

Research Article

Fuzzy Wavelet Neural Network Using a Correntropy Criterion for Nonlinear System Identification

Leandro L. S. Linhares,^{1,2} Aluisio I. R. Fontes,^{1,2} Allan M. Martins,³
Fábio M. U. Araújo,² and Luiz F. Q. Silveira²

¹Postgraduate Program in Electrical and Computer Engineering (PPgEEC), Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

²Department of Computer Engineering, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

³Department of Electrical Engineering, Federal University of Rio Grande do Norte, 59078-970 Natal, RN, Brazil

Correspondence should be addressed to Aluisio I. R. Fontes; aluisio.igor@gmail.com

Received 17 September 2014; Accepted 28 November 2014

Academic Editor: Yudong Zhang

Copyright © 2015 Leandro L. S. Linhares et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recent researches have demonstrated that the Fuzzy Wavelet Neural Networks (FWNNs) are an efficient tool to identify nonlinear systems. In these structures, features related to fuzzy logic, wavelet functions, and neural networks are combined in an architecture similar to the Adaptive Neurofuzzy Inference Systems (ANFIS). In practical applications, the experimental data set used in the identification task often contains unknown noise and outliers, which decrease the FWNN model reliability. In order to reduce the negative effects of these erroneous measurements, this work proposes the direct use of a similarity measure based on information theory in the FWNN learning procedure. The Mean Squared Error (MSE) cost function is replaced by the Maximum Correntropy Criterion (MCC) in the traditional error backpropagation (BP) algorithm. The input-output maps of a real nonlinear system studied in this work are identified from an experimental data set corrupted by different outliers rates and additive white Gaussian noise. The results demonstrate the advantages of the proposed cost function using the MCC as compared to the MSE. This work also investigates the influence of the kernel size on the performance of the MCC in the BP algorithm, since it is the only free parameter of correntropy.

1. Introduction

System identification is a modeling procedure where the mathematical representation of the input-output maps for dynamical systems can be obtained with the aid of experimental data. This procedure is a prominent alternative for the efficient modeling of complex systems without the need for using complex mathematical concepts. For this reason, this system identification plays an important role in some control engineering related tasks such as classification and decision making, monitoring, control, and prediction [1–8].

Artificial Neural Networks (ANNs) represent one of the most successful identification techniques used to model nonlinear dynamical systems [9]. This is due to their ability to learn by examples associated with intrinsic robustness

and nonlinear characteristics [10–13]. Recently, a wide variety of network structures have been used to model the input-output maps of nonlinear systems [5, 14, 15]. Multilayer Perceptron (MLP), Radial Basis Function (RBF) network, Neurofuzzy Hybrid Structures, for example, Adaptive Neurofuzzy Inference Systems (ANFIS), and Wavelet Neural Networks (WNN) are examples of ANNs commonly used in applications involving nonlinear systems [9, 13, 16, 17].

WNNs combine the flexibility of ANNs and the curve fitting ability of wavelet functions [18–20]. Besides, it can be improved in terms of extending the domain of validity by the addition of an extra layer of fuzzy structures to achieve the course delimitation of the universe of discourse, resulting in Fuzzy Wavelet Neural Networks (FWNNs) [5]. The architecture of the FWNN is very close to the traditional ANFIS [21],

although wavelets are used as membership functions (MFs) [22, 23], or in the consequent part of fuzzy rules, through the use of WNNs as local models. In literature, it is often possible to find several research works applying FWNN to deal with modeling, control, function approximation, and nonlinear system identification, among others [6, 24–28].

In [29], Linhares et al. evaluate an alternative FWNN structure to identify the nonlinear dynamics of a multisection liquid tank. The aforementioned proposed structure is similar to the ones presented by Yilmaz and Oysal [5], Abiyev and Kaynak [6], and Lu [24]. However, the FWNN presented in [29] uses only wavelets in the consequent fuzzy rules. The wavelets in each node of the FWNN consequent layer are weighted by the activation signals of the fuzzy rules. Therefore, the local models of such FWNN are solely represented by a set of wavelet functions, which differs from [5, 6, 24]. The results presented in [29] demonstrate that the modified FWNN structure maintains the generalization capability and also other important features presented by traditional FWNNs, despite the reduction in the complexity of these structures.

In practical applications, the experimental data set used in the identification procedure is often corrupted by unknown noise and outliers. The outliers are incorrect measurements which markedly deviate from the typical ranges of other observations [30]. The main source of the outliers comes from sporadic malfunctioning of sensors and equipments [31]. The presence of noise and outliers in experimental data negatively affects the performance and reliability of the dynamical model under identification, because it tries to fit such undesired measurements [30, 32, 33]. Despite the fact that there are many outlier detection methods presented in literature, many approaches are not able to detect all the outliers. Therefore, the resulting data obtained after the application of such methods may still be contaminated with outliers [30, 31].

Generally, the learning process of the neural networks is based on a given gradient method, for example, the classical error backpropagation (BP) algorithm which uses the Mean Squared Error (MSE) as its cost function. However, the applicability of MSE to obtain a model that represents an input-output relationship is optimal only if the probability distribution function (pdf) of the errors is Gaussian [34]. However, the error distribution in most cases is nongaussian and nonlinear [8]. In literature we can find some researches that demonstrate that the use of the Maximum Correntropy Criterion (MCC) replacing the traditional MSE is an effective approach to handle the problem of prediction and identification when the dynamical system has unknown noise and outliers [7, 8, 30, 35]. The correntropy evaluation allows the extraction of additional information from available data because such similarity measure takes into account all the moments of a probability distribution that are typically not observed by MSE [7].

In this work, the reliability of the FWNN recently proposed in [29] is evaluated when different percentages of outliers and noise contaminate the experimental data used to identify a nonlinear system. The aforementioned neural network is used to identify the dynamic relationship between

the input and output of a multisection liquid tank. In order to train the FWNN, the BP algorithm is used, although the traditional MSE cost function is replaced by the Maximum Correntropy Criterion using an adaptive adjustment of its kernel size, which is the free parameter of the MCC. The obtained models using each one of the quality measures are properly evaluated and compared. Despite the advantages of correntropy over MSE, little effort has been reported towards the application of correntropy to identify nonlinear systems using neural networks [7, 8]. The results presented in this work demonstrate that the FWNN architecture proposed in [29] is less sensitive to the presence of outliers and noise when it is trained using the MCC. In addition, this work also investigates the influence of the kernel size on the performance of the MCC in BP algorithm.

This paper is organized as follows. Section 2 presents the definition and the basic mathematical theory of the similarity measure of correntropy. Then, Section 3 describes the FWNN proposed in [29], which is applied in this work to identify an experimental nonlinear dynamical system considering the presence of outliers and noise. Section 4 presents the updating equations of BP algorithm, which are modified according to the MCC. Section 5 describes the proposed identification architecture in detail. Section 6 presents the multisection liquid tank under study, while the performance of FWNN models obtained using MSE and MCC cost functions is evaluated, considering the presence of both outliers and noise in experimental data. Finally, concluding remarks are given in Section 7.

