A New Fuzzy Inference System Applied to Time Series Forecasting

Kaike Sa Teles Rocha Alves
Graduate Program in Computational Modeling
Federal University of Juiz de Fora
Juiz de Fora, Brazil
kaike.alves@engenharia.ufjf.br

Rosangela Ballini
Institute of Economics
University of Campinas
Campinas, Brazil
ballini@unicamp.br

Eduardo Pestana de Aguiar
Department of Industrial and Mechanical Engineering
Federal University of Juiz de Fora
Juiz de Fora, Brazil
eduardo.aguiar@engenharia.ufjf.br

***Abstract*—Fuzzy systems are a class of machine learning introduced by Zadeh that combine accuracy and interpretability. This class of models consists of two main parts, the antecedent and the consequent. While the antecedent is responsible for modeling the inputs, the consequent concerns modeling the output. The literature reports two main types of fuzzy systems: Mamadani and Takagi-Sugeno. While Mamdani uses fuzzy sets in the consequent part, Takagi-Sugeno uses polynomial functions. Consequently, Mamdani provides better understandable models and Takagi-Sugeno more accurate ones. In this paper, we propose a new Takagi-Sugeno model. Still, instead of defining the rules based on the input, the proposed model designs the rules based on the output variation to capture linearities in the output and clusters them in the same rule. The model is applied in the regression problems of benchmark series and real datasets of power transformers. The performance of the proposed model is compared with the performance of classical models and evolving Fuzzy Systems. The results are evaluated using error metrics, the number of final rules, and runtime.**

***Index Terms*—Fuzzy Inference System, Takagi-Sugeno, time series forecasting, thermal modeling of power transformers**

I. Introduction

Zadeh [1], [2] inspired a new class of machine learning models known as fuzzy inference systems (FIS). FIS are well-known for their capability to provide accurate results that can be easily interpreted by the user. Such models have a basic structure that consists of an antecedent and a consequent part. While the antecedent part regards the modeling of the input, the consequent concerns the computation of the output. Two main types of FIS are found in the literature: Mamdani fuzzy inference systems (MFIS) and Takagi-Sugeno (TS). The only difference between the two models is that in the consequent part, MFIS uses fuzzy sets, so MFIS results tend to be more understandable [3]. On the other hand, TS models are based on gain scheduling, meaning that a collection of local linear approximations models a nonlinear system. Each region is fuzzily designed with a set of linear parameters. Furthermore, TS approaches can model a nonlinear system using a collection of local linear approximations called fuzzy regions. The final output is a weighted average of the outputs of each rule. Consequently, the TS models usually present computationally efficient and accurate solutions to a wide range of control problems [4].

Another advantage of using TS models is that they require just a few rules to describe highly nonlinear and complex systems. TS usually presents fewer rules than Mandani-based models. Unfortunately, compared to Mandani, TS is less understandable [5]. In problems where interpretability is important, TS models are not suggested as they use local model parameters identified from the data and, consequently, there's no guarantee of interpretability [6]. There is no unique method to design the TS fuzzy rules, but they all build the rule-base structure based on the input vector. Furthermore, many hybrid models are proposed to form improved rules using optimization algorithms, such as simulated annealing (SA), particle swarm optimization (PSO), and gravitational search algorithm (GSA) [7], [8]; reinforcement learning (RL) [9].

The evolving Takagi-Sugeno (eTS) model was proposed by Angelov and Filev [4]. The eTS approach is an evolving Fuzzy System(eFS), a class with its parameters and structure continually updated as new input/output pair enters the system, creating and excluding rules when necessary. The eTS algorithm inspired the proposal of other rule-based eFS approaches, such as Simpl_eTS [10], exTS [11], ePL [12], ePL+ [13], eMG [14], ePL-KRLS-DISCO [15], FLEXFIS [16].

