cases reported in Table 1, the standard deviation values indicated that the three approaches are stable, even considering distinct windows in time. Table 1 also presents the RMSE values obtained during the testing phases of each method. By analyzing the RMSE standard deviation of all methods, a small value was obtained, with this equal being to  $7.81 \times 10^{-4}$ . MAE also provided a low standard deviation value between the analyzed methods of  $3.88 \times 10^{-3}$ . Finally, MAPE values follow the trend of the RMSE. Taking this information into account, the performance analysis presented in the sequence of this article considered the coefficient of determination as the main factor.

The FCM clustering structure is widely discussed in the specialized literature, as can be seen in [9,13,14,18]. The method provided a balanced performance when both execution time and  $R^2$  were evaluated. In this case, the mean time was considered as one of the criteria assumed to select the best fuzzy inference structure. Due to these aspects, and the  $R^2$  values presented in Table 1, the method presented in the next subsection was chosen for future analysis. Additionally, distinct decomposition configurations based on wavelet packets will also be discussed. Assuming FCM clustering, Table 2 shows an evaluation of the time and algorithm forecasting performance according to the number of clusters.

**Table 2.** Evaluation of the number of clusters considering the FCM structure.

Cluster Number	Mean Time (s)	$R^2$			RMSE	MAE	MAPE	Std. Deviation
		Training	Validation	Testing				
5	<u>104.89</u>	0.9577	0.9598	0.9583	0.0146	$6.5 \times 10^{-5}$	0.8576	$3.0 \times 10^{-3}$
10	314.42	0.9618	0.9632	0.9622	0.0140	$1.2 \times 10^{-4}$	0.7639	
15	445.00	0.9640	0.9649	0.9642	0.0136	$1.6 \times 10^{-4}$	0.7574	
20	737.56	0.9650	0.9658	0.9652	0.0135	$8.1 \times 10^{-5}$	0.6211	
25	1118.56	0.9666	0.9669	0.9667	<u>0.0132</u>	$3.7 \times 10^{-5}$	0.5990	
30	1497.45	<u>0.9668</u>	<u>0.9671</u>	<u>0.9669</u>	<u>0.0132</u>	<u><math>5.1 \times 10^{-5}</math></u>	<u>0.5388</u>	

In terms of performance, it can be emphasized that the results obtained between 5 and 10 clusters. In this way, 10 clusters were used for comparison with respect to WPT configurations. In terms of execution time, a progressive increase can be observed with a proportional increase in the number of clusters. To illustrate the relation between the input (target) and the predicted (output) signals during the testing phase, Figure 5 shows the results for 500 samples considering one-step ahead forecasting, using 10 clusters.



**Figure 5.** Comparison between predicted and real data assuming an FCM structure for the testing phase.

The *RMSE*, *MAE* and *MAPE* values for the testing phase were smaller using more clusters, however the time required for convergence was longer. Again, small variations in terms of the number of clusters for *RMSE* and *MAE* were obtained.

### 5.2. Analysis of the Fuzzy C-Means Clustering Method

After defining the structure of the model, this section provides an evaluation of the fuzzy c-means clustering method. The results reported in this section employed the third, fourth and fifth levels of wavelet decomposition and three nodes. The underlined results represent the best results of each configuration, while results in bold indicate the global best results.

Table 3 presents the results for the training phase. The first number in column 1 indicates the decomposition level, while the second one represents the number of nodes.

**Table 3.** Training analysis for FCM clustering.

Levels_Nodes Number	$R^2$				
	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
3_1	0.9618	0.9616	<u>0.9677</u>	0.9612	0.9649
3_2	0.9665	0.9667	0.9666	<u>0.9682</u>	<u>0.9682</u>
4_1	0.9618	0.9616	<u>0.9683</u>	0.9611	0.9650
4_2	0.9629	0.9653	<b>0.9699</b>	0.9623	0.9694
5_1	0.9618	0.9617	<u>0.9684</u>	0.9604	0.9653
5_2	0.9672	<u>0.9686</u>	0.9663	0.9653	0.9632

The algorithm provided the best results considering four decomposition levels and two nodes. Validation results are presented in Table 4.

