DOI: https://doi.org/10.1109/tfuzz.2019.2922152

Probabilistic Forecasting With Fuzzy Time Series

Petrônio Candido de Lima Silva, Hossein Javedani Sadaei, Rosangela Ballini, *Member, IEEE*, and Frederico Gadelha Guimarães, *Member, IEEE*

***Abstract*—In recent years, the demand for developing low computational cost methods to deal with uncertainties in forecasting has been increased. Probabilistic forecasting is a class of forecasting in which the method provides intervals or probability distributions as outcomes of its forecasting. The aim of this paper is, therefore, proposing a new forecasting approach based on fuzzy time series (FTS) that takes advantage of fuzzy and stochastic patterns on data and is capable to deal with point, interval, and distribution forecasts. The method proposed was empirically tested with typical financial time series, and the results were compared with other standard FTS and statistical methods. The results show that the proposed method obtained accurate results and outperformed standard FTS methods. The proposed method also combines versatility, scalability, and low computational cost, making it useful on a wide range of application scenarios.**

***Index Terms*—Forecast uncertainty, fuzzy systems, fuzzy time series (FTS), probabilistic forecasting, time series analysis.**

I. Introduction

MANY scientific and engineering applications demand the measurement of forecasting uncertainty inherent in natural and economic processes. In these processes, the uncertainty can be intrinsic or extrinsic and is classified in two categories: 1) the epistemic uncertainty; and 2) the ontological uncertainty [1]. The epistemic uncertainty represents the vagueness, lack of information, and imprecision (aggregations, measurement errors, sensor calibration, and other unknown factors) and can be modeled by the fuzzy theory. The ontological uncertainty represents the randomness, and the stochastic behavior and can be modeled by the probability theory.

Despite these points, the majority of forecasting methods are concerned with one step ahead point forecasting without outputting uncertainty measures. When the n-step ahead forecasting is considered, the uncertainty increases and the accuracy and reliability of models is affected. This effect becomes even worse as the forecasting horizon increases.

This fact led to the development of methods for probabilistic forecasting [2] and interval forecasting [3], to deal with forecasting uncertainty by estimating the distributions of possible values instead of a unique point forecast. The interval forecasting can be viewed as a particular case of the probabilistic forecasting where the probabilities are assumed to be uniform inside the bounds of the interval forecast. However, traditional methods of probabilistic forecasting require the use of parametric models with distribution assumptions, as in Bayesian inference, or costly estimation techniques and Monte-Carlo simulations. Probabilistic forecasting has been used in areas such as weather forecasting (see [4] and [5]), electric load forecasting (see [6], [7], and [8]), wind power generation (see [9] and [10]), and hydrology forecasting (see [11]).

With the recent phenomenon of the Big Data, other concerns were added to the requirements of forecasting methods: scalability, low computational cost, and self-adaptability. Some of these are the characteristics of the fuzzy time series (FTS) methods [12], which have been drawing more attention and relevance in recent years due to many studies reporting their good accuracy compared with other models [13]. FTS is also an approach to deal with the epistemic uncertainty, as with the time-aggregation case of the financial time series. The fuzzification of data gives a more flexible representation to the individual measures, embracing the range of possible value fluctuations not covered by the single values.

FTS methods have been commonly used for the forecasting of university enrollments (see [12], [14], and [15]), stock markets (see [16]–[21]), tourism (see [22]), electric load (see [23] and [24]), seasonal time series (see [25]–[27]), among many others. There are still some gaps in FTS methods (see [28] and [29]) related with methodological problems, but many of them have been approached in more recent studies [30]. But, the most notable lack in these methods is the absence of probabilistic forecasting approaches.

Taking these FTS drawbacks into consideration, this research presents the probabilistic weighted (PWFTS), a new method family that exploits both types of uncertainty to capture time

Manuscript received July 16, 2018; revised November 22, 2018, March 14, 2019, and May 12, 2019; accepted May 31, 2019. Date of publication June 11, 2019; date of current version August 4, 2020. This work was supported by the Brazilian agency CAPES. The work of P. C. de Lima Silva was supported by the PBQS program of the Brazilian institution IFNMG-Campus Januária. The work of H. J. Sadaei and R. Ballini was supported by the CNPq. The work of F. G. Guimarães was supported by the Brazilian Agencies CAPES, CNPq, and FAPEMIG. *(Corresponding author: Frederico Gadelha Guimarães.)*

P. C. de Lima Silva is with the Instituto Federal do Norte de Minas Gerais, Januária 39480-000, Brazil, and also with the Machine Intelligence and Data Science Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte 31270-010, Brazil (e-mail: petronio.candido@gmail.com).

H. J. Sadaei is with the Machine Intelligence and Data Science Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte 31270-010, Brazil (e-mail: h.javedani@gmail.com).

R. Ballini is with the Department of Economic Theory, Institute of Economics, University of Campinas, Campinas 13083-857, Brazil (e-mail: ballini@eco.unicamp.br).

F. G. Guimarães is with the Department of Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte 31270-901, Brazil, and also with the and the Machine Intelligence and Data Science Laboratory, Universidade Federal de Minas Gerais, Belo Horizonte 31270-010, Brazil (e-mail: fredericoguimaraes@ufmg.br).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2019.2922152

series patterns and translate them into the rule-based knowledge system (the probabilistic weighted fuzzy logical relationship groups). PWFTS is capable of forecasting points for one step ahead, intervals for one or many steps ahead, and probabilistic distributions for many steps ahead. This method does not make any parametric assumptions about the data and is computationally cheaper. The rules generated by the method are easy to update, human-readable, and their parameters make it flexible for several use cases.

The following are the main contributions of this paper:

1) development of the PWFTS, which is a new nonparametric, data driven, and highly accurate forecasting method;
2) a new representation method for fuzzy temporal rules, with weights on the precedent and the consequent of the rules, reflecting its *a priori* and *a posteriori* empirical probabilities;
3) new defuzzification methods capable of producing probability distributions, prediction intervals, and point forecasts. This is the first FTS method in the literature, which integrates point, interval, and probabilistic forecasting in the same model.

The method proposed was empirically tested with typical financial time series, and the results were compared with other standard FTS and statistical methods. The results show that the proposed method obtained accurate results and outperformed standard FTS methods. The proposed method also combines versatility, scalability, and low computational cost, making it useful on a wide range of application scenarios.

To explain the details of the method, this paper is organized as follows: Section II presents the related literature of FTS and probabilistic forecasting; Section III introduces the PWFTS method for generating the rules and a method for point forecasting one step ahead; Section IV describes the extensions for the standard method. In Section V, an empirical analysis is employed to validate the proposed method and compare its performance with that of other methods. Finally, Section VI concludes this paper.

II. Literature Review

A. FTS Models

FTS models are nonparametric models introduced by Song and Chissom [12] and based on fuzzy set theory [31]. These methods are easy to implement and very flexible, providing ways to deal with numeric and nonnumeric data. Some of FTS methods produce compact and human readable models of the time series behavior by using fuzzy rules, which could, in principle, be used by business experts and researchers.

There are several categories of FTS methods, varying mainly by their order and time-variance. The order indicates how many time-delays (lags) are used in modeling the time series. Given the time series data F, the first-order models use $F(t-1)$ data to predict $F(t)$, and the high-order models use $F(t-1), F(t-2), \ldots, F(t-k)$ data to predict $F(t)$. Time varying models require the updates of the current model with time to produce new forecasts.

Song and Chissom [12] proposed the main steps of all FTS methods, but its computation demanded many matrix operations for each forecasting, making the process computationally expensive. Later, Chen [32] simplified Song and Chissom's algorithm by creating the fuzzy logical rule groups (FLRGs), making the forecasting process cheaper by avoiding the use of matrix manipulations. FLRGs represent the knowledge base (rule base) of the model and are human readable and easy to interpret. Both the methods are known as conventional FTS models.

The initial step of training an FTS model is the partitioning of the universe of discourse, U, that is, the range of values covered by the training data. This is certainly one of the most crucial steps in FTS methods due to their influence on the forecasting accuracy and model over-fitting. The number of intervals, the length of these intervals, and their midpoints are all the parameters that should be chosen carefully or indeed optimized. The partitioning scheme initially proposed in the conventional FTS is just the division of the data range in k equal length intervals, the same method adopted hereafter in this paper just for simplicity.

However, more accurate methods have been proposed in the literature. Huarng [33] proposed an empirical method to find the ideal partition lengths. Li *et al.* [34] applied the fuzzy c-means clustering algorithm to find the best partitions' midpoints. Cheng [35] used the entropy of data to find the best partitions' midpoints. Enayatifar *et al.* [36] used an evolutionary algorithm to achieve the optimal partitioning scheme.