In all those models, the rules are managed based on a distance metric or compatibility measure. When a new input vector enters the system, the model compares the input with all existing rules, and if the input is considered compatible with the current structure, the number of rules remains the same. Otherwise, the model will create a new rule. Consequently, the model's performance will be directly dependent on the metric of compatibility chosen. However, those metrics are applied to the input vector to form the rules. It implies that the input vectors will be close or similar but doesn't guarantee that the rules will have outputs with linear curves. Furthermore, the models usually have many hyperparameters [15].

To overcome such limitations, we propose a new model, called Semi-evolving Output Based model (SeOB), that clusters the input space based on the forecasted value and with just one hyperparameter. The proposed approach is called Semi evolving because it inherits some characteristics of eFSs, such as self-learning of parameters and adaptive consequent parameters, but it is not possible to update the rule-based structure. The proposed model defines the rules by estimating the variation between the actual and previous values so that we will have clusters with curves close to a linear distribution. The model is evaluated using three benchmark time series, the Mackey-Glass, nonlinear system identification, and the Lorenz Attractor. Finally, the model is applied to the datasets of a power transformer to predict the hot spot temperature. The performance is estimated regarding RMSE, NDEI, and MAE errors, the number of final rules, and the runtime.

The remainder of the paper is organized as follows: Section II describes the proposed model. Section III presents and discusses the results. Finally, Section IV

II. The Proposed Model

This paper proposes a new regression algorithm based on Takagi-Sugeno (TS) composed of an antecedent and consequent part. The antecedent part consists of the definition of the model's rules. On the other hand, the consequent part corresponds to the recursive update of consequent parameters. The remainder of the section explains in detail how the algorithm works.

A. The Antecedent Part

The antecedent part defines how the model forms the rules. This paper proposes a new mechanism to create the rules using the outputs instead of the inputs. When the model receives the data, the model computes the value of the consecutive outputs according to Equation (1).

$$\Delta y^k = y^k - y^{k-1} \tag{1}$$

where Δy^k represents the difference between two consecutive outputs, for $k = [2, 3, \ldots, n]$, n is the number of samples, y^k is the desired output for the kth sample, and y^{k-1} is the desired output for the previous sample.

This model is adequate for data streams since it computes the difference between two consecutive samples to form the rules. After calculating the Δy for all data samples, the model computes the intervals according to Equation (2).

$$LI = \frac{\max\{\Delta y^k\} - \min\{\Delta y^k\}}{R_{max}} \tag{2}$$

where LI is the length of each interval and R_{max} is an integer value greater or equal to 1, representing the only hyperparameter of the model. R_{max} defines the number of rules the model will create, i.e., if the user defines $R_{max} = 3$, the model will create three rules. R_{max} must be set considering the trade-off accuracy interpretability.

After calculating the length of each interval, the model labels the samples with the number of the rule that they will be included, as follows:

$$R^k = int\left(\frac{\Delta y^k \ - \min_k \{\Delta y^k\}}{LI}\right) \tag{3}$$

where R^k represents the rule that the kth sample pertains to. Furthermore, $LI \neq 0 \Leftrightarrow \max \neq \min$.

After those calculations, the model computes the mean and the standard deviation of the samples that pertain to the same rule and starts computing recursively the consequent parameters for each rule.

B. The Consequent Part

The parameters of the consequent part are estimated using the weighted Recursive Least Squares (wRLS). First, the model computes the normalized firing level of the ith rules as follows:

$$\lambda_i = \frac{f^i}{\sum_{j=1}^{R_{max}} f^j} \tag{4}$$

where f^i is the firing level of the ith rule, calculated according to Equation (5)

$$f^i = \prod_{l=1}^{p} \exp\left[\frac{1}{2}\frac{\left(v_{i,l} - x_l^k\right)^2}{\sigma_{i,l}^2}\right] \tag{5}$$

where $v_{i,l}$ is the lth element of the vector v_i (rule center) for the ith rule that represents the mean value of the attribute, $v_i = [v_{i,1}, \ldots, v_{i,p}]^T$, p is the dimension of the inputs, x_l^k is the lth attribute of the input vector (X^k) for iteration k, and $\sigma_{i,l}$ is the lth element of the vector σ_i for the ith rule that represents the standard-deviation of the attribute, for $\sigma_i = [\sigma_{i,1}, \ldots, \sigma_{i,p}]^T$.