**Table 4.** Validation analysis for FCM clustering.

Levels_Nodes Number	$R^2$				
	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
3_1	0.9632	0.9622	<u>0.9690</u>	0.9660	0.9609
3_2	0.9669	0.9669	<u>0.9672</u>	0.9663	0.9665
4_1	0.9632	0.9622	<u>0.9698</u>	0.9660	0.9609
4_2	0.9629	0.9656	<b>0.9705</b>	0.9615	0.9659
5_1	0.9631	0.9623	<u>0.9698</u>	0.9655	0.9613
5_2	0.9681	<u>0.9692</u>	0.9663	0.9654	0.9630

When validation results were evaluated, a similar condition when compared to the training phase was observed, where both the decomposition level and the number of nodes that provided the best results for training were replicated for validation. The same behavior was obtained during the testing phase (see details in Table 5).

**Table 5.** Testing analysis for FCM clustering.

Levels_Nodes Number	$R^2$				
	Data Set 1	Data Set 2	Data Set 3	Data Set 4	Data Set 5
3_1	0.9622	0.9618	<u>0.9680</u>	0.9626	0.9639
3_2	0.9666	0.9668	<u>0.9667</u>	0.9677	0.9678
4_1	0.9622	0.9617	<u>0.9687</u>	0.9625	0.9640
4_2	0.9629	0.9654	<b>0.9700</b>	0.9621	0.9685
5_1	0.9622	0.9619	<u>0.9688</u>	0.9618	0.9643
5_2	0.9674	0.9687	0.9663	0.9654	0.9630

The comparison among distinct data sets during testing showed that the algorithm is stable, presenting variations in performance smaller than 0.79%. The best overall result was obtained considering the FCM clustering method with 10 clusters, with four levels and two nodes for the Data Set 3. The complete statistical analysis is presented in Table 6, where the covariance is calculated considering the variation in terms of the number of nodes associated to each decomposition level.

**Table 6.** Statistical results for the testing analysis of FCM clustering.

Levels_Nodes Number	Mean $R^2$	RMSE	MAE	MAPE	Std. Deviation	Variance	Covariance
3_1	0.9637	0.0140	$1.2 \times 10^{-4}$	0.7714	$2.6 \times 10^{-3}$	$6.49 \times 10^{-6}$	$1.46 \times 10^{-7}$
3_2	<u>0.9671</u>	0.0064	$2.9 \times 10^{-6}$	0.2325	$5.6 \times 10^{-4}$	$3.17 \times 10^{-7}$	
4_1	0.9638	0.0139	$1.1 \times 10^{-4}$	0.7431	$2.9 \times 10^{-3}$	$8.27 \times 10^{-6}$	
4_2	0.9658	<u>0.0061</u>	$1.3 \times 10^{-4}$	<u>0.0368</u>	$3.5 \times 10^{-3}$	$1.19 \times 10^{-5}$	$6.42 \times 10^{-6}$
5_1	0.9638	0.0140	$1.2 \times 10^{-4}$	0.7728	$3.0 \times 10^{-3}$	$8.80 \times 10^{-6}$	
5_2	0.9662	0.0063	$6.5 \times 10^{-5}$	0.2540	$2.2 \times 10^{-3}$	$4.74 \times 10^{-6}$	$1.30 \times 10^{-6}$

The algorithm provided considerable small variance values, showing that WPT can efficiently reduce the effect of noise in the time series, providing a stable algorithm. The importance of evaluating more performance measures can be highlighted at this point, as for the RMSE, three distinct configurations provided the similar results, using two nodes. The fact of adding the  $R^2$  metric contributes to the selection of the best model, as already described in this paragraph. The MAE values obtained in this case helped to confirm that three levels and three nodes provided the best model configuration.

### 5.3. Benchmarking with Nonlinear Autoregressive Methods

Assuming the task of comparing the proposed approach with well-established methods for time series forecasting, in this section, we considered two more structures: a Nonlinear Autoregressive (NAR) model, and a Nonlinear AutoRegressive with Exogenous Input (NARX) model, both of which are based on Neural Networks technique [49].