The generation of FLRG from the fuzzyfied data in FTS model has, at least, two drawbacks: 1) the loss of rule's recurrence; and 2) their chronological order. Thus, at the forecasting process, a very recurrent pattern of data has the same importance as that of a unique occurrence pattern. Moreover, newer and older patterns also have the same weight in the forecast. To fix these drawbacks, Yu [37] proposed the *weighted FTS* (WFTS) model by including weights on FLRGs. These weights are monotonically increasing and have a smoothing effect, giving more importance to the most recent data in the forecasting process. The work in [15] and [21] has presented the *improved weighted FTS* (IWFTS) model and changed the way in which the weights are assigned to the right-hand side (RHS) of the rules in Yu's model. The main difference is that the weights are calculated by the recurrence of each rule, discarding the chronological order. The *exponentially weighted FTS* (EWFTS) method, proposed in [38] and [20], replaces the linear weight growth of the WFTS model by an exponential growth. Lee *et al.* [39] proposed a broad generalization of the weighted methods with the *polynomial FTS*. This method demands the coefficient fitting by using optimization techniques but is capable to approximate WFTS, IWFTS, and EWFTS methods.

Above methods are first-order models. Higher order methods include [18], [40], and [41], and approaches that deal with the order, lag, and rule optimization as in [42]. Seasonal models include [43]. There are also hybrid FTS approaches such as [40], which proposes a high-order multivariable FTS algorithm based on fuzzy clustering; [30] combines statistical ARFIMA models with FTS for forecasting of long-memory time series; [44] uses genetic algorithms and neural networks coupled with

FTS. Several recent works are using the F-Transform defined in [45] as a method to represent and decompose the time series signal, as can be found in [46], [47], and [43].

All these models have some common drawbacks. First, in the forecasting step, just one FLRG from the rule base is chosen for computing the result, based on the maximum membership between the input value and all the FLRGs. This does not take advantage of the "smoothing" effect of fuzzy methods, which demands mixing many sets according to their fuzzy membership values. Second, these models are point-based forecasters and give no uncertainty measures about their results.

The first works to make the relationship of probabilities with fuzzy sets came from Zadeh, [48], [49], who defined the fuzzy set probability as the expectation of the membership function. Also, Klement *et al.* [50] and Dubois and Prade [51] explore the relationships between the fuzzy membership functions and the probability measures based on the measure theory. These theoretical works form the base of the fuzzy stochastic FTS in Song, *et al.* [52], where three models were presented, but unfortunately there is no empirical analysis of their results. In [53], the multivariate stochastic fuzzy forecasting is proposed based on the exponential smoothing among the diverse variables. Gangwar and Kumar [54] proposed the probabilistic and intuitionistic FTS method, strongly based on data normality and explicit Gaussian process assumption.

The fuzzy stochastic methods presented depend on knowing the data distribution beforehand, and some of them depend on the data normality. These assumptions make them less general, restricting their usability.

B. Probabilistic Forecasting

Gneiting and Katzfuss [2] define probabilistic forecasting as *"the form of a predictive probability distribution over future quantities or events of interest."* This definition encloses two main forecasting types: 1) intervals and 2) probability butions.

Interval forecasts take the form $\mathbb{I} = [\underline{a}, \overline{b}]$, where $\underline{a}$ represents the lower bound and $\overline{b}$ the upper bound of uncertainty and usually represent the quantiles of an estimate $\hat{Y}(t)$ or the conditional variance $Var[y(t) \mid y(t-1), y(t-2), \ldots]$. A simple method for creating prediction intervals for generic forecasting models was proposed in [55], namely the mean-variance model. From the point forecast, $\mu = \mathbb{E}[Y_{t+1} \mid Y_t, Y_{t-1}, \ldots]$ with the variance of the residuals, $\sigma_\epsilon = \sqrt{VAR[\epsilon]}$, by assuming that these residuals follow $\epsilon \sim \mathcal{N}(0,1)$. The prediction interval is calculated as $\mathbb{I} = [\mu - z_{\alpha/2}\sigma_\epsilon \ , \ \mu + z_{\alpha/2}\sigma_\epsilon]$, and $z_{\alpha/2} = \Phi((1-\alpha)/2)$ is the standard normal distribution function.

The main probabilistic approach for interval forecasting is the quantile auto regression (QAR) [56], based on the quantile regression [57]. QAR approaches have been used in many application fields, for instance, energy load forecasting [6] and wind forecasting [9]. Each QAR model is fitted for a specific τ, the quantile value; hence, for a certain confidence level α, two QAR models are needed (α and $1-\alpha$). The independence of quantiles also allows one to create asymmetric interquantile intervals, if needed.

In the FTS literature, Silva *et al.* [58] proposed the *Interval FTS* (IFTS). In this method, the length of forecasting interval is the measure of the fuzzy uncertainty, where wide intervals mean high uncertainty, and thin intervals mean low uncertainty. The main drawback of this method is considering only the fuzzy uncertainty, discarding the stochastic behavior of the time series.

On the other hand, in the probability distribution forecast methods, there is a probability distribution function (PDf) $P(y(t))$, where $P : U \rightarrow [0,1]$, and U is the sample space of time series $Y(t)$, which associates an occurrence probability for each $y(t) \in Y(t)$. A common approach for probabilistic forecasting is the use of ensembles of models and forecast combination. This is not a new concept, see for instance [59], and start on the assumption that mixing different forecasting sources may improve overall forecasting. This is slightly close to the concept of the ensemble methods defined in [60] as *"an ensemble prediction system consists of multiple runs of numerical weather prediction models, which differ in the initial conditions."* Also, Leutbecher and Palmer [5] states that *"The ultimate goal of ensemble forecasting is to predict quantitatively the probability density."*

These ensembles based on Monte Carlo methods can be homogeneous (same method with different parameters) or hybrid (different methods with different parameters). The internal methods are executed several times, and the larger is the sample, the better are the approximations made. After n runs, the empirical distribution $P(Y_i)$ of the outputs is available. Initially, ensemble learning methods were developed to produce point forecasts as a combination of the individual models forecasts by a weighted average or more complex methods such as the Bayesian model averaging; see for instance [61]. Soon after, these methods were adapted for probabilistic forecasting as in [5], [62], and [63]. In [64], a methodology for electric load probabilistic forecasting is proposed, combining point forecasting and scenario-based probabilistic forecasting.

Several other nonparametric approaches are possible, for instance, machine learning models such as k-Nearest Neighbors (kNN) and kernel density estimators (KDEs). Both the approaches smooth the discrete values in a continuous function, the kernel function, which approximates the empirical distribution of data. A review of density estimation methods can be found in [65], and a specific study on estimation of the kernel's parameters can be found in [66].

Some other non-FTS based and soft-computing modeling approaches exist in the literature, to solve similar problems. For instance, the use of nature-inspired optimization algorithms to induce fuzzy type-2 rules and to generate prediction intervals is presented in [67]. The projectional neural networks are used in [68] to forecast complex nonlinear systems. Bayesian filters are used in [69] to create probabilistic forecasts. When compared with these and other approaches, some advantages of PWFTS become clear: It is a white box and data-driven method that does not require computationally expensive optimization methods for training its models or even making parametric assumptions on its forecasts.

Previous work of the present authors tackled the interval forecasting by using FTS, in [58], and by using general fuzzy rules based systems, in [70]. In [71], the authors proposed the *ensemble FTS*. In this method, the point forecasting results of individual FTS models are combined using a KDE to produce a probability density. The main drawback of this method is the computational cost involved in training and using the ensemble. The probabilistic forecasting was also tackled using an evolutionary algorithm in [10], using Monte Carlo simulation. However, these approaches are limited in their application when compared to the proposed method, whose flexibility allows it to be used in a wider range of applications.

III. Probabilistic Weighted Fuzzy Time Series

The PWFTS is a nonparametric and rule-based method that extends previous WFTS approaches. More importantly, the method revises methodological aspects to fix the known drawbacks in FTS methods. First, it proposes to separate the model training procedure of the forecasting procedure, also renaming the controversial terms in the FTS literature [such as fuzzy logical relationships (FLRs) and FLRG] to new ones. The proposed PWFTS, also, is a truly fuzzy method, fixing the maximum membership trick of several FTS methods and taking into account all the active fuzzy rules that have membership values greater than zero.

The core concept of PWFTS is the fuzzy empirical probability, used to compute the weights of the model, whose intuition is discussed in the next section.

A. Fuzzy Empirical Probabilities

The initial Zadeh's proposition of fuzzy probability, $P(A) = E[\mu_A]$, proposed in [48], demands the previous knowledge of the probability distribution over the universe of discourse. Since this distribution is unknown, an empirical distribution must take place. The simplest definition of empirical probability is the relative frequency of a discrete value or of a range of continuous values. The fuzzy theory provides a different look on the traditional probability theory because it affects the way the events are counted. On fuzzy sets, the notion of event is more complex because the same value can belong to several sets with different degrees of membership. In that case, instead of accounting the integral (i.e., unary) occurrence of the event, their partial occurrence is accounted as the membership value. This method is known by fuzzy frequency and was first developed in [72]. A related formulation can be found in [45], with the concept of F-Transform, which decomposes the original domain of the time series into fuzzy frequencies over the fuzzy sets. This decomposition can also recreate the time series using the inverse transform.

