

Article

Wavelet-Like Transform to Optimize the Order of an Autoregressive Neural Network Model to Predict the Dissolved Gas Concentration in Power Transformer Oil from Sensor Data

Francisco Elânio Bezerra ¹, Fernando André Zemuner Garcia ², Silvio Ikuyo Nabeta ³, Gilberto Francisco Martha de Souza ³, Ivan Eduardo Chabu ³, Josemir Coelho Santos ³, Shigueru Nagao Junior ³  and Fabio Henrique Pereira ^{1,2,3,*}

¹ Industrial Engineering Graduate Program, Universidade Nove de Julho—UNINOVE, São Paulo 01525-000, Brazil; elanio@uni9.pro.br

² Informatics and Knowledge Management Graduate Program, Universidade Nove de Julho—UNINOVE, São Paulo 01525-000, Brazil; fernando.z.garcia@gmail.com

³ Polytechnic School, Universidade de São Paulo—EPUSP, São Paulo 05508-010, Brazil; nabeta@pea.usp.br (S.I.N.); gfmsoza@usp.br (G.F.M.d.S.); ichabu@pea.usp.br (I.E.C.); josemir@pea.usp.br (J.C.S.); snjunior@usp.br (S.N.J.)

* Correspondence: fabiohp@uni9.pro.br; Tel.: +55-11-2633-9000

Received: 18 March 2020; Accepted: 5 April 2020; Published: 11 May 2020



Abstract: Dissolved gas analysis (DGA) is one of the most important methods to analyze fault in power transformers. In general, DGA is applied in monitoring systems based upon an autoregressive model; the current value of a time series is regressed on past values of the same series, as well as present and past values of some exogenous variables. The main difficulty is to decide the order of the autoregressive model; this means determining the number of past values to be used. This study proposes a wavelet-like transform to optimize the order of the variables in a nonlinear autoregressive neural network to predict the in oil dissolved gas concentration (DGC) from sensor data. Daubechies wavelets of different lengths are used to create representations with different time delays of ten DGC, which are then subjected to a procedure based on principal components analysis (PCA) and Pearson's correlation to find out the order of an autoregressive model. The representations with optimal time delays for each DGC are applied as input in a multi-layer perceptron (MLP) network with backpropagation algorithm to predict the gas at the present and future times. This approach produces better results than choosing the same time delay for all inputs, as usual. The forecasts reached an average mean absolute percentage error (MAPE) of 5.763%, 1.525%, 1.831%, 2.869%, and 5.069% for C₂H₂, C₂H₆, C₂H₄, CH₄, and H₂, respectively.

Keywords: dissolved gas analysis; power transformers; wavelet-like transform; autoregressive model

1. Introduction

The transformer is one of the most important devices in the electricity distribution process, and reliable power distribution depends largely on the failure-free operation of this equipment. The failure of the transformer during operation can bring a significant loss of revenue to the utility, possible environmental damage, explosion and fire risks, and expensive costs of repair or replacement [1,2]. In the case in which these devices fail, operational life expectancy and reliability may change over the years and electricity to consumers may be interrupted. Therefore, the analysis of the condition and maintenance of the transformer are extremely important to ensure stable reliability of electricity [1,3–5].

When the power transformer is in normal operation, the insulating oil and solid insulating material will gradually deteriorate and a small amount of gas will be decomposed, including mainly hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), carbon monoxide (CO), and carbon dioxide (CO_2). On condition of internal transformer failure occurring, the emergence speed of these gases is accelerated [6]. So, one of the most important tools for power transformer condition monitoring and internal fault diagnosis is the transformer oil gas chromatography test, known as dissolved gas analysis (DGA) [7–10].

Several studies have addressed the creation of power transformer condition monitoring systems based on DGA. Many techniques for predicting the concentration of gases have been proposed, such as wavelet least squares, support vector regression, neural network, deep learning, fuzzy model, and long short-term memory (LSTM), just to name a few.

In general, artificial intelligence techniques have been widely used to develop more accurate diagnostic tools based on DGA data [5,9–14]. In [9], for example, a new approach for diagnosing transformer failure was created based on gas rate and support vector machine (SVM). First, on the basis of the International Electrotechnical Commission Technical Committees (IEC-TC) 10 database, optimal dissolved gas rates are obtained by genetic algorithm designed for simultaneous DGA rate selection and SVM parameter optimization. In that work, three traditional methods were used: SVM DGA, backpropagation neural network (BPNN) DGA, and IEC criteria, and three-key IEC gas proportions with SVM and back propagation neural network were employed to compare accuracy. The SVM technique also served as a basis for the approaches in [11,13,15]. The authors in [13] have used the least squares support vector machine (LS-SVM) for dissolved gases forecasting (H_2 , CH_4 , C_2H_2 , C_2H_4 , and C_2H_6) and assessing incipient faults of transformer polymer insulation. Meanwhile, in [15], a new approach has been proposed to combine technical wavelet regression with LS-SVM for the prediction of dissolved gases in power transformers immersed in oil. In [10], the authors have used a fuzzy inference system (FIS) to determine absolute concentrations of free and dissolved transformer oil, total dissolved combustible gases, total combustible gases, proportions of some gases with each other, and gas rates increasing to detect the decomposition of transformer isolation papers. A similar approach has been proposed in [5], in which an adaptive neuro fuzzy inference (ANFIS) system was employed to estimate the transformer isolation degradation rate with the input variables H_2 (hydrogen), CH_4 (methane), N_2 (nitrogen), O_2 (oxygen), CO (carbon monoxide), CO_2 (carbon dioxide), C_2H_6 (ethane), C_2H_4 (ethylene), C_2H_2 (acetylene), and TDCG (total dissolved combustible gas).

In general, these numerous studies have used artificial intelligence techniques as regression to predict gas concentration or faults in power transformers. More specifically, the use of prediction models in connection with the wavelet transformed has been addressed in some recent works to improve the forecast [13–15]. Despite satisfactory results, those approaches may not be the most efficient in predicting future values of the variable of interest, especially for a multi-step ahead forecast. Several empirical studies show that learning long-term time dependencies can be difficult for gradient-descent algorithms, which are more effective, converge faster, and generalize better in nonlinear autoregressive neural network models than in other neural networks [14,16–20]. Autoregressive models based upon neural networks specify that the output variable depends, in a non-linear way, on its own past values and on a stochastic imperfectly predictable term. Thus, the prediction of future values of the output variable can be realized from its past and present values. Additionally, the prediction model can also consider present and past values of one or more auxiliary external variables, resulting in a nonlinear autoregressive model with exogenous variables.

