Universidade Federal de Juiz de Fora

Curso de Graduação em Engenharia de Produção

Kaike Sa Teles Rocha Alves

An Enhanced Set−Membership Evolving Participatory Learning With Kernel Recursive Least Squares Applied in Thermal Modeling of Power Transformers

Juiz de Fora

2019

Kaike Sa Teles Rocha Alves

An Enhanced Set−Membership Evolving Participatory Learning With Kernel Recursive Least Squares Applied in Thermal Modeling of Power Transformers

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheiro de Produção.

Orientador: Eduardo Pestana de Aguiar

Juiz de Fora

2019

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Sa Teles Rocha Alves, Kaike.
An Enhanced Set−Membership Evolving Participatory Learning With Kernel Recursive Least Squares Applied in Thermal Modeling of Power Transformers / Kaike Sa Teles Rocha Alves. – 2019.
31 f. : il.

Orientador: Eduardo Pestana de Aguiar
Trabalho de Conclusão de Curso (Graduação) – Universidade Federal de Juiz de Fora, . Curso de Graduação em Engenharia de Produção, 2019.

1. Enhanced Set−Membership. 2. evolving Fuzzy Systems. 3. Power transformer. 4. Modeling I. Pestana de Aguiar, Eduardo, orient. II. Título.

Kaike Sa Teles Rocha Alves

An Enhanced Set−Membership Evolving Participatory Learning With Kernel Recursive Least Squares Applied in Thermal Modeling of Power Transformers

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheiro de Produção.

Aprovado em 14 de novembro de 2019

BANCA EXAMINADORA

Eduardo Pestana de Aguiar

D. Sc, Eduardo Pestana de Aguiar - Orientador
Universidade Federal de Juiz de Fora

Eliane da S. Christo

D. Sc, Eliane da Silva Christo
Universidade Federal Fluminense

Fernando M A Nogueira

D. Sc, Fernando Marques de Almeida Nogueira
Universidade Federal de Juiz de Fora

D. Sc, Roberto Malheiros Moreira Filho
Universidade Federal de Juiz de Fora

To my parents, Gilterney and Rita

ACKLOWLEDGMENT

I thank Jesus Christ for all he has done for me and all his love.

To my parents, Gilterney and Rita, for giving me precious teachings and for supporting me in my career.

To my sisters, Keyla and Anna Klara, with I have learned much from.

To all the friends who have made and are still part of my life.

To my guide Eduardo Aguiar for advice me in this process.

To the professors of the Industrial Engineering Program from the Federal University of Juiz de Fora.

To all people who contributed to my progress.

"If you want to go quickly, go alone. If you want to go far, go together."

(AFRICAN PROVERB)

RESUMO

Um fator que afeta diretamente a vida útil de um transformador é a temperatura do ponto quente e seu monitoramento é importante para prevenir falhas, reduzir custos, manter a segurança e prover um serviço confiável aos clientes. Com o avanço da inteligência artificial, vários modelos foram sugeridos na literatura relacionados a modelagem de transformadores de energia. Tais modelos produzem resultados com alta precisão e menores erros. Neste trabalho, são propostos dois modelos de previsão, classificados como modelos nebulosos evolutivos, para prever a temperatura do ponto quente de um transformados. A primeira é a implementação do conceito Set−Membership no modelo evolving Participatory Learning with Kernel Recursive Least Squares. A segunda é uma combinação do modelo evolving Participatory Learning with Kernel Recursive Least Squares com o Enhanced Set−Membership, uma versão melhorada do Set−Membership. Tanto o Set−Membership quanto o Enhanced Set−Membership são implementadas para atualizar a taxa de mudança do índice de excitação, que é um parâmetro que controla a taxa em que regras são criadas. A atualização desse parâmetro é realizada em função do valor do erro. Um conjunto de dados coletados de um transformador experimental foi utilizado para avaliar o desempenho dos dois modelos propostos. Os resultados da proposta são comparados com o desempenho do evolving Participatory Learning with Kernel Recursive Least Squares e com outros modelos clássicos sugeridos na literatura. Os modelos propostos atingiram os erros mais baixos. Eles obtiveram uma redução média de cerca de 31.38% no erro RMSE, 31.48% no NDEI, e 39.31% no MAE. Além disso, eles terminaram as simulações com um número de regras finais competitivo. Os resultados sugerem que os modelos são abordagens eficientes para modelar dados complexos com alta precisão.

Palavras-chave: Enhanced Set−Membership. Sistemas evolutivos participativos. Transformador de energia. Modelagem.

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. With the advance of the artificial intelligence, several models have been suggested in the literature regarding power transformer's modeling. Such models produce results with high accuracy and lower errors. In this work, we propose two forecasting models classified as an evolving Fuzzy Systems for 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, called 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 the controls the creation of rules. This parameter update is a function of the error value. A dataset 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 other classical models suggested in the literature. The proposals achieved the lowest errors. The mean percentage reduction was about 31.38% in the RMSE error, 31.48% in the NDEI, and 39.31% in the MAE. Furthermore, the proposals obtained a competitive number of final rules. The results suggest that the models are efficient approaches to modeling complex data with high accuracy.

Key-words: Enhanced Set−Membership. Evolving Fuzzy Systems. Power transformers. Modeling.