Then, the model can estimate the consequent parameters as follows:

$$\theta_i^k = \theta_i^{k-1} + P_i^k x_e^k \lambda_i \left(y^k - (x_e^k)^T\theta\right) \tag{6}$$

where P_i^k is the $R^{p+1} \times R^{p+1}$ covariance matrix, calculated according to the Equation (7), and $x_e^k = [1, (x^k)^T]^T$.

$$P_i^k = P_i^{k-1} - \frac{\lambda_i P_i^{k-1} x_e^k (x_e^k)^T P_i^{k-1}}{1 + \lambda_i (x_e^k)^T P_i^{k-1} x_e^k} \tag{7}$$

where P_i is initialized with $1000 \times I_{(p+1)\times(p+1)}$, and $I_{(p+1)\times(p+1)}$ is the identity matrix $R^{p+1} \times R^{p+1}$.

Finally, the model's output is calculated as follows:

$$\hat{y}^k = \sum_{i=1}^{R_{max}} \lambda_i (x_e^k)^T \theta_i \tag{8}$$

III. Experimental Results

Three benchmark series are implemented to evaluate the performance of the proposed model: the Mackey-Glass, nonlinear system identification, and Lorenz Attractor, widely used in the literature. Finally, the model is applied to a real-world dataset from a power transformer. The root-mean-square error (RMSE), non-dimensional index error (NDEI), and mean absolute error (MAE) measures the error of the models, calculated according to Equations (9), (10), and (11), respectively.

$$RMSE = \sqrt{\frac{1}{T}\sum_{k=1}^{T}(y^k - \hat{y}^k)^2} \tag{9}$$

$$NDEI = \frac{RMSE}{std([y^1, ..., y^T])} \tag{10}$$

$$MAE = \frac{1}{T}\sum_{k=1}^{T}|y^k - \hat{y}^k| \tag{11}$$

where y^k is the kth actual value, $\hat{y}^k$ is the kth predicted value, T is the sample size, and $std()$ is the standard deviation function. The number of final rules is also presented. The hyperparameters are heuristically defined by computational experiments aiming to produce the lowest RMSE, NDEI, and MAE. Furthermore, the models' execution time is estimated in seconds, computing the mean runtime and the standard deviation of thirty simulations. All codes were executed using Python 3.9 in a PC device that has Intel Core i7-8565U, 1.99 GHz Turbo, and 8 GB RAM.

A. Lorenz Attractor Time Series

Lorenz [17] introduced a multivariate time series composed of three ordinary differential equations, known as the Lorenz Attractor. The expressions are described in Equations (12), (13), and (14).

$$\frac{dx}{dt} = \sigma(y - x) \tag{12}$$

$$\frac{dy}{dt} = x(\rho - z) - y \tag{13}$$

$$\frac{dz}{dt} = xy - \beta z \tag{14}$$

The parameters of Lorenz's time series were defined as $\sigma = 10$, $\beta = 2.667$, and $\rho = 28$, with initial conditions $x(0) = 0$, $y(0) = 1$, and $z(0) = 1.05$, to obtain a chaotic behavior. 10000 data samples were generated. The goal is to predict x^{k+1} using as input vector $[x^k, y^k, z^k]$ for any k value. The k value is set as $k \in [1, 10000]$, where the first 8000 data samples trained the models, and the last 2000 data samples tested them.

Table I shows the simulations' results. The proposed model obtained the best results among all the models concerning the errors. It also got one of the lowest number of rules. The eMG achieved the highest number of final rules and the highest error values. Not only SeOB achieved the lowest errors but also the lowest runtime among all models. Figure 1 depicts the graphic of the predictions of the proposed model.