Given the sample space $U \subset \mathbb{R}$ and the fuzzy sets A, the empirical probability of a fuzzy set $A_i \in A$ is the sum of their memberships $\mu_{A_i}(y)\ \forall y \in U$ divided by the sum of the partition functions Z_{A_i} of all fuzzy sets $A_j \in A$, as presented in (1). The partition function Z_{A_i} is the integral (or some approximation) of the membership function μ_{A_i} over the sample space U, or $Z_{A_i} = \int_U \mu_{A_i}(y)dy$, or a simple discrete approximation $Z_{A_i} = \sum_{y \in U} \mu_{A_i}(y)$. The intuition behind this formula is that the empirical probability $P(A_i)$ is evenly spread over the shape of the fuzzy membership function of A_i and that point y is a slice of this shape whose area is equal to value $\mu_{A_i}(y)$, and the area of μ_{A_i} is Z_{A_i}. One has the following:

$$P(A_i) = \frac{\sum_{y \in U} \mu_{A_i}(y)}{\sum_{A_j \in A} Z_{A_j}}. \tag{1}$$

$P(A_i)$ is measured from a sample of U, and this empirical value is an approximation of the real probability. This approximation is used in (2) to approximate the conditional probability of a value $y \in U$, given a fuzzy set $A_i \in A$. One has the following:

$$P(y|A_i) = P(A_i) \cdot \frac{\mu_{A_i}(y)}{Z_{A_i}}. \tag{2}$$

Using (1) and (2) and the law of total probability, the empirical probability $P(y)$ can be approximated as

$$P(y, A) = \sum_{A_i \in A} P(y|A_i) \cdot P(A_i). \tag{3}$$

The advantage of this approach is the convenience in obtaining $P(A_i)$ from a time series dataset Y. The accuracy of $P(A_i)$ will be determined by k, the number of partitions of the universe of discourse U.

B. Training Procedure

The training procedure is a seven-step method, illustrated in Fig. 1, to learn the temporal dynamics of the time series training data Y and represent it on a fuzzy-probabilistic model, namely the probabilistic weighted fuzzy temporal pattern group (PWFTPG). The steps of the method are listed in the following.

Step 1 *Define the universe of discourse:* Define U as the sample space of in-sample training data Y, such that $U = [\min(Y) - D_1, \max(Y) + D_2]$, where $[\min(Y), \max(Y)]$ is the range of in-sample data and D_1 and D_2 the numbers used to extrapolate this range; for instance, $D_1 = 0.1 \cdot \min(Y)$ and $D_2 = 0.1 \cdot \max(Y)$.

Step 2 *Partitioning:* Split U in k even length intervals u_i, for $i = 1, \ldots, k$, with midpoints mp_i.

Step 3 *Define the linguistic variable A:* Create k overlapping fuzzy sets A_i, with membership functions μ_{A_i}, related to an interval u_i, and midpoints mp_i. Each fuzzy set $A_i \in A$ is a linguistic term of the linguistic variable A.

Step 4 *Fuzzification:* Transform the original numeric time series Y into FTS F, whose each data point $f(t) \in F$ is a k-tuple with the membership value of $y(t)$ with respect to each linguistic term $A_i \in A$, such that

$$f(t) = [\mu_{A_1}(y(t)),\ \mu_{A_2}(y(t)),\ \ldots,\ \mu_{A_k}(y(t))]. \tag{4}$$

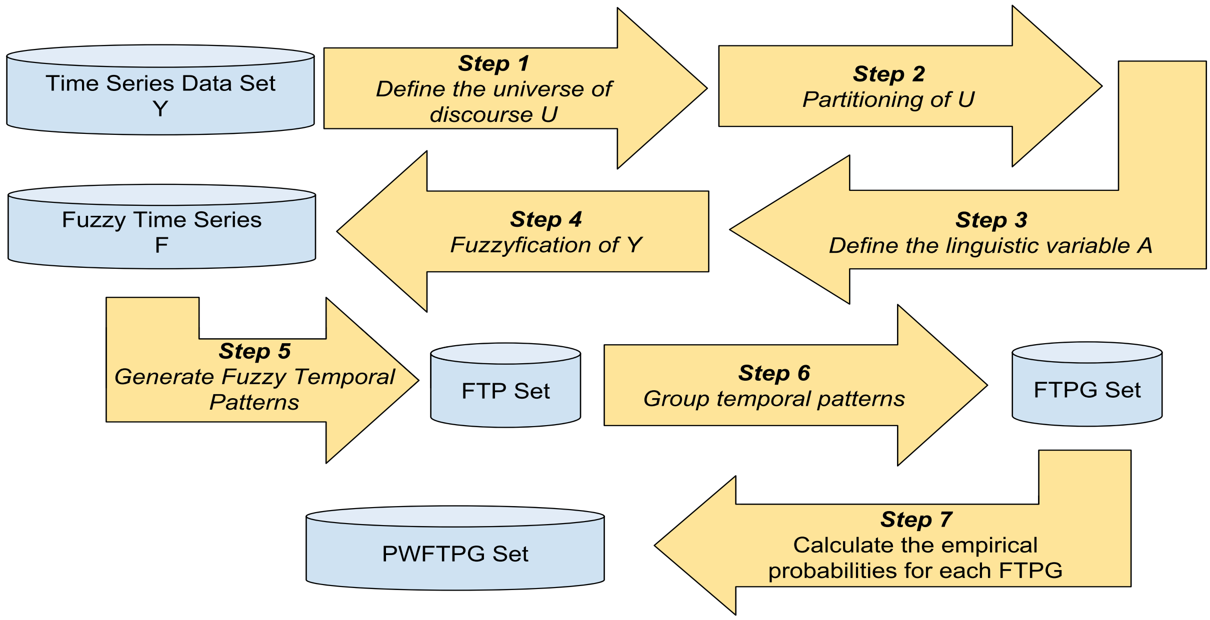

Fig. 1. PWFTS training procedure.

Step 5 *Generate the fuzzy temporal pattern (FTP) set:* The FTP[1] is a fuzzy rule with format $A_i \rightarrow A_k$ that indicates a temporal succession where the precedent [or the left-hand side (LHS)] is $A_i \in f(t)$, and the consequent (or the RHS) is $A_k \in f(t+1)$, for each possible pair of $A_i \times A_k$ of membership values greater than zero; i.e., $\{A_i \rightarrow A_k\}\ \forall A_i \in f(t) \mid \mu_{A_i}(y(t)) > 0$ and $\forall A_k \in f(t+1) \mid \mu_{A_k}(y(t+1)) > 0$. Therefore, $A_i \rightarrow A_k$ can be read as "IF $y(t)$ is A_i THEN $y(t+1)$ is A_k." As each $f(t) \in F$ is a sparse k-vector of membership values, there will be many possible fuzzy sets' combinations of two sequential vectors $f(t)$ and $f(t+1)$. Then, for each sequential pair on F, possibly more than one FTP will be generated.

Step 6 *Generate the fuzzy temporal pattern group (FTPG) set:* An FTPG[2] represents the set of all the FTPs with the same LHS and the union of their RHS, with the format $A_i \rightarrow A_k, A_j, \ldots$, where the LHS is $f(t) = A_i$ and the RHS is $f(t+1) \in \{A_k, A_j, \ldots\}$. Each FTPG can be understood as the set of possibilities that may happen at time $t+1$ (the consequent) when a certain set A_i is identified at time t (the precedent).

Step 7 *Calculate empirical probabilities:* The PWFTPG adds weights on the LHS and RHS that measure their fuzzy empirical probabilities. Each PWFTPG has the format as

$$P_i \cdot A_i \rightarrow w_{i0} \cdot A_0, \ldots, w_{ik} \cdot A_k$$

for $i = 1, \ldots, k$. Each weight P_i is the normalized sum of all the LHS memberships of all FTPs where the LHS is fuzzy set A_i, as in (1). P_i can be understood as the empirical *a priori* probability of fuzzy set A_i independent of time, or $P(A_i)$, such that $\sum_{i \in A} P_i = 1$. Each weight w_{ij} is associated with a fuzzy set A_j on the RHS of the FTP whose LHS is A_i, and it is the normalized sum of all the RHS memberships of all FTPs where the $LHS = A_i$ and $RHS = A_j$. Therefore, weight w_{ij} can be understood as the empirical conditional probability of the fuzzy set A_j on time $t+1$ when the fuzzy set A_i is identified on time t, or $P(A_j^{t+1} \mid A_i^t)$, such that $\sum_{j \in A} w_{ij} = 1$ for each A_i in the LHS.

The final PWFTPG set is shown in (5). Its size depends on the number of partitions k, and it could be represented in matrix form, but weights w_{ij} form a very sparse matrix, which justifies using optimized data structures for its representation. One has the following:

$$\begin{aligned} &P_1 \cdot A_1 \rightarrow w_{11} \cdot A_1, \ldots, w_{1k} \cdot A_k \\ &\cdots \quad \cdots \quad \cdots \\ &P_k \cdot A_k \rightarrow w_{k1} \cdot A_1, \ldots, w_{kk} \cdot A_k. \end{aligned} \tag{5}$$

The outcome of the training procedure is the PWFTPG set, and it represents the temporal dynamics of the original data. It is an empirical probability distribution of linguistic variable A over time series Y with sample space U, where each rule contains the unconditional probability $P(A_i) = P_i$ and conditional probabilities $P(A_j|A_i) = w_{ij}$, for $A_i, A_j \in A$.

