Electric Power Systems Research 184 (2020) 106334

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

An enhanced set-membership evolving participatory learning with kernel recursive least squares applied to thermal modeling of power transformers

Kaike Sa Teles Rocha Alves[a,*], Michel Hell[b], Fernando Luiz Cyrino Oliveira[c], Eduardo Pestana de Aguiar[a]

[a] Department of Industrial and Mechanical Engineering, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil
[b] Department of Electrical Engineering, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil
[c] Department of Industrial Engineering, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil

ARTICLE INFO

Keywords:
Enhanced set-membership
Evolving fuzzy systems
Power transformers
Thermal modeling

ABSTRACT

A factor that directly impacts the lifespan of a power transformer is the hot-spot temperature, and its monitoring is vital to prevent faults, reduce costs, keep the safety, and provide a reliable service to consumers. In this paper, we propose two forecasting models to predict the hot-spot temperature of power transformers. The first is the implementation of Set-Membership in the evolving Participatory Learning with Kernel Recursive Least Squares. And the second is a combination of the evolving Participatory Learning with Kernel Recursive Least Squares and the improved version of the Set-Membership concept, named Enhanced Set-Membership. Both Set-Membership and the Enhanced Set-Membership approaches are implemented to update the rate of change of the arousal index, which is a parameter that controls the creation of rules. A data set collected from an experimental transformer is adopted to evaluate the model's performance. The obtained results are compared with the performance of the original evolving Participatory Learning with Kernel Recursive Least Squares and with the performance of other classical models suggested in the literature. The proposals have lower errors and a competitive number of final rules, suggesting that the models are efficient approaches to modeling complex data with high accuracy.

1. Introduction

The power transformer is a critical equipment in power distribution [1]. It is responsible for stepped-up the voltage before to be transmitted over long distances to reduce waste, and stepped-down the voltage to provide the energy to consumers safely [2,3]. Due to the composition of a power transformer, it is the most expensive apparatus in energy distribution [4]. In the case of a power transformer's failure, when the recovering process is possible, it is slow and inefficient [5]. Thereof, monitoring is vital to prevent faults, reduce costs, keep the safety, and provide a reliable service to consumers [6,7]. The annual spent on power transformers' monitoring hardware will increase more than $ 642 million in eight years until 2020, according to [8,9], indicating the importance of the power transformers in power distribution.

Internal failures are about 10% of the total faults, and, among them, winding and bushing defects represent approximately 44% [10]. The bushing is a fragile component constituted of four parts: insulation, conductor, connection clamp, and accessories[11,12]. In the present work, we considered power transformers composed of Resin-bonded paper bushings (RBP) [13].

The principal factor in bushing failures is the hot-spot temperature, representing 32% of the total causes [11]. The hot-spot temperature is the highest temperature near to the top of the power transformers high-voltage (HV)/low-voltage (LV) windings [7,14] and represents the main limiting factor in the load capacity of the transformer [14], since increases in this temperature reduces the lifespan of the insulation and may determine the end life of the power transformer [15].

As the estimation of the hot-spot is a complex task, many models have been proposed in the literature with the purpose of estimating the hot-spot temperature of power transformers. Among them, the most commonly used in practice is the model based on the IEEE Standard C57.91-2011 [16] which is based on transient heating equations and specific thermal characteristics and parameters of power transformers. This deterministic model is imprecise due to assumed simplifications, and consequently, the power transformer must operate below the maximum capacity to prevent damages [7]. This conservative attitude

* Corresponding author.
E-mail addresses: kaike.alves@engenharia.ufjf.br (K.S.T.R. Alves), michel.hell@ufjf.edu.br (M. Hell), cyrino@puc-rio.br (F.L. Cyrino Oliveira), eduardo.aguiar@engenharia.ufjf.br (E.P. de Aguiar).

https://doi.org/10.1016/j.epsr.2020.106334
Received 7 November 2019; Received in revised form 20 February 2020; Accepted 12 March 2020
Available online 28 March 2020
0378-7796/

increases the operational cost to the company [17]. Therefore, more advanced techniques are necessary to optimize the use of the power transformer's capacity and its lifetime without put at risk its functionality and security [18].

In literature, we found applications of several works in power transformers monitoring. Paper [19] suggests a formulation to calculate the expected cost to repair a power transformer. In paper [19–21] was used artificial neural networks (ANN) for diagnostics and in [22] the ANN is proposed to analyze faults in dissolved gas-in-oil. Paper [23] suggests fuzzy sets and [18] presents neuro-fuzzy hybrids. Reference [17] introduces the use of participatory learning (PL) to train a hybrid neuro-fuzzy network, and in paper [24] is used as an evolving multivariable Gaussian (eMG). Cortez proposed a fault prognosis provided by an intelligent system based on cognitive systems [25] and a vector machine (SVM) was proposed in Bacha et al. [26], Ganyun et al. [27] to the same purpose. In paper [28], the use of a fuzzy expert system indicates the best moment to repair a power transformer based on its current state.

The main contributions of this work are summarized as follows:

- We introduce a novel forecasting model based on the Set-Membership (SM) filtering [29–31] to adjust the parameter that controls the rate of change of the arousal index in the evolving Participatory Learning with Kernel Recursive Least Squares (ePL-KRLS). This model was named Set-Membership evolving Participatory Learning with Kernel Recursive Least Squares (SM-ePL-KRLS).
- We implemented a new filtering strategy to update the rate of change of the arousal index in the ePL-KRLS. This proposal is so-called Enhanced Set-Membership evolving Participatory Learning with Kernel Recursive Least Squares (ESM-ePL-KRLS), which is an improved version of the SM-ePL-KRLS.
- We evaluate the performance of the proposed models in terms of errors and the number of final rules, using data set from thermal modeling of power transformers. Additionally, we compare the performance with other approaches suggested in the literature (evolving Multivariable Gaussian [32], Multilayer Perceptron (MLP) [33], Adaptative Neurofuzzy Inference System (ANFIS) [34], and IEEE Deterministic Model [16])

Our major conclusions are as follows:

- Both proposed models achieved the lowest errors, suggesting that these forecasting models can predict complex data with high accuracy.
- The proposed models obtained the lowest number of final rules in all simulations, indicating these models have a low computational cost.
- The monitoring of the hot-spot temperature by the introduced models is efficient to control the load current and improves the lifespan of the power transformers.

This work aims to propose two forecasting models able to predict the hot-spot temperature with high accuracy, and consequently, improving the load capacity and increasing the lifetime of a power transformer. Proposed models are classified as an evolving fuzzy system (eFS). Their main benefit is the adaptability of the model according to the data, which occurs in a continuous learning process through the creation and exclusion of rules [35].

The remainder of this paper is organized as follows. Section 2 presents the approach of the deterministic model. Section 3 details the proposed models. Section 4 presents the achieved results and discusses them. And finally, Section 5 presents the conclusions.