TABLE I: Simulations' results of Lorenz Attractor time series

Model	RMSE	NDEI	MAE	Rules	Runtime (s)
ARIMA [18]	0.0005543	0.0000677	0.0004575	-	-
eTS [4]	0.0002608	0.0000319	0.0002108	4	12.65 ± 0.66
Simpl_eTS [10]	0.0001561	0.0000191	0.0001103	**1**	5.75 ± 0.18
exTS [11]	0.0001467	0.0000179	0.0000304	7	23.01 ± 0.45
eMG [14]	0.5598119	0.0683843	0.3872295	56	225.88 ± 1.72
ePL+ [13]	0.00000156	0.0000019	0.0000110	**1**	9.52 ± 0.33
ePL-KRLS-DISCO [15]	0.0294889	0.0036022	0.0191308	17	277.43 ± 3.50
SeOB	$\mathbf{4.7 \times 10^{-8}}$	$\mathbf{5.8 \times 10^{-9}}$	$\mathbf{3.6 \times 10^{-8}}$	**1**	$\mathbf{3.89 \pm 0.19}$

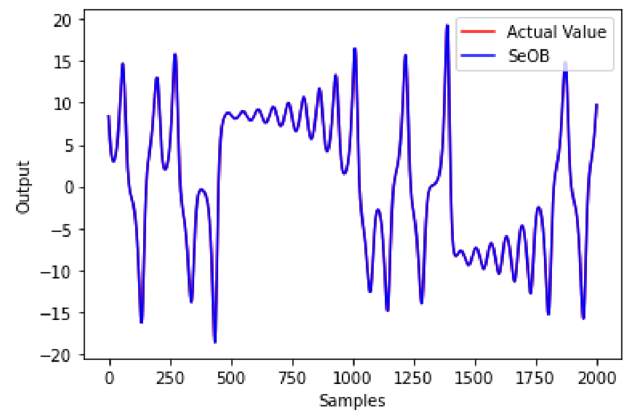

Fig. 1: Predictions of the proposed model for the Lorenz Attractor time series

B. Mackey–Glass Time Series Forecasting

Mackey and Glass [19] introduced a long-term time series, proposed as a model of white blood cell production, obtained through the following differential equation:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t-1) \tag{15}$$

where $x(0) = 1.2$ and $\tau = 17$.

The goal is to predict x^{k+85} using as input vector $[x^k, x^{k+6}, x^{k+12}, x^{k+18}]$ for any k value. The simulations were trained using 3000 data samples for $k \in [201, 3200]$, and then, 500 data samples were collected to test the model for $k \in [5001, 5500]$.

Table II shows the models' results. The PL-KRLS-DISCO model obtained the best error values, followed by the ePL-KRLS, then the eMG, with 19, 21, and 40 final rules, respectively. SeOB achieved lower errors than eTS and Simpl_eTS. Furthermore, SeOB performed simulations with the second-best runtime. The eTS model performed the best runtime. Figure 2 shows the graphic with the SeOB predictions.

TABLE II: Simulations' results of Mackey-Glass time series

Model	RMSE	NDEI	MAE	Rules	Runtime (s)
ARIMA [18]	0.1079377	0.4812567	0.0851530	-	-
eTS [4]	0.0835134	0.3723572	0.0676656	4	**4.92 ± 0.43**
Simpl_eTS [10]	0.0850095	0.3790278	0.0683714	**3**	8.80 ± 0.27
exTS [11]	0.0778960	0.3473112	0.0629018	6	8.06 ± 0.32
eMG [14]	0.0583901	0.2603414	0.0307220	40	76.30 ± 0.90
ePL+ [13]	0.0753687	0.3360384	0.0595750	25	65.17 ± 3.59
ePL-KRLS-DISCO [15]	**0.0025689**	**0.0114539**	**0.0014560**	19	52.23 ± 1.21
SeOB	0.0778081	0.3469191	0.0626661	8	6.40 ± 0.38