During training, three distinct classical approaches were considered: Levenberg–Marquardt (LM), Bayesian Regularization (BR), and Scaled Conjugate Gradient (SCG). Additionally, distinct configuration parameters were assumed: the number of hidden neurons (NHN), the number of regressors (ND), and the number of delayed outputs.

In the NAR network the calculation is based on Data Set 1, and in NARX networks, the calculation is based on the data relationship of Data Set 1 to Data Set 2. Data Set 2 represents values in a time

window ahead of Data Set 1. Table 7 was based on  $R^2$  and Table 8 on  $RMSE$ . These tables present the benchmark for all methods described above. Results were presented for network testing. For both hidden layers and regressors, amounts of 5, 10 and 15 were considered in the evaluation.

**Table 7.** Benchmark evaluation with nonlinear autoregressive methods based on  $R^2$ . NHN: number of hidden neurons; ND: number of regressors; LM: Levenberg–Marquardt; BR: Bayesian Regularization; SCG: Scaled Conjugate Gradient; NAR: Nonlinear Autoregressive; NARX: Nonlinear Autoregressive with Exogenous Input.

NHN	ND	NAR- $R^2$			NARX- $R^2$		
		LM	BR	SCG	LM	BR	SCG
05	05	<u>0.7441</u>	<u>0.7441</u>	0.7425	0.7409	0.7412	0.7365
	10	0.7877	<u>0.7879</u>	0.7869	0.7864	0.7869	0.7856
	15	0.8189	<u>0.8193</u>	0.8180	0.8173	0.8181	0.8171
10	05	0.7440	<u>0.7440</u>	0.7418	0.7411	0.7414	0.7367
	10	0.7874	<u>0.7880</u>	0.7856	0.7868	0.7876	0.7787
	15	0.8191	<u>0.8198</u>	0.8176	0.8176	0.8191	0.8162
15	05	0.7435	<u>0.7440</u>	0.7387	0.7410	0.7413	0.7392
	10	0.7879	<u>0.7879</u>	0.7845	0.7867	<u>0.7883</u>	0.7666
	15	0.8190	<b><u>0.8201</u></b>	0.8145	0.8175	0.8199	0.8141

**Table 8.** Benchmark evaluation with nonlinear autoregressive methods based on  $RMSE$ .

NHN	ND	NAR- $RMSE$			NARX- $RMSE$		
		LM	BR	SCG	LM	BR	SCG
05	05	0.0374	0.0377	0.0373	0.0374	0.0375	<u>0.0372</u>
	10	0.0350	0.0351	<u>0.0345</u>	<u>0.0345</u>	0.0348	<u>0.0352</u>
	15	0.0323	<u>0.0318</u>	<u>0.0350</u>	0.0323	0.0320	0.0330
10	05	0.0378	0.0379	0.0378	<u>0.0374</u>	0.0379	0.0376
	10	0.0352	0.0345	0.0344	0.0349	0.0347	<u>0.0343</u>
	15	0.0325	0.0327	0.0328	<u>0.0323</u>	0.0327	0.0325
15	05	0.0377	<u>0.0373</u>	0.0380	0.0389	0.0384	0.0375
	10	0.0348	0.0346	0.0350	0.0347	<u>0.0334</u>	0.0347
	15	0.0324	0.0325	0.0325	0.0327	<b><u>0.0323</u></b>	0.0328

In this analysis, NAR and NARX methods provided lower performance when compared to the proposed Wavelet Neuro-Fuzzy approach. In its best case, the NAR model reached 0.8201 in terms of  $R^2$  during the testing phase, which was much lower than the Wavelet Neuro-Fuzzy model, which reported 0.9700. The variation of the training method did not significantly impact the final results of both NAR and NARX models, as well as the number of hidden neurons. However, when the number of regressors was increased, an improvement in the performance associated with  $R^2$  values could be noticed. In this case, it is important to emphasize that, by increasing the number of regressors, the computational effort also increases. After 15 regressors, the maximum number of iterations (1000) was reached by both methods.