2. Problem formulation

Nowadays, the most used model in practice for the prediction of the hot-spot temperature of power transformers is the deterministic model proposed in the IEEE Standard C57.91-2011 [16]. This model consists of a series of differential equations whose calculation requires knowledge of load curves and operating conditions for which values usually conservative are fixed [36].

As presented in Hell et al. [18], the main steps of the hot-spot calculation made by the IEEE deterministic model can be summarized as follows:

The deterministic modeling starts by calculating the ultimate top oil rise ($\Delta\Theta_{TO,U}$), using the following Equation:

$$\Delta\Theta_{TO,U} = \Delta\Theta_{TO,R}\left[\frac{K^2R+1}{R+1}\right]^n \tag{1}$$

where $\Delta\Theta_{TO,R}$ is the rated top oil temperature rise over the environment, K is the load current, R is the ratio of load loss at rated-load to no-load loss at applicable tap position and n is the empirically derived value depending on the cooling.

Using Eq. (1) and the environment temperature (Θ_A), the increment in the top oil temperature (Θ_{TO}) is found by the differential equation expressed below:

$$\tau_{TO}\frac{d\Theta_{TO}}{dt} = [\Delta\Theta_{TO,U} + \Theta_A] - \Theta_{TO} \tag{2}$$

where τ_{TO} is the top oil rise time constant.

The next step consists of calculating the last hot-spot rise over top oil ($\Delta\Theta_{H,U}$), as follows:

$$\Delta\Theta_{H,U} = \Delta\Theta_{H,R}K^{2m} \tag{3}$$

where $\Delta\Theta_{H,R}$ is the rated hot-spot rise over top oil.

We calculate the increment in hot-spot rise above top oil temperature ($\Delta\Theta_H$) using the value of $\Delta\Theta_{H,U}$ obtained from Eq. (3), as the following differential equation:

$$\tau_H\frac{d\Delta\Theta_H}{dt} = \Delta\Theta_{H,U} - \Delta\Theta_H \tag{4}$$

where τ_H is the hot-spot rise time constant and $\Delta\Theta_H$ is the hot-spot rise above top oil temperature.

Finally, the hot-spot temperature is calculated as a function of Θ_{TO} and $\Delta\Theta_H$ according to Eq. (5), where these parameters are obtained from Eqs. (2) and (4) respectively.

$$\Theta_H = \Theta_{TO} + \Delta\Theta_H \tag{5}$$

As mentioned before, the model presented in this section presumes a series of simplification assumptions, and its use makes that the power transformer operates between 70% to 80% of their nominal capacity [7] which implies a loss of more than one power transformer at every five [17]. In this sense, the main goal of the proposed models is to introduce an algorithm able to predict the hot-spot with high accuracy and low computational cost in order to increase system operation margin mainly in the presence of overload conditions [37].

3. Proposed models

In this section, we introduce the proposed models. First, we discuss the ePL-KRLS algorithm. Then, we explain the Set-Membership (SM) concept. And finally, we present the Enhanced Set-Membership (ESM) filtering.

3.1. ePL-KRLS algorithm

The ePL-KRLS is a fuzzy evolving model based on Takagi-Sugeno (TS) rules [38,39]. This model clusters the input space according to the degree of similarity in the knowledge process [32,40]. This technique of clustering uses the concept of participatory learning (PL). The PL procedure is based on human learning [41,42]. Every cluster has a local output, obtained as a function of the consequent parameters associated with each cluster, that contributes as a weighted average to calculate

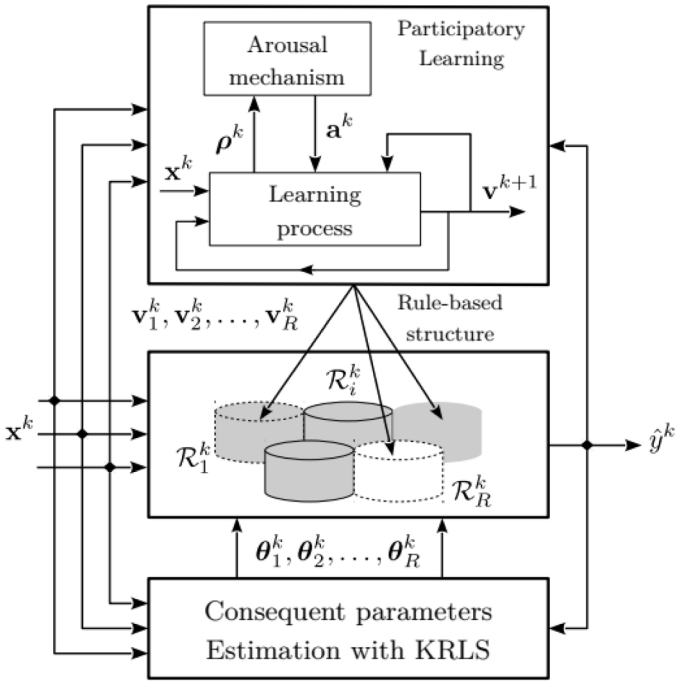

Fig. 1. Mechanism of learning of ePL-KRLS model [47].

the global output [43–45]. The ePL-KRLS estimates the consequent parameters using the kernel recursive least squares [46,47]. The learning structure of ePL-KRLS is represented in Fig. 1.

The compatibility index (ρ_i^k) and the arousal index (a_i^k) are calculated from Eqs. (6) and (7), respectively.

$$\rho_i^k = 1 - \frac{\|X^k - V_i^k\|}{m} \tag{6}$$

where $X^k = [x_1, ..., x_m]^T$ is input data at step k, m is dimension of X, V_i^k is the center at step k of ith rule.

$$a_i^k = a_i^{k-1} + \beta(1 - \rho_i^k - a_i^{k-1}) \tag{7}$$

where $\beta \in [0, 1]$ controls the growth rate of a_i^k.

The compatibility index is a measure of similarity between the new input vector and created rules. A result of the compatibility index equal to one represents the input vector has the maximum similarity with a created rule, and a result zero represents the minimum similarity. The interval of these variables is the follow: ρ_i^k, $a_i^k \in [0, 1]$, $\forall\ i = 1, ..., R^k$, where R is the number of rules at kth step.

The value of the arousal index indicates the need to create a new rule. The compatibility index is used to calculate the arousal index, which reduces the effect of outliers in the model. If the lowest arousal index is higher than a threshold, i.e., $\rho_i^k > \tau$, where $i = argmax_i \rho_i^k$ and $\tau = \beta$, then a new rule is created. Otherwise, the input vector is included in the most compatible rule, expressed in the following Equation:

$$\hat{V}_i^{k+1} = \hat{V}_i + \alpha(\rho_i^k)^{1-a_i^k}(X^k - \hat{V}_i^k) \tag{8}$$

where $10^{-5} < \alpha < 10^{-1}$ is the learning rate.