In this sense, the authors in [14] proposed a combination of a nonlinear autoregressive neural network model with the discrete wavelet transform, resulting in a high-accuracy multi-step ahead forecast of in-oil gas concentrations. The authors investigated the use of different wavelet functions and different time delays in the autoregression model, but they did not assess how different delays in external series can influence the values of the output series.

In fact, the definition of the optimal input and output delays is one of the main limitations of an autoregressive model. In general, in multidimensional models with n external variables, equal variables delays are adopted. This means that the prediction of the output value at time $t + 1$, $y(t + 1)$, is performed using the past outputs $y(t)$, $y(t - 1)$, $\dots$, $y(t - d_y)$ and the past observations $u_i(t)$, $u_i(t - 1)$, $\dots$, $u_i(t - d_{u_i})$ of the external variables u_i as inputs, $i = 1, \dots, n$. In addition, the adoption of many inputs can increase the complexity of the forecasting model and reduce its accuracy. Thus, some difficulties and limitations remain despite the advances, motivating research for new models to be conducted.

The investigation of the use of different time delays in external series that influence the output does not seem to have received the necessary attention, especially considering that there is a strong correlation between the concentrations of different gases and failures in transformers. This work seeks to contribute to overcome this limitation by proposing a wavelet-like transform to optimize the order of the factors in an autoregressive neural network model, with some exogenous variables, to predict the dissolved gas concentration in power transformer oil.

The main objective of this work is to determine the optimal delay for each input and for the output to create an autoregressive model with a reduced number of inputs and with competitive precision in relation to the literature. The hypothesis is that wavelet-like approximations of the external variables and the output variable incorporate the temporal memory of the autoregressive model. In addition, the selection of the best approximation for each variable determines the ideal delay for each input while reducing the size of the model, as each sample of the approximation is calculated considering a time window of the series.

Consequently, the contributions of the proposed approach can be stated as follows:

1. Development of an approach based on a wavelet-like transform that determines the optimal delay for each external variable and for the output variable in an autoregressive prediction model;
2. A prediction model with high precision as it focuses on the trend of the input signals from the noise-free approximations calculated by the wavelet transform;
3. Expansion of knowledge of the temporal relationship between gases underlying degradation process of the insulating oil and solid insulating material;
4. Reduction of the number of input variables in the autoregression model when using the approximations resulting from transformations with wavelets of different lengths, which already consider the time delay determined for each variable.

The remainder of this paper is organized as follows. The related theory is discussed in Section 2—dissolved gas-in-oil analysis in Section 2.1, discrete wavelet transform in Section 2.2, and nonlinear autoregressive exogenous model in Section 2.3. In Section 3, materials and methods will be presented, followed by the results in Section 4, discussion in Section 5, and finally the conclusion in Section 6.

2. Related Theory

2.1. Dissolved Gas-In-Oil Analysis

Power transformers are one of the most important devices for the electrical system, and more than 90% of transformers are immersed in oil [21]. Dissolved gas-in-oil (DGO) is a simple parameter used to monitor energized power transformers and assess the condition of power transformers (PTE) [22–24].

Because of the thermal and electrical stresses experienced by the insulation of operating transformers, paper and oil decomposition occurs, generating gases that dissolve in the oil and reduce its dielectric strength. Thus, concentrations of various gases dissolved in the transformer oil owing to the decomposition of the oil and paper insulation [22].

Depending on the composition and the level of gas concentration, the power transformer may fail. Usually, the gases that are present as dissolved gases in insulating oil in transformers are hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon dioxide (CO_2), carbon mono-oxide (CO), oxygen (O_2), and nitrogen (N_2) [1].

There are a couple of DGA-based interpretive methods for detecting power transformer failure that use a relation between two gases to determine a possible problem. The methods are gas key; IEC ratios; the graphical representation, IEC 60599, Duval; and Doernenburg, Rogers, among others [4,11,21]. As an example, in the following, we show three of these methods and their respective gas concentration levels. Table 1 shows the fault description for gas concentration, problem description, and normal and abnormal values for each gas concentration. In Table 2, we have fault diagnosis by the Dornenburg ratio method (R1, R2, R3, R4) and fault type, while Table 3 shows fault classification using IEC ratio codes (C_2H_2/C_2H_4 , CH_4/H_2 , C_2H_4/C_2H_6) and fault type [2,24].

Table 1. Fault description for gas concentration.

Chemical Formula	Normal	Abnormal	Problem Description
H ₂ (hydrogen)	<150 ppm	>1000 ppm	Electric discharge (corona effect, low partial discharge)
CH ₄ (methane)	<25 ppm	>80 ppm	Secondary indicator of an arc or serious overheating
N ₂ (nitrogen)	1%–10%	NA	-
O ₂ (oxygen)	0.03%	>0.5%	Transformer seal fault
CO (carbon monoxide)	<500 ppm	>1000 ppm	Cellulose decomposition
CO ₂ (carbon dioxide)	<10,000 ppm	>15,000 ppm	Cellulose decomposition
C ₂ H ₆ (ethane)	<10 ppm	>35 ppm	Secondary indicator of thermal fault
C ₂ H ₄ (ethylene)	<20 ppm	>100 ppm	Thermal fault (overheating local)
C ₂ H ₂ (acetylene)	<15 ppm	>70 ppm	Electric fault (arc, spark)

Table 2. Fault diagnosis by the Dornenburg ratio method.

Ratio R1 (CH ₄ /H ₂)	Ratio R2 (C ₂ H ₂ /C ₂ H ₄)	Ratio R3 (C ₂ H ₂ /CH ₄)	Ratio R4 (C ₂ H ₆ /C ₂ H ₂)	Fault Type
>1	<0.75	<0.3	>0.4	Thermal decomposition
<0.1	Insignificant	<0.3	>0.4	Corona
>0.1 and <1	>0.75	>0.3	<0.4	Arcing

Table 3. Fault classification using International Electrotechnical Commission (IEC) ratio codes.