Based on  $RMSE$  results, NAR and NARX methods continued to maintain inferior results when compared to the proposed Wavelet Neuro-Fuzzy approach; even when varying both the settings and the optimization model, the  $RMSE$  values provided by these methods were much higher than the Wavelet Neuro-Fuzzy model.

#### 5.4. State-of-the-Art Approaches and Comparisons

Huang, Oh and Pedrycz presented two studies in [13] and [14] comparing different techniques with FCM and wavelets. In the proposed evaluations, other techniques based on FCM also presented

small errors. The article presented in [13] exposed how hybrid algorithms provided superior results when compared to the application of isolated techniques. In [14], the FCM method was used for the premise calculation, while the consequence calculation was obtained by wavelet functions whose parameters were estimated with the aid of the least square method.

Other work based in FCM was presented by Yang and Liu in [9], where an application focusing on time series also presented interesting results. The proposed model was also based on feature extraction through wavelets. The application considered the technique proposed by [50] for noise detection in time series. Comparisons showed that this algorithm is superior to ANFIS and the classic Artificial Neural Networks approach.

In the works reported in [5,19,20], ANFIS was assumed for time series forecasting. Fu, Cheng, Yang, and Batista showed in [20] that ANFIS provided better prediction when compared to classical approaches. Additionally, an improved Wavelet-ANFIS was proposed and the results reached 98.5% in terms of accuracy assuming three association functions.

In [18], Damayanti compared ANFIS and fuzzy learning vector quantization (FLVQ). The author showed that FLVQ provided better results for image classification purposes when wavelet transformation was used.

The ANFIS method was also assumed in [15] considering two Gaussian association functions with WPT. In this study, ANFIS was adopted to classify different types of disturbance events in power quality. The method was assumed for fuzzy inference structure evaluation based on grid partition, the same evaluated in this research and reported in the first line of Table 1. Additionally, here, the method was compared to FCM and subtractive clustering. Moreover, in [15], promising results were obtained, and an accuracy of 99.56% was obtained for the classification task. In this case, it is important to emphasize that a considerable small data set was assumed, and the variability of the method was not evaluated. In this way, even providing interesting results, there is a lack of information about the algorithm's precision and robustness.

Similar to the previously mentioned work, Babayomi and Oluseyi obtained an accuracy of around 81% for location and prediction for 10 different types of faults [16]. In this case, just the ANFIS method was assumed considering grid partition.

## 6. Conclusions and Future Research

This article presented a complete approach for predicting electrical insulator conditions. This work was based on an experimental procedure for data acquisition using a contaminated insulator, which was removed during an inspection of an electrical system in the South Region of Brazil. Ultrasound equipment was used during the experiment and a data set was obtained. To predict the condition of the insulator, a hybrid neuro-fuzzy approach was adopted. The signal provided by the ultrasound apparatus was filtered assuming a Wavelet Packets Transform in order to improve the performance of the time series forecasting model. Additionally, three inference system structures were evaluated: grid partition, fuzzy c-means clustering, and subtractive clustering. Moreover, distinct parameters as the numbers of clusters, levels, and nodes were adjusted to improve the model performance.

The application of ANFIS for time series forecasting was shown to be a reasonable approach, considering both computational effort and performance. By assuming a larger number of clusters, a considerable increase in time (computational effort) was reported, whereas no significant improvement in the result was observed in terms of coefficient of determination.

In a specific evaluation associated with the algorithm configuration, the FCM clustering method showed balanced results in terms of training time and accuracy. This approach was successfully reported by other researchers and emphasized in this work.

The statistical analysis showed that the proposed approach provided low variability, even considering distinct data sets, confirming the method's robustness for this application. Additionally, it can be emphasized that the method robustness was improved by the application of Wavelet Packets Transform for noise reduction and feature extraction.

Contaminated insulators are reported by energy companies as a frequent problem. Taking into account the fact that most of the energy network uses aerial lines without coverage, the application of this technique for insulator monitoring can provide interesting information, whether they are going to reach failure in a future horizon or not.