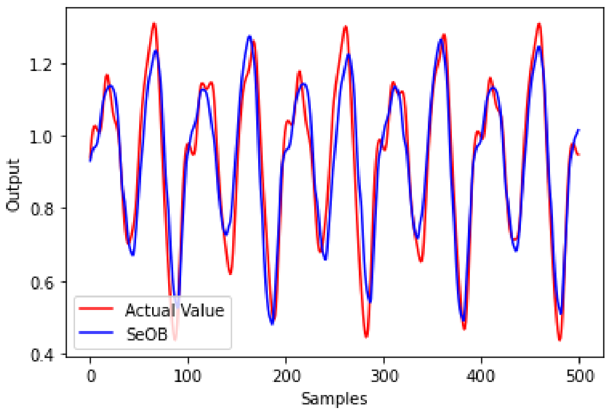

Fig. 2: Predictions of the proposed model for the Mackey–Glass time series

C. Nonlinear Dynamic System Identification

The classic nonlinear dynamic system identification, described in [20], is obtained through the following difference equation [21]:

$$y^k = \frac{y^{k-1}y^{k-2}(y^{k-1}-0.5)}{1+(y^{k-1})^2+(y^{k-2})^2} - u^{k-1} \tag{16}$$

where $u^k = sin(\frac{2\pi k}{25})$, and $y^0 = y^1 = 0$.

The goal is to predict y^k using as input vector $[y^{k-2}, y^{k-1}, u^{k-1}]$ for any k value. The k value is set as $k \in [2, 5201]$, where the first 5000 data samples trained the models, and the last 200 data samples tested them.

Table III shows the results of the simulations. The proposed model achieved lower errors than ARIMA with 8 rules. These results show that, although the model presents a simple structure, it can get accurate results. The ePL-KRLS-DISCO algorithm reached the best error results with 20 final rules. The eMG model performed the second-lowest errors with 25 final rules. SeOB performed the third-best runtime. The eTS model obtained the lowest runtime among the models. Figure 3 shows the graphic of the predictions of ePL-KRLS-DISCO.

D. Hot Spot Temperature Forecasting

And finally, the model is applied in the thermal modeling of the power transformer. Table IV presents the characteristics of the power transformer.

The aim is to predict the hot spot temperature using as inputs 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). Three datasets were collected for the simulations. Each one consists of measurements taken every 5 minutes for 24 hours. The first dataset corresponds to the first day of measurements. The second dataset corresponds to the second day and has no overload conditions. The last dataset has overload conditions and corresponds to the last day of measurements. The first dataset was implemented to train the models, and datasets 2 and 3 were implemented to test the model.

Table V shows the performance of the models for dataset 2. The proposed model obtained the lowest errors and just three final rules. The ePL+ performed the lowest number of final rules. The eMG model achieved the highest number of final rules. Table VI presents the simulation results in the presence of overload conditions. One can note that the proposed model also achieved the best error values. Furthermore, although the ePL+ performed the second-best error values in dataset 2, it obtained sixth in dataset 3. The eMG and ePL achieved the lowest number of final rules, and the Simp_eTS obtained the highest. SeOB achieved the best results with just 2 final rules.

TABLE III: Simulations' results of nonlinear time series

Model	RMSE	NDEI	MAE	Rules	Runtime (s)
ARIMA [18]	0.0453301	0.0413639	0.030261	-	-
eTS [4]	0.0082345	0.0075140	0.0048862	5	9.12 ± 0.46
Simpl_eTS [10]	0.0021659	0.0019764	0.0014770	19	30.01 ± 0.78
exTS [11]	0.0178349	0.0162744	0.0144450	**4**	9.25 ± 0.46
eMG [14]	1.2×10^{-7}	1.1×10^{-7}	3.7×10^{-8}	25	80.73 ± 0.25
ePL+ [13]	0.0216558	0.0197610	0.0124800	31	194.47 ± 1.19
ePL-KRLS-DISCO [15]	$\mathbf{6.0 \times 10^{-8}}$	$\mathbf{5.50 \times 10^{-8}}$	$\mathbf{1.3 \times 10^{-8}}$	20	46.43 ± 0.65
SeOB	0.0364501	0.0332608	0.0260871	8	10.12 ± 0.26