C ₂ H ₂ /C ₂ H ₄	CH ₄ /H ₂	C ₂ H ₄ /C ₂ H ₆	Fault Type
0	0	0	Normal aging, no fault
Insignificant	1	0	Partial discharge of low energy density
1	1	0	partial discharge of high energy density
1	0	1	Discharges of low energy
1	0	2	Discharges of high energy
0	0	1	Thermal fault of <150 °C
0	2	0	Thermal fault of ≥150 °C and ≤300 °C
0	2	1	Thermal fault of >300 °C and ≤700 °C
0	2	2	Thermal fault of >700 °C

2.2. Discrete Wavelet Transform

Wavelet transform (WT) is a widespread signal processing technique. In the last decades, several algorithms of compactly supported wavelet have been created by mathematical analysis and signal processing communities. In fact, several works attempt to motivate and explain the basic ideas behind wavelets, what makes them so successful in many applications in different areas, as well as some limitations [25–28].

Mostly, WT is used for trend analysis, correlation and coherence between two time series, cross-spectral analysis, and space-based verification wavelets, while there are limited applications related to forecasting [29].

Discrete wavelet transform (DWT) is any wavelet transformation for which wavelets are discretely sampled, meaning this is a discrete set of the wavelet scales and translations [29]. The basic idea of this technique is to exploit the correlation present in most real-life signals to build a sparse approximation [14]. So, DWT possesses many favorable properties that are useful for researchers in the time series data mining field [1].

The dependency on discrete one-dimensional wavelet transform is presented in Equations (1)–(3).

$$f(t) = \sum_k A_{m,k} \varphi_{m,k}(t) + \sum_{j=1}^m \sum_k D_{j,k} \psi_{j,k}(t) \quad (1)$$

$$H(\omega) = \sum_{n \in \mathbb{Z}} h_n e^{-in\omega} \quad (2)$$

$$G(\omega) = \sum_{n \in \mathbb{Z}} g_n e^{-in\omega} \quad (3)$$

where $H(\omega)$ is the transfer function of high-pass filter, $G(\omega)$ is the transfer function of low-pass filter, and they filter the signs of low and high frequency keeping important information from the original signal. h_n, g_n are the coefficients depending on the mother wavelet; $A_{m,k}$ is an approximated profile; $D_{j,k}$ is the detail profile; j, m are the decomposition level of the wavelet transform; $\psi_{j,k}(t)$ is mother wavelet and $\varphi_{m,k}(t)$ is the scaling function.

Daubechies wavelets (DW), *dbN*, offer a family of orthogonal transformations, where N refers to the number of vanishing moments, which generally vary from *db2* to *db22* and have the ability to accurately approximate constant and linear functions and a relatively simple form [1,27].

2.3. Nonlinear Autoregressive Exogenous Model

A time series is any set of observations organized in time; usually measurements are made at evenly spaced times, for example, daily pollution values, monthly temperature values, daily values of electricity consumption, and daily stock exchange indices, among others [30].

An exogenous nonlinear autoregressive model (NARX) in time series modelling is a nonlinear model that has exogenous inputs, in which the model relates past values of the same series and current and past values of the driving series (exogenous), that is, externally determined series that influence the series of interest, as defined in Equation (4) for l exogenous variables:

$$y_t = F(y_{t-1}, y_{t-2}, \dots, y_{t-d_y}, \dots, u_{1,t}, u_{1,t-1}, \dots, u_{1,t-d_{u1}}, \dots, u_{l,t}, u_{l,t-1}, \dots, u_{l,t-d_{ul}}) + \varepsilon \quad (4)$$

where y is the output variable; $u_i, i = 1, \dots, l$ are externally determined variables; d_{ui} is the order of variable u_i ; ε is the error term; and F is some non-linear function, such as a polynomial for example.

There is a trend toward the adoption of computational techniques and many effective attempts have been developed, such as the following: the authors of [14] have used NARX to predict gas concentration in power transformer oil. Meanwhile, the authors of [5] and [10] have created an autoregressive model using ANFIS models to detect and isolate, as well as perform transformer paper expected life estimation.

As can be seen, there are a couple of autoregressive models that have been used to assist in the prediction of DGA and the health status of the power transformer [7]. However, in all the researched papers, the order of the factors is always the same, that is, $d_y = d_{u1} = \dots = d_{ul}$ in Equation (4).

3. Materials and Methods

The proposed approach relies on a wavelet-like transform to optimize the order of the factor (gas concentrations) in a nonlinear autoregressive model with exogenous variables. This means to define the optimal order $d_y, d_{u1}, \dots, d_{ul}$ for each gas concentration.

Thus, the approach proposed has the following steps: step 1, gas concentration acquisition and data normalization; step 2, Kaiser–Meyer–Olkin (KMO) and Bartlett test; step 3, wavelet-like decomposition of gas concentration; step 4, Pearson’s correlation; step 5, standardized regression coefficients; step 6, a model using principal components analysis (PCA) to select the principal component; step 7, calculation of contribution rate for each wavelet decomposition level; and, finally, step 8, prediction using the best time delay as input in a multi-layer perceptron (MLP) network. All these steps are illustrated in Figure 1 and described in detail as follows.

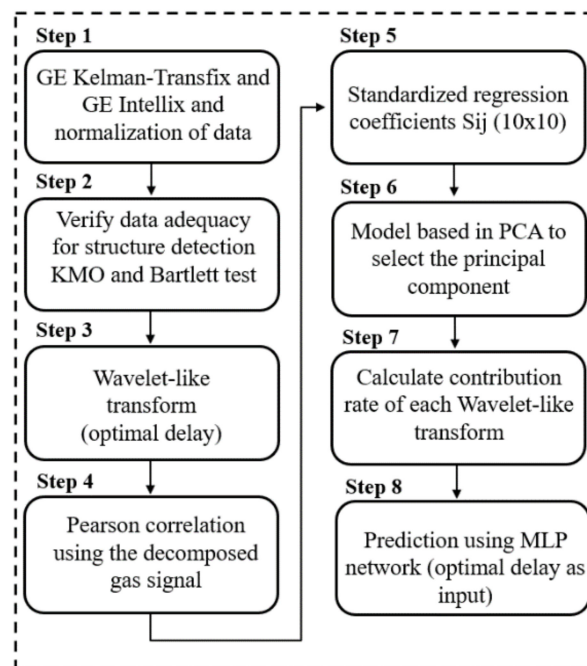


Figure 1. Model for selection and contribution rate of gases concentration and prediction. MLP, multi-layer perceptron; PCA, principal components analysis; KMO, Kaiser–Meyer–Olkin; GE, General Electric.