LIST OF FIGURES

Figure 1 – Mechanism of learning of ePL−KRLS model 17
Figure 2 – ESM−ePL−KRLS Algorithm 22
Figure 3 – Set−Memebership model 23

Figure 4 – Number of rules of the models 25
Figure 5 – Desired values and predictions 26

LIST OF TABLES

Table 1 – Characteristics of the experimental transformer 13

Table 2 – Results of proposed models and ePL−KRLS 24
Table 3 – Percentage reduction in the errors of the proposed models 24
Table 4 – Results of the MGN test 27
Table 5 – Comparing performance with other forecasting models 27

CONTENTS

1 INTRODUCTION 11
1.1 BACKGROUND OF THE STUDY 11
1.2 IMPORTANCE OF THE STUDY 12
1.3 SCOPE 13
1.4 STATEMENT OF THE OBJECTIVES 13
1.5 METHODOLOGY 13
1.6 WORK ORGANIZATION 14

2 PROBLEM FORMULATION 15

3 PROPOSED MODELS 17
3.1 ePL-KRLS ALGORITHM 17
3.2 SET-MEMBERSHIP (SM) 20
3.3 ENHANCED SET-MEMBERSHIP (ESM) 21

4 EXPERIMENTAL RESULTS 24

5 CONCLUSIONS 28

REFERENCES 29

1 INTRODUCTION

1.1 BACKGROUND OF THE STUDY

The power transformer is a critical equipment in power distribution. 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 (JAN; AFZAL; KHAN, 2015). Due to the composition of a power transformer, it is the most expensive apparatus in energy distribution. In the case of a power transformer's failure, when the recovering process is possible, it is slow and inefficient (WANG et al., 2007). Thereof, the monitoring is vital to prevent faults, reduce costs, keep the safety and provide a reliable service to consumers (BENGTSSON, 1996). The annual spent on power transformers' monitoring hardware will increase more than $ 642 million in eight years until 2020, according to (ARABUL; SENOL, 2018; METWALLY, 2011), 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 totalize approximately 44% (ALI et al., 2018). The bushing is a fragile component constituted of four parts: insulation, conductor, connection clamp, and accessories. In the present work, we considered power transformers composed of Resin-bonded paper bushings (RBP) (CHRISTINA et al., 2018).

The principal factor in bushing failures is the hot−spot temperature, representing 32% of the total causes. The hot−spot is the highest temperature on the winding, located near to the top of the power transformer (BÉRUBÉ; AUBIN; MCDERMID, 2006). The increase in the hot−spot temperature reduces the lifespan of the insulation. Its defect may determine the end life of the power transformer (RADAKOVIC; FESER, 2003). This temperature is the main limiting of loading capability (BÉRUBÉ; AUBIN; MCDERMID, 2006).

As the estimation of the hot−spot is a complex task, there exist many models. A classical model is the deterministic model using transient heating equations proposed by (BOARD, 1995). This model is imprecise due to assumed simplifications, and consequently, the power transformer must operate below the maximum capacity to prevent damages. This conservative attitude increases the operational cost to the company (BOARD, 1995; HELL; COSTA; GOMIDE, 2008). 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 (HELL; COSTA; GOMIDE, 2007).

In literature, we found applications of several works in power transformers monitoring. Paper (MIJAILOVIC, 2008) suggests a formulation to calculate the expected cost to repair a power transformer. In paper (HE; SI; TYLAVSKY, 2000; MIJAILOVIC,

2008; NGAOPITAKKUL; KUNAKORN, 2006) was used artificial neural networks (ANN) for diagnostics and in (CASTRO; MIRANDA, 2005) the ANN is proposed to analyze faults in dissolved gas−in−oil. Paper (ROIZMAN; DAVYDOV, 1999) suggests fuzzy sets and (HELL; COSTA; GOMIDE, 2007) presents neuro−fuzzy hybrids. Reference (HELL; COSTA; GOMIDE, 2008) introduces the use of participatory learning (PL) to train a hybrid neuro−fuzzy network and in paper (SOUZA et al., 2012) is used an evolving multivariable Gaussian (eMG). Cortez proposed a fault prognosis provided by an intelligent system based on cognitive systems (SICA et al., 2015) and a vector machine (SVM) was proposed in (BACHA; SOUAHLIA; GOSSA, 2012; GANYUN et al., 2005) to the same purpose. In paper (ŽARKOVIĆ; STOJKOVIĆ, 2017), 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 (AGUIAR et al., 2017; CLARKE; LAMARE, 2011; LI; WANG; JIANG, 2016) 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, runtime, and the number of final rules, using data set from thermal modeling of power transformers. Additionally, we compare the performance of the proposal with other approaches suggested in the literature (evolving Multivariable Gaussian (LEMOS; CAMINHAS; GOMIDE, 2010), Multilayer Perceptron (DUDA; HART; STORK, 2012), extended Tagaki−Sugeno (ANGELOV; ZHOU, 2006), and Deterministic Model (BOARD, 1995))

1.2 IMPORTANCE OF THE STUDY

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 ESM−ePL−KRLS obtained the same number of final rules and a competitive runtime comparing with the ePL−KRLS. The SM−ePL−KRLS achieved a lower

number of final rules than ePL−KRLS, but the runtime increased considerably. These results suggest that the computational cost of ESM−ePL−KRLS is lower than of SM−ePL−KRLS.

- The results indicate that monitoring the hot−spot temperature by the introduced models is efficient to control the load current and improves the lifespan of the power transformers.

1.3 SCOPE

The assessment of the proposed models is performed using data from an experimental transformer provided by (GALDI et al., 2000). Measurements of the temperature and the Hall effect are collected using sensors inside the transformer. Table 1 presents the characteristics of the transformer.

Table 1 – Characteristics of the experimental transformer

Copper losses	776 W
Factory year	MACE/1987
Iron losses	195 W
Nameplate rating	25 kVA
Tank dimensions	$64 \times 16 \times 80\ cm^3$
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

Source: (HELL; COSTA; GOMIDE, 2008)

1.4 STATEMENT OF THE OBJECTIVES

This work aims to propose two forecasting models able to predict complex data with high accuracy. The 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 (KASABOV; FILEV, 2006).