C. Probabilistic Forecasting Procedure

The forecasting procedure is a four-step procedure listed in this section, which takes as input a sample $y(t) \in U$ and uses the PWFTPG model learned in the previous section to generate the output, which depends on the type of forecasting.

[1]This nomenclature is adopted in the replacement of the FLRs used in [12], to avoid misunderstandings with the terms "logical" and "relationship" with their classical meanings in fuzzy theory literature.

[2]In the replacement of FLRG used in [73].

In this section, a probability distribution $P(\cdot|y(t), A)$, for all $y(t+1) \in U$ will be computed using a mixture distribution approach to transform each PWFTPG probability into a continuous distribution, as described in (7).

A mixture distribution is defined as $P(y) = \sum \omega_j \cdot \pi_j(y)$, where $\pi_j : U \to [0,1]$ are specific PDFs and ω_j a weight associated with each PDF, such that $\sum \omega_j = 1$. Given an input value $y(t) \in Y$ and the PWFTPG set, the probability distribution for each $y(t+1) \in U$ is given by (7), where ω_j is replaced by probability $P(y(t)|A_i)$, the LHS probability, given the input value, and distribution π_j is replaced by probability $P(y(t+1)|A_j, A_i)$ $\forall A_j \in RHS$.

Looking back to (2), it is pretty obvious that

$$\sum_{i=1}^{k} P(y(t)|A_i) < 1.$$

Once $P(A_i)$ is the probability of the whole fuzzy set and $y(t)$ just a small slice of it, it does not comply with the $\sum \omega_j = 1$ restriction of the mixture distribution. To work around this issue, $P(y(t)|A_i)$ is rescaled to

$$\frac{P(y(t)|A_i)}{\sum_{j \in A} P(y(t)|A_j)}.$$

Therefore, considering $x = y(t+1)$

$$P(x|y(t), A) = \sum_{i=1}^{k} \frac{P(y(t)|A_i)\left(\sum_{j=1}^{k} P(x|A_j, A_i)\right)}{\sum_{j=1}^{k} P(y(t)|A_j)} \tag{6}$$

$$P(x|y(t), A) = \sum_{i=1}^{k} \frac{P_i \frac{\mu_{A_i}(y(t))}{Z_{A_i}} \left(\sum_{j=1}^{k} w_{ij} \frac{\mu_{A_j}(x)}{Z_{A_j}}\right)}{\sum_{j=1}^{k} P_j \frac{\mu_{A_j}(y(t))}{Z_{A_j}}}. \tag{7}$$

The complete forecasting procedure is presented in the following.

Step 1 *Fuzzification:* For a given input value $y(t) \in Y$, find the fuzzyfied values $f(t) = \{A_i \mid \mu_{A_i}(y(t)) > 0\}$.

Step 2 *Pattern matching:* Locate all the PWFTPGs whose LHS is $f(t)$.

Step 3 *Forecast:* The distribution of $f(t+1)$ is given by the RHS sets of each PWFTPG matched.

Step 4 *Defuzzification:* Build the probability distribution $P(y(t+1)|y(t))$ $\forall y(t+1) \in U$ applying (7).

The PWFTS overall forecasting method is summarized in Algorithm 1, which takes as input the forecasting type, a sample $y(t) \in U$, and uses the PWFTPG model $\mathcal{M}$ learned in the previous section to generate the output, which depends on the type of forecasting (probabilistic, interval, or point forecasting).

Algorithm 1: PWFTS Forecasting Method.

```
   input : M: the PWFTPG model
           y(t) ∈ U: the input sample
           type: the forecasting type
1  begin
2      f(t) ← {A_i | μ_{A_i}(y(t)) > 0} ;     // Fuzzification
3      active_rules ← ∅;
4      foreach PWFTPG ∈ M do
5          if f(t) matches the LHS then  // Rule Matching
6              active_rules ← this PWFTPG
7          end
8      end
9      if type ='probabilistic' then
10         P ← ∅;
11         foreach x ∈ U do
12             Apply Eq. (7) to compute P(y(t + 1))
13         end
14     else if type ='interval' then
15         Apply Eq. (10) to compute 𝕀(t + 1)
16     else
17         Apply Eq. (12) to compute y(t + 1)
18     end
19 end
```

D. Interval Forecasting Procedure

With $P(\cdot|y(t))$, it is possible to build a cumulative density function $F(\cdot|y(t))$ and use it to construct the quantile function $Q(\tau) : [0,1] \to U$ as

$$Q(\tau) = \min\{x \in U \mid F(x|y(t)) = \tau\} \tag{8}$$

where $\tau \in [0,1]$ is the desired quantile. Then, for a certain confidence level $\alpha \in [0,1]$, it is possible to compute an interquantile interval $\mathbb{I}_f = [Q(\alpha), Q(1-\alpha)]$.

However, the above method demands the computation of the whole PDF. A simpler and fast heuristic for generating intervals is to compute the maximum uncertainty interval, i.e., the interval over the bounds of each fuzzy set. Each PWFTPG will be represented by an interval $\mathbb{I}$ whose bounds are the expectation of the bounds of its RHS fuzzy sets, such that $\underline{A_i}$ and $\overline{A_i}$ represent the lower and upper bounds of fuzzy set A_i, and $\mathbb{E}[A_i]$ is the expectation of the PWFTPG where the LHS is A_i. The forecasting interval $\mathbb{I}(t+1)$, then, is the sum of these expectations weighted by $P(y(t)|A_i)$ probabilities, as in (10). One has the following:

$$\mathbb{I}_i = [\underline{\mathbb{E}[A_i]}\,,\ \overline{\mathbb{E}[A_i]}]$$

$$\underline{\mathbb{E}[A_i]} = \sum_{A_j \in A_i^{\text{RHS}}} w_{ij} \cdot \underline{A_j}$$

$$\overline{\mathbb{E}[A_i]} = \sum_{A_j \in A_i^{\text{RHS}}} w_{ij} \cdot \overline{A_j} \tag{9}$$

$$\mathbb{I}(t+1|y(t), A) = \left[\frac{\sum_{i=1}^{k} P(y(t)|A_i) \cdot \underline{\mathbb{I}_i}}{\sum_{i=1}^{k} P(y(t)|A_i)},\ \frac{\sum_{i=1}^{k} P(y(t)|A_i) \cdot \overline{\mathbb{I}_i}}{\sum_{i=1}^{k} P(y(t)|A_i)}\right]. \tag{10}$$

Equation (10) is used in line 15 of Algorithm 1 to obtain the interval forecast $\mathbb{I}(t+1)$.

E. Point Forecasting Procedure

To produce point forecasts from the existing distribution $P(\cdot|y(t))$, it is only needed to apply the expectation operator, such that $y(t+1) = \mathbb{E}[P(\cdot|y(t))]$. This is also computationally expensive due to the computation of $P(\cdot|y(t))$. A simple heuristic for producing point forecasts is to compute the expectation $\mathbb{E}[A_i]$ of each PWFTPG, as presented in (11), where mp_j is the midpoint of each fuzzy set $A_j \in RHS$. The expectation $\mathbb{E}[A_i]$ for each PWFTPG A_i is constant and can be precomputed. The final forecast $y(t+1)$, then, is the sum of these expectations weighted by $P(y(t)|A_i)$ probability, as shown in (12). One has the following:

$$\mathbb{E}[A_i] = \sum_{j \in A_i^{\text{RHS}}} w_{ij} \cdot mp_j \tag{11}$$

$$y(t+1|y(t), A) = \sum_{i=1}^{k} \frac{P(y(t)|A_i) \cdot \mathbb{E}[A_i]}{\sum_{i=1}^{k} P(y(t)|A_i)}. \tag{12}$$

Equation (12) is used in line 17 of Algorithm 1 and to defuzzify the point forecast $y(t+1)$.

IV. PWFTS Extensions

A. Multistep Ahead Forecasting

The forecasting procedures listed in Section III are one step ahead methods. To extend the forecasting procedures to multistep ahead forecasting, an iterative approach is adopted, in which the $(t+1)$th step is computed with the previous presented methods, and its output is fed back as an input to the next m steps. From the $(t+2)$th step onward, let $s \in [t+2, t+m]$ be the new time indexer. The simpler approach is to perform the point forecast of $y(s+1)$ with input $y(s)$.

The interval procedure requires a few more modifications. Given the input $\mathbb{I}(s)$, the same interval forecasting procedure will be executed with inputs $\underline{\mathbb{I}(s)}$ and $\overline{\mathbb{I}(s)}$, producing two new intervals $\underline{\mathbb{I}(s+1)}$ and $\overline{\mathbb{I}(s+1)}$. Then, the final forecasting interval will be $\mathbb{I}(s+1) = [\min\{\underline{\mathbb{I}(s+1)}\}, \max\{\overline{\mathbb{I}(s+1)}\}]$.