The local output is calculated as follows:

$$\hat{y}_i = f_i(X, \theta_i) = \sum_{j=1}^{n_i} \theta_{ij} \kappa(d_{ij}, X) \tag{9}$$

where θ_i is the consequent parameters of ith rule, d_{ij} is the jth element of the local dictionary in ith rule and $\kappa\langle ., . \rangle$ is the Gaussian-Kernel function shown in Eq. (10) [48].

$$\kappa\langle X^i, X^j \rangle = exp\left(-\frac{||X^i - X^j||^2}{2\nu^2}\right) \tag{10}$$

where ν is positive and represents the size of the kernel. The ν is found minimizing the error function using Eq. (11) and it is based on a recursive Levenberg-Marquardt model [49].

$$\nu_i^k = \nu_i^{k-1} + P_i^k \nabla_i^k \tilde{e}_i^k \tag{11}$$

where P_i^k is calculated from Eq. (12), ∇_i^k is the gradient of error expressed in (13) and $\tilde{e}^k = y^k - \hat{y}^k$ is the error, which y^k is the actual value and $\hat{y}^k$ is the predicted value.

$$P_i^k = \left[P_i^{k-1} - \frac{P_i^{k-1}\nabla_i^k[\nabla_i^k]^T P_i^{k-1}}{1 + [\nabla_i^k]^T P_i^{k-1}\nabla_i^k}\right] \tag{12}$$

$$\nabla_i^k = -\left[\frac{\partial \tilde{e}_i^k}{\partial \nu_{i1}^{k-1}}, ..., \frac{\partial \tilde{e}_i^k}{\partial \nu_{in_1}^{k-1}}\right]^T = \Gamma_i^N(X^k)\begin{bmatrix} \theta_{i1}^{k-1} - \frac{\|X^k - d_{i1}^k\|^2}{(\nu_{i1}^{k-1})^3} k(X^k, d_{i1}^k) \\ \vdots \\ \theta_{in_1}^{k-1} - \frac{\left\|X^k - d_{in_1}^k\right\|^2}{(\nu_{in_1}^{k-1})^3} k\left(X^k, d_{in_1}^k\right) \end{bmatrix} \tag{13}$$

where $P_i^1 = \Omega I$, $\Omega \in]0, 1000[$ and $\nu_{i1}^1 = 0.5$.

The model output is computed from local outputs as follows:

$$\hat{y} = \sum_{i=1}^{R} \hat{y}_i \Gamma_i^N(X) \tag{14}$$

where $\Gamma_i(X)$ is the normalized firing degree expressed in Eq. (15).

$$\Gamma_i(X) = \frac{\mathcal{A}_i(X)}{\sum_{j=1}^{R} \mathcal{A}_j(X)} \tag{15}$$

where $\mathcal{A}_i$ is the fuzzy set of ith rule.

The mechanism of clusters the input space and calculate the global output according to the importance degree improve the precision of the model and better predicts nonlinear data. The next step is to calculate the consequent parameter from Eq. (16).

When the model creates a new rule, the initialization of the variables are as following: $a_i^k = 0$, $D_i^k = X^k$ and the Eq. (16) are used to calculate the consequent parameter.

$$\theta_i^k = [\lambda + \kappa(X^k, X^k)]^{-1} y^{k-1} \tag{16}$$

where $\lambda \in [10^{-5}, 10^{-2}]$ is a parameter of regularization.

Otherwise, the consequent parameter is updated using Eq. (17). When Eq. (18) is satisfied, if the addition of the input vector to the local dictionary reduces the error, the model makes this inclusion. This technique of sparcification aims to reduce the computational cost and is called novelty criterion [50,51].

$$\theta_i^k = \begin{bmatrix} \theta_i^{k-1} - z_i^k [r^{k_i}]^{-1} \tilde{e}_i^k \\ [r_i^k]^{-1} \tilde{e}_i^k \end{bmatrix} \tag{17}$$

$$\min_{\forall d_{ij} \in \mathcal{D}_i^k} \|X^k - d_{ij}^k\| \geq \delta \tag{18}$$

To eliminate redundant rules, the value of compatibility measure, obtained from Eq. (19), is calculated at the end of each iteration. If $\rho_{ij}^k > \gamma$, $\forall\ i \neq j$, where $\gamma = 1 - \beta$, then the rules i and j are merged according Eq. (20).

$$\rho_{ij}^k = 1 - \sum_{l=1}^{m} \frac{|v_{il}^k - v_{jl}^k|}{m} \tag{19}$$

$$V_i^k = \frac{V_i^k + V_j^k}{2} \tag{20}$$

3.2. Set-membership (SM)

The Set-Membership (SM) is an adaptative algorithm that adjusts a chosen parameter as a function of the model errors. The updating is

performed comparing the error with a default value. If the error value is higher than a threshold (γ_{bar}), then the rate of change of the arousal index (β) increases to improve the model learning. Otherwise, β is zeroed to reduce the computational cost. The SM is a filtering proposed by Clarke and de Lamare [30] to limit the increase of the error, reduce the computational complexity, and improve the capacity of convergence [29]. The mechanism to update β performs, as shown below:

$$\beta = \begin{cases} 1 - \frac{\gamma_{bar}}{|\tilde{e}^k|}, & if \quad |\tilde{e}^k| > \gamma_{bar} \\ 0, \quad \text{otherwise} \end{cases} \tag{21}$$

where $\tilde{e}_i^k$ is the error at ith iteration.

To β not becomes less than zero, an inferior limit (IL) prevents it.

There are alternatives proposed calculations of SM. Eq. (22) demonstrate an alternative calculation of the SM using average errors [29,30].

$$\beta = \begin{cases} 1 - \frac{\gamma_{bar}}{\sum_{q=1}^{k} \frac{1}{k} |\tilde{e}^q|}, & if \quad |\tilde{e}^k| > \gamma_{bar} \\ 0, \quad \text{otherwise} \end{cases} \tag{22}$$

where $\sum_{q=1}^{k} \frac{1}{k} |\tilde{e}^q|$ is the average error at step k.

3.3. Enhanced set-membership (ESM)

The ESM is an improvement of the SM. Its mechanism of work is to adjust the β as a function of the error. However, instead of β becomes zero as the SM, the ESM reduces the chosen parameter when the error is lower than γ_{bar} as presented below:

$$\beta^k = \begin{cases} \beta^{k-1} + \frac{|\tilde{e}^k|}{10^{gr} \times \gamma_{bar}}, & if \quad |\tilde{e}^k| > \gamma_{bar} \\ \beta^{k-1} - \frac{|\tilde{e}^k|}{10^{dr} \times \gamma_{bar}}, & \text{otherwise} \end{cases} \tag{23}$$

where $gr, dr \in \mathbb{Z}$ are the rate of parameter increase and decrease, respectively.