1.5 METHODOLOGY

The collection of data occurred every five minutes for 24 hours. The simulations consist of predicting the hot−spot temperature using as input: the load current (K), the top oil temperature (Θ_{TO}), and one step delayed−load−current as suggested by (HELL; COSTA; GOMIDE, 2008).

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 Equations (1.1), (1.2) and (1.3) respectively.

$$RMSE = \sqrt{\frac{1}{T}\sum_{k=1}^{T}(y^k - \hat{y}^k)^2} \tag{1.1}$$

$$NDEI = \frac{RMSE}{std([y^1, ..., y^T])} \tag{1.2}$$

$$MAE = \frac{1}{T}\sum_{k=1}^{T}|y^k - \hat{y}^k| \tag{1.3}$$

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

The processing time (seconds) and the number of rules estimates the computational cost (VIEIRA; GOMIDE; BALLINI, 2018).

1.6 WORK ORGANIZATION

This work is organized as follows:

- Section 2 presents the approach of the deterministic model.
- Section 3 discusses the ePL-KRLS model and the SM filtering. And finally, it introduces the ESM concept.
- Section 4 discusses the errors of the predictions, the processing time, and the number of rules of the simulations. Additionally, a statistical test validates the performance of the proposal.
- Section 5 presents the main conclusions of this work.

2 PROBLEM FORMULATION

The deterministic model proposed in (BOARD, 1995) is detailed in this section. The deterministic model starts by calculating the ultimate top oil rise ($\Delta\Theta_{TO,U}$), using the Equation (2.1).

$$\Delta\Theta_{TO,U} = \Delta\Theta_{TO,R}\left[\frac{K^2R+1}{R+1}\right]^n \tag{2.1}$$

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 Equation (2.1) and the environment temperature (Θ_A), the increment in the top oil temperature (Θ_{TO}) is found by the differential equation expressed in (2.2).

$$\tau_{TO}\frac{d\Theta_{TO}}{dt} = [\Delta\Theta_{TO,U} + \Theta_A] - \Theta_{TO} \tag{2.2}$$

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:

$$\Delta\Theta_{H,U} = \Delta\Theta_{H,R}K^{2m} \tag{2.3}$$

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 Equation (2.3), as the differential equation (2.4).

$$\tau_H\frac{d\Delta\Theta_H}{dt} = \Delta\Theta_{H,U} - \Delta\Theta_H \tag{2.4}$$

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 Equation 2.5, where these parameters are obtained from Equations (2.2) and (2.4) respectively.

$$\Theta_H = \Theta_{TO} + \Delta\Theta_H \tag{2.5}$$

The power transformer operates between 70% to 80% of their nominal capacity by using the deterministic model presented in the current section. It implies a loss of more than one power transformer at every five (HELL; COSTA; GOMIDE, 2008). 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.

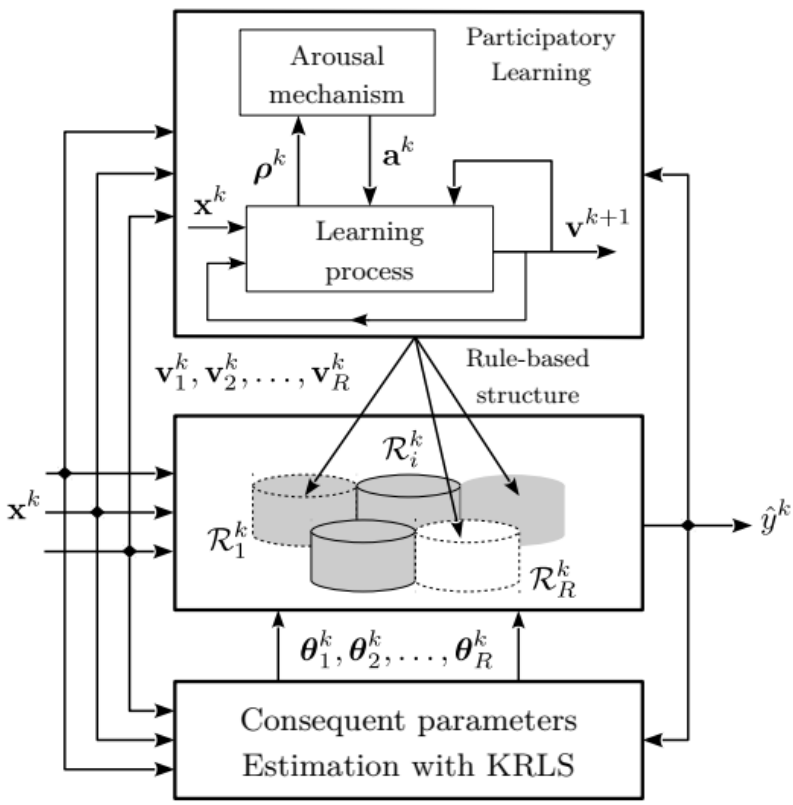
3 PROPOSED MODELS

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 (TAKAGI; SUGENO, 1985). This model clusters the input space according to the degree of similarity in the knowledge process (LEMOS; CAMINHAS; GOMIDE, 2010). This technique of clustering uses the concept of participatory learning (PL). The PL procedure is based on human learning (YAGER, 1990; LIMA et al., 2010). 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 the global output. The ePL−KRLS estimates the consequent parameters using the kernel recursive least squares (ANGELOV; FILEV, 2004; VIEIRA; GOMIDE; BALLINI, 2018). The learning structure of ePL−KRLS is represented in Figure 1.