Step 1:

Usually, interpretation techniques such as Duval triangle are applied to the information on the concentration of gases in the transformer oil, which is collected using an equipment such as Morgan Calisto, Luman Sense Smart DGA, General Electric (GE) Transfix, Qualitrol DGA 150, or others [8].

Initially, this work collected a set of 190 historical oil-dissolved gas data from a transformer equipped with a GE Kelman-Transfix (GE—General Electric, Sao Paulo, Brazil) and GE Intellix BMT 330 (GE—General Electric, Sao Paulo, Brazil). In this stage, the variables pointed out by [10–13] are C_2H_2 , C_2H_4 , C_2H_6 , CO , CO_2 , CH_4 , O_2 , and H_2 . However, H_2O and combined gas concentrations were added as input, resulting in ten variables. Before the next step, all the data were normalized between 0 and 1.

Step 2:

The KMO test is applied to verify the measure adequacy sampling for each variable in the model [31] and the Bartlett test to test the hypothesis that the correlation matrix is an identity matrix, which would indicate that variables are unrelated, and thus unsuitable for structure detection [32].

KMO (1977) is a criterion for identifying whether a factor analysis model being used is adequately fitted to the data, testing the overall consistency of the data [31]. Meanwhile, Bartlett’s sphericity test is a technique created by Maurice Stevenson Bartlett in 1937, which indicates the strength of the relationship between variables.

Step 3:

At this stage, DWT is used in two forms. In the first one, each gas concentration is decomposed keeping level of decomposition in 1 while changing the wavelet from *db2*, *db4*, ..., to *db20*, in order to create smooth approximations of the original gas concentration using the low frequency filters. Additionally, the wavelet transform is applied in the gas concentrations in reverse chronological order so that each sample of the approximation is created with values passed from the original signal.

Considering m samples from a time series in reverse chronological order, that is, the most recent samples at the beginning, $S = (s_t, s_{t-1}, \dots, s_{t-k}, \dots, s_{t-m+k}, \dots, s_{t-m+2}, s_{t-m+1})$, and a low pass wavelet filter H of length k , $H = (h_0, h_1, \dots, h_{k-1})$, $k \ll m$, Equation (5) defines the application of the transform to the signal S to create an approximation $S'_{dbk} = (s'_t, s'_{t-1}, \dots, s'_{t-(m/2)+1})$ with time delay $k-1$, as proposed in this work,

$$s'_{t-j} = \sum_{i=0}^{k-1} h_i s_{t-i-jk}, \quad j = 0, \dots, \frac{m}{2} - 1 \quad (5)$$

Approximations S'_{dbk} , $k = 2, 4, \dots, 20$, with half the length of the original signal, $m/2$, for each Daubechies wavelets from *db2* to *db20*, are created, resulting in 10 approximations for each time series S . Here, we have 190 samples of each gas concentration.

Step 4:

Unlike the authors of [33], who have used Pearson's correlation coefficient between the constant characteristic parameter and the candidate of the variable characteristic parameters to verify the concentration of gas that presents the best correlation to electrical faults, this work uses the Pearson's correlation to calculate a relationship between the various approximations created for gas concentrations with different time delays (wavelets of different lengths). Thus, this step results in a matrix X with 110 columns and 190 rows, such that the 110 columns represent the time $t, t-2$, through $t-20$ of each gas concentration, which generates 110 input variables.

Steps 5 and 6:

In these steps, we apply PCA in the matrix A created from the relation between inputs X_j (gas concentrations delayed at time $t-2$ to $t-20$ according to wavelet-like transform) and output Y_i (a gas concentration in time instant t). So, the values of A are calculated as standardized regression coefficients a_{ij} (Equation (6)) for each input and output, describing the relationship between the concentration of a given gas and the approximations created for all other gases in different time delays generated by the wavelet transform. Therefore, a square matrix is created for each gas concentration, in which the PCA is applied to select the main components that represent at least 99% of the original data variation, generating a supervised PCA (SPCA), according to [34,35].

$$a_{ij} = \frac{X_j^T Y_i}{\sqrt{X_j^T X_j}} \quad (6)$$

Step 7:

The contribution of each time delay is calculated as follows: $C_i(j) = \sum_{i=1}^p (A_j^T A'_j) \lambda_i$, in which A represents the input data, λ are corresponding eigenvalues, A' is the representation of A in the principal component space, and p is the most important principal component [35].

Step 8:

An MLP neural network is trained with the Levenberg–Marquardt backpropagation algorithm with 100 epochs, 1 input layer, 1 hidden layer, and 1 output layer. The neurons in the hidden layer

were used following two approaches—the first one following [36], which propose a method using $N_h = 2^n - 1$, and the second following [37], proposing $N_h = (4n^2 + 3) / (n^2 - 8)$, where N_h corresponds to the best neurons numbers and n is the number of input parameters.

Unlike [3], we normalize the input data between -1 and 1 for applying a population-based metaheuristic algorithm to optimize the structure of the MLP neural network with back propagation algorithm. We propose using the optimal time delays made with the wavelet as input in an MLP with a backpropagation algorithm.

In order to test the temporal relationship between gases underlying the degradation process of the insulating oil and solid insulating material, five gas concentrations were chosen, as the main methods basically used by the IEC ratios and Rogers and Dornenburg ratios are C_2H_2 , C_2H_4 , C_2H_6 , CH_4 , and H_2 to identify possible power transformer faults.

Figure 2 shows an example of the neural network architecture to predict gas concentration C_2H_4 , where the optimal time delays for C_2H_2 , C_2H_6 , CH_4 , and O_2 are selected according Pearson's correlation and PCA.

Regarding the output, 95 samples related to odd days were selected to create the matrix A , as well as to train and test the forecasting model, as it is necessary to put the input and output data with the same length.