An inferior limit (IL) and a superior limit (SL) are predefined to limit β to improve the results of the predictions, and to β does not achieve values inconsistently. In other words, $\beta \in [IL, SL]$, where $IL \geq 0$, $SL \leq 1$ and $IL \leq SL$.

Finally, we update the dependent parameters of β as follow: $\tau = \beta$ and $\gamma = 1 - \beta$. The ESM-ePL-KRLS algorithm is shown in Fig. 2.

The algorithm of SM-ePL-KRLS can be observed by replacing Algorithm 2 (see Fig. 3) for the algorithm shown in Fig. 2.

4. Experimental results

To evaluate the effectiveness of the method proposed in this work, the models presented in previous sections are applied in the estimation of the hot-spot temperature of a real transformer whose characteristics are shown in Table 1.

The data sets collected from this transformer are the same presented in Galdi et al. [36] and were obtained through a measurement system composed by three fiber-optical-based temperature sensors and a hall-effect current sensor. The first two temperature sensors (S1 and S2) were inserted in the spacer between the disks at the top of the high-voltage and low-voltage windings, as shown in Fig. 4. The aggregation of the values obtained from these two sensors provides a measure of the actual value of the transformer's hot-spot temperature (Θ_H). The third temperature sensor (S3) was inserted at the top of the tank and provides the actual value of top oil temperature (Θ_{TO}). The hall-effect current sensor (S4) provides the actual value of the load current (K). Fig. 4 shows the location of these sensors in the experimental transformer used in this work.

In our experiments, two data sets composed of the records of the temperatures and load current acquired from each sensor in an interval of 24 h with a 5-min sample rate were used to evaluate the proposed models. To cover all the transformers operating conditions, two different load conditions were considered: *i*) Data set 1: without overload and *ii*) Data set 2: with overload. Figs. 5 and 6 shows the behavior of the hot-spot and top-oil temperatures for a given load current for these two data sets.

Therefore, the purpose of the proposed models is to estimate the hot-spot temperature from the load current and the top-oil temperature. Different studies [17,36] and experimental trial and error tests indicate that the relevant model inputs for this case are the load current (K), the top oil temperature (Θ_{TO}) and one step delayed load current (q^{-1} K, where q^{-1} is the delay operator). This choice has shown to reduce the model sensitivity concerning fluctuations in the thermal parameters, which can vary considerably from one transformer to another [52].

The root mean squared error (RMSE), non-dimensional index error (NDEI), and mean absolute error (MAE) are error measures used to evaluate the precision of the models. The formulas of RMSE, NDEI, and MAE are shown in Eqs. (24)–(26) respectively.

$$RMSE = \sqrt{\frac{1}{T} \sum_{k=1}^{T} (y^k - \hat{y}^k)^2} \tag{24}$$

$$NDEI = \frac{RMSE}{std([y^1, ..., y^T])} \tag{25}$$

$$MAE = \frac{1}{T} \sum_{k=1}^{T} |y^k - \hat{y}^k| \tag{26}$$

where $\hat{y}^k$ is the kth forecasted value, y^k the kth actual value and T is the sample size.

Additionally, the number of final rules was used to estimate the computational cost of the evaluated models [47].

Thus, initially, we used the two models proposed in this work, the SM-ePL-KRLS and ESM-ePL-KRLS models, as well as the original ePL-KRLS model, as presented in Section 3, in the estimation's problem of the hot-spot temperature for the two data sets previously presented. Also, to test the evolving characteristics of the proposed models, we have also implemented another evolving model shown in the literature, the evolving Multivariable Gaussian (eMG) model proposed in Souza et al. [24].

The parameters of ePL-KRLS, SM-ePL-KRLS and ESM-ePL-KRLS are defined as follows: $\alpha = 0.01$, $\beta = 0.18$, $\gamma = 0.18$, $\tau = 0.82$, $\sigma = 0.05$, $\lambda = 10^{-4}$ and $v_0 = 0.5$. In addition, the SM-ePL-KRLS has the following parameters: $\gamma_{bar} = 0.011$ for the data set 1 and $\gamma_{bar} = 0.01106$ for the data set 2. And the ESM-ePL-KRLS has the following parameters: the $IL = 0$, $SL = 0.6$, $gr = 1$, $dr = 0$, and $\gamma_{bar} = 0.01120$ for the data set 1 and $IL = 0.01$, $SL = 0.35$, $gr = 1$, $dr = 0$, and $\gamma_{bar} = 0.0172$ for the data set 2. The adopted γ_{bar} was chosen as the best result among 1439 simulations starting at $\gamma_{bar} = 0.00001$ and finishing at $\gamma_{bar} = 0.07196$. Tables 2 and 3 shows the results of the simulations for all implemented evolving models.

As can be seen in Tables 2 and 3 the proposed models achieved the best results in terms of estimation accuracy than the other models. In particular, the ESM-ePL-KRLS model showed higher accuracy with a competitive computational cost if compared to all other models when applied to the data set 2, i.e., in the presence of an overload condition in the transformer's operation. It is worth mentioning that one of the main applications of the hot-spot temperature estimation is concerned in the computing the load capability rate in real-time to increase system operation margin in the presence of overload conditions. This makes the ESM-ePL-KRLS model emerge as a promising alternative in the solution of the proposed problem.

The better performance of the proposals is a consequence of the updating of β, τ and γ according to the error increase or decrease. The adjustment of β is a function of the magnitude of the error, limiting the maximum error, and implying more ability to treat nonlinear data. The rate of creation of new rules increases if the error increases more than γ_{bar}. Otherwise, the number of rules tends to decrease.

Algorithm 1: ePL-KRLS

Input: $X^k, k = 1, \ldots, n$

Initialization: $V_1^1 = X^1, \mathcal{D}_1^1 = X^1, \nu_{11}^1 = 0.5, R = 1, a_1^1 = 0$

for $k = 2, 3, \ldots, n$ **do**

  for $i = 1, 2, \ldots, R$ **do**

    $\rho_i^k = 1 - \frac{\|X^k - V_i^k\|}{m}$

    $a_i^k = a_i^{k-1} + \beta(1 - \rho_i^k - a_i^{k-1})$

    $\Gamma_i^N(X^k) = \frac{\Gamma_i(X^k)}{\sum_{l=1}^{R} \Gamma_l(X^k)}$

    $\hat{y}_i = f_i(X, \theta_i) = \sum_{j=1}^{|\mathcal{D}_i|} \theta_{ij} \kappa(d_{ij}, X)$