Figure 1 – Mechanism of learning of ePL−KRLS model



Source: (VIEIRA; GOMIDE; BALLINI, 2018)

The compatibility index (ρ_i^k) and the arousal index (a_i^k) are calculated from Equations (3.1) and (3.2), respectively.

$$\rho_i^k = 1 - \frac{\left\|X^k - V_i^k\right\|}{m} \tag{3.1}$$

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 of a rule at step k of $i-$th rule.

$$a_i^k = a_i^{k-1} + \beta(1 - \rho_i^k - a_i^{k-1}) \tag{3.2}$$

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. The model normalizes the input vector X^k before the simulation, and therefore, ρ ranges from 0 to 1. 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: $\rho_i^k, a_i^k \in [0, 1], \forall i = 1, ..., R^k$, where R is the number of rules at $k-$th 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 Equation (3.3).

$$\hat{V}_i^{k+1} = \hat{V}_i + \alpha(\rho_i^k)^{1-a_i^k}(X^k - \hat{V}_i^k) \tag{3.3}$$

where $10^{-5} < \alpha < 10^{-1}$ is the learning rate defined in (VIEIRA; GOMIDE; BALLINI, 2018).

The local output is calculated according to Equation (3.4).

$$\hat{y}_i = f_i(X, \theta_i) = \sum_{j=1}^{n_i} \theta_{ij} \kappa(d_{ij}, X) \tag{3.4}$$

where θ_i is the consequent parameters of $i-$th rule, d_{ij} is the $j-$th element of the local dictionary in $i-$th rule and $\kappa\langle .,. \rangle$ is the Gaussian$-$Kernel function shown in Equation (3.5) (SCHOLKOPF; SMOLA, 2001).

$$\kappa\langle X^i, X^j \rangle = exp\left(-\frac{\|X^i - X^j\|^2}{2\nu^2}\right) \tag{3.5}$$

where ν is positive and represents the size of the kernel. The ν is found minimizing the error function using Equation (3.6) and it is based on a recursive Levenberg−Marquardt model (NGIA; SJÖBERG; VIBERG, 1998).

$$\nu_i^k = \nu_i^{k-1} + P_i^k \nabla_i^k \tilde{e}_i^k \tag{3.6}$$

where P_i^k is calculated from Equation (3.7), ∇_i^k is the gradient of error expressed in (3.8) 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.

$$P_i^k = \left[P_i^{k-1} - \frac{P_i^{k-1} \nabla_i^k [\nabla_i^k]^T P_i^{k-1}}{1 + [\nabla_i^k]^T P_i^{k-1} \nabla_i^k} \right] \tag{3.7}$$

$$\nabla_i^k = -\left[\frac{\partial \tilde{e}_i^k}{\partial v_{i1}^{k-1}}, \ldots, \frac{\partial \tilde{e}_i^k}{\partial v_{in_1}^{k-1}} \right]^T = \Gamma_i^N(X^k) \begin{bmatrix} \theta_{i1}^{k-1} - \frac{\left\| X^k - d_{i1}^k \right\|^2}{(v_{i1}^{k-1})^3} k(X^k, d_{i1}^k) \\ \vdots \\ \theta_{in_1}^{k-1} - \frac{\left\| X^k - d_{in_1}^k \right\|^2}{(v_{in_1}^{k-1})^3} k(X^k, d_{in_1}^k) \end{bmatrix} \tag{3.8}$$

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

The model output is computed from local outputs as the Equation (3.9).

$$\hat{y} = \sum_{i=1}^{R} \hat{y}_i \Gamma_i^N(X) \tag{3.9}$$

where $\Gamma_i(X)$ is the normalized firing degree expressed in Equation (3.10).

$$\Gamma_i(X) = \frac{\mathcal{A}_i(X)}{\sum_{j=1}^{R} \mathcal{A}_j(X)} \tag{3.10}$$

where $\mathcal{A}_i$ is the fuzzy set of i-th rule (VIEIRA; GOMIDE; BALLINI, 2018).

The mechanism of clusters the input space and calculate the global output according to the importance degree improves the precision of the model and better predicts nonlinear data. The next step is to calculate the consequent parameter from Equation (3.11).

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 Equation (3.11) are used to calculate the consequent parameter.

$$\theta_i^k = [\lambda + \kappa(X^k, X^k)]^{-1} y^{k-1} \tag{3.11}$$

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

Otherwise, the consequent parameter is updated using Equation (3.12). When Equation (3.13) 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 (LIU; PRINCIPE; HAYKIN, 2011; RICHARD; BERMUDEZ; HONEINE, 2008).

$$\theta_i^k = \begin{bmatrix} \theta_i^{k-1} - z_i^k [r_i^k]^{-1} \tilde{e}_i^k \\ [r_i^k]^{-1} \tilde{e}_i^k \end{bmatrix} \tag{3.12}$$

$$\min_{\forall d_{ij} \in \mathcal{D}_i^k} \left\| X^k - d_{ij}^k \right\| \geq \delta \tag{3.13}$$

To eliminate redundant rules, the value of compatibility measure, obtained from Equation (3.14), 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 Equation (3.15).

$$\rho_{ij}^k = 1 - \sum_{l=1}^{m} \frac{\left| v_{il}^k - v_{jl}^k \right|}{m} \tag{3.14}$$

The shorter the distance between two rules, the higher the value of ρ. A ρ equal to 1 indicates the maximum compatibility between two rules and the ρ of zero, the minimum compatibility.

$$V_i^k = \frac{V_i^k + V_j^k}{2} \tag{3.15}$$

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; LAMARE, 2011) to limit the increase of the error, reduce the computational complexity and improve the capacity of convergence (AGUIAR et al., 2017). The mechanism to update β performs as shown in Equation (3.16).

$$\beta = \begin{cases} 1 - \frac{\gamma_{bar}}{\left|\tilde{e}^k\right|}, & if \quad \left|\tilde{e}^k\right| > \gamma_{bar} \\ 0, & otherwise \end{cases} \tag{3.16}$$