2. Correntropy

Correntropy is a generalized similarity measure between two arbitrary scalar random variables X and Y defined by [36]

$$V_{\sigma}(X, Y) = E_{XY} [k_{\sigma}(X, Y)] = \int \int k(x, y) p_{X,Y}(x, y) dx dy, \quad (1)$$

where $p_{X,Y}$ is the joint probability distribution, $E[\cdot]$ is the expectation operator, and $k_{\sigma}(\cdot, \cdot)$ is a symmetric positive definite kernel. In this work, $k_{\sigma}(\cdot, \cdot)$ is a Gaussian kernel given as

$$k_{\sigma}(x_j, x_i) = K_{\sigma}(x_j - x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x_j - x_i)^2}{2\sigma^2} \right], \quad (2)$$

where σ is the variance defined as the kernel size. The kernel size may be interpreted as the resolution for which correntropy measures similarity in a space with characteristics of high dimensionality [36].

By applying a Taylor series expansion to the Gaussian function in (1) and assuming that all the moments of the joint pdf are finite, such equation becomes

$$V_{\sigma}(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k \sigma^{2k} k!} E[(X - Y)^{2k}]. \quad (3)$$

In practice, the joint pdf in (1) is unknown and only a finite amount of data $\{(x_i, y_i)\}_{i=1}^N$ is available, leading to the sample correntropy estimator defined by

$$\hat{v}_\sigma(X, Y) = \frac{1}{n} \sum_{i=1}^N k_\sigma(x_i, y_i). \quad (4)$$

Correntropy involves all the even moments of difference between X and Y . Compared with MSE $E[(X - Y)^2]$ which is a quadratic function in the joint input space, correntropy includes second-order and higher-order statistical information [37]. However, for sufficiently large values of σ , the second-order moment is predominant and the measure approaches correlation [38].

Nowadays, correntropy has been successfully used in a wide variety of applications where the signals are non-Gaussian or nonlinear, for example, automatic modulation classification [39], classification systems of pathological voices [40], and principal component analysis (PCA) [41].

2.1. Maximum Correntropy Criterion for Model Estimation. The correntropy concept can be extended to the model estimation. The variable X can be considered as a mathematical expression of the unknown function $f(X, w)$, where X is an input set $X = \{x_i \in R^m\}_{i=1, \dots, N}$ and the model parameters are $w = [w_1, \dots, w_m]^T$, which approximates the dependence on an output set $Y = \{y_i \in R^m\}_{i=1, \dots, N}$ [42].

Therefore, it is possible to determine the optimal solution for the MCC from (4) as [43]

$$\begin{aligned} \arg \max_w J(w) &= \frac{1}{M} \sum_{i=1}^M k_\sigma(f(X, w), Y) \\ &= \frac{1}{M} \sum_{i=1}^N K_\sigma(e_i), \end{aligned} \quad (5)$$

where $e_i = f(X, w) - y_i$ and $i = 1, \dots, N$, which are the errors generated by the model during the supervised learning for each of the N training samples. It is worth mentioning that such criterion is used as the cost function of the BP algorithm to adjust the parameters of the FWNN.

One of the advantages of using correntropy in system identification lies in the robustness of such measure against impulsive noise due to the use of the Gaussian kernel in (5), which is close to zero; that is, $k_\sigma(f(X, w), Y) \approx 0$ when $f(X, w)$ or Y is an outlier. Correntropy is positive and bounded, and it gives $0 < \hat{v}_\sigma(X, Y) \leq 1/\sqrt{2\pi}\sigma$ for the Gaussian kernel.

The Gaussian variance (also called kernel size) is a free parameter that must be selected by the user [38]. Therefore, when the correntropy is estimated, the resulting values depend on the selected kernel size. In addition, the kernel size of correntropy influences the nature of the performance surface, presence of local optima, rate of convergence, and robustness to impulsive noise during adaption [37, 43]. If the training data size is not large enough, the kernel size must be chosen considering tradeoffs between outlier rejection and estimation efficiency [44].

Some approaches can be employed to determine the kernel size, for example, the statistical method [45], Silverman's rule [46], cross validation techniques [47, 48], and shape of the prediction error distribution [44]. This work uses an adaptive kernel size algorithm [42], which is given by

$$\sigma = \frac{\max |e_i|}{2\sqrt{2}}, \quad i = 1, \dots, N. \quad (6)$$

In order to assess the improved performance of an adaptive kernel size over fixed ones, Section 6 is supposed to show how the error evolves during the FWNN training for different values of the kernel size.

3. Fuzzy Wavelet Neural Networks

3.1. Brief Review. Wavelets are obtained by scaling and translating a special function $\psi(x)$ localized in both time/space and frequency called mother wavelet, which can be defined in such a way to serve as a basis to describe other functions. Wavelets are extensively used in the fields of signal analysis, identification and control of dynamical systems, computer vision, and computer graphics, among other applications [49–52]. Given $\psi(x)$, the corresponding family of wavelets is obtained by

$$\Psi_j(\mathbf{x}) = |\mathbf{d}_j|^{-1/2} \psi\left(\frac{\mathbf{x} - \mathbf{t}_j}{\mathbf{d}_j}\right), \quad \mathbf{d}_j \neq 0, \quad j = 1, 2, \dots, k, \quad (7)$$

where $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ and $\Psi_j(\mathbf{x})$ is obtained from $\psi(\mathbf{x})$ by scaling it by a factor $\mathbf{d}_j = \{d_{1j}, d_{2j}, \dots, d_{nj}\}$ and translating it by $\mathbf{t}_j = \{t_{1j}, t_{2j}, \dots, t_{nj}\}$.

A WNN is a nonlinear regression structure that can represent input-output maps by combining wavelets with appropriate scalings and translations [53]. The output of a WNN is determined as follows:

$$y = \sum_{j=1}^k \mathbf{w}_j \Psi_j(\mathbf{x}) = \sum_{j=1}^k \mathbf{w}_j |\mathbf{d}_j|^{-1/2} \psi\left(\frac{\mathbf{x} - \mathbf{t}_j}{\mathbf{d}_j}\right), \quad (8)$$

where \mathbf{w}_j are the synaptic weights, \mathbf{x} is the input vector, and \mathbf{d}_j and \mathbf{t}_j are parameters characterizing the wavelets.

In a concise manner, the purpose of FWNNs is to incorporate WNNs into the ANFIS structure in order to obtain faster convergence and better approximation capabilities, eventually with a greater number of parameters to be adjusted. The fuzzy rules allow tackling the uncertainties, while wavelets contribute to improving the accuracy in the process of approximating input-output maps [6].

3.2. FWNN Architecture. A particular instance of FWNN proposed in [29] is applied in this work to identify a real nonlinear system, investigating its performance and reliability when the experimental data set is corrupted by unknown noise and outliers. In this FWNN architecture, the consequent part of its fuzzy rules is described only by wavelet functions. It differs from other structures such as those proposed in [5, 6, 24]. The basic architecture of the FWNN can be seen in Figure 1 and its layers are described as follows.