In addition, as an alternative approach to the use of a neuro-fuzzy system for time series forecasting, some authors are assuming deep learning techniques for the same purpose, as presented in [51,52]; for example, the Long Short-Term Memory (LSTM) method. Taking into account the fact that, in most studies, no comparisons were performed between these algorithms [53], this approach can be suggested as interesting future work when considering the same data sets. Finally, the future of this research will be focused on the development of hardware capable of detecting defective insulators early. Additionally, by associating failure classification presented in [44], and time series forecasting as discussed in this work, a more elaborated method to predict distinct types of failures in electrical insulators can be developed. The idea for future works is to combine both models focusing on the development of a specialized system capable of both to predict and classify failures as cracks, contamination among others.

**Author Contributions:** To develop the analysis presented in this article, a laboratory test and a computer analysis were performed. The laboratory test was performed by S.F.S. at the Regional University of Blumenau under the supervision of L.H.M. The computer analysis was also performed by S.F.S. within their Ph.D. in Electrical Engineering of Santa Catarina State University under the guidance of A.N. R.Z.F. and L.d.S.C. helped to elaborate the manuscript and provided sustainable analysis requirements before the submission. R.B.G. and W.G.B. assisted in the development of the research foundation. All authors have read and agreed to the published version of the manuscript.

**Funding:** National Council of Scientific and Technologic Development of Brazil—CNPq (Grants number: 304783/2017-0-PQ, 303908/2015-7-PQ, and 404659/2016-0-Univ), and PRONEX 'Fundação Araucária' 042/2018 for financial support of this work.

**Acknowledgments:** The authors are thankful to the Coordination of Superior Level Staff Improvement (CAPES), awarding a doctoral scholarship to one of the authors.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Stefenon, S.F.; de Oliveira, J.R.; Coelho, A.S.; Meyer, L.H. Diagnostic of Insulators of Conventional Grid Through LabVIEW Analysis of FFT Signal Generated from Ultrasound Detector. *IEEE Lat. Am. Trans.* **2017**, *15*, 884–889. [[CrossRef](#)]
2. Zhong, J.; Li, Y.; Yijia Cao, Y.; Sidorov, D.; Panasetsky, D. A Uniform Fault Identification and Positioning Method of Integrated Energy System. *Energy Syst. Res.* **2018**, *1*, 14–24.
3. Deng, H.; He, Z.; Chen, L. Ultrasonic Guided Wave-based Detection of Composite Insulator Debonding. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 3586–3593. [[CrossRef](#)]
4. Aghay Kaboli, S.H.; Al Hinai, A.; Al-Badi, A.; Charabi, Y.; Al Saifi, A. Prediction of Metallic Conductor Voltage Owing to Electromagnetic Coupling Via a Hybrid ANFIS and Backtracking Search Algorithm. *Energies* **2019**, *12*, 3651. [[CrossRef](#)]
5. Gil, P.; Oliveira, T.; Palma, L. Adaptive Neuro-Fuzzy Control for Discrete-Time Nonaffine Nonlinear Systems. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 1602–1615. [[CrossRef](#)]
6. Atuahene, S.; Bao, Y.; Ziggah, Y.Y.; Gyan, P.S.; Li, F. Short-Term Electric Power Forecasting Using Dual-Stage Hierarchical Wavelet- Particle Swarm Optimization- Adaptive Neuro-Fuzzy Inference System PSO-ANFIS Approach Based On Climate Change. *Energies* **2018**, *11*, 2822. [[CrossRef](#)]
7. Chena, S.; Zoub, X.; Gunawana, G.C. Fuzzy Time Series Forecasting Based on Proportions of Intervals and Particle Swarm Optimization Techniques. *Inf. Sci.* **2019**, *500*, 127–139. [[CrossRef](#)]
8. Martínez-Álvarez, F.; Troncoso, A.; Asencio-Cortés, G.; Riquelme, J.C. A Survey on Data Mining Techniques Applied to Electricity-Related Time Series Forecasting. *Energies* **2015**, *8*, 13162–13193. [[CrossRef](#)]
9. Yang, S.; Liu, J. Time-Series Forecasting Based on High-Order Fuzzy Cognitive Maps and Wavelet Transform. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 3391–3402. [[CrossRef](#)]