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

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

There are alternatives proposed calculations of SM. Equation (3.17) demonstrate an alternative calculation of the SM using average errors (AGUIAR et al., 2017; CLARKE; LAMARE, 2011).

$$\beta = \begin{cases} 1 - \frac{\gamma_{bar}}{\sum_{q=1}^{k} \frac{1}{k} |\tilde{e}^q|}, & if \quad \left|\tilde{e}^k\right| > \gamma_{bar} \\ 0, & otherwise \end{cases} \tag{3.17}$$

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 in Equation (3.18).

$$\beta = \begin{cases} \beta + \frac{\left|\tilde{e}^k\right|}{10^{gr} \times \gamma_{bar}}, & if \quad \left|\tilde{e}^k\right| > \gamma_{bar} \\ \beta - \frac{\left|\tilde{e}^k\right|}{10^{dr} \times \gamma_{bar}}, & otherwise \end{cases} \tag{3.18}$$

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 inconsistently values. 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 Figure 2.

Figure 2 – ESM−ePL−KRLS Algorithm

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**

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

      Remove V_j^k

Algorithm 2: ESM

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

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

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

else

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

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

  $\beta = IL$

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

  $\beta = SL$

$\tau = \beta$

$\gamma = 1 - \beta$

Source: Personal collection

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

Figure 3 – Set−Memebership model

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**

  $\beta = 0$

$\tau = \beta$

$\gamma = 1 - \beta$

Source: Personal collection

4 EXPERIMENTAL RESULTS

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, $\gamma_{bar} = 0.01986$ for the SM−ePL−KRLS. And for the ESM−ePL−KRLS, the $IL = 0.01$, $SL = 0.35$, $gr = 1$, $dr = 1$, and $\gamma_{bar} = 0.0205$. The adopted γ_{bar} was chosen as the best result among 1439 simulations starting at $\gamma_{bar} = 0.00001$ and finishing at $\gamma_{bar} = 0.07196$.

Table 2 shows the results of the simulations of the ePL−KRLS and the proposed models. The ESM−ePL−KRLS achieves the lowest errors, indicating it is more precise. The SM−ePL−KRLS obtains the lower errors than ePL−KRLS and is competitive concerning the SM−ePL−KRLS. All the proposed models obtain lower errors than ePL−KRLS.

Table 2 – Results of proposed models and ePL−KRLS

Algorithm	γ_{bar}	RMSE	NDEI	MAE	Rules	Time (s)
ESM−ePL−KRLS	0.0205	**0.0162**	**0.1312**	**0.0120**	2.00	0.4375
SM−ePL−KRLS	0.0199	0.0177	0.1433	0.0124	**1.00**	1.71870
ePL−KRLS	-	0.0247	0.2003	0.0201	2.00	**0.3998**

Source: Personal collection

Table 3 presents the percentage decrease in the errors of the proposed models concerning the errors of the original one. The error that achieved the highest percentage decrease was the MAE error. It indicates that the implementation of the SM and the ESM filtering have more impact on the reduction of the absolute error than in the reduction of the squared error.

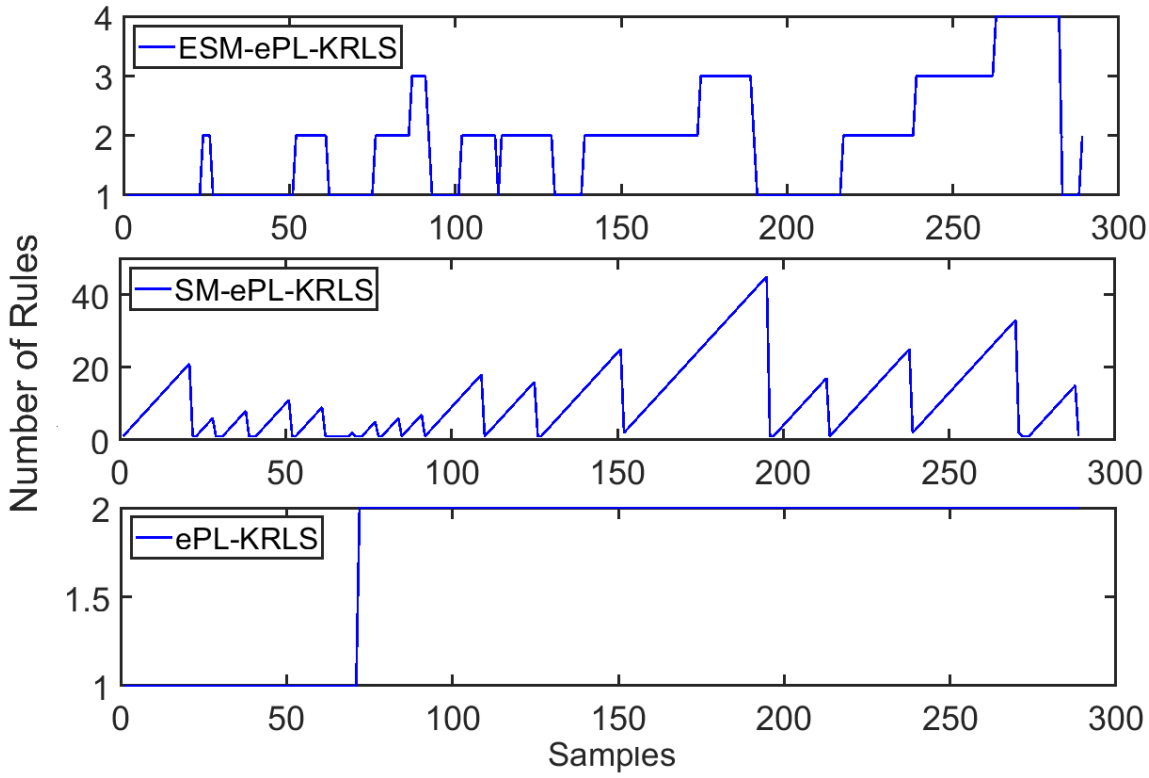
Table 3 – Percentage reduction in the errors of the proposed models