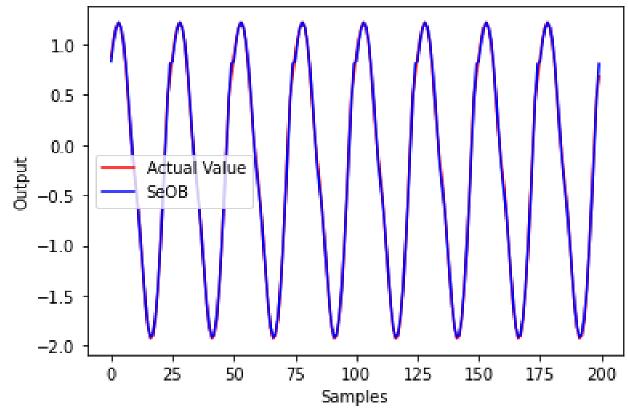

Fig. 3: Predictions of the proposed model for the nonlinear dynamic system identification

TABLE IV: Characteristics of the power 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

SeOB performed simulations with the lowest runtime in both datasets.

TABLE V: Simulations' results of the power transformer - dataset 2

Model	RMSE	NDEI	MAE	Rules	Runtime (s)
ARIMA [18]	0.3201851	5.0613831	0.3015114	-	-
eTS [4]	0.0236377	0.3736580	0.0164963	4	0.57 ± 0.04
Simpl_eTS [10]	0.0184061	0.2909584	0.0144366	15	2.06 ± 0.11
exTS [11]	0.0208743	0.3299738	0.0147444	6	0.92 ± 0.06
eMG [14]	0.0207822	0.3285185	0.0138175	19	2.67 ± 0.14
ePL+ [13]	0.0162218	0.2564288	0.0132873	**1**	0.40 ± 0.03
ePL-KRLS-DISCO [15]	0.0233095	0.3684689	0.0179699	2	2.70 ± 0.13
SeOB	**0.0095076**	**0.1502933**	**0.0061835**	2	**0.27 ± 0.00**

TABLE VI: Simulations' results of the power transformer - dataset 3

Model	RMSE	NDEI	MAE	Rules	Runtime (s)
ARIMA [18]	0.6066052	2.9606032	0.5587693	-	-
eTS [4]	0.1810075	0.8834273	0.1257706	4	0.57 ± 0.06
Simpl_eTS [10]	0.1630808	0.7959338	0.1196723	15	2.06 ± 0.11
exTS [11]	0.1668536	0.8143476	0.1224999	5	0.81 ± 0.01
eMG [14]	0.1699642	0.8295290	0.1267486	**1**	2.28 ± 0.00
ePL+ [13]	0.1692664	0.8261234	0.1223584	**1**	0.41 ± 0.03
ePL-KRLS-DISCO [15]	0.0933427	0.4555695	0.0736821	3	2.34 ± 0.20
SeOB	**0.0443827**	**0.2166146**	**0.0251311**	2	**0.28 ± 0.2**

