Solar Energy Forecasting With Fuzzy Time Series Using High-Order Fuzzy Cognitive Maps

Omid Orang
Machine Intelligence and Data Science (MINDS) Lab
Department of Electrical Engineering
Universidade Federal de Minas Gerais, UFMG
31270-000 Belo Horizonte, Brazil
ORCID 0000-0002-4077-3775

Rodrigo Silva
Department of Computer Science
Universidade Federal de Ouro Preto, UFOP
35400-000 Ouro Preto, Brazil
rodrigo.silva@ufop.edu.br

Petrônio Cândido de Lima e Silva
Instituto Federal do Norte de Minas Gerais, IFNMG
Januaria, Brazil
https://orcid.org/0000-0002-1202-2552

Frederico Gadelha Guimarães
Machine Intelligence and Data Science (MINDS) Lab
Department of Electrical Engineering
Universidade Federal de Minas Gerais, UFMG
31270-000 Belo Horizonte, Brazil
https://orcid.org/0000-0001-9238-8839

***Abstract*—Various studies indicate that Fuzzy Time Series (FTS) methods can obtain high accuracy in a variety of forecasting applciations. However, weighted FTS methods tend to show superiority in contrast to weightless ones. This study exploits the use of Fuzzy Cognitive Map (FCM) technique to generate the rules in the knowledge base for the FTS forecasting method. The proposed hybrid method, named HFCM-FTS, combines High Order Fuzzy Cognitive Maps (HFCM) and High Order Fuzzy Time Series (HOFTS), where the weight matrices associated with the state transitions are learned via the genetic algorithm from the data. The objective of FCM is to find the weight matrices that model the causal relations among the concepts defined in the Universe of Discourse. As a case study, we consider solar energy forecasting with public data for Brazilian solar stations from the year 2012 to 2015. The proposed HFCM-FTS is compared with HOFTS, Weighted High Order FTS (WHOFTS), and Probabilistic Weighted FTS (PWFTS) methods. The experiments also cover the influence of three modeling elements on the accuracy of the presented model including the number of concepts, activation function, and bias. The results show that the HFCM-FTS is able to achieve the best results with a low number of concepts.**

***Index Terms*—Fuzzy Cognitive Map, Fuzzy Time Series, Time series forecasting.**

I. Introduction

Time series forecasting is a very active and challenging field of research that plays a pivotal role in our daily life nowadays due to its applications in various research disciplines such as economics, management, energy, medicine etc. For instance, energy forecasting is amongst the most crucial tasks in energy management systems [1], [2], where accurate predictions can affect energy reliability, energy market and other fields related to producing and dispatching energy. Therefore, forecasting with a high level of accuracy has direct impact on making sensible decisions in many applications and situations.

A wide range of forecasting algorithms exist from traditional methods to soft computing ones. To complement the repertoire of forecasting methods, in 1993, Song and Chissom [3]–[5] introduced the concept of Fuzzy Time Series (FTS), where the values are represented by fuzzy sets. Various studies indicate that FTS models increased the accuracy of forecasting, with the advantage of explainability, flexibility and readability in contrast to some of the traditional methods.

In recent decades, the penetration of different FTS methods motivated researchers to deal with some forecasting problems. In [6], an improved forecasting model is proposed based on the weighted fuzzy relationship matrix combined with a Particle Swarm Optimization (PSO) adaptation for enrollments. In [7], a multi-variate FTS forecasting method is presented based on fuzzy clustering techniques and fuzzy interpolation technique for predicting the temperature and the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX). In [8] it is proposed a model based on the FTS and fuzzy variation groups for forecasting the TAIEX. In [9], Chen and Kao presented a combined model based on FTS, PSO and support vector machines for TAIEX forecasting. In [10], a FTS forecasting method is proposed based on fuzzy logical relationships and similarity measures. In [11], a ratio-based length of intervals is introduced to improve the FTS forecasting. In [12] a short-term load forecasting method is presented based on fuzzy time series, seasonality and long memory process. In [13] fuzzy time series are used for very short-term solar forecasting. In [14] advanced fuzzy time series are applied to short term load forecasting.

There is a vast literature on FTS forecasting methodologies, all of them consist of stages including: defining the universe

This work was supported by the Brazilian agencies CAPES, CNPq and FAPEMIG.

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

of discourse, partitioning, data fuzzification, inference, and defuzzification of the forecast [15], [16]. The number of fuzzy sets in the partitioning step of the universe of discourse plays a crucial role in the accuracy of the model. Different partitioning methods have been proposed, such as Grid partitioning [17], entropy partitioning [18], [19] or Fuzzy C-Means clustering techniques [20], [21]. Furthermore, some evolutionary algorithms such as Genetic algorithm (GA) or PSO have also been applied to generate unequal intervals [22], [23]. The order of the model acts as a memory of the system, it indicates the amount of past information (lags) that is used to predict the future event. It is another effective factor in FTS forecasting accuracy. The number of lags depends on the type of time series and it should be considered that increasing the number of lags leads to slower learning process.

Despite the various FTS forecasting methods, whether the model is weighted or weightless, the accuracy of the model will be affected significantly. Weighted Fuzzy Time Series (WFTS) methods consider weights on Fuzzy Logical relation groups (FLRGs). Obtained results by weighted methods confirm an obvious improvement in comparison to the weightless methods based on the literature. For instance, Silva [16] shows the superiority of using different WFTS methods in contrast to weightless ones. Fuzzy Cognitive Map (FCM) presents itself as a promising alternative, since it is a kind of weighted method, which was introduced by Kosko [24] according to the cognitive map theory, applied for explaining and analyzing the dynamic behavior of complex systems [25]. Qualitative modeling, ease of perception, high ability to dealing with uncertainties, capability to represent nonlinear and causal behaviors, flexibility and explainability can be accounted as interesting properties of FCMs [26].