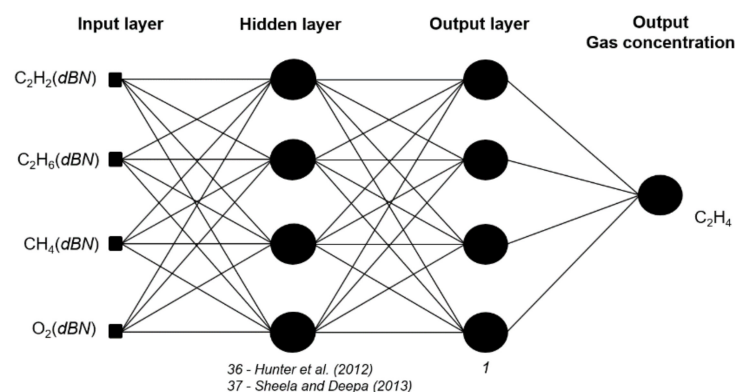


Figure 2. Gas concentration C_2H_4 decomposed by Wavelet $db2$ to $db20$.

Similar experiments were carried out for the other gases: when the output is C_2H_2 , the inputs are the approximations of C_2H_4 , C_2H_6 , CH_4 , and O_2 with their respective optimal delays defined by the proposed approach; when the output is C_2H_6 , the inputs are the wavelet approximations of C_2H_2 , C_2H_4 , CH_4 , and O_2 ; and so on.

4. Results

Firstly, we evaluated the results of the KMO and Bartlett test. In Table 4, the KMO test indicated 0.743, while at the same time, the Bartlett test indicated 0; for this reason, these data are suitable for data structure detection, while the Bartlett test indicates that a factor analysis may be useful with your data.

Table 4. Kaiser–Meyer–Olkin (KMO) and Bartlett sphericity test.

KMO and Bartlett Test		
KMO sampling adequacy measure		0.743
Aprox. Square-Root		418.644
Bartlett's sphericity test	Gl	45
	Sig.	0

The next stage shows the resulting of selection and contribution rate of decomposition of each variable. Table 5 shows the contribution rate per gases concentration (normalized).

Each variable presents a different importance rate and first order, C_2H_6 has Wavelet $db20$, which means that all gas concentration have to delay in time instant $t - 20$, while $db8$ has more impact in CH_4 , $db18$ in O_2 , and so on (see Table 5).

Table 5. Importance order and rate of each wavelet order for gas concentration.

Importance Order	Wavelet-Like Order	Gas	Importance Rate	Importance Order	Wavelet Order	Gas	Importance Rate
1	$db20$	C_2H_6	1.000	11	$db4$	GC	0.636
2	$db8$	CH_4	0.858	12	$db20$	H_2	0.581
3	$db18$	O_2	0.848	13	$db8$	C_2H_6	0.575
4	$db20$	CH_4	0.803	14	$db8$	H_2	0.572
5	$db16$	O_2	0.793	15	$db20$	C_2H_4	0.568
6	$db12$	CO_2	0.791	16	$db20$	H_2O	0.539
7	$db10$	CO_2	0.779	17	$db16$	GC	0.53
8	$db6$	CH_4	0.776	18	$db18$	GC	0.529
9	$db14$	O_2	0.692	19	$db14$	H_2	0.507
10	$db2$	GC	0.644	20	$db20$	CO	0.495

The level of correlation of the time delays for each gas concentration in is shown Table 6, wherein the values marked in bold and italics are the time delays that have higher correlation with C_2H_2 , C_2H_6 , C_2H_4 , H_2 , and CH_4 .

The following shows the results of the neural network prediction for two gases concentrations using 8 and 15 neurons in the hidden layer, as the methods DGA IEC ratios, as well as the Rogers and Dornenburg ratios, basically use the following to analyze the potential problems in power transformers: CH_4 gas, H_2 , C_2H_2 , C_2H_4 , and C_2H_6 .

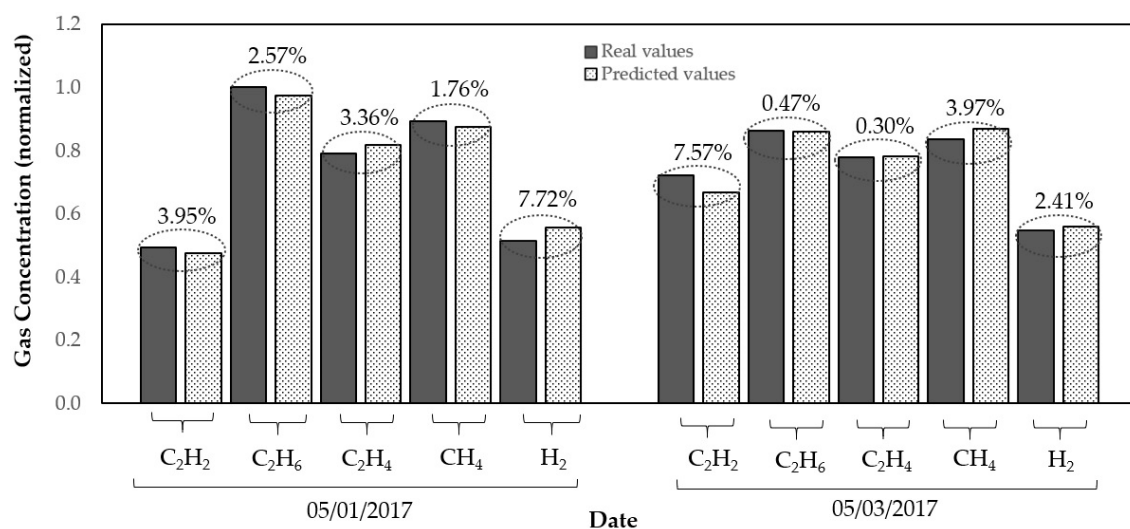
Table 6. Correlation level of time delays.