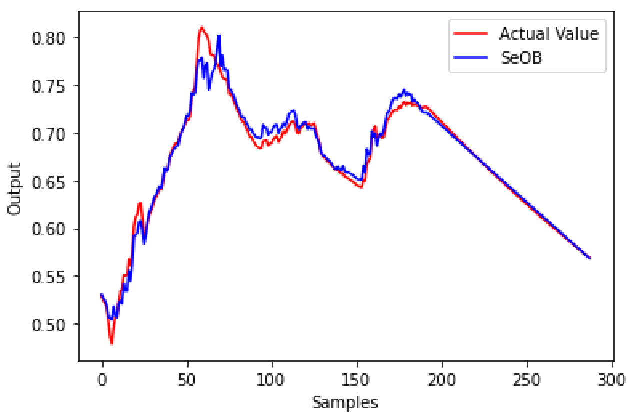

Fig. 4: Estimation of hot spot temperature without overload condition

E. Discussion

In this paper, a new model is proposed to time series forecasting. The model is applied to three benchmark series and thermal modeling from a real power transformer. Among the benchmark datasets, the proposed model achieves the lowest errors and runtime in the Lorenz Attractor, a dataset well-known as chaotic. In this series, the model uses exogenous variables. Consequently, the results suggest that the proposed model is suitable to be applied in time series forecasting of high chaotic series using as predictors the exogenous variables. Furthermore, as the model presents a simple clustering structure, and in Lorenz's example, it created just one rule, the model obtains the lowest execution time in this simulation.

On the other hand, concerning the Mackey-Glass time series, the proposed model obtained the fourth-lowest errors

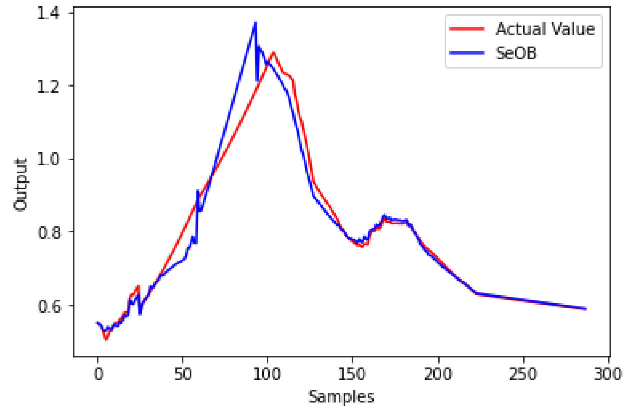

Fig. 5: Estimation of hot spot temperature with overload condition

and the second-lowest runtime. Mackey-Glass is a series with chaotic behavior that uses as predictors the same delayed series. In this case, the best results were achieved by ePL-KRLS-DISCO, with errors approximately 97% lower than SeOB but with a runtime approximately 8 times higher. Moreover, the second-best model concerning the errors achieves error values approximately 25% lower than SeOB with runtime about 12 times higher. The last benchmark series is called nonlinear. It doesn't present a chaotic behavior but simulates a non-linear series. The proposed model performed the second-highest errors but the third-lowest runtime and number of final rules. The best performance in the nonlinear series is obtained by ePL-KRLS-DISCO, followed by the eMG model.

Finally, SeOB achieved the best error values and runtime to predict the hot-spot temperature of power transformers, indicating that the proposed model is suitable to be implemented in control areas, presenting faster and more reliable outputs. Still, the simplicity of the model's structure makes it a light and fast algorithm required in many control areas. Another advantage of the proposed model is the number of hyper-parameters. SeOB needs just one hyper-parameter to be set: the number of rules. The number of hyper-parameters is an issue in many machine learning models, as it demands more time to find the best combination of hyper-parameters that leads to the best results possible.

IV. Conclusions

This paper proposes a new machine learning model for time series forecasting. The Takagi-Sugeno inspires the introduced model but introduces new concepts that make it robust and can be applied in many control areas. One of the main advantages of this model is that it has just one hyperparameter, making it also suitable for people starting in machine learning. The model was tested using the Lorenz Attractor, Mackey-Glass, nonlinear time series, and datasets of a real-world power transformer. The model obtained satisfactory results in the Mackey-Glass and nonlinear and achieved the lowest errors in the Lorenz Attractor and the power transformers' datasets.

As the demand for mechanisms that can guarantee the safe operation of critical systems has increased, the model proposed in this paper is indicated to be implemented as a control tool for its reliability and simplicity. The correct operation of these systems helps to prevent unexpected failures due to ageing or deterioration and, consequently, reduces costs. In future works, we suggest testing SeOB performance in the presence of noise.