FCMs are constructed based on nodes and edges in a graph structure. The nodes in the graph represent fuzzy concepts and weights among concepts display the causal relationship among these concepts. Concepts have a value in $[0, 1]$ which indicates the amount of activation for each concept. The weights among concepts in range $[-1, 1]$ express to what extend concepts affect each other based on sign and value. In order to update the value of each node, the transfer function is used to ensure that the value of each node remains in $[0, 1]$. The main focus of FCM is related to learning methods, which are classified as supervised learning approach, defined in order to adjust the weight matrix [27], [28]. The learning methods are divided into three main groups: Hebbian based, population based and hybrid approaches, which have been described in full details in [25]

The FCM can be used to generate the rules in the knowledge base for the FTS method. The learning capability of FCM can be exploited to model the time series data automatically and discover not only the rules but also the weights of these rules. Once these are obtained, FTS methods can be used to produce a forecast. This approach for generating the rules in FTS by employing FCM learning model is novel in the literature. The contribution of this paper is adopting a hybrid method HFCM-FTS for time series prediction. The hybrid method is designed by combining High Order Fuzzy Cognitive Maps (HFCM) and High Order Fuzzy Time Series (HOFTS), where the weight matrix associated with each fuzzy rule is learned via the genetic algorithm. The motivation of the proposed method is to include GA as FCM learning method to minimize the Root Mean Square Error (RMSE). The SONDA dataset (see details in section IV-A) was used in our study and the goal is predicting solar radiation with the proposed univariate model. The obtained results are also compared with High Order FTS (HOFTS), Weighted High Order FTS (WHOFTS) and Probabilistic Weighted FTS (PWFTS) methods.

The structure of this paper is organized as follows. Section II presents a brief description of Fuzzy Time Series (FTS) and Fuzzy Cognitive Map (FCM). Section III provides information about the proposed forecasting model, combining FCM and FTS. Section IV reports experimental results and discussion. Finally, Section V provides the conclusions of the paper and ideas for future research.

II. Background

A. Fuzzy Cognitive Maps

Fuzzy Cognitive Maps (FCM) are derived from Axelrod cognitive maps [29]. FCM was proposed as a powerful tool to analyze complex nonlinear systems [24]. FCM consists of a graph representation of some system behavior, consisting of nodes, known as concepts or fuzzy sets, and edges, representing the causal relationships between the concepts. FCMs are defined by the 4-tuple $(\mathbf{C}, \mathbf{W}, \mathbf{a}, f)$, where $\mathbf{C} = [c_1, \ldots, c_n]$ is the set of n concepts, which are the variables (and the nodes of the graph) that compose the system. At any time t, each concept $c_i \in \mathbf{C}$ has an activation $a_i \in [0, 1]$, and $\mathbf{a} = (a_1, \ldots, a_n)$ is the state vector containing the set of all concept activations. $\mathbf{W}$ is a $n \times n$ weight matrix, which represents the connections among concepts, such that:

$$\mathbf{W} = \begin{pmatrix} w_{11} & \ldots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \ldots & w_{nn} \end{pmatrix} \quad (1)$$

The weight w_{ij} between any two concepts can be positive, negative or zero. It represents how concepts influence each other. If w_{ij} is positive, an increment in concept c_i will produce an increment on concept c_j. If w_{ij} is negative, an increment on c_i has a reverse effect on c_j, and when w_{ij} is zero, this indicates that there is no relationship between concepts c_i and c_j.

Finally, f is the activation function and the most common transfer functions are the bivalent, the trivalent, the hyperbolic tangent and the sigmoid functions. Kosko's activation rule [30] is shown in (2):

$$a_i(t+1) = f\left(\sum_{j=1}^{n} w_{ji} a_j(t)\right) \quad (2)$$

w_{ji} highlights the value of the causal relationship between concepts c_j and c_i, whereas a_i^t represents the state value of concept c_i at time step t. It should be noted that, at each time step, the FCM generates a state vector which contains all concept activations. The state vector $\mathbf{a}(t) = (a_1(t), \ldots, a_n(t))$ refers to the activation level of all concepts at time t. The next state can be written as:

$$\mathbf{a}(t+1) = f(\mathbf{W} \cdot \mathbf{a}(t)) \tag{3}$$

In (3), the activation level of each node at time $t+1$ just relies on the activation level of all nodes at time t. In order to improve the capability of the FCM to describe the dynamic behavior of complex systems, one can define high-order FCM according to:

$$\mathbf{a}(t+1) = f\left(\mathbf{w}^0 + \mathbf{W}^1\mathbf{a}(t) + \ldots + \mathbf{W}^k\mathbf{a}(t-k+1)\right) \tag{4}$$

As can be seen from (4), the state value of node i at time $t+1$ depends on the state value of all nodes at $t, t-1, \ldots, t-k+1$. Vector $\mathbf{w}^0$ is a bias term. $\mathbf{a}(t+1)$ is the response vector of high order FCM for state vectors $\mathbf{a}(t)$, $\mathbf{a}(t-1)$, ..., $\mathbf{a}(t-k+1)$.

In the above formula, f refers to the activation function. Different types of transfer functions e.g. the sigmoid function and the Rectified Linear Units (ReLU) can be adopted. Besides, each node can have its own activation allowing the creation of complex networks.

Various types of learning methods have been proposed for the FCM and their goal is to find the weights of the model based on historical data. Some popular methods are the Hebbian based learning [28] and evolutionary algorithms such as Particle Swarm Optimization (PSO) [31], Artificial Bee Colony algorithm (ABC) [32], Genetic Algorithm (GA) [33] etc.