  $\hat{y} = \sum_{i=1}^{R} \hat{y}_i \Gamma_i^N(X)$

  // The SM or ESM is called here

  if $\min |a^k| > \tau$ **then**

    $R = R + 1$

    $V_R^{k+1} = X^k$

    $\mathcal{D}_R^k = X^k$

    $\nu_{R1}^k = 0.5$

    $\theta_i^k = [\lambda + \kappa(X^k, X^k)]^{-1} y^{k-1}$

  else

    $\hat{V}_i^{k+1} = \hat{V}_i + \alpha(\rho_i^k)^{1-a_i^k}(X^k - \hat{V}_i^k)$

    if $\min_{(\forall d_{ij} \in \mathcal{D}_i^k)} \|X^k - d_{ij}^k\| \geq \delta$ & $e^k < \tilde{e}^k$ **then**

      $\mathcal{D}_i^k = \mathcal{D}_i^k \cup X^k$

    $v_i^k = v_i^{k-1} + P_i^k \nabla_i^k \tilde{e}_i^k$

    $\theta_i^k = \begin{bmatrix} \theta_i^{k-1} - z_i^k [r_i^k]^{-1} \tilde{e}_i^k \\ [r_i^k]^{-1} \tilde{e}_i^k \end{bmatrix}$

  for $i, j = 1, 2, \ldots, c, i \neq j$ **do**

    if $\rho_{ij}^k > \gamma$ **then**

      $V_i^k = \frac{V_i^k + V_j^k}{2}$

      $RemoveV_j^k$

Algorithm 2: ESM

$\tilde{e}^k = y^k - \hat{y}^k$

if $\tilde{e}^k > \gamma_{bar}$ **then**

  $\beta^k = \beta^{k-1} + \frac{|\tilde{e}^k|}{10^{gr} \times \gamma_{bar}}$

else

  $\beta^k = \beta^{k-1} - \frac{|\tilde{e}^k|}{10^{dr} \times \gamma_{bar}}$

if $\beta^k < IL$ **then**

  $\beta^k = IL$

if $\beta^k > SL$ **then**

  $\beta^k = SL$

$\tau = \beta^k$

$\gamma = 1 - \beta^k$

Fig. 2. ESM-ePL-KRLS algorithm.

In order to statistically validate the performance of the proposed models we also perform the Morgan-Granger-Newbold test (MGN) introduced in Diebold and Mariano [53].

The statistical test is performed as follow:

$$MGN = \frac{\hat{\rho}_{sd}}{\sqrt{\frac{1 - \hat{\rho}_{sd}^2}{n-1}}} \tag{27}$$

where $\hat{\rho}_{sd}$ is the correlation coefficient between s and d, with $s = r_1 + r_2$, $d = r_1 - r_2$, r_1 is the residual of model 1 and r_2 is the residual of model 2.

This statistical test is a student's $t-$distribution with $n - 1$ degrees of freedom. Tables 4 and 5 present the results of the tests for the two presented data sets, considering a significant level (α) of 5%. If the p-value is lower than α, we reject the null hypothesis, which assumes the models have equal accuracy. In this way, it is possible to see in Tables 4 and 5 that both proposed models, SM-ePL-KRLS and ESM-ePL-KRLS, showed a statistically proven better accuracy than ePL-KRLS.

To prove the efficiency of evolving models in the estimation of the hot-spot temperature of power transformers, they were also compared with other non-evolving (fixed structure) models described in the literature. These models include the deterministic model based on IEEE Standard C57.91-2011 (IEEE-DM) describe in Section 2, a model based on a Multi-layer Perceptron Neural Network (MLP) and a model based on an Adaptative Neurofuzzy Inference System (ANFIS) [34].

In the deterministic modeling (IEEE-DM) the experimental transformer characteristic parameters used in this work were the following:

$R = 4$, $\Delta\Theta_{H,R} = 5\,^{\circ}$C, $\Delta\Theta_{TO,R} = 54\,^{\circ}$ C, $\Theta_{H,R} = 80\,^{\circ}$ C

$\Theta_{A,R} = 21\,^{\circ}$ C, $q = 0.8$, $m = 0.8$, $\tau_{TO} = 3$ h, $\tau_H = 0.1$ h

The MLP neural network was implemented with a single hidden layer with 4 neurons trained with the backpropagation algorithm. The ANFIS model was implemented with four fuzzy sets for each input variable and four fuzzy rules generated by means of the fuzzy c-means

Algorithm 3: SM

$\tilde{e}^k = y^k - \hat{y}^k$
if $\tilde{e}^k > \gamma_{bar}$ **then**
  $\beta = 1 - \frac{\gamma_{bar}}{|\tilde{e}^k|}$
else
  $\beta = 0$
if $\beta < 0$ **then**
  $\beta = 0$
$\tau = \beta$
$\gamma = 1 - \beta$

Fig. 3. SM model.

Table 1
Characteristics of the experimental power transformer.

Copper losses	776 W
Factory year	MACE/1987
Iron losses	195 W
Nameplate rating	25 kVA
Tank dimensions	64 × 16 × 80 cm^3
Top oil temperature rise at full load	73.1 °C
Type of cooling	ONAN
$V_{primary}/V_{secondary}$	10 kV / 380 kV
Weight of core and coil assembly	136 kg
Weight of oil	62 kg

clustering procedure [54]. Tables 6 and 7 show the results obtained for all models implemented in this work.

In Figs. 7 and 8 graphical depictions of the results obtained in this work are presented. For visualization purposes only the models proposed in this work and the IEEE-DM model, which is the most used model in practice for the prediction of the hot-spot temperature, are shown in the Figure.

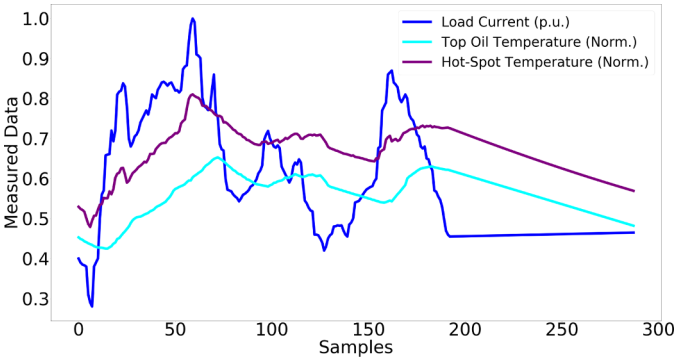

Fig. 5. Data set 1: no overload condition.

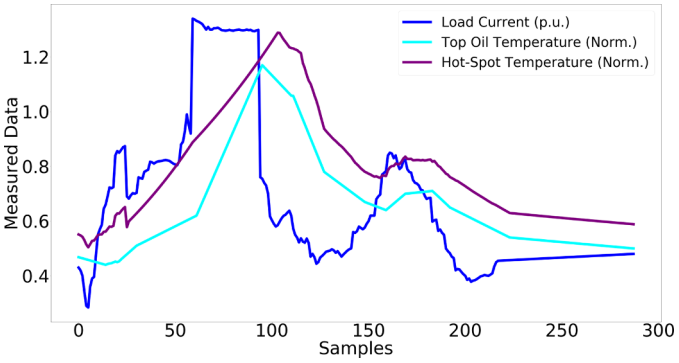

Fig. 6. Data set 2: with overload condition.