Gas Concentration (Delayed)	C_2H_2	C_2H_6	C_2H_4	H_2	CH_4
$C_2H_2(db2)$	0.01254	0.00116	<i>0.01904</i>	0.00029	0.00608
$C_2H_2(db4)$	0.00449	0.00116	0.01796	0.00109	<i>0.00689</i>
$C_2H_2(db6)$	0.00000	0.00032	0.01232	<i>0.00548</i>	0.00490
$C_2H_2(db20)$	<i>0.01769</i>	<i>0.00563</i>	0.00000	0.00137	0.00029
$C_2H_6(db4)$	0.01103	0.01440	0.03349	0.01061	<i>0.00314</i>
$C_2H_6(db6)$	<i>0.02190</i>	<i>0.02045</i>	<i>0.04203</i>	<i>0.01232</i>	0.00281
$C_2H_4(db8)$	0.00922	0.00017	0.00073	<i>0.00292</i>	0.00026
$C_2H_4(db10)$	0.00865	0.00003	0.00130	0.00044	0.00044
$C_2H_4(db12)$	<i>0.01082</i>	0.00010	<i>0.00240</i>	0.00000	<i>0.00185</i>
$C_2H_4(db20)$	0.00706	<i>0.00144</i>	0.00130	0.00036	0.00010
$H_2(db2)$	0.00410	0.00130	0.00410	<i>0.00504</i>	0.00102
$H_2(db6)$	0.00068	<i>0.00348</i>	0.00384	0.00026	0.00109
$H_2(db8)$	<i>0.00130</i>	0.00250	0.00240	0.00048	0.00017
$H_2(db20)$	0.01300	0.00130	<i>0.00608</i>	0.00176	<i>0.00270</i>
$CH_4(db4)$	0.00005	0.00281	0.00036	<i>0.00336</i>	0.00044
$CH_4(db10)$	<i>0.00922</i>	0.00058	0.00032	0.00023	0.00144
$CH_4(db14)$	0.00044	0.00000	0.00006	0.00012	<i>0.00281</i>
$CH_4(db20)$	0.00336	<i>0.00336</i>	<i>0.00212</i>	0.00160	0.00020

The results presented in Table 7 show us an average MAPE for two days of 1.525% for C_2H_6 and 1.831% for C_2H_4 . Meanwhile, Figure 3 compares the predicted values with the real values for the five gas concentrations. As can be seen, the selection of the optimal time delay in gas concentration can improve prediction accuracy, when comparing predictions with the input variables at the same time $t - 2$ and $t - 4$ (Table 7).

Table 7. Predicted values with and without selection of best delay time. MAPE, mean absolute percentage error.

Number of Neurons	Inputs/Date	Gas Concentration C ₂ H ₆			Gas Concentration C ₂ H ₄		
		05/01/2017	05/03/2017	Average MAPE%	05/01/2017	05/03/2017	Average MAPE%
8 neurons	Real	1	0.864	-	0.791	0.779	-
	Selection of time delay	0.972	0.756	7.645	0.789	0.657	7.891
	$t - 2$	0.842	0.849	8.811	0.749	0.753	4.308
	$t - 4$	0.769	0.889	13	0.786	0.847	4.64
15 neurons	Selection of time delay	0.974	0.86	1.525	0.817	0.781	1.831
	$t - 2$	0.818	0.919	12.294	0.794	0.805	1.909
	$t - 4$	0.83	0.864	8.492	0.65	0.995	22.777

**Figure 3.** Real and predicted values of the gas concentration for two days.

5. Discussion

This study aimed to study the optimal time delay of each gas concentration impacting the gases H₂, CH₄, C₂H₂, C₂H₄, and C₂H₆ (Tables 6 and 7), in which a DGA technique subsequently be used to detect the defect in the power transformer.

The approach using a wavelet-like transform and SPCA shows the contribution rate of different time delays of each gas concentration, which differs from the proposal of recent works, such as, for example [13,14]. In [14], for example, despite testing different wavelet functions and different delays, all models adopted the same time delay for external variables. Here, the approach shows the rate and order of importance and wavelet-like order for ten gas concentrations (Table 5), indicating that *db20* ($t - 20$), *db8* ($t - 8$), and *db18* ($t - 18$) are the three most important time delays for the gas concentrations C₂H₆, CH₄, and O₂, respectively. This result shows that the effect that a given gas suffers from other gases varies differently over time for each gas.

We have used Pearson's correlation to consideration the impact of each time delay as using different time delays $t - 2$ to $t - 20$ in each gas concentration, showing, for example, that to predict the concentrations of C₂H₂, the best time delays for the other gas concentrations are as follows: $t - 12$ for C₂H₄, $t - 6$ for C₂H₆, $t - 10$ for CH₄, and $t - 8$ for H₂. It is important to highlight that a traditional autoregressive model that adopts the same delay for all variables would not have identified this relationship. In addition, this is a very important result for calibrating monitoring systems, as it indicates that any variation in C₂H₄, for example, will take about 12 units of time to reflect on the concentration of C₂H₂. A similar analysis applies to other gases.

A similar kind of relationship of different gases has been studied in [38] and [33]. In [38], the authors have studied a correlation between the five gas concentrations, by applying the value of grey relational grade to reveal the relationships between gas features. Those authors show that the grey relation analysis is efficient in selecting and removing redundant features from the set of input variables. However, it does not consider any time delay in sampling the input series of gas concentrations. On the other hand, the authors in [33] have used correlation coefficients of gas concentration CO as a constant characteristic parameter for the correlation of time series analysis and H₂, CH₄, C₂H₂, C₂H₄, and C₂H₆ as characteristic variable parameters to be used to distinguish electrical faults from thermal faults.

However, approaches based on autoregressive models apply the same order for all input variables and do not take into account the time delay relationship between gas concentrations. Notwithstanding, we have seen that the optimal selection of the time delay for each concentration of gas affects the output.

Regarding forecast accuracy, this approach shows some better predictions than [33,38,39] (see Table 8).

Table 8. Comparison of predicted gas concentrations.

Authors	Approach	Average MAPE(%)				
		C ₂ H ₂	C ₂ H ₄	C ₂ H ₆	CH ₄	H ₂
Wang et al., 2015	Time series correlation	38.900	42.100	22.200	42.100	11.100
Lin et al., 2018	LSTM_DBN Network	2.450	1.450	2.100	0.260	1.890
Lu et al., 2018	ANN, SVM, LSSVM and Gaussian process regression	6.433	7.375	5.913	5.500	6.313
Zhang et al., 2018	RBFINN	4.310	5.670	5.520	3.940	4.640
	LSSVM (RBF)	3.960	5.420	2.330	1.690	3.130
Liu et al., 2019	Wavelet Least SVM and Imperialist Competition Algorithm	4.168	0.1684	1.993	0.9675	1.854
This approach	Wavelet-like transform/MLP neural network	5.763	1.831	1.525	2.869	5.069

It is important to highlight the low computational cost of the proposed model, because it takes a matter of seconds to run. In the example above regarding the prediction of the C₂H₂ concentration, instead of input 12, 6, 10, and 8 passed values of the gases C₂H₄, C₂H₆, CH₄, and H₂, respectively; according to in Equation (4), we simply use the corresponding four approximations created by the wavelet-like transform for each exogenous gas.