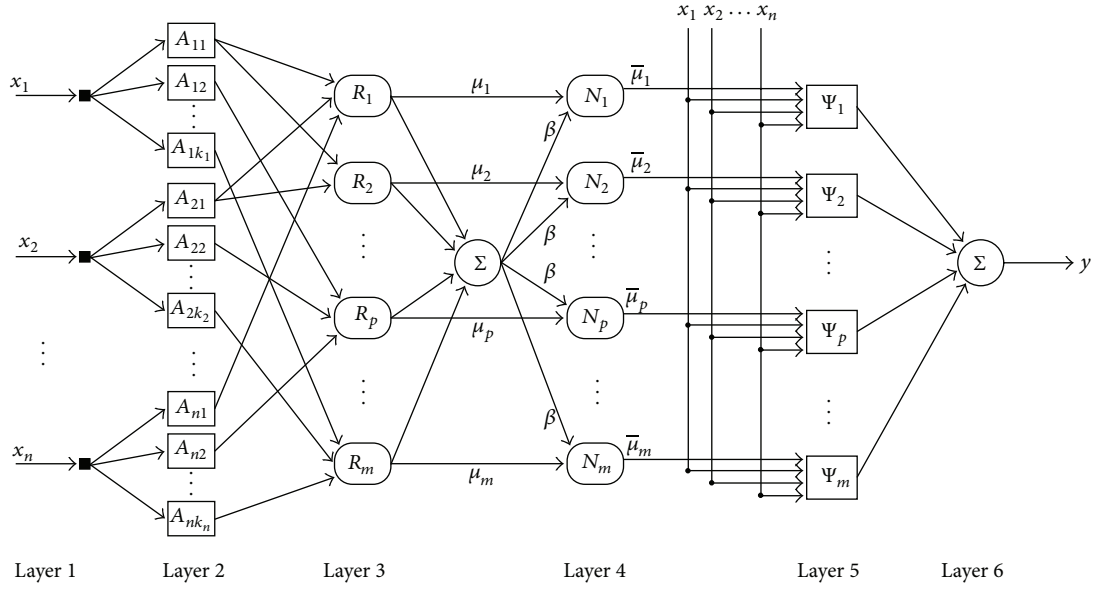


FIGURE 1: Graphical representation of the FWNN architecture.

Layer 1. The input layer just transfers the input signal vector $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ to the next layer.

Layer 2. In the fuzzification layer, the membership functions are parameterized to match the specific requirements of a variety of applications. For instance, a Gaussian membership function can be described by the following equation:

$$A_{qr}(x_q) = \exp \left[-\frac{1}{2} \left(\frac{x_q - a_{qr}}{b_{qr}} \right)^2 \right], \quad (9)$$

where for $q = 1, 2, \dots, n$ and $r = 1, 2, \dots, k_q$, A_{qr} would be associated with the r th membership function appearing in a given rule and evaluated for the q th component of the input vector. The adjustable parameters are a_{qr} and b_{qr} , representing the center and width of the membership function, respectively.

Layer 3. This is the inference layer. Assuming that there are m rules, where R_i is a given rule and $i = 1, 2, \dots, m$, each rule is supposed to produce and output μ_i by aggregating A_{qr} using a T-norm. The output of the p th rule in this layer is

$$\mu_p = \prod_{q=1}^n A_{qr}(x_q), \quad p = r_1, r_2, \dots, r_n, \quad (10)$$

where $r_1 = 1, \dots, k_1, r_2 = 1, \dots, k_2, r_n = 1, \dots, k_n$.

All the rule outputs of this layer are added up to the summation node located between Layers 3 and 4. The output β of this node is later used in the normalization stage.

Layer 4. In the normalization layer, the normalization factor for the output of the i th rule, denoted by $\bar{\mu}_i$ is given by

$$\bar{\mu}_i = \frac{\mu_i}{\beta} = \frac{\mu_i}{\sum_{i=1}^m \mu_i}, \quad i = 1, 2, \dots, m. \quad (11)$$

Layer 5. This is the consequent layer of the FWNN. In this work, the Mexican Hat family of wavelets is adopted as in [5, 6, 54]. Its mathematical representation is given by

$$\psi(x) = \frac{1}{\sqrt{d}} \left(1 - \left(\frac{x-t}{d} \right)^2 \right) \exp \left[-0.5 \left(\frac{x-t}{d} \right)^2 \right]. \quad (12)$$

The inputs of the wavelet layer are the normalized weights $\bar{\mu}_i$ and the input vector $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, while the outputs of this layer represented by f_j are given by

$$f_j = \bar{\mu}_j \Psi_j, \quad j = 1, 2, \dots, m, \quad (13)$$

$$f_j = \bar{\mu}_j \sum_{i=1}^n \left(1 - T_{ij}^2 \right) \exp \left[-0.5 T_{ij}^2 \frac{1}{\sqrt{d_{ij}}} \right],$$

where the term $T_{ij} = (x_i - t_{ij})/d_{ij}$, $d_{ij} > 0$ is adopted to simplify the mathematical notation and n is the number of wavelet functions in a node of Layer 5.

Layer 6. In the output layer, all signals from the wavelet neurons are summed up as follows:

$$y = \sum_{j=1}^m f_j. \quad (14)$$

By observing Figure 1 and considering (9) to (14), it is possible to notice that the FWNN related parameters are located in the second and fifth layers. The membership functions and wavelet functions are adjusted according to the application using any learning algorithm, such as BP algorithm.

4. Error Backpropagation Algorithm with MCC

The classical BP algorithm is the learning algorithm used in this work to adjust the free parameters of the FWNN models. According to [54], this algorithm is probably the most frequently used technique to train a FWNN. Despite its functionality, it presents some shortcomings such as the fact that it may get stuck on a local minimum of the error surface and that the training convergence rate is generally slow [55–57]. However, it is well known that the use of wavelet functions in neural network structures reduces such inconveniences [6, 58].

A neural system should be designed to present a desired behavior; hence, it is necessary to define a cost function for this task. It provides an evaluation of the quality of the solution obtained by the neural model [59]. The gradient based learning algorithms, such as the BP algorithm, require the differentiation of the chosen cost function with respect to the adjustable parameters of the FWNN model. Therefore, it is necessary to obtain the partial derivatives of the chosen cost function with respect to parameters d_{ij} and t_{ij} of the wavelets and parameters a_{qr} and b_{qr} of the membership functions A_{qr} .

Typically, MSE is the cost function used with BP algorithm [10]. Such classical cost criterion is replaced by MCC in this work in order to increase the reliability of the FWNN model when the identified dynamical system presents outliers and noise. When using MCC, the main goal is to maximize the correntropy similarity measure between two random process variables. In the FWNN learning procedure, such variables are the desired output y_d and the estimated output \hat{y} provided by the FWNN model. Considering the estimation error of the FWNN model given by $e = y_d - \hat{y}$, maximizing the MCC is equivalent to minimizing

$$\mathcal{E}_{av} = \frac{1}{N} \sum_{n=1}^N \left[\frac{1}{\sqrt{2\pi}\sigma} - K_\sigma(e_n) \right], \quad (15)$$

where N is the number of samples in the experimental data. Equation (15) corresponds to the cost function used during the minimization process of the BP algorithm applied to adjust the parameters of the FWNN models. As such parameters are adjusted sequentially, (16) defines the instantaneous correntropy used to update the wavelet functions and membership functions parameters of the FWNN after each training pair is presented to this network. Consider

$$\mathcal{E} = \frac{1}{\sqrt{2\pi}\sigma} - K_\sigma(e_n). \quad (16)$$

By differentiating \mathcal{E} with respect to d_{ij} and t_{ij} , it gives

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial d_{ij}} &= -\frac{e_n K_\sigma(e_n) \bar{\mu}_i \left(\exp[-0.5T_{ij}^2] (3T_{ij}^2 - T_{ij}^4) \right)}{\sigma^2 \sqrt{d_{ij}^3}}, \\ \frac{\partial \mathcal{E}}{\partial t_{ij}} &= -\frac{e_n K_\sigma(e_n) \bar{\mu}_i \left(\exp[-0.5T_{ij}] (3T_{ij} - T_{ij}^3) \right)}{\sigma^2 \sqrt{d_{ij}^3}}. \end{aligned} \quad (17)$$