Table 2
Results of evolving models - Data set 1: no overload condition.

Algorithm	γ_{bar}	RMSE	NDEI	MAE	Rules
eMG [24]	–	0.01180	0.18560	0.00900	**1**
ePL-KRLS	–	0.01367	0.21529	0.01010	**1**
SM-ePL-KRLS	0.01100	0.01032	0.16244	0.00756	**1**
ESM-ePL-KRLS	0.01120	**0.01026**	**0.16158**	**0.00750**	**1**

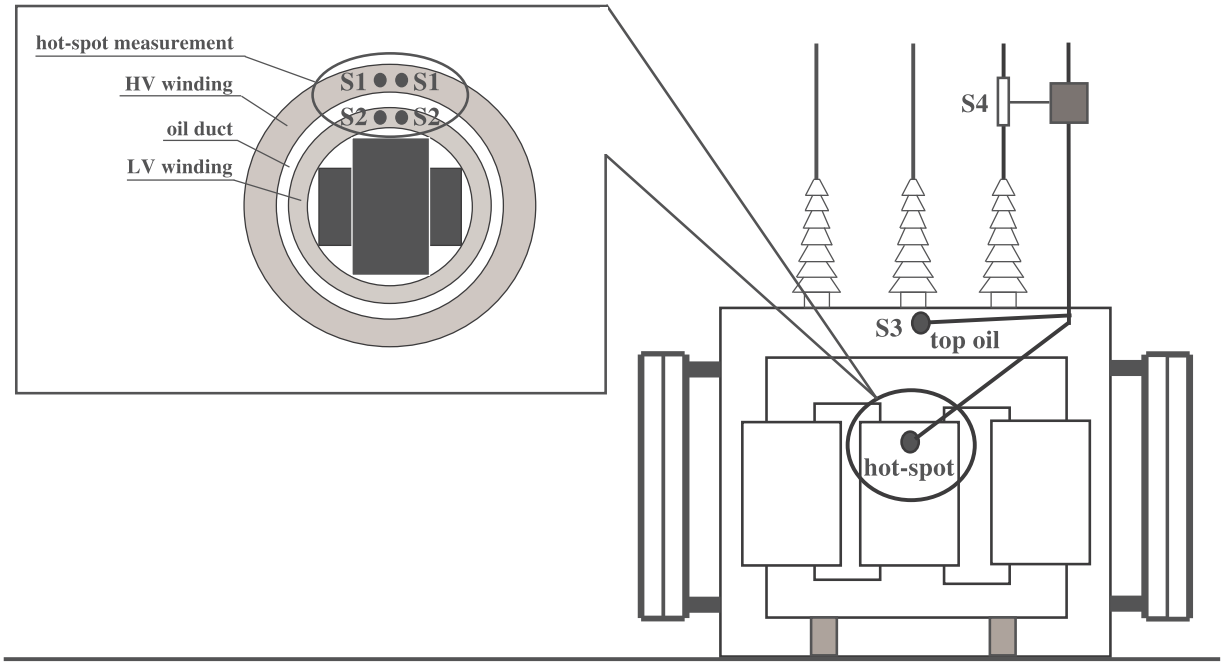

Fig. 4. Sensor's location in the experimental transformer.

Table 3
Results of evolving models - Data set 2: with overload condition.

Algorithm	γ_{bar}	RMSE	NDEI	MAE	Rules
eMG [24]	–	0.0303	0.1444	0.0184	2
ePL-KRLS	–	0.03304	0.16146	0.02420	2
SM-ePL-KRLS	0.01106	0.02647	0.12933	0.01870	**1**
ESM-ePL-KRLS	0.01720	**0.02519**	**0.12308**	**0.01855**	**1**

Table 4
Results of the *MGN* test - Data set 1.

Model 1 × Model 2	*MGN*	*p*-value
ESM-ePL-KRLS × ePL-KRLS	5.5747	0.0000
SM-ePL-KRLS × ePL-KRLS	5.4579	0.0000

Table 5
Results of the *MGN* test - Data set 2.

Model 1 × Model 2	*MGN*	*p*-value
ESM-ePL-KRLS × ePL-KRLS	5.2614	0.0000
SM-ePL-KRLS × ePL-KRLS	5.02844	0.0000

Table 6
Comparing performance with non-evolving models - Data set 1.

Algorithm	RMSE	NDEI	MAE	Rules
Deterministic Model (IEEE-DM) [16]	1.0245	16.1089	0.7524	–
Multilayer Perceptron (MLP) [33]	0.0467	0.7336	0.0343	4
Adapt. Neurofuzzy Inf. Sys. (ANFIS) [34]	0.0124	0.1952	0.0091	4
Evolving Multivariable Gaussian (eMG) [24]	0.0118	0.1856	0.0090	1
ePL-KRLS	0.0137	0.2153	0.0101	1
SM-ePL-KRLS	0.0103	0.1624	0.0076	1
ESM-ePL-KRLS	**0.0103**	**0.1616**	**0.0075**	**1**

Table 7
Comparing performance with non-evolving models - Data set 2.

Algorithm	RMSE	NDEI	MAE	Rules
Deterministic Model (IEEE-DM) [16]	0.4005	1.9446	0.2769	–
Multilayer Perceptron (MLP) [33]	0.0317	0.1539	0.0219	4
Adapt. Neurofuzzy Inf. Sys. (ANFIS) [34]	0.0481	0.2340	0.0333	4
Evolving Multivariable Gaussian (eMG) [24]	0.0303	0.1444	0.0184	2
ePL-KRLS	0.0330	0.1615	0.0242	2
SM-ePL-KRLS	0.0265	0.1293	0.0188	1
ESM-ePL-KRLS	**0.0252**	**0.1231**	**0.0186**	**1**

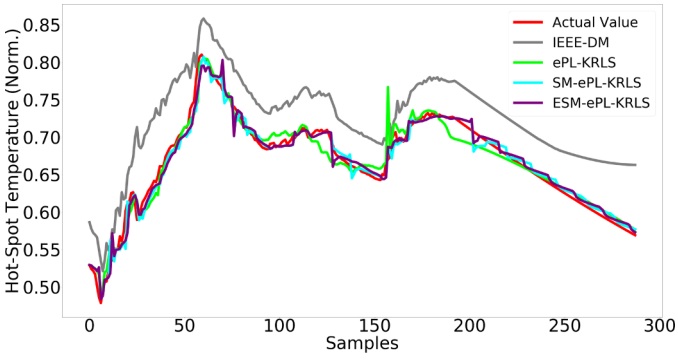

Fig. 7. Hot-spot estimation - Data set 1.