Finally, the probabilistic forecasting for $P(\cdot|y(s))$, given the input, will change to (13), instead of (7). This equation replaces $P(y(s)|A_i)$ for the previous probability distribution $P(\cdot|y(s))$. One has the following:

$$P(y(s+1)|y(s)) = \sum_{i=1}^{k} \frac{P(y(s)|y(s-1), A_i)}{\sum_{j=1}^{k} P(y(s)|y(s-1), A_j)} \times \left(\sum_{z=1}^{k} P(y(s+1)|A_z, A_i) \right). \tag{13}$$

B. High-Order Models

The PWFTS method described in Section III is a first-order method; i.e., it just needs $y(t)$ to forecast $y(t+1)$, while high-order models use m time lags. To extend the standard approach to high order, a modification in Step 5 of the training procedure is needed to adapt the FTPs and FTPGs to store m fuzzy sets on their LHS.

Once the fuzzyfied value $f(t)$ has multiple fuzzy sets (with different membership values greater than zero), a set of fuzzyfied values $f(t-m), \ldots, f(t)$ must be represented with all possible combinations between the fuzzy sets of each lag, such as $f(t-m) \times f(t-m-1) \times \cdots \times f(t)$, where $\times$ represents the Cartesian product operator.

In Step 5, the FTPs will have the format $A_i^m, A_i^{m-1}, \ldots, A_i^0 \rightarrow A_j$, which can be read as "IF $f(t-m)$ is A_i^m AND $f(t-m-1)$ is A_i^{m-1} AND ... AND $f(t)$ is A_i^0 THEN $f(t+1)$ is A_j." In Step 6, the high-order FTPGs gather all the high-order FTPs with the same LHS.

In Step 7, P_i weight is replaced by P_{LHS} that aggregates the μ_{LHS} memberships of each FTPG for the samples. Given a sample $y(t-m), \ldots, y(t) \in Y$ with m lags, their membership grades with a FTPG is the product T-norm between all the memberships of the LHS. One has the following:

$$\mu_{\text{LHS}}(y(t-m), \ldots, y(t)) = \bigcap_{i=m}^{0} \mu_{A_i}(y(t-i)). \tag{14}$$

In the forecasting procedure, Step 1 requires a sample with m lags that will generate m fuzzyfied values. In Step 2, all the combinations between the fuzzy sets of each fuzzyfied lag will be the LHS of the affected PWFTPGs. In Step 3, in (7), (10), and (12), the empirical conditional probability $P(y(t)|A_i)$ will be replaced by $P(y(t-m), \ldots, y(t)|LHS)$, the empirical conditional probability of sample $y(t-m), \ldots, y(t)$, given the LHS of the PWFTPG. One has the following:

$$P(y(t-m) \ldots y(t)|LHS) = P_{\text{LHS}} \frac{\mu_{\text{LHS}}(y(t-m) \ldots, y(t))}{\sum_{A_i \in \text{LHS}} Z_{A_i}}. \tag{15}$$

V. Computational Experiments

To measure the performance of the proposed models, three well-known financial time series data [the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX), Standard & Poor's 500 (S&P 500), and National Association of Securities Dealers Automated Quotations (NASDAQ) datasets] were selected, each of them with 5000 instances. A rolling window cross-validation methodology [74] was applied, using a working set of 1000 instances, 800 for training (80%), 200 for testing (20%), and a sliding increment of 200 instances, totaling 23 experiments, and all the measurements were performed out of sample.

The TAIEX[3] is a well-known economic time series data commonly used in the FTS literature (see, for instance, [37], [75], [76], [17], [77], [30], etc.). This dataset is sampled from 1995 to 2014 time window and has the averaged daily index by business day. This is a stationary time series dataset whose augmented Dickey–Fuller (ADF) statistic is -2.65, where the critical value for $\alpha = 0.05$ is -2.86.

[3][Online]. Available: http://www.twse.com.tw/en/products/indices/Index_Series.php. Accessed on May 23, 2016.

The NASDAQ8–Composite Index (NASDAQ ÎXIC)[4] is an economical index already used in the FTS literature (see [75], [28], [78], and [20]). The historical data were sampled from 2000 to 2016 time window and have the averaged daily index by business day. This is a stationary time series dataset whose ADF statistic is 0.04, where the critical value for $\alpha = 0.05$ is -2.86.

The S&P500[5] is a market index composed by 500 assets quoted on the New York Stock Exchange and the NASDAQ. This dataset was already used in [78] and [77] and contains the averaged daily index, by business day, from 1950 to 2017 with 16 000 instances. This is a stationary dataset whose ADF statistic is 0.00, where critical value for $\alpha = 0.05$ is -2.86.

For all these datasets, PWFTS models were obtained for orders $n \in \{1, 2, 3\}$ and several different numbers of partitions. The TAIEX and NASDAQ datasets were tested with the original data and with differentiation transformed data, generally obtaining more accurate results. For saving space, only the best results of each method were reported in this paper. Therefore, all the methods are compared by taking the best performing model of each method, instead of fixing the same hyper-parameters for all the methods.

In the following sections, the benchmarks for point, interval, and distribution forecasts will be presented, where the evaluation metrics are discussed, and the presented models are compared with other models in the FTS literature.

The hypothesis testing procedures adopted best practices discussed in [79]–[81]. The Friedman aligned ranks test [82] nonparametric procedure was adopted to test the equality of the means, where the null hypothesis H_0 stands for the equality of all means and the inability to distinguish between the methods, and alternative hypothesis H_1 stands for the difference of the means and the distinguishability between the models. The paired *post hoc* procedure adopted was the Finner test [83], in a one-versus-all design, where the PWFTS method is taken as control method. In Finner test, the null hypothesis H_0 stands for the equality between the control and the test methods, and the alternative hypothesis H_1 stands for the significant difference between the control and test methods. All the tests adopted the significance level $\alpha = .05$ and were performed on STAC framework [84], and all FTS methods were tested with the pyFTS library[6] [85].

In order to contribute to the replication of all the results in this paper, the data of the instances, full results, and all the source codes for this paper are provided as the supplementary material. All this will also be made available online.[7]

In the experiments with other FTS models, the universe of discourse was partitioned in a grid scheme, where all the partitions have the same length. Each model was trained and tested with the number of partitions on the [10,100] range for original data and on the [3, 30] range for differentiated data. The models trained with differentiated data generally perform better than the ones trained with original data, and only their best results were selected.

[4][Online]. Available: http://www.nasdaq.com/aspx/flashquotes.aspx?symbol=IXIC&selected=IXIC. Access in May 23, 2016.

[5][Online]. Available: https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC. Accessed on Mar. 19, 2017.

[6][Online]. Available: https://pyfts.github.io/pyFTS/. Accessed in Jul. 1, 2018.

[7][Online]. Available: https://github.com/petroniocandido/PWFTS

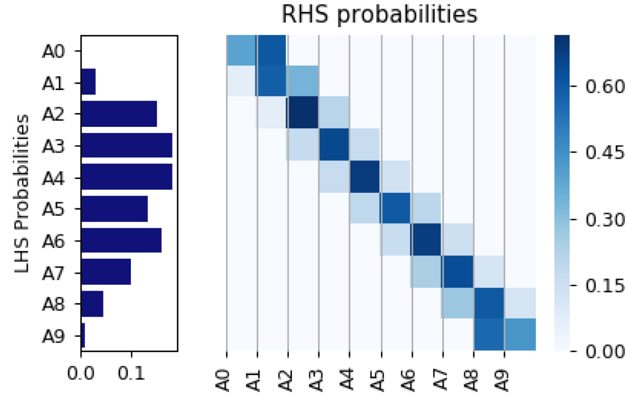

Fig. 2. Example PWFTPG set.

A. Method Parameterization

All FTS methods are affected mainly by the partitioning of the universe of discourse and data transformations, and their performances also depend on the stationarity of the data. There is a tradeoff between the complexity and accuracy directly related with the partitioning scheme. Less partitions make the model simpler but also less accurate, and decreasing the partition number can lead to under-fitting. Increasing the number of partitions also increases the accuracy until the model is lead to over-fitting. The best fit, generally, is data dependent and is achieved through grid search. However, transformations that make data stationary and homoscedastic can reduce or minimize this effect.

To assess the sensitivity of the proposed method to these parameters, several experiments were performed, where the number of partitions has more impact for nontransformed data and the overall best partition number is 35. With the transformed data, less partitions produce better accuracy, though the difference is not remarkable. The Occam's razor principle can be adopted to choose the simplest partitioning scheme. It is also important to note that the number of partitions affects also the length of the prediction intervals and the probabilistic forecasting.