6. Conclusions

This work presented an approach based on the wavelet transform to determine the ideal time delay for each gas concentration and use it as a regression model in an MLP network. The application of a wavelet-like transform creates sparse approximations of gas concentrations with different time delays, and enables us to define an accurate and computationally efficient prediction model. The prediction model has high precision as it focuses on the trend of the input signals from the noise-free approximations calculated by the wavelet transform. Moreover, it reduces the number of input variables in the autoregression model when using the approximations resulting from transformations with wavelets of different lengths, which already consider the time delay determined for each variable.

In addition to simply assessing the relationship of different gases, the proposal makes it possible to identify the relationship between a specific gas and delayed approximations of all other gases. This is a fundamental result for monitoring systems, as it indicates that variations in different gases take different times to be reflected in the concentrations of other gases under observation. As an example, it was observed that variations in ethane affect acetylene concentrations more quickly than variations in ethylene. So, the approach extends the discussion of the temporal relation between gas concentrations, providing an expansion of knowledge of the temporal relationship between gases underlying degradation process of the insulating oil and solid insulating material.

Thus, wavelet-like approximations of the external variables were an effective tool to incorporate temporal memory of the autoregressive model and to determine the ideal delay for each input. As future research, we suggest investigating other feature selection techniques to evaluate the results of Pearson's correlation and PCA.

Author Contributions: F.E.B. and F.H.P., conceptualization, methodology, validation, formal analysis, investigation, data creation, and writing—original draft preparation; F.H.P., F.A.Z.G., S.I.N., and G.F.M.d.S., supervision and writing—review and editing; F.H.P. and S.I.N., project administration; I.E.C., J.C.S., and S.N.J., validation and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by INVESTCO–UHE Lajeado, under the Brazilian Electricity Regulatory Agency - ANEEL R&D Program.

Acknowledgments: The authors F.E.B., F.A.Z.G. and F.H.P. would like to thank *Universidade Nove de Julho* for the support and the scholarship granted to the first two of them.

Conflicts of Interest: The authors declare no conflict of interest. The sponsors had no role in the design, execution, interpretation, or writing of the study.

Nomenclature

ANFIS	Adaptive Neuro Fuzzy Inference
BPNN	Backpropagation Neural Network
C ₂ H ₂	Acetylene Gas
C ₂ H ₄	Ethylene Gas
C ₂ H ₆	Ethane Gas
CH ₄	Methane Gas
CO	Carbon Monoxide Gas
CO ₂	Carbon Dioxide Gas
Db	Daubechies Wavelets
DGA	Dissolved Gas Analysis
DGC	Dissolved Gas Concentration
DWT	Discrete Wavelet Transform
FIS	Fuzzy Inference System
GE	General Electric
H ₂	Hydrogen Gas
IEC	International Electrotechnical Commission
KMO	Kaiser–Meyer–Olkin
LS-SVM	Vector Machine Least Squares Support
LSTM	Long Short-Term Memory
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
NARX	Nonlinear Autoregressive Exogenous
PCA	Principal Components Analysis
PTE	Power Transformers
SPCA	Sparse Principal Components Analysis
SVM	Support Vector Machine
TDCG	Total Dissolved Combustible Gas
W-LSSVR	Vector machine least squares support wavelet regression
WT	Wavelet transform
$\psi_{j,k}(t)$	Mother Wavelet
$\varphi_{m,k}(t)$	Scaling Function
$H(\omega)$	Transfer Function of High-Pass Filter
$G(\omega)$	Transfer Function of Low-Pass Filter
h_n	High-Pass Coefficients
g_n	Low-Pass Coefficients
$A_{m,k}$	Approximated Profile

$D_{j,k}$	Detail Profile
j, m	Decomposition Level of the Wavelet Transform

References

1. Tang, W.H.; Wu, Q.H. *Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence*; Springer Publishing Company, Incorporated: London, UK, 2011.
2. Chakravorti, S.; Dey, D.; Chatterjee, B. *Recent Trends in the Condition Monitoring of Transformers: Theory, Implementation and Analysis*; Springer: London, UK, 2013. [\[CrossRef\]](#)
3. Dong, H.; Yang, X.; Li, A.; Xie, Z.; Zuo, Y. Bio-Inspired PHM Model for Diagnostics of Faults in Power Transformers Using Dissolved Gas-in-Oil Data. *Sensors* **2019**, *19*, 845. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Faria Junior, H.; Costa, J.; Olivas, J. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. *Renew. Sustain. Energy Rev.* **2015**, *46*. [\[CrossRef\]](#)
5. Prasajo, R.A.; Diwyacitta, K.; Gumilang, H. Transformer Paper Expected Life Estimation Using ANFIS Based on Oil Characteristics and Dissolved Gases (Case Study: Indonesian Transformers). *Energies* **2018**, *10*, 1135. [\[CrossRef\]](#)
6. Jun, L.; Lei, S.; Yingjie, Y.; Gehao, S.; Da, X.; Xiuchen, J. Prediction Method for Power Transformer Running State Based on LSTM_DBN Network. *Energies* **2018**, *11*, 1–14.
7. Cheng, L.; Yu, T. Dissolved Gas Analysis Principle Based Intelligent Approaches to Fault Diagnosis and Decision Making of Large Oil-Immersed Power Transformers: A Survey. *Preprints* **2018**, 2018040109. [\[CrossRef\]](#)
8. El-Hoshy, S.; Gouda, O.; El-Tamaly, H. Condition Assessment of Power Transformers Based on Dissolved Gas Analysis. *IET Gener. Transm. Distrib.* **2019**, *13*. [\[CrossRef\]](#)
9. Li, J.; Qiaogen, Z.; Wang, K.; Wang, J.; Zhou, T.; Zhang, Y. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 1198–1206. [\[CrossRef\]](#)
10. Noori, M.; Effatnejad, R.; Hajihosseini, P. Using Dissolved Gas Analysis Results to Detect and Isolate the Internal Faults of Power Transformers by Applying a Fuzzy Logic Method. *IET Gener. Transm. Distrib.* **2017**, *11*. [\[CrossRef\]](#)
11. Bacha, K.; Souahlia, S.; Gossa, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electr. Power Syst. Res.* **2012**, *83*, 73–79. [\[CrossRef\]](#)
12. Lin, C.H.; Chen, J.L.; Huang, P.Z. Dissolved gases forecast to enhance oil-immersed transformer fault diagnosis with grey prediction-clustering analysis. *Expert Syst.* **2011**, *28*, 123–137. [\[CrossRef\]](#)
13. Liu, J.; Zheng, H.; Zhang, Y.; Li, X.; Fang, J.; Liu, Y.; Liao, C.; Li, Y.; Zhao, J. Dissolved Gases Forecasting Based on Wavelet Least Squares Support Vector Regression and Imperialist Competition Algorithm for Assessing Incipient Faults of Transformer Polymer Insulation. *Polymers* **2019**, *11*, 85. [\[CrossRef\]](#)
14. Pereira, F.H.; Bezerra, F.E.; Junior, S.; Santos, J.; Chabu, I.; Souza, G.F.M.; Micerino, F.; Nabeta, S.I. Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations. *Energies* **2018**, *11*, 1691. [\[CrossRef\]](#)
15. Zheng, H.; Zhanga, Y.; Liua, J.; Weia, H.; Zhaod, J.; Liao, R. A novel model based on wavelet LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. *Electr. Power Syst. Res.* **2018**, *155*. [\[CrossRef\]](#)
16. Diaconescu, E. The use of NARX neural networks to predict chaotic timeseries. *Wseas Trans. Comput. Res.* **2008**, *3*, 182–191.
17. Lin, T.; Horne, B.G.; Tino, P.; Giles, C.L. Learning long-term dependencies in NARX recurrent neural networks. *IEEE Trans. Neural Netw.* **1996**, *7*, 1329–1351. [\[PubMed\]](#)
18. Gao, Y.; Meng, J.E. NARMAX time series model prediction: Feedforward and recurrent fuzzy neural network approaches. *Fuzzy Sets Syst.* **2005**, *150*, 331–350. [\[CrossRef\]](#)
19. Lin, T.; Giles, C.L.; Horne, B.G.; Kung, S.Y. A Delay Damage Model Selection Algorithm for NARX Neural Networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2719–2730.
20. Siegelmann, H.T.; Horne, B.G.; Giles, C.L. Computational capabilities of recurrent NARX neural networks. *IEEE Trans. Syst. Man Cybern. Part B* **1997**, *27*, 208–215. [\[CrossRef\]](#)