Acknowledgment

The authors acknowledge the Federal University of Juiz de Fora for essential support during this work and the anonymous referees for their valuable comments. The authors also thank the financial support of CNPq (grant 433389/2018-4), FAPEMIG (grant APQ-02922-18), and CAPES - Finance Code 001.

References

[1] L. ZADEH, "Fuzzy algorithms," *Information and Control*, vol. 12, pp. 94–102, 1968.

[2] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 1, pp. 28–44, 1973.

[3] V. Ojha, A. Abraham, and V. Snášel, "Heuristic design of fuzzy inference systems: A review of three decades of research," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 845–864, 2019.

[4] P. P. Angelov and D. P. Filev, "An approach to online identification of takagi-sugeno fuzzy models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 484–498, 2004.

[5] D. Kukolj and E. Levi, "Identification of complex systems based on neural and takagi-sugeno fuzzy model," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 272–282, 2004.

[6] T. A. Johansen, R. Shorten, and R. Murray-Smith, "On the interpretation and identification of dynamic takagi-sugeno fuzzy models," *IEEE Transactions on Fuzzy systems*, vol. 8, no. 3, pp. 297–313, 2000.

[7] S. Vrkalovic, T.-A. Teban, and I.-D. Borlea, "Stable takagi-sugeno fuzzy control designed by optimization," *International Journal of Artificial Intelligence*, vol. 15, no. 2, pp. 17–29, 2017.

[8] S.-H. Tsai and Y.-W. Chen, "A novel identification method for takagi–sugeno fuzzy model," *Fuzzy Sets and Systems*, vol. 338, pp. 117–135, 2018.

[9] E. Zander, B. van Oostendorp, and B. Bede, "Reinforcement learning with Takagi-Sugeno-Kang fuzzy systems," *Complex Engineering Systems*, vol. 3, no. 2, 2023.

[10] P. Angelov and D. Filev, "Simpl_ets: A simplified method for learning evolving takagi-sugeno fuzzy models," in *The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ'05.*, pp. 1068–1073, IEEE, 2005.

[11] P. Angelov and X. Zhou, "Evolving fuzzy systems from data streams in real-time," in *2006 International Symposium on Evolving Fuzzy Systems*, pp. 29–35, IEEE, 2006.

[12] E. Lima, M. Hell, R. Ballini, and F. Gomide, "Evolving fuzzy modeling using participatory learning," *Evolving Intelligent Systems: Methodology and Applications*, pp. 67–86, 2010.

[13] L. Maciel, F. Gomide, and R. Ballini, "An enhanced approach for evolving participatory learning fuzzy modeling," in *2012 IEEE Conference on Evolving and Adaptive Intelligent Systems*, pp. 23–28, IEEE, 2012.

[14] A. Lemos, W. Caminhas, and F. Gomide, "Multivariable gaussian evolving fuzzy modeling system," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 91–104, 2010.

[15] K. S. T. R. Alves and E. P. de Aguiar, "A novel rule-based evolving fuzzy system applied to the thermal modeling of power transformers," *Applied Soft Computing*, vol. 112, p. 107764, 2021.

[16] E. D. Lughofer, "Flexfis: A robust incremental learning approach for evolving takagi–sugeno fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1393–1410, 2008.

[17] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[18] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[19] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[20] L. Wang and J. Yen, "Extracting fuzzy rules for system modeling using a hybrid of genetic algorithms and kalman filter," *Fuzzy Sets and Systems*, vol. 101, no. 3, pp. 353–362, 1999.

[21] L.-X. Wang and J. M. Mendel, "Back-propagation fuzzy system as nonlinear dynamic system identifiers," in *[1992 Proceedings] IEEE International Conference on Fuzzy Systems*, pp. 1409–1418, IEEE, 1992.