B. Fuzzy Time Series

The Fuzzy Time series was first introduced by Song and Chissom [3], [4], [34] as a method that employs Fuzzy Sets [35] to transform numerical time series into linguistic time series and then model its behavior (generally using membership matrices or rules). The trained model is later used to forecast unseen values of the time series.

Given a univariate time series $Y \in \mathbb{R}^1$, and its individual values $y(t) \in Y$, for $t = 0, 1, \ldots, T$, Each membership function $\mu_{A_i} : \mathbb{R} \to [0, 1]$ is responsible to assess the membership of a value $y(t) \in Y$ to a fuzzy set $A_i \in \tilde{A}$. There are many possible shapes for Membership functions but the most common are the triangular, trapezoidal, sigmoidal and Gaussian functions.

There are several categories of FTS methods, varying mainly by their order Ω and time-variance. The order is the number of time-delays (lags) that are used in modeling the time series. Given the time series data Y, First Order models need just $y(t-1)$ data to forecast $y(t)$, while Higher Order models require $y(t-1), \ldots, F(t-\Omega)$ data to forecast $y(t)$. The time variance indicates if an FTS model changes along the time. If the model is static, it is called Time-Invariant, otherwise Time-Variant. This work focuses only on Time-Invariant methods.

TABLE I
FTS HYPERPARAMETERS

Alias	Parameter
$k \in \mathbb{N}^+$	Number of partitions (fuzzy sets)
$\mu : U \to [0, 1]$	Membership function, measures the membership of a value $y \in U$ to a fuzzy set
$\Omega \in \mathbb{N}^+$	Order, the number of past lags used by the model

The pioneer FTS work in [3] used weight matrices to store the knowledge learned from the data. In [34] rule models were introduced as a way to improve the readability and computational cost of the FTS approaches. Following the approach proposed in [34], an FTS model $\mathcal{M}$ is trained with the Training Procedure, which contains three main steps:

A) *Partitioning*: the Universe of Discourse (UoD) U is delimited by the known bounds of Y, such that $U = [\min(Y), \max(Y)]$. U is then divided in k overlapping intervals and for each interval a fuzzy set A_i is defined with its membership function μ_{A_i} and midpoint c_i. A linguistic variable $\tilde{A}$ is defined upon U, where the linguistic terms are fuzzy sets $A_i \in \tilde{A}$, for $i = 1 \ldots k$.
B) *Fuzzification*: transform the original numeric time series Y into a fuzzy time series F, where each data point $f(t) \in F$ is an $1 \times k$ array with the fuzzified values of $y(t) \in Y$ with respect to the linguistic terms $A_i \in \tilde{A}$.
C) *Knowledge Extraction*: analyze the sequential terms $f(t-1), f(t)$ in F, grouping the patterns in rules $A_i \to A_j, A_k, \ldots$ which can be read as "IF $f(t)$ IS A_i THEN $f(t+1)$ IS $A_j, A_k, \ldots$".

The trained FTS model $\mathcal{M}$ contains both the linguistic variable $\tilde{A}$ and the rules. Given a sample $y(t)$, the model can be used to forecast $y(t+1)$ with the following procedure:

A) *Fuzzification*: transform the numeric sample $y(t) \in Y$ into a fuzzy sample $f(t)$, where $f(t) \in \tilde{A}$.
B) *Rule Matching*: Find the rules that contain the fuzzy set $f(t)$ in its precedent, the set of selected rules is R. The activation $\mu_r \in R$ of the rules are given by the membership values in $f(t)$.
C) *Defuzzification*: For each rule $r \in R$, first find the midpoint of each rule by summing the midpoints of each fuzzy set of the consequent:

$$mp_r = \sum_{i \in consequent} c_i \tag{5}$$

Then find the numerical value by weight average the midpoints of each rule by its activation:

$$y(t+1) = \frac{\sum_{r \in R} \mu_r \cdot mp_r}{\sum_{r \in R} \mu_r} \tag{6}$$

There are several models available in the literature that extended the classical Song and Chissom method [3], [4], such as High Order FTS (HOFTS) proposed in [13], and Weighted