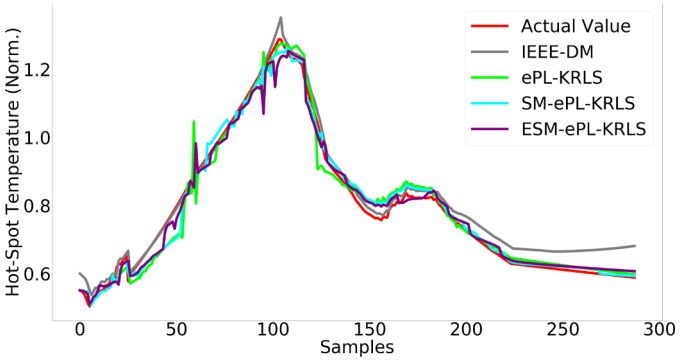

Fig. 8. Hot-spot estimation - Data set 2.

As can be seen in Tables 6 and 7, the proposed models maintain a better performance than their non-evolving counterparts. In the same way as seen previously, the prediction accuracy of the ESM-ePL-KRLS model surpasses all others in the presence of an overload condition in the operation of the transformer. Once again, this model appears as a promising alternative to integrate decision-making systems to assist during the operational planning of the electrical system.

It is worth mentioning that although the data presented in this work have been collected in experimental equipment, the proposed model can be easily applied to real-world transformers through small modifications in this equipment. These modifications include the installation of a hall effect sensor to measure the load current and the insertion of only one temperature fiber-optical based sensor in the transformer's inspection cover to measure the top oil temperature.

In addition to these modifications being non-invasive and having a low implementation cost, many transformers in operation already have such sensors installed, which makes the proposed approach highly applicable to the real-world problem of estimating the hot-spot temperature of power transformers.

5. Conclusions

In this paper, two forecasting models are suggested to deal with the problem of power transformers' hot-spot temperature estimation: the ESM-ePL-KRLS and SM-ePL-KRLS. These models were tested with data sets collected from a real experimental transformer, where two load conditions were considered: with and without an overload condition.

The evaluation of these models was measured in terms of error and number of final rules and showed that both models have better accuracy and lower computational cost when compared to other evolving and fixed-structure (non-evolving) models suggested in the literature. Another benefit of the introduced models is that their structure makes the knowledge process continuous and more adaptable as the data changes than its evolving counterparts. In addition, a MGN statistical test supports that the proposed models have better accuracy than the original ePL-KRLS model.

In particular, the ESM-ePL-KRLS model shown higher accuracy if compared to all others in the presence of an overload condition in the operation of the transformer. Besides, the number of rules of the SM-ePL-KRLS model presented a considerable variation and reached a high number during the evolving process, indicating a high computational cost. These facts suggest that the ESM-ePL-KRLS model appears as a better choice to integrate a decision support tool to assist the operational planning of the electrical power system.

Future studies include the integration of the proposed method into a system that will predict, from the hot-spot temperature, the effects of an eventual overload during operation on the transformer's residual life and the implementation of a mechanism to update the value of the Enhanced Set-Membership parameters: γ_{bar}, *gr* and *dr*. This implementation will make the model more flexible according to the input

data varies along time and consequently, it will improve the performance of the hot-spot temperature's estimation. We hope to address these issues in the near future.

CRediT authorship contribution statement

Kaike Sa Teles Rocha Alves: Project administration, Validation, Writing - original draft. **Michel Hell:** Conceptualization, Methodology, Data curation, Writing - review & editing. **Fernando Luiz Cyrino Oliveira:** Conceptualization, Methodology, Supervision. **Eduardo Pestana de Aguiar:** Conceptualization, Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge FAPEMIG, FAPERJ, CAPES, CNPq and Federal University of Juiz de Fora for financial support and the anonymous reviewers for their comments and useful suggestions that led to the improvement of this work.