Algorithm	RMSE decrease (%)	NDEI decrease (%)	MAE decrease (%)
ESM−ePL−KRLS	34.41	34.50	40.30
SM−ePL−KRLS	28.34	28.46	38.31
Mean	31.38	31.48	39.31

Source: Personal collection

The ESM−ePL−KRLS has the same number of final rules as ePL−KRLS and the SM−ePL−KRLS has the lowest number of final rules. The processing time of ESM−ePL−KRLS is competitive concerning ePL−KRLS, but the processing time of SM−ePL−KRLS increased considerably. The increase in SM−ePL−KRLS execution time is a consequence of the high number of rules created over the simulation, achieving more than forty rules, suggesting an increase in the computational cost. Figure 4 shows the graphic of the number of rules during simulations for the three models.

Figure 4 – Number of rules of the models

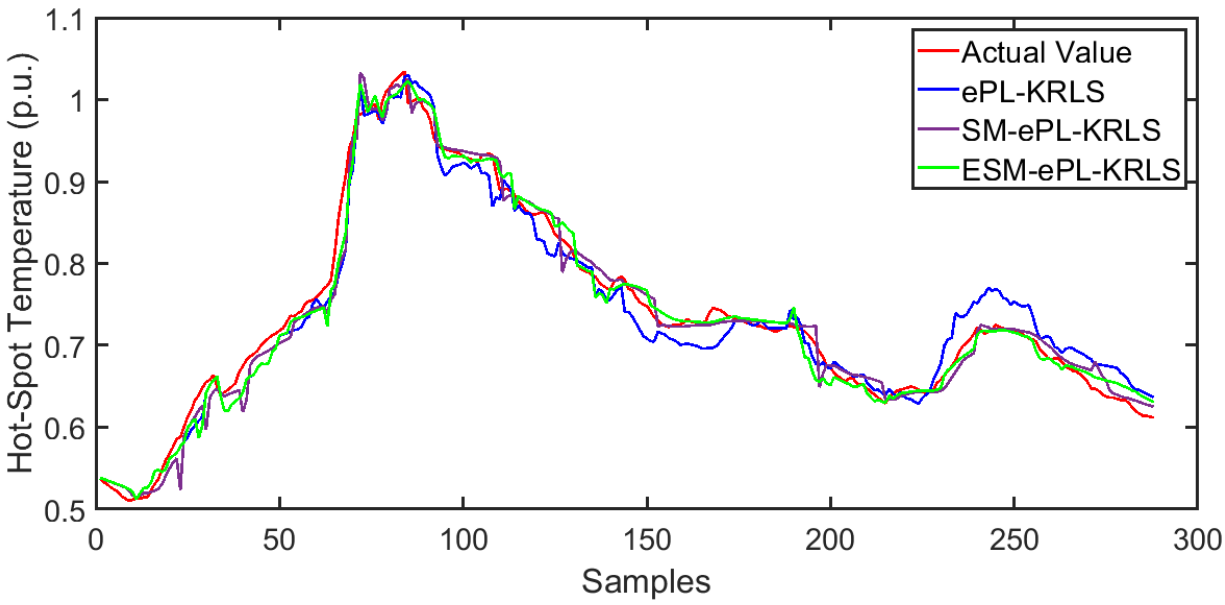


Source: Personal collection

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. It can see in Figure 4 that models introduced in this work are more flexible in the creation and merging of rules. The rate of creation of new rules increases if the error increases more than γ_{bar}. Otherwise, the number of rules tends to decrease.

Figure 5 shows the graphics of the desired value and the predictions. At some points in the graphic, the output of the original model is far from the actual value, while the output from the proposed models remains close to the desired value. It can be noted in samples 150 and 250.

Figure 5 – Desired values and predictions



Source: Personal collection

The Morgan−Granger−Newbold test (MGN) proposed by (DIEBOLD; MARIANO, 1995) validates statistically the performance of the models. The statistical test is performed as Equation (4.1).

$$MGN = \frac{\hat{\rho}_{sd}}{\sqrt{\frac{1-\hat{\rho}_{sd}^2}{n-1}}} \tag{4.1}$$

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. The expression (1) describes the hypothesis test.

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases} \tag{4.2}$$

where μ_1 is the error of population 1 and μ_2 is the error of population 2.

Table 4 presents the results of the tests, 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. Both proposed models, SM−ePL−KRLS and ESM−ePL−KRLS, have better accuracy than ePL−KRLS.

Table 4 – Results of the MGN test

Model 1 x Model 2	MGN	$p-$value
ESM−ePL−KRLS x ePL−KRLS	8.5767	0.0000
SM−ePL−KRLS x ePL−KRLS	6.2966	0.0000

Source: Personal collection

The models discussed in this work are compared with other forecasting models as shown in Table 5. The ESM−ePL−KRLS achieved the lowest errors. And the lowest number of final rules was performed by SM−ePL−KRLS and eMG.

Table 5 – Comparing performance with other forecasting models

Algorithm	RMSE	NDEI	MAE	Rules
ESM−ePL−KRLS	**0.0162**	**0.1312**	**0.0120**	2
SM−ePL−KRLS	0.0177	0.1433	0.0124	**1**
ePL−KRLS	0.0247	0.2003	0.0201	2
evolving Multivariable Gaussian (eMG) (LEMOS; CAMINHAS; GOMIDE, 2010)	0.0220	0.1785	0.0171	**1**
Multilayer Perceptron (MLP) (DUDA; HART; STORK, 2012)	0.0327	0.1278	0.0158	4
extended Tagaki−Sugeno (xTS) (ANGELOV; ZHOU, 2006)	0.0259	0.1015	0.1570	8
Deterministic Model (DM)	0.4005	1.5671	0.3846	-

Source: Amended from (SOUZA et al., 2012)

5 CONCLUSIONS

Two forecasting models are suggested in this work: the SM−ePL−KRLS and ESM−ePL−KRLS. The evaluation of the introduced models is measured in terms of error, the number of final rules and the runtime, using data from an experimental power transformer.