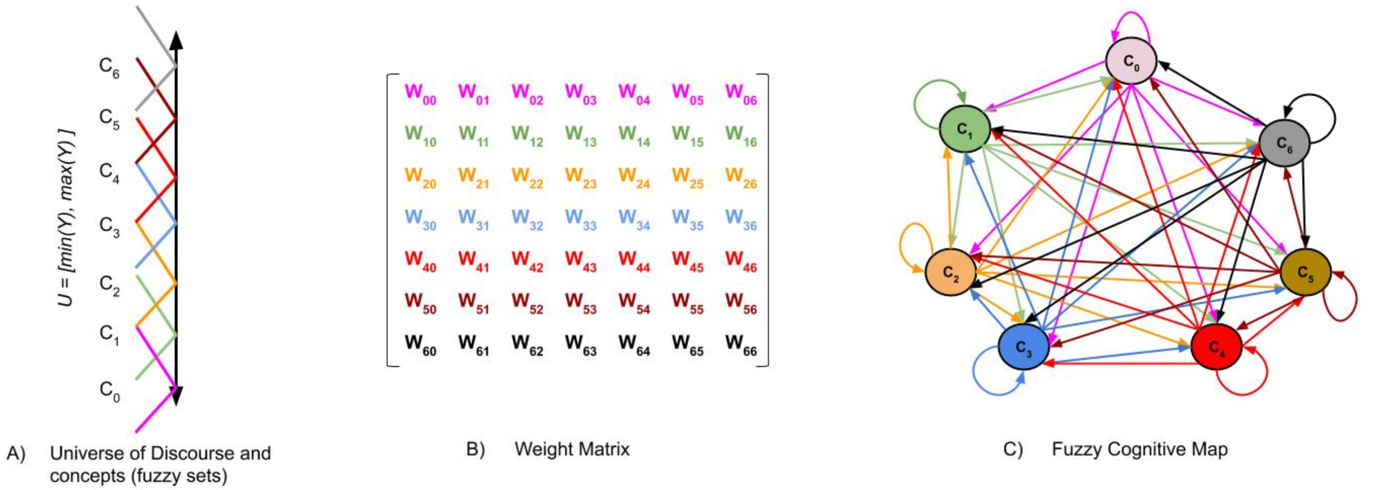

Fig. 1. Structure of the fuzzy cognitive map.

FTS (WFTS) [36], which adds weights to the rules and modifies the way that the rule midpoint is calculated. In [16] the Weighted High Order FTS (WHOFTS) is proposed, which extends the HOFTS model by including weights in its rules. In [37] a new kind of weighted method is proposed, namely the Probabilistic Weighted FTS (PWFTS), including weights in the precedent and the consequent of the rules. These weights represent conditional probabilities and the PWFTS was the first approach to integrate point, interval and probabilistic forecasting for one to many steps ahead. The most important hyperparameters of an FTS model are summarized in Table I. In [38] the authors develop an evolutionary and distributed method for hyperparameter tunning of FTS models.

There are, in the literature, several extensions that embrace other aspects of forecasting using FTS methods. A comprehensive review of FTS methods can be found in [16] while the review of concepts, applications and future directions was recently provided in [15]. Multivariate FTS models can be found in [39] and [40]. An approach for dealing with non-stationary data is proposed in [41].

III. Proposed HFCM-FTS Method

In this work, a new hybrid forecasting method is introduced by combining High Order Fuzzy Cognitive Map (HFCM) and High Order Fuzzy Time Series (HOFTS), termed as HFCM-FTS. The architecture of this model consists of different stages including: partitioning, fuzzification, FCM learning method, defuzzification and accuracy measurement.

Figure 1 illustrates the structure of the FCM. Figure 1-A shows a given partition of the universe of discourse and the definition of seven concepts, which is also the number of fuzzy sets, because in our model the number of concepts and fuzzy sets are considered the same. As shown in Figure 1-C, FCM is a collection of concepts and causal interactions among these concepts. In particular, FCM is a graph model of a system in which the nodes represent the concepts and the weight matrix represents the relationship among these concepts. According to Figure 1-B, this weight matrix is a square connection matrix. For an FCM with k concepts the weight matrix is of size $k \times k$. Each weight saves the corresponding relationship among the source and the target concept, that is, for example w_{05} presents the relation between source concept c_0 and target concept c_5.

It is clear from Figure 1 that the objective of FCM is to find the weight matrix for the concepts defined in the Universe of Discourse. These weights should be learned from the data. The HFCM-FTS method is divided in two procedures: the Training Procedure, detailed in Section III-A, and the Forecasting Procedure, detailed in Section III-B and depicted in Figure 2. In the next sections, the training and forecasting procedures are detailed.

A. Training Procedure

The aim of the Training Procedure is to create the linguistic variable C with k concepts, and find the set of Ω weight matrices $\mathbf{W}^t$, given a crisp training set Y and the activation function f informed by the user. The steps of the method are listed below:

1) **Partitioning**: Split the Universe of Discourse, $U = [\min(Y), \max(Y)]$ into k even length and overlapped intervals, and for each interval a fuzzy set C_i (i.e. concept) is defined with a membership function μ_{C_i}. The group of the k concepts form the linguistic variable C, such that $C_i \in C, \forall i = 1..k$, as shown in Figure 1. Here the grid partitioning is used to generate versions of the FCM with $k = \{5, 10, 20\}$ and triangular membership function μ_{C_i}.
2) **Weight Matrix Learning**: Each matrix $\mathbf{W}^t$, for $t = 0..\Omega$, is a $k \times k$ matrix where $w_{ij}^t \in \mathbb{R}$ is the weight between the concepts C_i and C_j at the time lag t, and Ω is the order of the model. These weight matrices are trained using Genetic Algorithm (GA) where the genotype is real encoded and simply contains the list of all w_{ij}^t matrix values.