Now, differentiating \mathcal{E} with respect to a_{qr} and b_{qr} , it gives

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial a_{qr}} &= \frac{\partial \mathcal{E}}{\partial y} \frac{\partial y}{\partial \Psi} \frac{\partial \Psi}{\partial \bar{\mu}} \frac{\partial \bar{\mu}}{\partial \mu} \frac{\partial \mu}{\partial A} \frac{\partial A}{\partial a_{qr}}, \\ \frac{\partial \mathcal{E}}{\partial b_{qr}} &= \frac{\partial \mathcal{E}}{\partial y} \frac{\partial y}{\partial \Psi} \frac{\partial \Psi}{\partial \bar{\mu}} \frac{\partial \bar{\mu}}{\partial \mu} \frac{\partial \mu}{\partial A} \frac{\partial A}{\partial b_{qr}}, \end{aligned} \quad (18)$$

where

$$\begin{aligned} \frac{\partial A_{qr}}{\partial b_{qr}} &= \begin{cases} A_{qr} \frac{(x_q - b_{qr})}{a_{qr}^2}, & \text{if } x_q \neq b_{qr}, \\ 0, & \text{if } x_q = b_{qr}, \end{cases} \\ \frac{\partial A_{qr}}{\partial b_{qr}} &= \begin{cases} A_{qr} \frac{(x_q - b_{qr})^2}{a_{qr}^3}, & \text{if } x_q \neq b_{qr}, \\ 0, & \text{if } x_q = b_{qr}. \end{cases} \end{aligned} \quad (19)$$

Following the delta rule mentioned in [10], the parameters of the proposed FWNN are updated as follows:

$$\begin{aligned} d_{ij}(k+1) &= d_{ij}(k) - \eta \frac{\partial \mathcal{E}}{\partial d_{ij}}, \\ t_{ij}(k+1) &= t_{ij}(k) - \eta \frac{\partial \mathcal{E}}{\partial t_{ij}}, \\ a_{qr}(k+1) &= a_{qr}(k) - \eta \frac{\partial \mathcal{E}}{\partial a_{qr}}, \\ b_{qr}(k+1) &= b_{qr}(k) - \eta \frac{\partial \mathcal{E}}{\partial b_{qr}}, \end{aligned} \quad (20)$$

where η is the learning rate. For the training algorithm initialization, wavelets and membership functions parameters are set with random numbers from a uniform distribution.

The replacement of the traditional MSE by MCC inserts another learning parameter to BP algorithm. As already explained, the success of the correntropy is based on the appropriate adjustment of the kernel size of its Gaussian functions. This new parameter influences the nature of the performance surface, presence of local optima, rate of convergence, and robustness. Therefore, if an unsuitable kernel size is chosen, the expected improved performance of the MCC will not be confirmed [60]. For this reason, an adaptive kernel method is applied in this work (see (6)) to adjust the kernel size over the learning epochs.

5. Proposed Identification Architecture

The proposed architecture adopted in this work identifies the dynamic relationship between the input and output of a multisection tank for water storage. The system is evaluated when the experimental data used during the identification task is corrupted with noise and outliers. The proposed architecture is based on the series-parallel identification scheme described in [13], with small modifications due to the experimental data set characteristic and the learning

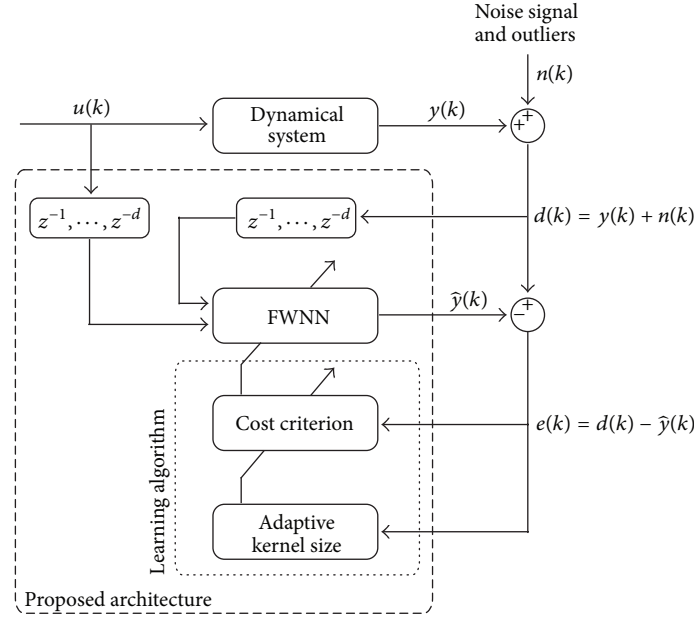


FIGURE 2: Proposed identification architecture.

procedure used to adjust the parameters of the FWNN model. Figure 2 presents a schematic diagram of the proposed identification architecture in this work.

The inputs of the FWNN model are past values of input signal $u(k)$ and the system output when corrupted with noise and outliers $d(k) = y(k) + n(k)$, while the estimated output is given by $\hat{y}(k)$. The work developed in [9] shows that well-known linear modeling structures, such as FIR (Finite Impulse Response), ARX (AutoRegressive, eXogenous input), ARMAX (AutoRegressive, Moving Average, eXogenous input), OE (Output Error), and SSIF (State Space Innovations Form) may be extended by using nonlinear functions or representations, thus leading to the nonlinear modeling structures NFIR, NARX, NARMAX, NOE, and NSSIF. This concept is used to define the inputs of the FWNN models obtained in this study.

According to [9], the advantage of a NARX model is that none of its regressors depends on past outputs of the model, which ensures that the predictor remains stable. This is particularly important in the nonlinear case since the stability issue in this particular case is much more complex than in linear systems. Considering that the inputs of the FWNN models in this work are described exactly as the regression vector of the NARX modeling structure, they inherit important characteristics from such structure. Figure 3 shows more details on the FWNN inputs in accordance with the NARX structure, where λ , τ , and α are constants that define a model of order α and delay τ .