A simple PWFTPG model, with the 10 even-length partition scheme, is shown in Fig. 2, where the empirical probabilities distribution for the TAIEX dataset is represented. One has the following:

$$(0.004)A0 \rightarrow (0.4)A0, (0.6)A1$$
$$(0.032)A1 \rightarrow (0.1)A0, (0.6)A1, (0.3)A2$$
$$(0.15)A2 \rightarrow (0.1)A1, (0.7)A2, (0.2)A3$$
$$(0.181)A3 \rightarrow (0.2)A2, (0.6)A3, (0.2)A4$$
$$(0.182)A4 \rightarrow (0.15)A3, (0.7)A4, (0.15)A5$$
$$(0.134)A5 \rightarrow (0.2)A4, (0.6)A5, (0.2)A6$$
$$(0.16)A6 \rightarrow (0.15)A5, (0.7)A6, (0.15)A7$$

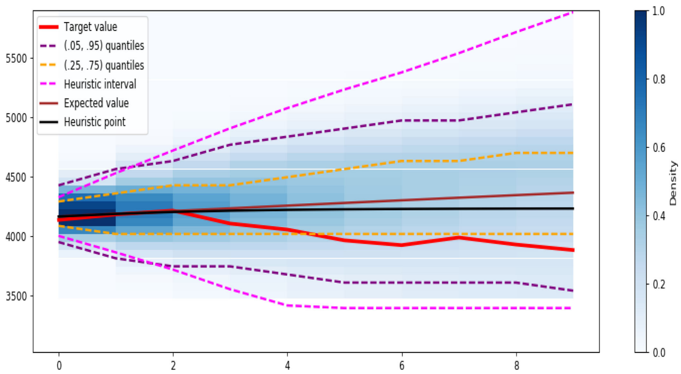

Fig. 3. Sample of multiple step forecasting for the TAIEX dataset.

$$(0.101)A7 \rightarrow (0.25)A6, (0.6)A7, (0.15)A8$$
$$(0.045)A8 \rightarrow (0.25)A7, (0.6)A8, (0.15)A9$$
$$(0.01)A9 \rightarrow (0.6)A8, (0.4)A9. \quad (16)$$

The forecasting horizon affects the accuracy; that is, the error and the uncertainty both increase as the forecasting horizon grows, see Fig. 3. The point forecasts tend to the unconditional expected value of the model as the horizon increases, and the prediction intervals tend to reach the bounds of the universe of discourse.

The relation between all kinds of forecasting provided by the model is illustrated in Fig. 4. For a given sample of the TAIEX dataset, the probabilistic forecast for each point was calculated, and with the probability distribution, the interval quantile forecasts for $\tau = [0.05, 0.95]$ and $\tau = [0.25, 0.75]$ and the point interval with the expected value of the distribution were generated. Additionally, the heuristic methods for interval and point forecasts were calculated. Visually, the probabilistic forecast is well suited around the target data. The heuristic intervals fall between the $[0.05, 0.95]$ and $[0.25, 0.75]$ quantile intervals, and the expected value and the heuristic point forecasts are similar with the residual difference between them.

B. Probabilistic Forecasting

To evaluate the probabilistic forecasting methods, the continuous ranked probability score (CRPS) was chosen. The CRPS, defined in [86], is a proper measure for probabilistic forecasts. It provides a direct way to benchmark probabilistic forecast since it is expressed in the same unit as the observed variable and is a generalization of the mean absolute error (MAE). Thus, the perfect score for the CRPS, as in the MAE, is 0.

The chosen methods for comparison were the ARIMA mean-variance model [55], Ensemble FTS [87], kNN with KDE (kNN + KDE) [88], [89] and QAR [56], [57]. For the ARIMA model, the standard Box–Jenkins methodology was employed for the identification of the model's parameters, and the probability distribution was built from the prediction intervals with $\alpha \in \{0.05, 0.15, 0.25, 0.35, 0.45\}$. Ensemble FTS was trained with the main FTS methods, for orders 1, 2, and 3. The kNN + KDE method uses $k = 30$, with the Epanechnikov kernel and bandwidth $h = 0.55$. The QAR model was tested with orders 1 and 2, and the probability distribution was built from the prediction intervals with $\alpha \in \{0.05, 0.15, 0.25, 0.35, 0.45\}$.

The experiments results of the best performing models are presented in Table I. The equality of means statistical analysis rejected the null hypothesis with the test statistic of 11.31693 and p-value of 0.02322. The results of *post hoc* tests in Table II show significant differences between PWFTS and k-NN and ensemble FTS. Although PWFTS obtained better average rank than those of ARIMA and QAR, the test did not show significant difference between them.

C. Interval Forecasting

Usually three metrics are used to evaluate prediction intervals: *1) coverage rate*; *2) calibration;* and *3) sharpness* [9], [90]. The coverage refers to the statistical consistency between the forecasts and the observations, and it measures which proportion of the observations are inside the prediction interval, in which the ideal value is 100%. The properties of sharpness and resolution refer to the concentration of the predictive distribution, or how wide and variable are the intervals and refer uniquely to the forecasts. The sharpness is the average size of the intervals and the resolution the variability of the intervals.

While small values of sharpness are desirable, meaning a compact interval, wide values of resolution are better, meaning the capability of the model to adapt the length of interval with the increase in uncertainty. There are no absolute reference values for sharpness and resolution, which depend on the statistical properties of the data. Empirically, when the sharpness is reduced to make the intervals thinner and more precise, the risk of reducing the coverage increases, and that is why the resolution is important.

However, using three separate metrics makes the analysis of interval forecasters more complex. The most common option in these cases is the Winkler score [91], which encompasses the three characteristics in only one measure. Given a target value y and a prediction interval $I = [\underline{L}, \overline{U}]$ with nominal probability $(1 - \alpha)$, the Winkler score is defined in (17), where $\delta = \overline{U} - \underline{L}$. The score value is the interval width, but it increases when the target value is not covered by the interval and the penalty is proportional to the error, given the nominal probability. Lower values, therefore, represent better prediction intervals. The mean score is defined in (18), where n is the sample size

$$S(\alpha, y, I) = \begin{cases} \delta & \text{if } \underline{L} \leq y \leq \overline{U} \\ \delta + 2(\underline{L} - y)/\alpha & \text{if } y < \underline{L} \\ \delta + 2(y - \overline{U})/\alpha & \text{if } \overline{U} < y \end{cases} \quad (17)$$

$$S_\alpha = \sum_{i=1}^{n} S(\alpha, y_i, I_i). \quad (18)$$

The chosen benchmarking methods were the ARIMA mean-variance model, IFTS [58], and QAR. For the ARIMA model, the standard Box–Jenkins methodology was employed for the identification of the model's parameters, and the prediction intervals with $\alpha = 0.05$. The QAR model was tested with orders 1 and 2 and prediction intervals with $\alpha = 0.05$.

The experiments results of the best performing models are presented in Table III. The equality of means statistical analysis

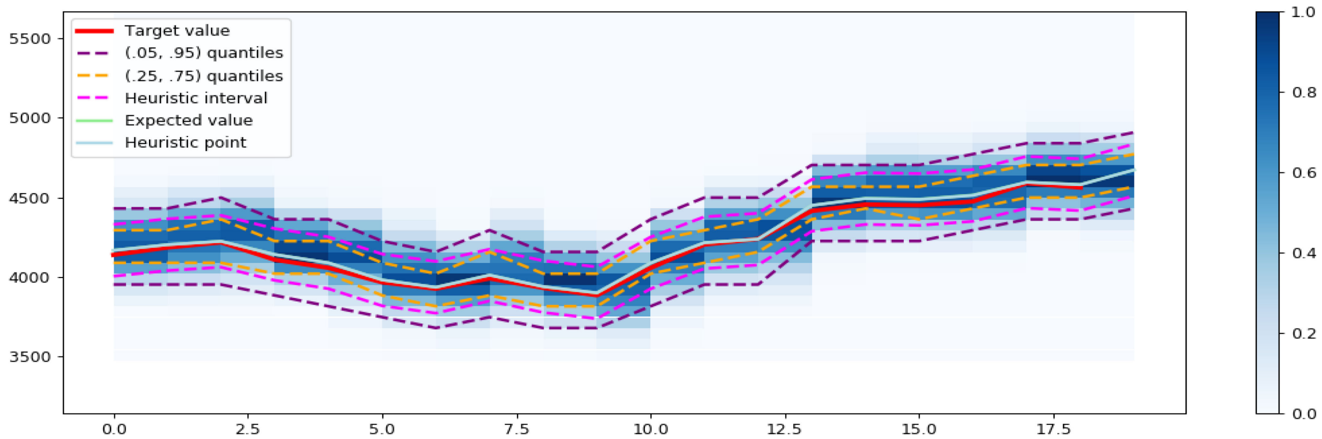

Fig. 4. All PWFTS forecasting methods applied on a TAIEX sample.