10. Zhukov, A.V.; Sidorov, D.N.; Foley, A.M. Random Forest Based Approach for Concept Drift Handling. In Proceedings of the International Conference on Analysis of Images, Social Networks and Texts, AIST 2016: Analysis of Images, Social Networks and Texts, Yekaterinburg, Russia, 7–9 April 2017; Volume 661, pp. 69–77.
11. Liu, F.; Li, R.; Dreglea, A. Wind Speed and Power Ultra Short-Term Robust Forecasting Based on Takagi–Sugeno Fuzzy Model. *Energies* **2019**, *12*, 3551. [[CrossRef](#)]
12. Zhao, J.; Lin, C. Wavelet-TSK-Type Fuzzy Cerebellar Model Neural Network for Uncertain Nonlinear Systems. *IEEE Trans. Fuzzy Syst.* **2019**, *27*, 549–558. [[CrossRef](#)]
13. Huang, W.; Oh, S.; Pedrycz, W. Hybrid Fuzzy Wavelet Neural Networks Architecture Based on Polynomial Neural Networks and Fuzzy Set/Relation Inference-Based Wavelet Neurons. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 3452–3462. [[PubMed](#)]
14. Huang, W.; Oh, S.; Pedrycz, W. Fuzzy Wavelet Polynomial Neural Networks: Analysis and Design. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 1329–1341. [[CrossRef](#)]
15. Alkhraijah, M.M.; Abido, M.A. Power Quality Classification Using Neuro Fuzzy Logic Inference System. In Proceedings of the IEEE-GCC Conference and Exhibition (GCCCE), Manama, Bahrain, 8–11 June 2017; pp. 1–4.
16. Babayomi, O.; Oluseyi, P.; Ofodile, N.A.; Keku, G. Fault Diagnosis in an Extra-High Voltage Power Line. In Proceedings of the IEEE PES Power Africa, Accra, Ghana, 27–30 June 2017; pp. 311–316.
17. BasantaK, P.; Ray, P.K.; Rout, P.K.; Kiran, A. Fault Detection and Classification Using Wavelet Transform and Neuro Fuzzy System. In Proceedings of the International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 1–3 March 2018; pp. 1–5.
18. Damayanti, A. Fuzzy Learning Vector Quantization, Neural Network and Fuzzy Systems for Classification Fundus Eye Images with Wavelet Transformation. In Proceedings of the International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 1–2 November 2017; pp. 331–336.
19. Karri, C.; Durgam, R.; Raghuram, K. Electricity Price Forecasting in Deregulated Power Markets using Wavelet-ANFIS-KHA. In Proceedings of the International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, India, 28–29 September 2018; pp. 982–987.
20. Fu, Z.; Cheng, J.; Yang, M.; Batista, J. Prediction of Industrial Wastewater Quality Parameters Based on Wavelet De-noised ANFIS Model. In Proceedings of the IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 301–306.
21. Nozaki, K.; Ishibuchi, H.; Tanaka, H. Adaptive Fuzzy Rule-based Classification Systems. *IEEE Trans. Fuzzy Syst.* **1996**, *4*, 238–250. [[CrossRef](#)]
22. Chiu, S. Method and Software for Extracting Fuzzy Classification Rules by Subtractive Clustering. In Proceedings of the North American Fuzzy Information Processing, Berkeley, CA, USA, 19–22 June 1996; pp. 461–465.
23. Li, E.; Wang, L.; Song, B.; Jian, S. Improved Fuzzy C-Means Clustering for Transformer Fault Diagnosis Using Dissolved Gas Analysis Data. *Energies* **2018**, *11*, 2344. [[CrossRef](#)]
24. Picolotto Corso, M.; Frizzo Stefenon, S.; Couto, V.F.; Lopes Cabral, S.H.; Nied, A. Evaluation of Methods for Electric Field Calculation in Transmission Lines. *IEEE Lat. Am. Trans.* **2018**, *16*, 2970–2976. [[CrossRef](#)]
25. Stefenon, S.F.; Americo, J.P.; Meyer, L.H.; Grebogi, R.B.; Nied, A. Analysis of the Electric Field in Porcelain Pin-Type Insulators via Finite Elements Software. *IEEE Lat. Am. Trans.* **2018**, *16*, 2505–2512. [[CrossRef](#)]
26. Park, K.C.; Motai, Y.; Yoon, J.R. Acoustic Fault Detection Technique for High-Power Insulators. *IEEE Trans. Ind. Electron.* **2017**, *64*, 9699–9708. [[CrossRef](#)]
27. Ghosh, R.; Chatterjee, B.; Chakravorti, S. A Novel Leakage Current Index for the Field Monitoring of Overhead Insulators Under Harmonic Voltage. *IEEE Trans. Ind. Electron.* **2018**, *65*, 1568–1576. [[CrossRef](#)]
28. Dong, M.; Wang, B.; Ren, M.; Zhang, C.; Zhao, W.; Albarracín, R. Joint Visualization Diagnosis of Outdoor Insulation Status with Optical and Acoustical Detections. *IEEE Trans. Power Deliv.* **2019**, *34*, 1221–1229. [[CrossRef](#)]
29. Stefenon, S.F.; Meyer, L.H. *Inspection of Electrical Distribution Network*, 1st ed.; LAP LAMBERT Academic Publishing: Saarbrücken, Germany, 2015.
30. Stefenon, S.F.; Meyer, L.H.; Molina, F.H. Analysis of the Ultrasound Emitted from Defective Insulators. In Proceedings of the International Conference on Condition Monitoring and Diagnosis, Jeju, Korea, 21–25 September 2014.