Figure 2 illustrates that the FWNN model parameters are updated according to the error signal $e(k) = d(k) - \hat{y}(k)$, by using a learning algorithm, for example, the BP algorithm. By adopting the MCC as its respective cost criterion, the learning algorithm is applied to the FWNN model. As it was previously explained in Section 2, the success of the MCC also

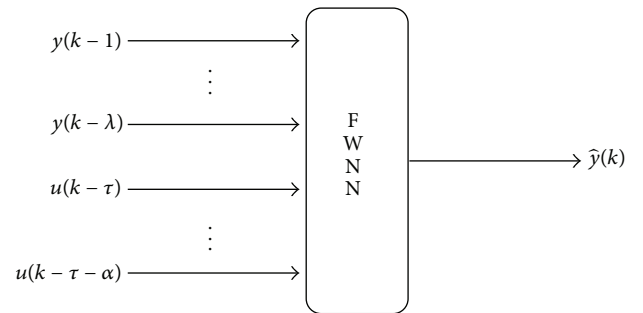


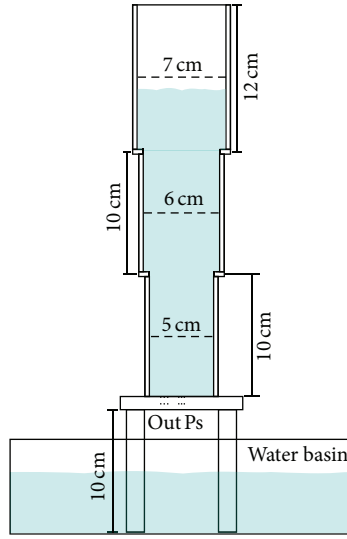
FIGURE 3: NARX model structure using FWNN.

depends on the correct choice of the kernel size. Therefore, the adaptive method described by (6) is used in this work to adjust the kernel size during the learning epochs.

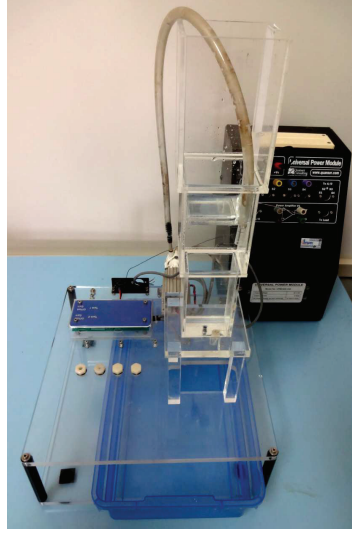
6. Experiments and Results

In order to evaluate the performance of the FWNN when the traditional MSE cost criterion of the error backpropagation algorithm is replaced by MCC, the aforementioned neural network is used to identify a real dynamical system, considering that its experimental data is corrupted by noise and outliers.

6.1. Multisection Liquid Tank. The multisection liquid tank consists of an acrylic tank for containment of liquids with three abrupt changes in its cross-sectional area, as it can be seen in Figure 4. The liquid tank was originally designed for educational purposes in order to be used in studies of identification and control of dynamical systems [61]. It was also used in [29] to evaluate the performance of



(a) Schematic diagram of the multisection liquid tank



(b) Multisection liquid tank

FIGURE 4: Three-section liquid tank with distinct cross-sectional areas.

the alternative FWNN structure employed in this work. In addition to the acrylic tank structure, the system is composed by a water reservoir, a water pump, a pressure sensor, an electronic power driver, and an electronic interface with A/D (analog-to-digital) and D/A (digital-to-analog) converters.

The nonlinearity presented in the liquid flow output, which is due to the different pressure levels at the tank base in accordance with the height of the liquid column, can be clearly noticed in the aforementioned dynamical system. Besides, the distinct cross-sectional areas make such nonlinearity even more evident. It is worth mentioning that the abrupt transitions between the tank sections are also responsible for discontinuities. The whole system can be seen as a set of three coupled nonlinear systems, since each tank section has its own dynamic behavior.

6.2. System Identification. Initially, in order to collect the experimental data set used during the learning and testing phase of the identified FWNN models in this work, the water pump is excited with an APRBS (Amplitude Modulated Pseudorandom Binary Sequence) and the water level inside the tank is registered at a sample rate of $T_s = 2$ Hz. For the generation of the persistent excitation signal, the following parameters are considered: the minimum hold time $T_h = 10$ s, minimal amplitude $A_{\min} = 0$ V, and maximum amplitude $A_{\max} = 15$ V. Since only positive values of voltage are considered in this case study, the pump only operates in order to shift the liquid from the reservoir to the multisection tank.

After the system excitation, the collected data is corrupted with additive white Gaussian noise and two different percentages of outliers (1% and 3%). The resulting data are divided into two sets comprising approximately 80% and 20% of the total amount. The first set is used to train the FWNN model and the second one is used during the testing phase.

The whole data set is normalized to fit within the range $[0, 1]$ in order to avoid numerical problems during the FWNN learning procedure. Since the multisection tank is a first-order nonlinear system and also considering Figure 3, the inputs of the FWNN models are defined with $\lambda = \alpha = 1$ and $\tau = 0$. Thus, $u(k)$, $u(k-1)$, and $y(k)$ are defined as inputs to the FWNN models to predict $y(k+1)$.

The BP algorithm presented in Section 4 is used to adjust the parameters d_{ij} , t_{ij} , a_{qr} , and b_{qr} of the FWNN. After a trial-and-error procedure the learning rate $\eta = 0.0001$ was found as a good choice to identify the multisection tank. It is worth mentioning that the results presented in this work were obtained after 350 learning epochs.

Figure 5 presents the model validation when 1% of the original experimental data set is corrupted with outliers and additive white Gaussian noise is inserted. In this figure, the tank water level in cm is in function of the sample time step, where each time step is equivalent to 0.5 seconds, defined by the sample rate $T_s = 2$ Hz. The terms FWNN-MCC and FWNN-MSE are used to identify the FWNN models obtained using MCC and MSE as cost criterion of BP algorithm, respectively. It is evident that FWNN-MCC has the best performance due to the use of the higher-order statistical information. On the other hand, the FWNN-MSE model based on second-order moments presents some problems to efficiently identify the input-output dynamic relationship of the multisection tank at some points of the validation curve. The presence of outliers in the experimental data has a significant negative impact on the FWNN model when the MSE criterion is used in the learning procedure, once the error due to the outliers is increased by a square rate. The same behavior is not observed in FWNN-MCC when 1% of the experimental data are corrupted by outliers because the outliers power is weighted by the Gaussian kernel.

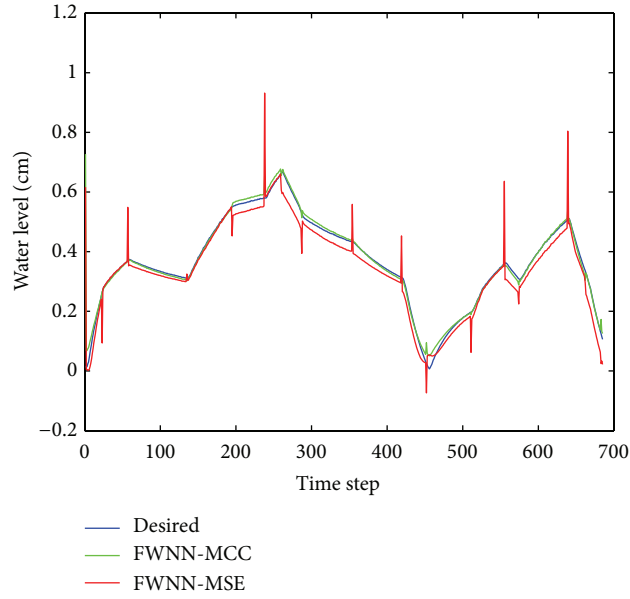


FIGURE 5: Validation of the FWNN-MCC with 1% outlier rate in experimental data.