The ESM−ePL−KRLS and SM−ePL−KRLS achieved RMSE, NDEI, and MAE lower than ePL−KRLS, suggesting that those models are more precise. The percentage mean reduction of the proposals were about 31.38% in the RMSE error, 31.48% in the NDEI, and 39.31% in the MAE.

The number of final rules of SM−ePL−KRLS is lower than the ePL−KRLS. However, the number of rules during the simulation presented a considerable variation and reached a high number of rules, indicating a higher computational cost. And consequently, the runtime of the SM−ePL−KRLS increased. The execution time and the number of final rules of ESM−ePL−KRLS is competitive. Another benefit from introduced models is that its structure makes the process of knowledge to be continuous and more adaptable than the ePL−KRLS as the data change.

An MGN test supports that proposed models have better accuracy than ePL−KRLS. The proposed models achieved lower errors and a lower number of final rules than other classical models suggested in the literature.

Since the lifetime of a power transformer is related to its hot−spot temperature, the results presented in this work suggest that the implementation of proposed models to control the hot−spot temperature protect the power transformer of premature failures, improves the use of capacity, reduces costs and guarantees reliable service.

REFERENCES

AGUIAR, E. P. de et al. Set-membership type-1 fuzzy logic system applied to fault classification in a switch machine. *IEEE Transactions on Intelligent Transportation Systems*, IEEE, v. 18, n. 10, p. 2703–2712, 2017.

ALI, E. et al. Power transformer differential protection using current and voltage ratios. *Electric Power Systems Research*, Elsevier, v. 154, p. 140–150, 2018.

ANGELOV, P.; ZHOU, X. Evolving fuzzy systems from data streams in real-time. In: IEEE. *2006 International Symposium on Evolving Fuzzy Systems*. [S.l.], 2006. p. 29–35.

ANGELOV, P. P.; FILEV, D. P. An approach to online identification of takagi-sugeno fuzzy models. *IEEE Transactions on Systems, Man, and Cybernetics, Part B - Cybernetics*, IEEE, v. 34, n. 1, p. 484–498, 2004.

ARABUL, A. Y.; SENOL, I. Development of a hot-spot temperature calculation method for the loss of life estimation of an onan distribution transformer. *Electrical Engineering*, Springer, v. 100, n. 3, p. 1651–1659, 2018.

BACHA, K.; SOUAHLIA, S.; GOSSA, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electric Power Systems Research*, Elsevier, v. 83, n. 1, p. 73–79, 2012.

BENGTSSON, C. Status and trends in transformer monitoring. *IEEE Transactions on Power Delivery*, IEEE, v. 11, n. 3, p. 1379–1384, 1996.

BÉRUBÉ, J.; AUBIN, J.; MCDERMID, W. Transformer winding hot spot-temperature determination. In: *Fifth Annual Weidmann ACTI Technical Conf.* [S.l.: s.n.], 2006. p. 1–10.

BOARD, I. Ieee guide for loading mineral-oil-immersed transformers. *IEEE Std C*, v. 57, p. 1–112, 1995.

CASTRO, A. R. G.; MIRANDA, V. An interpretation of neural networks as inference engines with application to transformer failure diagnosis. *International Journal of Electrical Power & Energy Systems*, Elsevier, v. 27, n. 9-10, p. 620–626, 2005.

CHRISTINA, A. et al. Causes of transformer failures and diagnostic methods–a review. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 82, p. 1442–1456, 2018.

CLARKE, P.; LAMARE, R. C. de. Low-complexity reduced-rank linear interference suppression based on set-membership joint iterative optimization for ds-cdma systems. *IEEE Transactions on Vehicular Technology*, IEEE, v. 60, n. 9, p. 4324–4337, 2011.

DIEBOLD, F. X.; MARIANO, R. S. Com paring predictive accu racy. *Journal of Business and Economic Statistics*, v. 13, n. 3, p. 253–263, 1995.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern classification*. [S.l.]: John Wiley & Sons, 2012.

GALDI, V. et al. Neural diagnostic system for transformer thermal overload protection. *IEE Proceedings - Electric Power Applications*, IET, v. 147, n. 5, p. 415–421, 2000.

GANYUN, L. et al. Fault diagnosis of power transformer based on multi-layer svm classifier. *Electric Power Systems Research*, Elsevier, v. 74, n. 1, p. 1–7, 2005.

HE, Q.; SI, J.; TYLAVSKY, D. J. Prediction of top-oil temperature for transformers using neural networks. *IEEE Transactions on Power Delivery*, IEEE, v. 15, n. 4, p. 1205–1211, 2000.

HELL, M.; COSTA, P.; GOMIDE, F. Recurrent neurofuzzy network in thermal modeling of power transformers. *IEEE Transactions on Power Delivery*, IEEE, v. 22, n. 2, p. 904–910, 2007.

HELL, M.; COSTA, P.; GOMIDE, F. Participatory learning in power transformers thermal modeling. *IEEE Transactions on Power Delivery*, IEEE, v. 23, n. 4, p. 2058–2067, 2008.

JAN, S. T.; AFZAL, R.; KHAN, A. Z. Transformer failures, causes & impact. In: *International Conference Data Mining, Civil and Mechanical Engineering*. [S.l.: s.n.], 2015. p. 49–52.