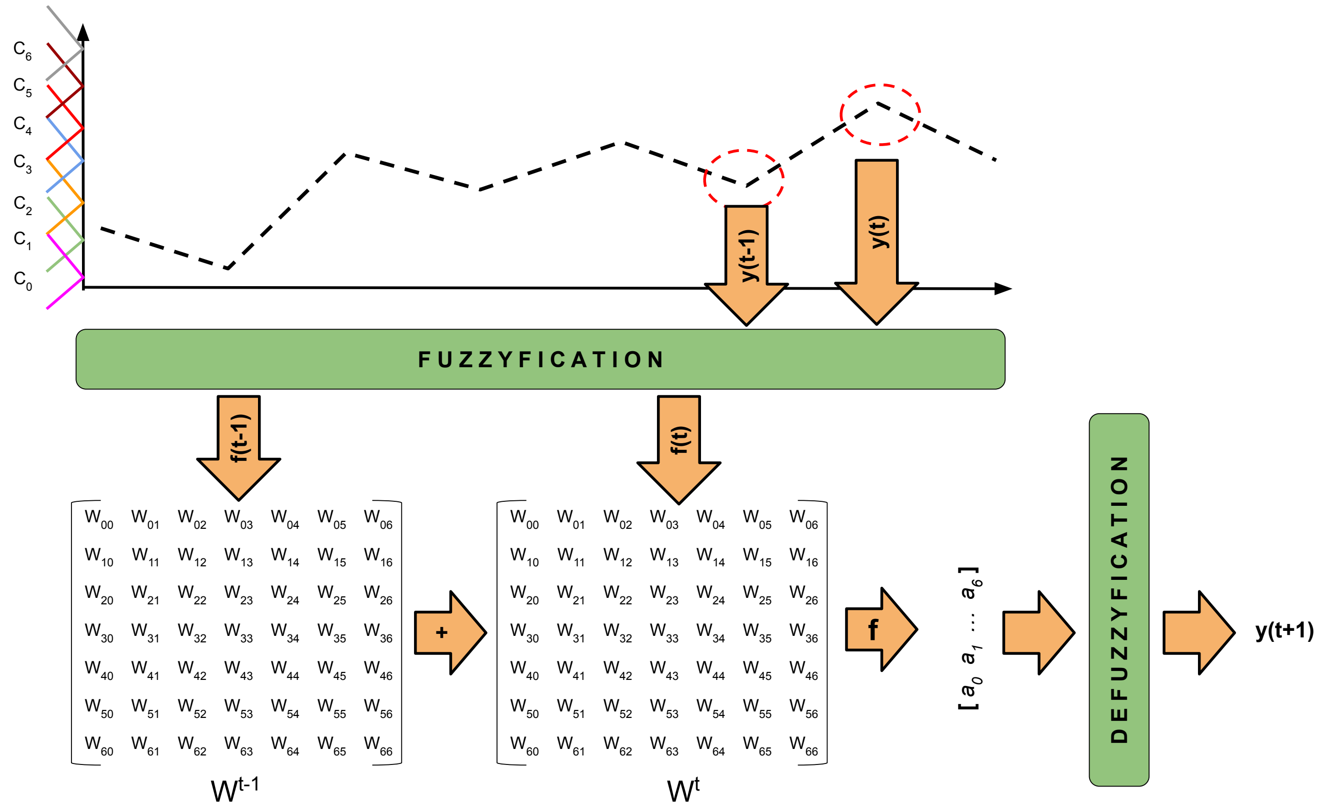

Fig. 2. The structure of the proposed HFCM-FTS method.

The fitness of each genotype is the forecasting error, using the Root Mean Squared Error (RMSE) metric, using the training sample Y and the weight matrices represented by the genotype in the Forecasting method presented in Section III-B. Initially a population P with random individuals is created, such that $w_{ij}^t \sim \mathcal{N}(0, 0.1)$, $\forall i, j = 0..k, t = 0..\Omega$.

After the evaluation of the individuals, a percentage P_{sel} of the individuals are selected using tournament. Then a percentage P_{cross} of the population undersgoes crossover and a percentage P_{mut} is chosen for random mutation. This process is repeated for G generations. This iterative process aims to minimize the fitness function (the RMSE) for the given training set Y.

It is worth noting that in the present work only a second order FCM ($\Omega = 2$) was considered. Thus there are two weight matrices for times t and $t-1$ as shown in Figure 2.

B. Forecasting Procedure

The aim of the Forecasting Procedure is to estimate the crisp value $y(t+1)$ given the linguistic variable C, the weight matrices $\mathbf{W}^t$, the activation function f and a crisp input Y. The steps of the method are listed below:

1) **Fuzzification**: Given the crisp input sample Y with size T, each instance $y(t) \in Y, t = 1..T$, is transformed into an activation vector $\mathbf{a}(t)$ such that $a_i(t) = \mu_{C_i}(y(t))$, $\forall C_i \in C$, that is, each value $a_i(t) \in \mathbf{a}(t)$ corresponds to the membership degree of $y(t)$ to the concept C_i.
2) **Activation**: The state value for each concept in time $t+1$ can be defined by the following formula:

$$\mathbf{a}(t+1) = f\left(\mathbf{w}^0 + \sum_{j=1}^{\Omega} \mathbf{W}^j \cdot \mathbf{a}(t-j+1)\right) \quad (7)$$

3) **Defuzzification**: After calculating the activation level of each concept, in this step, the defuzzification is carried out to produce the forecasts. Equation (8) defines the forecast produced by HFCM-FTS:

$$\hat{y}(t+1) = \frac{\sum_{i=1}^{k} a_i(t+1) \cdot mp_i}{\sum_{i=1}^{k} a_i(t+1)} \quad (8)$$

where $a_i(t+1)$ is the activation calculated from the previous step for each concept at time $t+1$ and mp_i is the center of each concept C_i. With this equation the forecast value at time $t+1$ is obtained in numeric terms.

In order to evaluate the performance of the proposed model, the Root Mean Square Error (RMSE), described in (9), is employed as the accuracy metric.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i(t) - \hat{y}_i(t))^2} \quad (9)$$

IV. Computational Experiments

A. Dataset

In order to test the utility of the proposed method, in this section, we apply it to solar radiation time series data[1]. Figure 3 shows the referred time series taken from the SONDA Project at http://sonda.ccst.inpe.br/.

SONDA - *Sistema de Organização Nacional de Dados Ambientais* (Brazilian National System of Environmental Data Organization), is a governmental project which groups environmental data (solar radiance, wind speed, precipitation, etc) from INPE - Instituto Nacional de Pesquisas Espaciais (Brazilian Institute of Space Research).

In this experiment, 8,000 samples have been used, with a sliding window of 2,000 samples in the cross validation method. 80% of the window for training and 20% for test.