TABLE I
CRPS Results for Probabilistic Forecasts

Dataset	Model	Order	Part.	CRPS AVG	CRPS STD	Rank
TAIEX	ARIMA	1,1,0	-	0.059	0.021	2
	Ensemble FTS	3	3	0.365	0.070	4
	k-NN	1	-	0.515	0.033	5
	PWFTS	1	9	0.058	0.017	1
	QAR	1	-	0.092	0.033	3
NASDAQ	ARIMA	1,1,0	-	0.094	0.050	2
	Ensemble FTS	3	3	0.360	0.062	4
	k-NN	1	-	0.477	0.025	5
	PWFTS	1	9	0.080	0.033	1
	QAR	2	-	0.127	0.073	3
S&P 500	ARIMA	1,1,0	-	0.079	0.038	2
	Ensemble FTS	3	3	0.416	0.075	4
	k-NN	1	-	0.488	0.024	5
	PWFTS	1	9	0.074	0.032	1
	QAR	2	-	0.115	0.051	3

TABLE II
CRPS Post Hoc Tests Using PWFTS as Control Method

Comparison	Statistic	Adjusted p-value	Result
PWFTS vs kNN	3.10376	0.00762	H0 is rejected
PWFTS vs Ensemble FTS	2.28218	0.04445	H0 is rejected
PWFTS vs QAR	1.46059	0.18740	H0 is accepted
PWFTS vs ARIMA	0.45644	0.64808	H0 is accepted

TABLE III
Mean Winkler Score Results for 95% Prediction Intervals

Dataset	Model	Order	Part.	Winkler Score AVG	Winkler Score STD	Rank
TAIEX	ARIMA	1,1,0	-	568.43	312.47	1
	IFTS	3	10	921.35	285.32	3
	PWFTS	1	10	751.04	478.12	2
	QAR	2	-	2624.56	917.99	4
NASDAQ	ARIMA	1,1,0	-	203.69	139.88	1
	IFTS	2	10	272.51	92.05	2
	PWFTS	1	10	304.37	218.42	3
	QAR	1	-	824.18	287.67	4
S&P 500	ARIMA	1,1,0	-	97.56	72.87	1
	IFTS	1	10	103.56	29.78	2
	PWFTS	1	10	135.09	100.03	3
	QAR	2	-	302.80	101.40	4

TABLE IV
RMSE Results for Point Forecasts

Dataset	Model	Order	Part.	RMSE AVG	RMSE STD	Rank
TAIEX	ARIMA	1,1,0	-	87.32	28.15	3
	CFTS	1	31	95.08	31.11	9
	EWFTS	1	7	88.70	30.80	6
	FTS	1	13	90.73	26.60	7
	HOFTS	3	5	91.06	30.55	8
	Hwang	3	5	105.54	42.75	10
	IWFTS	1	5	87.65	29.09	5
	Naïve	1	-	127.24	40.68	11
	PWFTS	3	3	83.72	27.63	1
	TWFTS	1	7	85.35	28.44	2
	WFTS	1	5	87.43	29.12	4
NASDAQ	ARIMA	1,1,0	-	28.53	11.07	3
	CFTS	1	10	31.89	11.02	8
	EWFTS	1	10	29.01	11.02	6
	FTS	1	10	31.94	10.12	9
	HOFTS	3	7	30.47	10.89	7
	Hwang	2	10	36.11	14.79	10
	IWFTS	1	10	28.70	10.91	4
	Naïve	1	-	41.24	15.15	11
	PWFTS	1	10	28.21	10.89	2
	TWFTS	1	10	27.94	10.83	1
	WFTS	1	10	28.71	10.89	5
S&P 500	ARIMA	1,1,0	-	11.79	4.00	8
	CFTS	1	7	11.65	3.55	7
	EWFTS	1	5	11.46	4.10	5
	FTS	1	31	10.42	3.580	2
	HOFTS	3	5	12.00	3.24	9
	Hwang	3	29	12.11	4.26	10
	IWFTS	1	5	11.50	3.89	6
	Naïve	1	-	16.18	5.22	11
	PWFTS	1	3	10.21	3.50	1
	TWFTS	1	9	10.78	3.81	3
	WFTS	1	5	11.44	3.93	4

accepted the null hypothesis with the test statistic of 6.94595 and p-value of 0.07364; hence, no multiple *post hoc* tests were employed. The results show the equilibrium of PWFTS intervals, laying between QAR and ARIMA prediction intervals, where QAR has wider and less reliable intervals and ARIMA has thinner and more precise ones.

TABLE V
RMSE POST HOC TESTS USING PWFTS AS CONTROL METHOD

Comparison	Statistic	Adjusted p-value	Result
PWFTS vs Naive	2.91318	0.03521	H0 is rejected
PWFTS vs Hwang	2.65986	0.03848	H0 is rejected
PWFTS vs CFTS	2.19544	0.09073	H0 is rejected
PWFTS vs HOFTS	1.22438	0.46407	H0 is accepted
PWFTS vs FTS	1.18216	0.46407	H0 is accepted
PWFTS vs EWFTS	0.67552	0.68433	H0 is accepted
PWFTS vs IWFTS	0.59108	0.68493	H0 is accepted
PWFTS vs WFTS	0.50664	0.69418	H0 is accepted
PWFTS vs ARIMA	0.50664	0.69418	H0 is accepted
PWFTS vs TWFTS	0.08444	0.93271	H0 is accepted

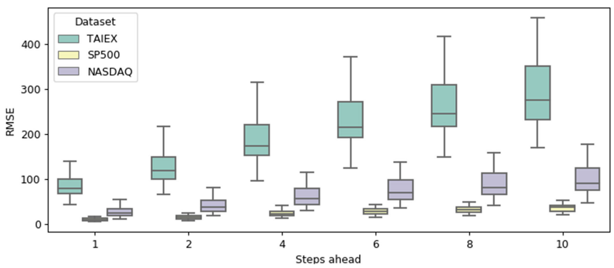

Fig. 5. Point forecasting error by prediction horizon.

D. Point Forecasting

The standard accuracy metric used to evaluate point forecasting methods is the root mean squared error (RMSE), described in (19), where y are the target values, $\hat{y}$ the forecast values, and n sample size. One has the following:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}. \quad (19)$$

The chosen benchmark models were the Naïve method (just repeat the last value), ARIMA, traditional FTS [12], conventional FTS (CFTS) [32], WFTS [37], IWFTS [15], EWFTS [28], high-order FTS (HOFTS) [41], Hwang [92], and trend weighted FTS (TWFTS) [93]; these last ones are the most cited FTS methods in the literature.

The experiments results of the best performing models are presented in Table IV. The equality of means statistical analysis rejected the null hypothesis with the test statistic of 22.65328 and p-value of 0.01210. The results of multiple *post hoc* tests in Table V show significant differences between PWFTS and Naïve, and Hwang and CFTS, with no significant difference between the others, though PWFTS has the best average rank among all the competing methods. It can be seen from Fig. 5 that it is possible to note how the error grows as the forecasting horizon increases, using the best PWFTS model for each dataset.

VI. CONCLUSION

In the time series forecasting research field, dealing with uncertainties is somehow mandatory, yet many of the forecasting methods are only concerned with point forecasting. The point forecasting methods have as their main general drawback the inability to measure the uncertainty of their results, and depending on the field of application, this is crucial information. Although FTS methods represent a growing field, there is also a gap of probabilistic forecasting methods based on FTS.

The aim of this paper was to propose a new univariate and time-invariant FTS method—the PWFTS, which is a data-driven approach, which splits the universe of discourse of a time series in overlapping fuzzy sets, represents their temporal patterns as fuzzy rules, and associates with them an empirical probability—based on the proposed concept of fuzzy frequency. The PWFTS rule model—the PWFTPG—describes fuzzy and stochastic behavior of time series and combines them to produce forecasts. This model is used to produce probability densities, prediction intervals, and point forecasting. Extensions for high-order models and multiple step ahead forecasting were also proposed.

Computational experiments were performed to evaluate the accuracy of the proposed model by using three financial time series: 1) TAIEX; 2) S&P 500; and 3) NASDAQ. The model's accuracy is directly related with the number of partitions of the universe of discourse, but if the data are preprocessed with a differentiation data transformation, this effect is almost null. PWFTS was compared with traditional statistical methods, such as ARIMA, QAR, kNN, and KDE, and with other standard FTS methods for probabilistic, interval, and point forecasting one step ahead. The proposed method for probabilistic forecasting tied with QAR and ARIMA regarding the CRPS metric and outperformed kNN + KDE and ensemble FTS. For interval forecasting, the performance of the method tied with ARIMA, QAR, regarding point forecasting.

The proposed PWFTS method extends FTS methods to deal with interval and probabilistic forecasting applications, which is a major contribution of this paper. Moreover, PWFTS improves on former FTS methods in the literature by considering the concept of fuzzy frequency and empirical probabilities in the generation of the rule knowledge base. The proposed method improves the previous FTS methods by aggregating probabilistic and interval forecasting capabilities into a single model, being useful for a wide range of applications and user needs.

A. Main Contributions

This paper uses the concept of fuzzy frequency to approximate the empirical probabilities of fuzzy sets, given sample crisp data. With these empirical probabilities, it is possible to calculate the probability of a crisp number, given a fuzzy set, and, therefore, a probability distribution for the original data described by a fuzzy linguistic variable.