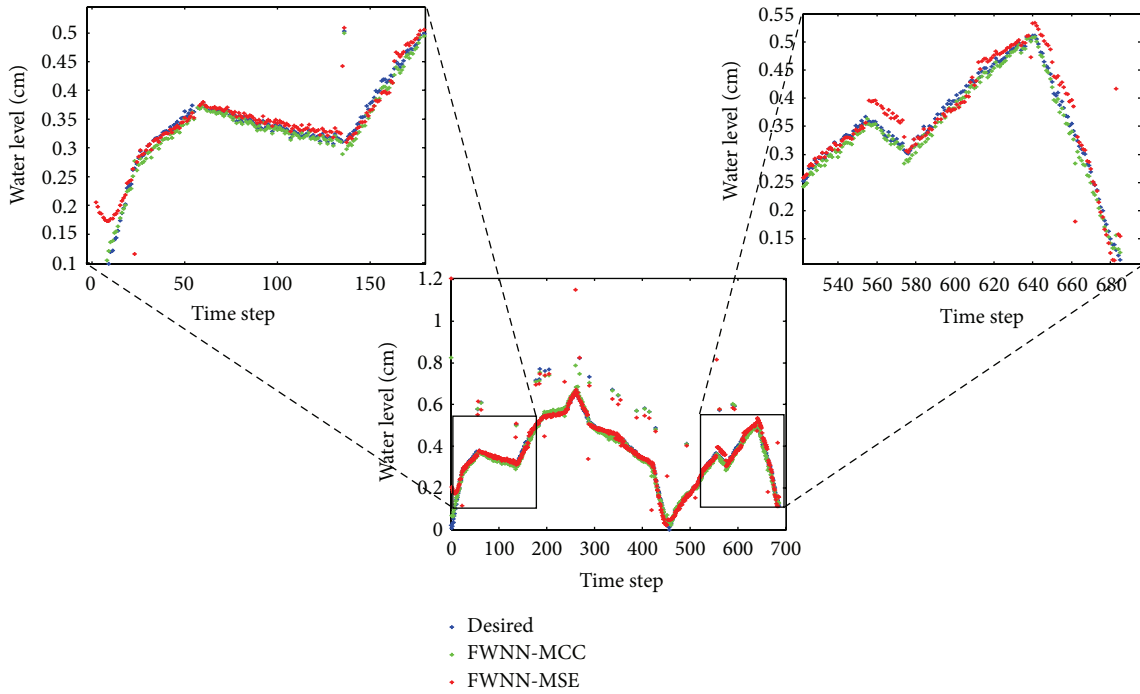


FIGURE 6: Validation of the FWNN-MCC with 3% outlier rate in experimental data.

Figure 6 shows the model validation when 3% of the original experimental data is corrupted with outliers and additive white Gaussian noise is inserted. In Figure 6, only the validation points are plotted to allow the better visualization of outliers and its respective effects in the FWNN-MCC and FWNN-MSE models. Two regions are highlighted in Figure 6, thus demonstrating the improvement of the FWNN-MCC model. Both models present problems at some points, although the performance of FWNN-MCC one is

improved in the identification of the multisection tank dynamics, as it also seems to be less sensitive to outliers and noise than FWNN-MSE model. It is noteworthy that MCC has intrinsic robustness due to the local estimation produced by the kernel size.

It is also important to mention that the correntropy criterion has a free parameter, that is, the kernel size, which is at the core of the learning process [38]. An adaptive kernel is applied in this work to improve the performance of

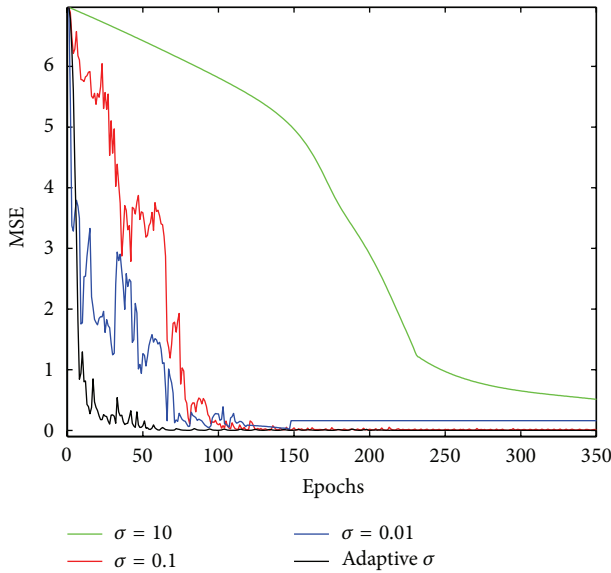


FIGURE 7: Evolution of MSE for different kernel sizes.

the FWNN learning procedure performance. Figure 7 shows MSE obtained over the 350 epochs, for three different fixed kernel sizes, that is, 0.01, 0.1, and 10, and also using the adaptive kernel. The adaptive kernel size method mathematically described by (6) has the highest convergence rate and the best performance in the attenuation of outliers and noise.

Figure 8 presents the behavior of the adaptive kernel size during the learning stage of the FWNN-MCC model when the experimental data is composed by 1% and 3% of outliers. During the initial epochs of the BP algorithm, the kernel size is quite oscillatory. However, the behavior of the kernel size becomes more stable as it comes to the hundredth epoch.

7. Conclusions

This work has analyzed the performance of a FWNN when applied to identify a real nonlinear dynamical system in the presence of unknown noise and outliers. Such erroneous measurements in experimental data reduce the reliability of the identified model, once it tries to fit some behaviors that are not part of the dynamical system. The most common learning techniques applied to adjust the FWNN parameters in identification applications are methods based on gradient that use the MSE as their cost function. This paper has then proposed the replacement of this traditional evaluation measure by a similarity measure based on information theory denominated correntropy. Therefore, the MCC was used in this paper as the cost function of the error backpropagation algorithm in order to reduce the negative effects of the unknown noise and outliers. The results have demonstrated that the FWNN-MCC models based on the MCC cost function represent the input-output dynamics of the multisection liquid tank more properly, being also less sensitive to outliers and noise than the FWNN-MSE models. This work also has investigated the influence of the kernel size on the performance of the MCC in the BP algorithm, since it is a free

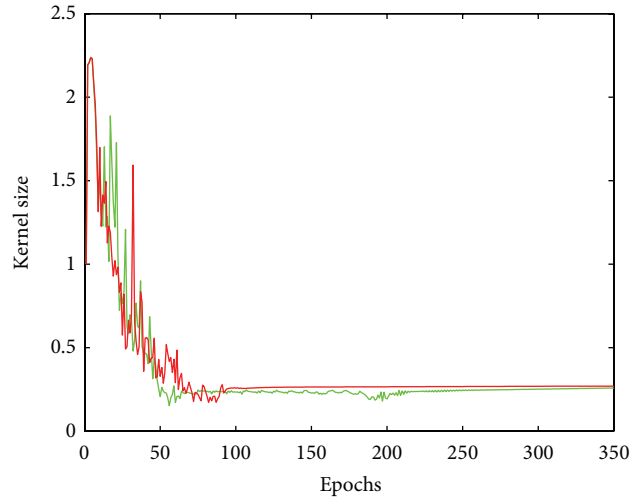


FIGURE 8: Adaptive kernel size behavior.

parameter of correntropy. The addition of this new parameter in the learning procedure of the FWNN can be considered a disadvantage of the proposed architecture, mainly because the MCC is very dependent on its proper adjustment. Within this context, the adopted adaptive kernel has shown to be more efficient if compared to the case when this parameter remains fixed during the whole FWNN learning process. The adaptive kernel size method has improved the convergence rate of the backpropagation algorithm and contributed to attenuating the effects of the outliers and noise. Due to the use of the BP algorithm, the proposed architecture is susceptible to local minima falls, limiting the correntropy action to remove the outliers.