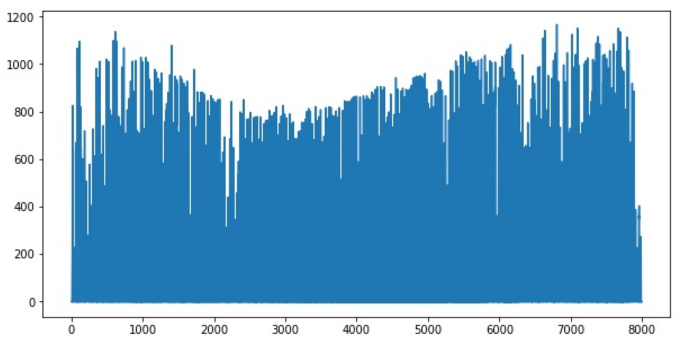

Fig. 3. Solar radiance time series

The number of concepts (partitions), the type of activation function and the presence of bias terms are the main parameters affecting the accuracy of the proposed model. Thus, in sections IV-B and IV-C, we assess the importance of the bias term and test different types of activation functions, respectively. Finally, in Section IV-D, we compare the proposed method with state-of-the-art fuzzy time series methods.

B. The influence of the bias term

To assess the effect of the bias term, we execute the proposed methodology with 5, 10 and 20 concepts, with and without the bias terms. In this experiment, the sigmoid was used as activation function and since the learning method is stochastic (a GA), each configuration was tested 15 times.

Table II shows the number of parameters for each configuration, the best, the average, and the worst RMSE of the 15 independent runs. It can be seen that adding bias improves the accuracy of the proposed model. Without the bias term a null activation vector at time $t-1$ can produce activations greater than 0 at time t which, in turn, should not make sense. It must be noticed, however, that the number of parameters increases when compared the models without bias.

[1]Available at: https://query.data.world/s/2bgegjggydd3venttp3zlosh3wpjqj accessed on April, 4, 2020

C. The influence of the activation function

In Table III, we compare the accuracy of the model using different activation functions. In addition to varying the activation function, we vary the number of concepts, running the model with 5, 10, 20 sets. In this experiment, all the models include a bias term.

The results indicate that when the model uses ReLUs, on average, the accuracy is slightly better.

D. Comparison with other Fuzzy Time Series methods

In this section, we compare the proposed method with other fuzzy time series forecasting methods available in the pyFTS library [42]. In particular, the proposed method is tested against HOFTS (High Order Fuzzy Time Series), WHOFTS (Weighted High Order Fuzzy Time Series), and PWFTS (Probabilistic Weighted Fuzzy Time Series). In [37], the PWFTS was shown to be the best performing method in a multitude of experiments against many statistic forecasting methods, other FTS, and machine learning methods. Since GA was used as a learning method, and it is a stochastic optimization algorithm, the HFCM-FTS method was executed 15 times, with the GA hyper-parameters defined according to Table V.

Table IV presents the results considering bias terms, the ReLU activation function and sliding window cross validation for various numbers of concepts (partitions) in comparison with the aforementioned methods.

The results show the superiority of HFCM-FTS, which is a mixture of HOFTS and FCM, over HOFTS. This indicates that adding FCM to HOFTS indeed improves the methodology. The results also show that the proposed method is superior to all the other methods when the number of sets was 5. On the other hand, as the number of concepts increases the HFCM-FTS performance deteriorates. Meanwhile, the performance of the other methods improves as the number of concepts increases.

We hypothesize that the decrease in the performance of HFCM-FTS is due to the lack of convergence of the learning algorithm. We suspect that the GA settings were not adequate to train the model for such a high number of parameters.

Nevertheless, the results of HFCM-FTS were quite competitive with the other methods. It is noticeable that the best results achieved by WHOFTS and PWFTS required a partition with 20 concepts, with a higher number of rules, while the average results of HFCM-FTS with 5 concepts were significantly better.

V. Conclusion and future works

The core contribution of this paper is the design of a hybrid method based on Fuzzy Cognitive Maps (FCM) applied to time series forecasting. This method, called HFCM-FTS, consists of a High Order Fuzzy Cognitive Map (HFCM) and a High Order Fuzzy Time Series (HOFTS). A Genetic Algorithm (GA) is used as a learning method to find the weight matrices used to assess the relationship between concepts. The paper also covers the influence of three elements on the accuracy of the

TABLE II

THE EFFECT OF THE BIAS ON THE ACCURACY OF THE PROPOSED HFCM-FTS METHOD WITH SIGMOID ACTIVATION FUNCTION

Number of concepts	Parameters without bias	Best RMSE without bias	Average RMSE without bias	Worst RMSE without bias	Parameters with bias	Best RMSE with bias	Average RMSE with bias	worst RMSE with bias
5	50	146.332	152.939	158.968	60	**111.207**	**119.565**	**132.957**
10	200	183.894	193.658	198.827	220	**141.576**	**162.576**	**184.484**
20	800	223.537	231.972	240.436	840	**172.900**	**183.355**	**196.890**

TABLE III

PERFORMANCE OF THE PROPOSED HFCM-FTS METHOD USING DIFFERENT ACTIVATION FUNCTIONS WHEN BIAS HAS BEEN CONSIDERED

Number of concepts	Best Value with ReLU	Average RMSE with ReLU	Worst RMSE with ReLU	Best RMSE with sigmoid	Average RMSE with sigmoid	Worst RMSE with sigmoid	Best RMSE with tgnh	Average RMSE with tgnh	Worst RMSE with tgnh
5	**105.886**	**119.191**	**131.393**	111.207	119.565	132.957	110.802	128.726	139.687
10	**136.60**	**157.401**	**171.147**	141.567	162.576	184.484	153.032	164.789	188.700
20	**166.922**	**182.044**	**189.306**	172.900	183.355	196.890	173.340	183.828	199.511