31. Stefenon, F.S.; Meyer, L.H.; Molina, F.H. Real Time Automated Diagnosis of Insulating System Employing Ultrasound Inspection. In Proceedings of the XXIII International Conference on Electricity Distribution, Lyon, France, 15–18 June 2015.
32. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; Wiley & Sons: Hoboken, NJ, USA, 2015; pp. 2–44.
33. Suykens, J.A.; Vandewalle, J.P.; Moor, B.L. *Artificial Neural Networks for Modelling and Control of Non-linear Systems*; Springer Science & Business Media: London, UK, 1996; pp. 37–54.
34. Ahmadipour, M.; Hizam, H.; Lutfi Othman, M.; Amran Mohd Radzi, M. An Anti-Islanding Protection Technique Using a Wavelet Packet Transform and a Probabilistic Neural Network. *Energies* **2018**, *11*, 2701. [[CrossRef](#)]
35. Almalki, M.M.; Hatziadoniu, C.J. Classification of Many Abnormal Events in Radial Distribution Feeders Using the Complex Morlet Wavelet and Decision Trees. *Energies* **2018**, *11*, 546. [[CrossRef](#)]
36. Gan, M.; Wang, C.; Zhu, C. Construction of Hierarchical Diagnosis Network Based on Deep Learning and its Application in the Fault Pattern Recognition of Rolling Element Bearings. *Mech. Syst. Signal Process.* **2016**, *72–73*, 92–104. [[CrossRef](#)]
37. Hong, Y.-Y.; Wei, Y.-H.; Chang, Y.-R.; Lee, Y.-D.; Liu, P.-W. Fault Detection and Location by Static Switches in Microgrids Using Wavelet Transform and Adaptive Network-Based Fuzzy Inference System. *Energies* **2014**, *7*, 2658–2675. [[CrossRef](#)]
38. Xie, H.; Sivakumar, B.; Boonstra, T.W.; Mengersen, K. Fuzzy Entropy and Its Application for Enhanced Subspace Filtering. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 1970–1982. [[CrossRef](#)]
39. Elena Dragomir, O.; Dragomir, F.; Stefan, V.; Minca, E. Adaptive Neuro-Fuzzy Inference Systems as a Strategy for Predicting and Controlling the Energy Produced from Renewable Sources. *Energies* **2015**, *8*, 13047–13061. [[CrossRef](#)]
40. Verstraete, J. The Spatial Disaggregation Problem: Simulating Reasoning Using a Fuzzy Inference System. *IEEE Trans. Fuzzy Syst.* **2017**, *25*, 627–641. [[CrossRef](#)]
41. Rezaeian, M.H.; Esmaeili, S.; Fadaeinedjad, R. Generator Coherency and Network Partitioning for Dynamic Equivalencing Using Subtractive Clustering Algorithm. *IEEE Syst. J.* **2018**, *12*, 3085–3095. [[CrossRef](#)]
42. Yang, M.; Nataliani, Y. A Feature-Reduction Fuzzy Clustering Algorithm Based on Feature-Weighted Entropy. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 817–835. [[CrossRef](#)]