The further research work will focus on the following items: (1) analyzing the application of the MCC associated with different algorithms in order to train the FWNN architecture to avoid the outliers harmful effects. The meta-heuristic algorithms such as Genetic Algorithm, Particle Swarm Optimization, and Bat Algorithm are good options since they are less sensitive to local minima than the BP algorithm; (2) including and comparing different adaptive kernel methods to improve the functionality of the MCC; (3) applying the proposed architecture to identify reliable dynamical models to be used in advanced control strategies, such as the predictive controllers; (4) evaluating the feasibility to apply the FWNN-MCC as an inferential system to estimate chemical compositions, calibrate sensors [62], and fault diagnosis, among others.

Acronyms

A/D:	Analog-to-digital
ANFIS:	Adaptive Neurofuzzy Inference System
ANN:	Artificial Neural Network
APRBS:	Amplitude Modulated Pseudorandom Binary Sequence

ARMAX:	AutoRegressive, Moving Average, eXogenous input model
ARX:	AutoRegressive, eXogenous input model
BP:	Error backpropagation algorithm
D/A:	Digital-to-analog
FIR:	Finite Impulse Response model
FWNN:	Fuzzy Wavelet Neural Network
FWNN-MCC:	FWNN obtained using MCC
FWNN-MSE:	FWNN obtained using MSE
MCC:	Maximum Correntropy Criterion
MF:	Membership function
MLP:	Multilayer Perceptron network
MSE:	Mean Squared Error
NARMAX:	Nonlinear AutoRegressive, Moving Average, eXogenous input model
NARX:	Nonlinear AutoRegressive, eXogenous input model
NFIR:	Nonlinear Finite Impulse Response model
NOE:	Nonlinear Output Error model
NSSIF:	Nonlinear State Space Innovations Form model
OE:	Output Error model
PCA:	Principal component analysis
RBF:	Radial Basis Function network
SSIF:	State Space Innovations Form model
WNN:	Wavelet Neural Network.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] X. Lu, K. L. V. Iyer, K. Mukherjee, and N. C. Kar, "A dual purpose triangular neural network based module for monitoring and protection in Bi-directional off-board level-3 charging of EV/PHEV," *IEEE Transactions on Smart Grid*, vol. 3, no. 4, pp. 1670–1678, 2012.
- [2] Z. Tian and M. J. Zuo, "Health condition prediction of gears using a recurrent neural network approach," *IEEE Transactions on Reliability*, vol. 59, no. 4, pp. 700–705, 2010.
- [3] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 2, pp. 310–321, 2013.
- [4] Z. Shao, Y. Zhan, and Y. Guo, "Fuzzy neural network-based model reference adaptive inverse control for induction machines," in *Proceedings of the International Conference on Applied Superconductivity and Electromagnetic Devices (ASEMD '09)*, pp. 56–59, Chengdu, China, 2009.
- [5] S. Yilmaz and Y. Oysal, "Fuzzy wavelet neural network models for prediction and identification of dynamical systems," *IEEE Transactions on Neural Networks*, vol. 21, no. 10, pp. 1599–1609, 2010.
- [6] R. H. Abiyev and O. Kaynak, "Fuzzy wavelet neural networks for identification and control of dynamic plants—a novel structure and a comparative study," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 8, pp. 3133–3140, 2008.
- [7] R. J. Bessa, V. Miranda, and J. Gama, "Entropy and correntropy against minimum square error in offline and online three-day ahead wind power forecasting," *IEEE Transactions on Power Systems*, vol. 24, no. 4, pp. 1657–1666, 2009.
- [8] H. Qu, W. Ma, J. Zhao, and T. Wang, "Prediction method for network traffic based on maximum correntropy criterion," *China Communications*, vol. 10, no. 1, pp. 134–145, 2013.
- [9] O. Nelles, *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, Springer, Berlin, Germany, 2001.
- [10] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, NJ, USA, 1999.
- [11] Y. Zhang and L. Wu, "Crop classification by forward neural network with adaptive chaotic particle swarm optimization," *Sensors*, vol. 11, no. 5, pp. 4721–4743, 2011.
- [12] Y. Zhang, S. Wang, G. Ji, and P. Phillips, "Fruit classification using computer vision and feedforward neural network," *Journal of Food Engineering*, vol. 143, pp. 167–177, 2014.
- [13] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.
- [14] A. Banakar and M. F. Azeem, "Identification and prediction of nonlinear dynamical plants using TSK and wavelet neuro-fuzzy models," in *Proceedings of the 3rd IEEE Conference on Intelligent Systems*, pp. 617–620, London, UK, September 2006.
- [15] T. Kara and I. Eker, "Experimental nonlinear identification of a two mass system," in *Proceedings of the IEEE Conference on Control Applications (CCA '03)*, vol. 1, pp. 66–71, Istanbul, Turkey, 2003.
- [16] M. O. Efe, "A comparison of ANFIS, MLP and SVM in identification of chemical processes," in *Proceedings of the IEEE Control Applications & Intelligent Control*, pp. 689–694, St. Petersburg, Russia, 2009.
- [17] X. Xue, J. Lu, and W. Xiang, "Nonlinear system identification with modified differential evolution and RBF networks," in *Proceedings of the 5th IEEE International Conference on Advanced Computational Intelligence (ICACI '12)*, pp. 332–335, Nanjing, China, October 2012.
- [18] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 889–898, 1992.
- [19] S. A. Billings and H.-L. Wei, "A new class of wavelet networks for nonlinear system identification," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 862–874, 2005.
- [20] F.-J. Lin, P.-H. Shen, and Y.-S. Kung, "Adaptive wavelet neural network control for linear synchronous motor servo drive," *IEEE Transactions on Magnetics*, vol. 41, no. 12, pp. 4401–4412, 2005.
- [21] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.
- [22] L. Zhao, J.-P. Zhang, J. Yang, and Y. Chu, "Software reliability growth model based on fuzzy wavelet neural network," in *Proceedings of the 2nd International Conference on Future Computer and Communication (ICFCC '10)*, pp. V1664–V1668, IEEE, Wuhan, China, May 2010.
- [23] J.-R. Song and H.-B. Shi, "Dynamic system modeling based on wavelet recurrent fuzzy neural network," in *Proceedings of the 7th International Conference on Natural Computation (ICNC '11)*, vol. 2, pp. 766–770, Shanghai, China, July 2011.
- [24] C. H. Lu, "Wavelet fuzzy neural networks for identification and predictive control of dynamic systems," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 7, pp. 3046–3058, 2011.