TABLE IV

COMPARISON OF THE ACCURACY OF THE PROPOSED HFCM-FTS BY CONSIDERING RELU ACTIVATION FUNCTION AND BIAS WEIGHTS WITH OTHER METHODS

Number of concepts	HFCM-FTS Parameters	HFCM-FTS Best Value	HFCM-FTS Average Value	HFCM-FTS Worst Value	HOFTS	WHOFTS	PWFTS
5	60	105.886	119.191	131.393	368.42	182.98	149.79
10	220	136.60	157.401	171.147	288.51	146.15	134.91
20	840	166.922	182.044	189.306	204.18	134.43	134.22

TABLE V

HYPER-PARAMETERS FOR GA LEARNING ALGORITHM

GA Hyperparameters	*Pcruz*	*Pmut*	*Population*	*Generation*
	0.5	0.3	50	30

presented model including the number of concepts, activation function, and bias.

The results are still limited and more experiments are required in order to assess the efficacy of the method. Nevertheless, in the given scenario the proposed method presented competitive results when compared to other fuzzy time series methods in particular when the number of concepts is limited. It is important to highlight that keeping the number of concepts small may positively affect the readability and interpretability of the model.

A. Model Limitations and Challenging

Despite the success of the HFCM-FTS model when the number of concepts is low, the proposed training methodology will get very time consuming as this number increases. For instance, when the number of concepts is 30 the GA will have to solve an 1860 variable problem. Increasing this number by only 10, from 30 to 40 concepts, the GA will have to solve a 3280 variable problem. Besides, it is certain that this issue will be more acute if the model is used to predict a multivariate FTS. Thus, to tackle the aforementioned problem, finding an alternative to GA, which is not very efficient, is highly recommended.

B. Future Work

Improving the learning method is key to the success of the method. In addition to that, future research challenges may involve implementing multivariate and probabilistic HFCM-FTS forecasting methods.

REFERENCES

[1] V. N. Coelho, I. M. Coelho, B. N. Coelho, M. W. Cohen, A. J. Reis, S. M. Silva, M. J. Souza, P. J. Fleming, and F. G. Guimarães, "Multi-objective energy storage power dispatching using plug-in vehicles in a smart-microgrid," *Renewable Energy*, vol. 89, pp. 730 – 742, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960148115305000

[2] H. J. Sadaei, P. C. [de Lima e Silva], F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365 – 377, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544219304852

[3] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series — part i," *Fuzzy Sets and Systems*, vol. 54, no. 1, pp. 1 – 9, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016501149390355L

[4] Q. Song and B. Chissom, "Forecasting enrollments with fuzzy time series—part ii," *Fuzzy Sets and Systems*, vol. 62, pp. 1–8, 02 1994.

[5] Q. Song, R. P. Leland, and B. S. Chissom, "Fuzzy stochastic fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 88, no. 3, pp. 333 – 341, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165011496000772

[6] Y.-L. Huang, S.-J. Horng, T.-W. Kao, R.-S. Run, J.-L. Lai, R.-J. Chen, I.-H. Kuo, and K. Khan, "An improved forecasting model based on the weighted fuzzy relationship matrix combined with a pso adaptation for enrollments," *International Journal of Innovative Computing, Information and Control*, vol. 7, 07 2011.

[7] S.-M. Chen and Y.-C. Chang, "Multivariable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques," *Information Sciences - ISCI*, vol. 180, pp. 4772–4783, 12 2010.

[8] S. Chen and C. Chen, "Taiex forecasting based on fuzzy time series and fuzzy variation groups," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 1–12, Feb 2011.

[9] S.-M. Chen and P.-Y. Kao, "Taiex forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines," *Information Sciences: an International Journal*, vol. 247, pp. 62–71, 10 2013.

[10] S.-H. Cheng, S.-M. Chen, and W.-S. Jian, "Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures," *Information Sciences*, vol. 327, pp. 272–287, 01 2016.

[11] K.-H. Huarng and T. Yu, "Ratio-based lengths of intervals to improve fuzzy time series forecasting," *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics*, vol. 36, pp. 328–40, 05 2006.

[12] H. J. Sadaei, F. G. Guimarães, C. J. da Silva, M. H. Lee, and T. Eslami, "Short-term load forecasting method based on fuzzy time series, seasonality and long memory process," *International Journal of Approximate Reasoning*, vol. 83, pp. 196 – 217, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888613X17300464

[13] C. A. Severiano, P. C. L. Silva, H. J. Sadaei, and F. G. Guimarães, "Very short-term solar forecasting using fuzzy time series," in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, July 2017, pp. 1–6. [Online]. Available: http://doi.org/10.1109/FUZZ-IEEE.2017.8015732

[14] G. Costa Silva, J. L. Silva, A. Lisboa, D. Vieira, and R. Saldanha, "Advanced fuzzy time series applied to short term load forecasting," 11 2017.

[15] M. Bose and K. Mali, "A novel data partitioning and rule selection technique for modeling high-order fuzzy time series," *Applied Soft Computing*, vol. 63, pp. 87 – 96, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494617306749

[16] P. C. L. Silva, "Scalable models for probabilistic forecasting with fuzzy time series," Ph.D. dissertation, UFMG, 09 2019. [Online]. Available: http://doi.org/10.5281/zenodo.3374641

[17] "Heuristic models of fuzzy time series for forecasting," *Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 369 – 386, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165011400000932

[18] C.-H. Cheng, J.-W. Wang, and C.-H. Li, "Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix," *Expert Syst. Appl.*, vol. 34, pp. 2568–2575, 2008.