References

[1] H. Ma, T.K. Saha, C. Ekanayake, D. Martin, Smart transformer for smart grid-intelligent framework and techniques for power transformer asset management, IEEE Trans. Smart Grid 6 (2) (2015) 1026–1034.
[2] H. de Faria Jr, J.G.S. Costa, J.L.M. Olivas, A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis, Renew. Sustain. Energy Rev. 46 (2015) 201–209.
[3] S.T. Jan, R. Afzal, A.Z. Khan, Transformer failures, causes & impact, International Conference Data Mining, Civil and Mechanical Engineering, (2015), pp. 49–52.
[4] G. Rigatos, P. Siano, Power transformers condition monitoring using neural modeling and the local statistical approach to fault diagnosis, Int. J. Electr. Power Energy Syst. 80 (2016) 150–159.
[5] D. Wang, C. Mao, J. Lu, S. Fan, F. Peng, Theory and application of distribution electronic power transformer, Electr. Power Syst. Res. 77 (3–4) (2007) 219–226.
[6] C. Bengtsson, Status and trends in transformer monitoring, IEEE Trans. Power Deliv. 11 (3) (1996) 1379–1384.
[7] P. Daponte, D. Grimaldi, A. Piccolo, D. Villacci, A neural diagnostic system for the monitoring of transformer heating, Measurement 18 (1) (1996) 35–46.
[8] A.Y. Arabul, I. Senol, Development of a hot-spot temperature calculation method for the loss of life estimation of an ONAN distribution transformer, Electr. Eng. 100 (3) (2018) 1651–1659.
[9] I.A. Metwally, Failures, monitoring and new trends of power transformers, IEEE Potentials 30 (3) (2011) 36–43.
[10] E. Ali, A. Helal, H. Desouki, K. Shebl, S. Abdelkader, O. Malik, Power transformer differential protection using current and voltage ratios, Electr. Power Syst. Res. 154 (2018) 140–150.
[11] R. Murugan, R. Ramasamy, Understanding the power transformer component failures for health index-based maintenance planning in electric utilities, Eng. Fail. Anal. 96 (2019) 274–288.
[12] R.M.A. Velásquez, J.V.M. Lara, Bushing failure in power transformers and the influence of moisture with the spectroscopy test, Eng. Fail. Anal. 94 (2018) 300–312.
[13] A. Christina, M. Salam, Q. Rahman, F. Wen, S. Ang, W. Voon, Causes of transformer failures and diagnostic methods–a review, Renew. Sustain. Energy Rev. 82 (2018) 1442–1456.
[14] J. Bérubé, J. Aubin, W. McDermid, Transformer winding hot spot-temperature determination, Fifth Annual Weidmann ACTI Technical Conf, (2006), pp. 1–10.
[15] Z. Radakovic, K. Feser, A new method for the calculation of the hot-spot temperature in power transformers with ONAN cooling, IEEE Trans. Power Deliv. 18 (4) (2003) 1284–1292.
[16] IEEE guide for loading mineral-oil-immersed transformers and step-voltage regulators, IEEE Std C57.91-2011 (Revision of IEEE Std C57.91-1995) (2012) 1–123.
[17] M. Hell, P. Costa, F. Gomide, Participatory learning in power transformers thermal modeling, IEEE Trans. Power Deliv. 23 (4) (2008) 2058–2067.
[18] M. Hell, P. Costa, F. Gomide, Recurrent neurofuzzy network in thermal modeling of power transformers, IEEE Trans. Power Deliv. 22 (2) (2007) 904–910.
[19] V. Mijailovic, Method for effects evaluation of some forms of power transformers preventive maintenance, Electr. Power Syst. Res. 78 (5) (2008) 765–776.
[20] Q. He, J. Si, D.J. Tylavsky, Prediction of top-oil temperature for transformers using neural networks, IEEE Trans. Power Deliv. 15 (4) (2000) 1205–1211.
[21] A. Ngaopitakkul, A. Kunakorn, Internal fault classification in transformer windings using combination of discrete wavelet transforms and back-propagation neural networks, Int. J. Control Autom. Syst. 4 (3) (2006) 365–371.
[22] A.R.G. Castro, V. Miranda, An interpretation of neural networks as inference engines with application to transformer failure diagnosis, Int. J. Electr. Power Energy Syst. 27 (9–10) (2005) 620–626.
[23] O. Roizman, V. Davydov, Neuro-fuzzy computing for large power transformers monitoring and diagnostics, 18th International Conference of the North American Fuzzy Information Processing Society - NAFIPS (Cat. No. 99TH8397), IEEE, 1999, pp. 248–252.
[24] L. Souza, A.P. Lemos, W.M. Caminhas, W. Boaventura, Thermal modeling of power transformers using evolving fuzzy systems, Eng. Appl. Artif. Intell. 25 (5) (2012) 980–988.
[25] F.C. Sica, F.G. Guimarães, R. de Oliveira Duarte, A.J. Reis, A cognitive system for fault prognosis in power transformers, Electr. Power Syst. Res. 127 (2015) 109–117.
[26] K. Bacha, S. Souahlia, M. Gossa, Power transformer fault diagnosis based on dissolved gas analysis by support vector machine, Electr. Power Syst. Res. 83 (1) (2012) 73–79.
[27] L. Ganyun, C. Haozhong, Z. Haibao, D. Lixin, Fault diagnosis of power transformer based on multi-layer SVM classifier, Electr. Power Syst. Res. 74 (1) (2005) 1–7.
[28] M. Žarković, Z. Stojković, Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics, Electr. Power Syst. Res. 149 (2017) 125–136.
[29] E.P. de Aguiar, M.d.A. Fernando, M.M. Vellasco, M.V. Ribeiro, Set-membership type-1 fuzzy logic system applied to fault classification in a switch machine, IEEE Trans. Intell. Transp. Syst. 18 (10) (2017) 2703–2712.
[30] P. Clarke, R.C. de Lamare, Low-complexity reduced-rank linear interference suppression based on set-membership joint iterative optimization for DS-CDMA systems, IEEE Trans. Veh. Technol. 60 (9) (2011) 4324–4337.
[31] Y. Li, Y. Wang, T. Jiang, Sparse-aware set-membership NLMS algorithms and their application for sparse channel estimation and echo cancelation, AEU Int. J. Electron. Commun. 70 (7) (2016) 895–902.
[32] A. Lemos, W. Caminhas, F. Gomide, Multivariable gaussian evolving fuzzy modeling system, IEEE Trans. Fuzzy Syst. 19 (1) (2010) 91–104.
[33] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.
[34] J.-S. Jang, Anfis: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man. Cybern. 23 (3) (1993) 665–685.
[35] N. Kasabov, D. Filev, Evolving intelligent systems: methods, learning, & applications, 2006 International Symposium on Evolving Fuzzy Systems, IEEE, 2006, pp. 8–18.
[36] V. Galdi, L. Ippolito, A. Piccolo, A. Vaccaro, Neural diagnostic system for transformer thermal overload protection, IEE Proc. Electric Power Appl. 147 (5) (2000) 415–421.
[37] A.T. Azar, S. Vaidyanathan, Computational Intelligence Applications in Modeling and Control, Springer, 2015.
[38] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Syst. Man. Cybern. (1) (1985) 116–132.
[39] M. Sugeno, T. Takagi, Fuzzy identification of systems and its applications to modelling and control, Read. Fuzzy Sets Intell. Syst. 15 (1) (1993) 387–403.
[40] I. Škrjanc, J.A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, F. Gomide, Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: a survey, Inf. Sci. 490 (2019) 344–368.
[41] R.R. Yager, A model of participatory learning, IEEE Trans. Syst. Man. Cybern. 20 (5) (1990) 1229–1234.
[42] E. Lima, M. Hell, R. Ballini, F. Gomide, Evolving fuzzy modeling using participatory learning, Evol. Intell. Syst. (2010) 67–86.
[43] A. Khosravi, R. Koury, L. Machado, J. Pabon, Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms, J. Clean. Prod. 176 (2018) 63–75.
[44] A.O. Ayres, F.J. Von Zuben, Multitask learning applied to evolving fuzzy-rule-based predictors, Evol. Syst. (2019) 1–16.
[45] S.B. Roh, S.K. Oh, W. Pedrycz, K. Seo, Z. Fu, Design methodology for radial basis function neural networks classifier based on locally linear reconstruction and conditional fuzzy c-means clustering, Int. J. Approx. Reason. 106 (2019) 228–243.
[46] P.P. Angelov, D.P. Filev, An approach to online identification of Takagi-Sugeno fuzzy models, IEEE Trans. Syst. Man. Cybern. Part B 34 (1) (2004) 484–498.
[47] R. Vieira, F. Gomide, R. Ballini, Kernel evolving participatory fuzzy modeling for time series forecasting, 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2018, pp. 1–9.
[48] B. Scholkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2001.
[49] L.S. Ngia, J. Sjöberg, M. Viberg, Adaptive neural nets filter using a recursive Levenberg-Marquadt search direction, Asilomar Conference on Signals, Systems, and Computers, 1 Citeseer, 1998, pp. 697–701.
[50] W. Liu, J.C. Principe, S. Haykin, Kernel Adaptive Filtering: A Comprehensive Introduction, 57 John Wiley & Sons, 2011.
[51] C. Richard, J.C.M. Bermudez, P. Honeine, Online prediction of time series data with kernels, IEEE Trans. Signal Process. 57 (3) (2008) 1058–1067.
[52] D. Villacci, G. Bontempi, A. Vaccaro, M. Birattari, The role of learning methods in the dynamic assessment of power components loading capability, IEEE Trans. Ind. Electron. 52 (1) (2005) 280–289.
[53] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econ. Stat. 20 (1) (2002) 134–144.
[54] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithm, Kluwer Academic Publishers, 1981.