KASABOV, N.; FILEV, D. Evolving intelligent systems: methods, learning, & applications. In: IEEE. *2006 International Symposium on Evolving Fuzzy Systems*. [S.l.], 2006. p. 8–18.

LEMOS, A.; CAMINHAS, W.; GOMIDE, F. Multivariable gaussian evolving fuzzy modeling system. *IEEE Transactions on Fuzzy Systems*, IEEE, v. 19, n. 1, p. 91–104, 2010.

LI, Y.; WANG, Y.; JIANG, T. Sparse-aware set-membership nlms algorithms and their application for sparse channel estimation and echo cancelation. *AEU-International Journal of Electronics and Communications*, Elsevier, v. 70, n. 7, p. 895–902, 2016.

LIMA, E. et al. Evolving fuzzy modeling using participatory learning. *Evolving Intelligent Systems: Methodology and Applications*, Wiley Online Library, p. 67–86, 2010.

LIU, W.; PRINCIPE, J. C.; HAYKIN, S. *Kernel adaptive filtering: a comprehensive introduction*. [S.l.]: John Wiley & Sons, 2011. v. 57.

METWALLY, I. A. Failures, monitoring and new trends of power transformers. *IEEE Potentials*, IEEE, v. 30, n. 3, p. 36–43, 2011.

MIJAILOVIC, V. Method for effects evaluation of some forms of power transformers preventive maintenance. *Electric Power Systems Research*, Elsevier, v. 78, n. 5, p. 765–776, 2008.

NGAOPITAKKUL, A.; KUNAKORN, A. Internal fault classification in transformer windings using combination of discrete wavelet transforms and back-propagation neural networks. *International Journal of Control, Automation, and Systems*, v. 4, n. 3, p. 365–371, 2006.

NGIA, L. S.; SJÖBERG, J.; VIBERG, M. Adaptive neural nets filter using a recursive levenberg-marquadt search direction. In: CITESEER. *Asilomar Conference on Signals, Systems, and Computers*. [S.l.], 1998. v. 1, p. 697–701.

RADAKOVIC, Z.; FESER, K. A new method for the calculation of the hot-spot temperature in power transformers with onan cooling. *IEEE Transactions on Power Delivery*, IEEE, v. 18, n. 4, p. 1284–1292, 2003.

RICHARD, C.; BERMUDEZ, J. C. M.; HONEINE, P. Online prediction of time series data with kernels. *IEEE Transactions on Signal Processing*, IEEE, v. 57, n. 3, p. 1058–1067, 2008.

ROIZMAN, O.; DAVYDOV, V. Neuro-fuzzy computing for large power transformers monitoring and diagnostics. In: IEEE. *18th International Conference of the North American Fuzzy Information Processing Society - NAFIPS (Cat. No. 99TH8397)*. [S.l.], 1999. p. 248–252.

SCHOLKOPF, B.; SMOLA, A. J. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. [S.l.]: MIT press, 2001.

SICA, F. C. et al. A cognitive system for fault prognosis in power transformers. *Electric Power Systems Research*, Elsevier, v. 127, p. 109–117, 2015.

SOUZA, L. et al. Thermal modeling of power transformers using evolving fuzzy systems. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 25, n. 5, p. 980–988, 2012.

TAKAGI, T.; SUGENO, M. Fuzzy identification of systems and its applications to modeling and control. *IEEE transactions on systems, man, and cybernetics*, IEEE, n. 1, p. 116–132, 1985.

VIEIRA, R.; GOMIDE, F.; BALLINI, R. Kernel evolving participatory fuzzy modeling for time series forecasting. In: IEEE. *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. [S.l.], 2018. p. 1–9.

WANG, D. et al. Theory and application of distribution electronic power transformer. *Electric Power Systems Research*, Elsevier, v. 77, n. 3-4, p. 219–226, 2007.

YAGER, R. R. A model of participatory learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, IEEE, v. 20, n. 5, p. 1229–1234, 1990.

ŽARKOVIĆ, M.; STOJKOVIĆ, Z. Analysis of artificial intelligence expert systems for power transformer condition monitoring and diagnostics. *Electric Power Systems Research*, Elsevier, v. 149, p. 125–136, 2017.

ANEXO A – TERMO DE AUTENTICIDADE

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA

Termo de Declaração de Autenticidade de Autoria

Declaro, sob as penas da lei e para os devidos fins, junto à Universidade Federal de Juiz de Fora, que meu Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Produção é original, de minha única e exclusiva autoria. E não se trata de cópia integral ou parcial de textos e trabalhos de autoria de outrem, seja em formato de papel, eletrônico, digital, áudio-visual ou qualquer outro meio.
Declaro ainda ter total conhecimento e compreensão do que é considerado plágio, não apenas a cópia integral do trabalho, mas também de parte dele, inclusive de artigos e/ou parágrafos, sem citação do autor ou de sua fonte.
Declaro, por fim, ter total conhecimento e compreensão das punições decorrentes da prática de plágio, através das sanções civis previstas na lei do direito autoral[1] e criminais previstas no Código Penal[2], além das cominações administrativas e acadêmicas que poderão resultar em reprovação no Trabalho de Conclusão de Curso.

Juiz de Fora, 14 de novembro de 2019.

Kaike Sa Teles Rocha Alves	201249048
NOME LEGÍVEL DO ALUNO (A)	Matrícula
Kaike Sa Teles Rocha Alves	115.857.626-95
ASSINATURA	CPF

[1] LEI N° 9.610, DE 19 DE FEVEREIRO DE 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências.
[2] Art. 184. Violar direitos de autor e os que lhe são conexos: Pena - detenção, de 3 (três) meses a 1 (um) ano, ou multa.