[19] S.-M. Chen and S.-W. Chen, "Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups and the probabilities of trends of fuzzy logical relationships," *IEEE transactions on cybernetics*, vol. 45, 07 2014.

[20] S.-T. Li, Y.-C. Cheng, and S.-Y. Lin, "A fcm-based deterministic forecasting model for fuzzy time series," *Comput. Math. Appl.*, vol. 56, no. 12, p. 3052–3063, Dec. 2008. [Online]. Available: https://doi.org/10.1016/j.camwa.2008.07.033

[21] S. Askari and N. Montazerin, "A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering," *Expert Syst. Appl.*, vol. 42, no. 4, p. 2121–2135, Mar. 2015. [Online]. Available: https://doi.org/10.1016/j.eswa.2014.09.036

[22] W. Zhang, S. Zhang, and S. Zhang, "Two-factor high-order fuzzy-trend fts model based on bso-fcm and improved ka for taiex stock forecasting," *Nonlinear Dynamics*, 06 2018.

[23] W. Zhang, S. Zhang, S. Zhang, D. Yu, and N. Huang, "A novel method based on fts with both ga-fcm and multifactor bpnn for stock forecasting," *Soft Comput.*, vol. 23, no. 16, p. 6979–6994, Aug. 2019. [Online]. Available: https://doi.org/10.1007/s00500-018-3335-2

[24] B. Kosko, "Fuzzy cognitive maps," *International Journal of Man-Machine Studies*, vol. 24, no. 1, pp. 65 – 75, 1986. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020737386800402

[25] E. I. Papageorgiou and J. L. Salmeron, "A review of fuzzy cognitive maps research during the last decade," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 1, pp. 66–79, 2012.

[26] M. van Vliet, K. Kok, and A. Veldkamp, "Linking stakeholders and modellers in scenario studies: The use of fuzzy cognitive maps as a communication and learning tool," *Futures 42 (2010) 1*, vol. 42, 02 2010.

[27] G. Felix, G. Nápoles, R. Falcon, W. Froelich, K. Vanhoof, and R. Bello, "A review on methods and software for fuzzy cognitive maps," *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1707–1737, Oct 2019. [Online]. Available: https://doi.org/10.1007/s10462-017-9575-1

[28] J. L. Salmeron, T. Mansouri, M. R. S. Moghadam, and A. Mardani, "Learning fuzzy cognitive maps with modified asexual reproduction optimisation algorithm," *Knowledge-Based Systems*, vol. 163, pp. 723 – 735, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705118304854

[29] R. Axelrod, Ed., *Structure of Decision: The Cognitive Maps of Political Elites*. Princeton Legacy Library, 1976.

[30] W. Lu, J. Yang, X. Liu, and W. Pedrycz, "The modeling and prediction of time series based on synergy of high-order fuzzy cognitive map and fuzzy c-means clustering," *Knowledge-Based Systems*, vol. 70, pp. 242 – 255, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705114002615

[31] K. Parsopoulos, E. Papageorgiou, and M. Vrahatis, "A first study of fuzzy cognitive maps learning using particle swarm optimization," 12 2002.

[32] E. Yesil, C. Öztürk, M. F. Dodurka, and A. Sakalli, "Fuzzy cognitive maps learning using artificial bee colony optimization," *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, 2013.

[33] N. H. Mateou, M. Moiseos, and A. S. Andreou, "Multi-objective evolutionary fuzzy cognitive maps for decision support," in *2005 IEEE Congress on Evolutionary Computation*, vol. 1, Sep. 2005, pp. 824–830 Vol.1.

[34] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy sets and systems*, vol. 81, no. 3, pp. 311–319, 1996.

[35] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[36] H.-K. Yu, "Weighted fuzzy time series models for TAIEX forecasting," *Physica A: Statistical Mechanics and its Applications*, vol. 349, no. 3, pp. 609–624, 2005.

[37] P. C. Silva, H. J. Sadaei, R. Ballini, and F. G. Guimarães, "Probabilistic forecasting with fuzzy time series," *IEEE Transactions on Fuzzy Systems*, 2019. [Online]. Available: http://doi.org/10.1109/TFUZZ.2019.2922152

[38] P. Silva, P. Lucas, H. Sadaei, and F. Guimarães, "Distributed evolutionary hyperparameter optimization for fuzzy time series," *IEEE Transactions on Network and Service Management*, vol. PP, pp. 1–1, 03 2020.

[39] P. C. d. L. Silva, P. d. O. e Lucas, and F. G. Guimarães, "A distributed algorithm for scalable fuzzy time series," in *International Conference on Green, Pervasive, and Cloud Computing*. Springer, 2019, pp. 42–56. [Online]. Available: http://doi.org/10.1007/978-3-030-19223-5_4

[40] P. C. L. Silva, C. A. Severiano Jr., M. A. Alves, M. W. Cohen, and F. G. Guimarães, "A new granular approach for multivariate forecasting," in *Computational Neuroscience*, V. R. Cota, D. A. C. Barone, D. R. C. Dias, and L. C. M. Damázio, Eds. Cham: Springer International Publishing, 2019, pp. 41–58. [Online]. Available: http://doi.org/10.1007/978-3-030-36636-0_4

[41] M. A. Alves, P. C. d. L. e Silva, C. A. Severiano, G. L. Vieira, F. G. Guimarães, and H. J. Sadaei, "An extension of nonstationary fuzzy sets to heteroskedastic fuzzy time series." in *ESANN*, 2018.

[42] P. C. de Lima e Silva et al, "pyfts: Fuzzy time series for python," May 2019. [Online]. Available: https://doi.org/10.5281/zenodo.597359