43. Havens, T.C.; Bezdek, J.C.; Leckie, C.; Hall, L.O.; Palaniswami, M. Fuzzy c-Means Algorithms for Very Large Data. *IEEE Trans. Fuzzy Syst.* **2012**, *20*, 1130–1146. [[CrossRef](#)]
44. Stefenon, F.S.; Grebogi, R.B.; Freire, R.Z.; Nied, A.; Meyer, L.H. Optimized Ensemble Extreme Learning Machine for Classification of Electrical Insulators Conditions. *IEEE Trans. Ind. Electron.* **2019**. Available online: <https://ieeexplore.ieee.org/document/8758449> (accessed on 8 December 2019).
45. Barzegar, R.; Moghaddam, A.A.; Adamowski, J.; Ozga-Zielinski, B. Multi-Step Water Quality Forecasting Using a Boosting Ensemble Multi-Wavelet Extreme Learning Machine Model. *Stoch. Environ. Res. Risk Assess.* **2018**, *32*, 799–813. [[CrossRef](#)]
46. Ülke, V.; Sahin, A.; Subasi, A. A Comparison of Time Series and Machine Learning Models for Inflation Forecasting: Empirical Evidence from the USA. *Neural Comput. Appl.* **2018**, *30*, 1519–1527. [[CrossRef](#)]
47. Wang, Q.; Li, S.; Li, R. Forecasting Energy Demand in China and India: Using Single-linear, Hybrid-linear, and Non-linear Time Series Forecast Techniques. *Energy* **2018**, *161*, 821–831. [[CrossRef](#)]
48. Freire, R.Z.; Dos Santos, G.H.; Coelho, L.S. Hygrothermal Dynamic and Mould Growth Risk Predictions for Concrete Tiles by Using Least Squares Support Vector Machines. *Energies* **2017**, *10*, 1093. [[CrossRef](#)]
49. Jawad, M.; Ali, S.M.; Khan, B.; Mehmood, C.A.; Farid, U.; Ullah, Z.; Usman, S.; Fayyaz, A.; Jadoon, J.; Tareen, N.; et al. Genetic Algorithm-based Non-linear Auto-regressive with Exogenous Inputs Neural Network Short-term and Medium-term Uncertainty Modelling and Prediction for Electrical Load and Wind Speed. *J. Eng.* **2018**, *2018*, 721–729.
50. Wu, K.; Liu, J. Robust Learning of Large-scale Fuzzy Cognitive Maps via the Lasso from Noisy Time Series. *Knowl.-Based Syst.* **2016**, *113*, 23–38. [[CrossRef](#)]
51. Zhao, H.; Sun, S.; Jin, B. Sequential Fault Diagnosis Based on LSTM Neural Network. *IEEE Access* **2018**, *6*, 12929–12939. [[CrossRef](#)]

52. Lu, W.; Li, Y.; Cheng, Y.; Meng, D.; Liang, B.; Zhou, P. Early Fault Detection Approach with Deep Architectures. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 1679–1689. [[CrossRef](#)]
53. Stefenon, S.F.; Silva, M.C.; Bertol, D.W.; Meyer, L.H.; Nied, A. Fault Diagnosis of Insulators From Ultrasound Detection Using Neural Networks. *J. Intell. Fuzzy Syst.* **2019**, *37*, 6655–6664. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).