21. Chen, Z.; Zhang, X.; Xiong, H.; Chen, D.; Cheng, H.; Tang, J.X.; Tian, Y.; Xiao, S. Dissolved Gas Analysis in Transformer Oil Using Pt-Doped WSe₂ Monolayer Based on First Principles Method. *IEEE Access* **2019**, *7*, 72012–72019. [[CrossRef](#)]
22. Abu Bakar, N.; Abu-Siada, A.; Islam, S. A review of dissolved gas analysis measurement and interpretation techniques. *IEEE Electr. Insul. Mag.* **2014**, *30*, 39–49. [[CrossRef](#)]
23. CIGRE. *Guide for Transformer Maintenance*; CIGRE: Paris, France, 2011; Volume 445.
24. Bustamante, S.; Manana, M.; Arroyo, A.A.; Castro, P.; Laso, A.; Martínez, R. Dissolved Gas Analysis Equipment for Online Monitoring of Transformer Oil: A Review. *Sensors* **2019**, *19*, 4057. [[CrossRef](#)] [[PubMed](#)]
25. Daubechies, I. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **1988**, *41*, 909–996. [[CrossRef](#)]
26. Daubechies, I.; Sweldens, W. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.* **1998**, *4*, 247–269. [[CrossRef](#)]
27. Gogolewski, D. Influence of the edge effect on the wavelet analysis process. *Measurement* **2020**, *152*, 107314. [[CrossRef](#)]
28. Jensen, A.; la Cour-Harbo, A. *The Discrete Wavelet Transform, Ripples in Mathematics*; Springer: Berlin, Germany, 2001; p. 257.
29. Rhif, M.; Abbes, A.; Farah, I.; Martinez, B.; Sang, Y.-F. Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review. *Appl. Sci.* **2019**, *9*, 1347. [[CrossRef](#)]
30. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; HoldenDay: New Jersey, NJ, USA, 2015; p. 756.
31. Cerny, C.A.; Kaiser, H.F. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivar. Behav. Res.* **1977**, *12*, 43–47. [[CrossRef](#)]
32. Bartlett, M.S. Properties of sufficiency and statistical tests. *Proc. R. Stat. Soc.* **1937**, *160*, 268–282.
33. Wang, X.; Li, Q.; Yang, R.; Li, C.; Zhang, Y. Diagnosis of solid insulation deterioration for power transformers with dissolved gas analysis-based time series correlation. *IET Sci. Meas. Technol.* **2015**, *9*, 393–399. [[CrossRef](#)]
34. De Oliveira, J.L., Jr.; Da Rocha, J.C.F.; Guimaraes, A.M.; Da Fonseca, A.F. A PCA and SPCA based procedure to variable selection in agriculture. *Revista Brasileira de Computação Aplicada* **2015**, *7*, 30–41. [[CrossRef](#)]
35. Schimit, P.H.; Pereira, F.H. Disease spreading in complex networks: A numerical study with Principal Component Analysis. *Expert Syst. Appl.* **2018**, *97*, 41–50. [[CrossRef](#)]
36. Hunter, D.; Yu, H.; Pukish III, M.S.; Kolbusz, J.; Wilamowski, B.M. Selection of proper neural network sizes and architectures: A comparative study. *IEEE Trans. Ind. Inform.* **2012**, *8*, 228–240. [[CrossRef](#)]
37. Sheela, K.G.; Deepa, S.N. Review on methods to fix number of hidden neurons in neural networks. *Math. Prob. Eng.* **2013**, *2013*, 425740. [[CrossRef](#)]
38. Lu, S.X.; Lin, G.; Que, H.; Li, M.J.J.; Wei, C.H.; Wang, J.K. Grey relational analysis using Gaussian process regression method for dissolved gas concentration prediction. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 1313–1322. [[CrossRef](#)]
39. Liu, C.; Zhang, H.; Xie, Z.; Wang, Z.; Chen, L.; Lin, X.; Ding, H.; Li, Z.; Xiong, W. Combined Forecasting Method of Dissolved Gases Concentration and Its Application in Condition-Based Maintenance. *IEEE Trans. Power Deliv.* **2018**, *34*, 1269–1279. [[CrossRef](#)]