- [25] D. W. C. Ho, P.-A. Zhang, and J. Xu, "Fuzzy wavelet networks for function learning," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 1, pp. 200–211, 2001.
- [26] A. Ebadat, N. Noroozi, A. A. Safavi, and S. H. Mousavi, "New fuzzy wavelet network for modeling and control: the modeling approach," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 8, pp. 3385–3396, 2011.
- [27] S. H. Mousavi, N. Noroozi, A. A. Safavi, and A. Ebadat, "Modeling and control of nonlinear systems using novel fuzzy wavelet networks: the output adaptive control approach," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 9, pp. 3798–3814, 2011.
- [28] S. Ganjefar and M. Tofighi, "A fuzzy wavelet neural network stabilizer design using genetic algorithm for multi-machine systems," *Przegląd Elektrotechniczny*, vol. 89, no. 5, pp. 19–25, 2013.
- [29] L. L. S. Linhares, J. M. Araújo Jr., F. M. U. Araújo, and T. Yoneyama, "A nonlinear system identification approach based on fuzzy wavelet neural network," *Journal of Intelligent and Fuzzy Systems*, 2014.
- [30] Y. Liu and J. Chen, "Correntropy-based kernel learning for nonlinear system identification with unknown noise: an industrial case study," in *Proceedings of the 10th IFAC Symposium on Dynamics and Control of Process Systems (DYCOPS '13)*, vol. 10, pp. 361–366, Mumbai, India, December 2013.
- [31] J. C. Munoz and J. Chen, "Removal of the effects of outliers in batch process data through maximum correntropy estimator," *Chemometrics and Intelligent Laboratory Systems*, vol. 111, no. 1, pp. 53–58, 2012.
- [32] T. Söderström, "System identification for the errors-in-variables problem," *Transactions of the Institute of Measurement and Control*, vol. 34, no. 7, pp. 780–792, 2012.
- [33] S. Khatibisepehr and B. Huang, "Dealing with irregular data in soft sensors: Bayesian method and comparative study," *Industrial and Engineering Chemistry Research*, vol. 47, no. 22, pp. 8713–8723, 2008.
- [34] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK, 1995.
- [35] V. Miranda, C. Cerqueira, and C. Monteiro, "Training a FIS with epso under an entropy criterion for wind power prediction," in *Proceedings of the International Conference on Probabilistic Methods Applied to Power Systems (PMAPS '06)*, pp. 1–8, Stockholm, Sweden, 2006.
- [36] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, 2006.
- [37] S. Zhao, B. Chen, and J. C. Principe, "Kernel adaptive filtering with maximum correntropy criterion," in *Proceedings of the International Joint Conference on Neural Network (IJCNN '11)*, pp. 2012–2017, San Jose, Calif, USA, August 2011.
- [38] J. C. Principe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*, Springer, 2010.
- [39] A. I. Fontes, A. D. M. Martins, L. F. Silveira, and J. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Systems with Applications*, vol. 42, no. 1, pp. 1–8, 2014.
- [40] A. I. R. Fontes, P. T. V. Souza, A. D. D. Neto, A. M. Martins, and L. F. Q. Silveira, "Classification system of pathological voices using correntropy," *Mathematical Problems in Engineering*, vol. 2014, Article ID 924786, 7 pages, 2014.
- [41] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1485–1494, 2011.
- [42] Y. Liu and J. Chen, "Correntropy-based kernel learning for nonlinear system identification with unknown noise: an industrial case study," in *Proceedings of the 10th IFAC Symposium on Dynamics and Control of Process Systems*, pp. 361–366, December 2013.
- [43] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-Gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.
- [44] S. Zhao, B. Chen, and J. C. Principe, "An adaptive kernel width update for correntropy," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '12)*, pp. 1–5, Brisbane, Australia, 2012.
- [45] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.
- [46] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, vol. 3, CRC Press, New York, NY, USA, 1986.
- [47] A. W. Bowman, "An alternative method of cross-validation for the smoothing of density estimates," *Biometrika*, vol. 71, no. 2, pp. 353–360, 1984.
- [48] D. W. Scott and G. R. Terrell, "Biased and unbiased cross-validation in density estimation," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1131–1146, 1987.
- [49] N. Terzija and H. McCann, "Wavelet-based image reconstruction for hard-field tomography with severely limited data," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1885–1893, 2011.
- [50] A. Bouzida, O. Touhami, R. Ibtouen, A. Belouchrani, M. Fadel, and A. Rezzoug, "Fault diagnosis in industrial induction machines through discrete wavelet transform," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 9, pp. 4385–4395, 2011.
- [51] J. Gao, Z. Leng, Y. Qin, Z. Ma, and X. Liu, "Short-term traffic flow forecasting model based on wavelet neural network," in *Proceedings of the 25th Chinese Control and Decision Conference (CCDC '13)*, pp. 5081–5084, Guiyang, China, May 2013.
- [52] J. J. Cordova, W. Yu, and X. Li, "Haar wavelet neural networks for nonlinear system identification," in *Proceedings of the 6th IEEE Multi-Conference on Systems and Control (MSC '12)*, pp. 276–281, Dubrovnik, Croatia, October 2012.
- [53] R. K. Galvão, V. M. Bécerra, J. M. Calado, and P. M. Silva, "Linear-wavelet networks," *International Journal of Applied Mathematics and Computer Science*, vol. 14, no. 2, pp. 221–232, 2004.
- [54] M. Davanipoor, M. Zekri, and F. Sheikholeslam, "Fuzzy wavelet neural network with an accelerated hybrid learning algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 3, pp. 463–470, 2012.
- [55] V. V. J. Rajapandian and N. Gunaseeli, "Modified standard back propagation algorithm with optimum initialization for feed forward neural networks," *International Journal of Imaging Science and Engineering*, vol. 1, no. 3, pp. 86–89, 2007.
- [56] T. Kathirvalavakumar and P. Thangavel, "A modified backpropagation training algorithm for feedforward neural networks," *Neural Processing Letters*, vol. 23, no. 2, pp. 111–119, 2006.

- [57] S. Abid, R. Fnaicch, and M. Najim, "A fast feedforward training algorithm using a modified form of the standard backpropagation algorithm," *IEEE Transactions on Neural Networks*, vol. 12, no. 2, pp. 424–430, 2001.
- [58] M. Davanipoor, M. Zekri, and F. Sheikholeslam, "The preference of fuzzy wavelet neural network to ANFIS in identification of nonlinear dynamic plants with fast local variation," in *Proceedings of the 18th Iranian Conference on Electrical Engineering (ICEE '10)*, pp. 605–609, Isfahan, Iran, May 2010.
- [59] E. Soria, J. D. Martin, and P. J. G. Lisboa, "Classical training methods," in *Metaheuristic Procedures for Training Neural Networks*, E. Alba and R. Marti, Eds., pp. 25–32, Springer, New York, NY, USA, 2006.
- [60] A. Singh and J. C. Principe, "Information theoretic learning with adaptive kernels," *Signal Processing*, vol. 91, no. 2, pp. 203–213, 2011.
- [61] C. A. G. Fonseca, *Estrutura ANFIS Modificada para identificação e Controle de Plantas com Ampla Faixa de Operação e não Linearidade Acentuada [Ph.D. thesis]*, Universidade Federal do Rio Grande do Norte (UFRN), Natal, Brazil, 2012.
- [62] J. M. de Araújo Jr., J. M. P. de Menezes, A. A. M. de Albuquerque, O. D. M. Almeida, and F. M. U. de Araújo, "Assessment and certification of neonatal incubator sensors through an inferential neural network," *Sensors*, vol. 13, no. 11, pp. 15613–15632, 2013.