The main contribution of PWFTS is to combine versatility, accuracy, scalability, and human readability. PWFTS is a versatile, data-driven, nonparametric approach that integrates point, interval, and probabilistic forecasting for one or multiple steps ahead, for first or higher orders. The measured accuracy shows its compatibility with, when it is not better than, standard approaches in the literature.

The PWFTPG rule model is human-readable, easy to understand, and interchangeable, which allows its assessment by experts and also by nontechnical people. The PWFTPG rule set can be viewed as the conditional probability distribution of the fuzzy sets, and its visualization can even be used for data description and comprehension tasks. PWFTS method can be easily adapted for distributed environments, which can scale its training for high data volumes. Additionally, due to its empirical probabilities based on fuzzy frequencies, the model is easy to update when new data are presented.

B. Method Limitations

The presented method lacks the abilities on heteroscedastic time series, demanding data preprocessing (as Box–Cox transformation) to deal with this kind of data. The generated model is time invariant and, despite being easily trained, needs to be frequently updated to follow new data behaviors. The choice of parameters (order and partitioning scheme) demands a heuristic search, and this is highly recommended to achieve the best results.

C. Future Work

Future works include the extensions of the proposed method to 1) support multivariate and spatial–temporal time series, especially by using clustering and granular approaches; 2) increase the robustness of the method to nonstationarity and concept-drift with incremental/online learning; 3) implement the distributed training and execution to support big data; 4) employ the proposed method in the forecasting of chaotic time series; and 5) develop a new fuzzy Markov chain method based on the proposed empirical fuzzy probabilities and defuzzification methods.

References

[1] V. Georgescu, "Joint propagation of ontological and epistemic uncertainty across risk assessment and fuzzy time series models," *Comput. Sci. Inf. Syst.*, vol. 11, no. 2, pp. 881–904, 2014. [Online]. Available: http://doi.org/10.2298/CSIS121215048G

[2] T. Gneiting and M. Katzfuss, "Probabilistic forecasting," *Annu. Rev. Statist. Appl.*, vol. 1, pp. 125–151, 2014. [Online]. Available: http://doi.org/10.1146/annurev-statistics-062713-085831

[3] C. Chatfield, "Calculating interval forecasts," *J. Bus. Econ. Statist.*, vol. 11, no. 2, pp. 121–135, 1993. [Online]. Available: http://doi.org/10.2307/1391361

[4] C. Fraley, A. E. Raftery, and T. Gneiting, "Probabilistic weather forecasting in R," *R J.*, vol. 3, pp. 55–63, Jun. 2011. [Online]. Available: http://doi.org/10.1198/jasa.2009.ap07184

[5] M. Leutbecher and T. N. Palmer, "Ensemble forecasting," *J. Comput. Phys.*, vol. 227, no. 7, pp. 3515–3539, 2008. [Online]. Available: http://doi.org/10.1016/j.jcp.2007.02.014

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, pp. 914–938, 2016. [Online]. Available: http://doi.org/10.1016/j.ijforecast.2015.11.011

[7] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, pp. 896–913, 2016. [Online]. Available: http://doi.org/10.1016/j.ijforecast.2016.02.001

[8] Y. Liu, B. Wang, H. Zhan, Y. Fan, Y. Zha, and Y. Hao, "Simulation of nonstationary spring discharge using time series models," *Water Resour. Manage.*, vol. 31, no. 15, p. 4875–4890, 2017. [Online]. Available: http://doi.org/10.1007/s11269-017-1783-6

[9] P. Pinson, G. Kariniotaki, H. A. Nielsen, T. S. Nielsen, and H. Madsen, "Properties of quantile and interval forecasts of wind generation and their evaluation," presented at Eur. Wind Energy Conf. Exhib., Athens, Greece, 2006. [Online]. Available: http://localhost/pubdb/p.php?4250

[10] G. G. Netto *et al.*, "A hybrid evolutionary probabilistic forecasting model applied for rainfall and wind power forecast," in *Proc. IEEE Conf. Evolving Adapt. Intell. Syst.*, 2016, pp. 73–78.

[11] F. Laio and S. Tamea, "Verification tools for probabilistic forecasts of continuous hydrological variables," *Hydrol. Earth Syst. Sci.*, vol. 11, pp. 1267–1277, 2007. [Online]. Available: https://doi.org/10.5194/hess-11-1267-2007

[12] Q. Song and B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets Syst.*, vol. 54, no. 3, pp. 269–277, 1993.

[13] P. Singh, "A brief review of modeling approaches based on fuzzy time series," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 2, pp. 397–420, 2017.

[14] Q. Song and B. S. Chissorn, "Forecasting enrollments with fuzzy time series—Part II," *Fuzzy Sets Syst.*, vol. 62, pp. 1–8, 1994.

[15] Z. Ismail and R. Efendi, "Enrollment forecasting based on modified weight fuzzy time series," *J. Artif. Intell.*, vol. 4, no. 1, pp. 110–118, 2011.

[16] H. J. Sadaei, R. Enayatifar, M. H. Lee, and M. Mahmud, "A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting," *Appl. Soft Comput. J.*, vol. 40, pp. 132–149, 2016.

[17] M. H. Lee, H. J. Sadaei, and Suhartono,"Improving TAIEX forecasting using fuzzy time series with BoxCox power transformation," *J. Appl. Statist.*, vol. 40, no. 11, pp. 2407–2422, 2013.

[18] M.-Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," *Future Gener. Comput. Syst.*, vol. 37, pp. 461–467, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.future.2013.09.025

[19] B. Sun *et al.*, "Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series," *Neurocomputing*, vol. 151, pp. 1528–1536, 2015.

[20] F. M. Talarposhti, H. J. Sadaei, R. Enayatifar, F. G. Guimarães, M. Mahmud, and T. Eslami, "Stock market forecasting by using a hybrid model of exponential fuzzy time series," *Int. J. Approx. Reasoning*, vol. 70, pp. 79–98, 2016.

[21] R. Efendi, Z. Ismail, and M. M. Deris, "Improved weight fuzzy time series as used in the exchange rates forecasting of US dollar to ringgit Malaysia," *Int. J. Comput. Intell. Appl.*, vol. 12, no. 1, 2013, Art. no. 1350005. [Online]. Available: http://doi.org/10.1142/S1469026813500053

[22] M. H. Lee and H. Javedani, "A weighted fuzzy integrated time series for forecasting tourist arrivals," in *Proc. Int. Conf. Inform. Eng. Inf. Sci.*, 2011, pp. 206–217.

[23] Z. Ismail, R. Efendi, and M. M. Deris, "Application of fuzzy time series approach in electric load forecasting," *New Math. Natural Comput.*, vol. 11, no. 3, pp. 229–248, 2015.

[24] T. Silva and M. Stabile, *Monitoramento e Pesquisa Em MÍDias Sociais: Metodologias, Aplicaçes E Inovaçes*, 1st ed. São Paulo, Brazil: Uva Limão, 2016.

[25] Q. Song, "Seasonal forecasting in fuzzy time series," *Fuzzy Sets Syst.*, vol. 107, pp. 235–236, 1999.

[26] F.-M. Tseng, G.-H. Tzeng, and H.-C. Yu, "Fuzzy seasonal time series for forecasting the production value of the mechanical industry in Taiwan," *Technol. Forecasting Soc. Change*, vol. 60, pp. 263–273, 1999.

[27] P.-T. Chang, "Fuzzy seasonality forecasting," *Fuzzy Sets Syst.*, vol. 90, no. 1, pp. 1–10, 1997. [Online]. Available: https://doi.org/10.1016/S0165-0114(96)00138-8

[28] H. J. Sadaei, "Improved models in fuzzy time series for forecasting," Ph.D. dissertation, Fac. Sci., Universiti Teknologi Malaysia, Johor Bahru, Malaysia, 2013.

[29] V. Georgescu, "Fuzzy time series estimation and prediction: Criticism, suitable new methods and experimental evidence," *Stud. Inform. Control*, vol. 19, no. 3, pp. 229–242, 2010.

[30] H. J. Sadaei, R. Enayatifar, F. G. Guimarães, M. Mahmud, and Z. A. Alzamil, "Combining ARFIMA models and fuzzy time series for the forecast of long memory time series," *Neurocomputing*, vol. 175, pp. 782–796, 2016. [Online]. Available: https://doi.org/10.1016/j.neucom.2015.10.079

[31] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965. [Online]. Available: https://doi.org/10.1016/S0019-9958(65)90241-X

[32] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets Syst.*, vol. 81, no. 3, pp. 311–319, 1996. [Online]. Available: https://doi.org/10.1016/0165-0114(95)00220-0

[33] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy Sets Syst.*, vol. 123, no. 3, pp. 387–394, 2001.

[34] 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, pp. 3052–3063, 12 2008. [Online]. Available: https://doi.org/10.1016/j.camwa.2008.07.033

[35] C.-H. H. Cheng, J.-R. R. Chang, and C.-A. A. Yeh, "Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost," *Technol. Forecasting Soc. Change*, vol. 73, no. 5, pp. 524–542, 2006. [Online]. Available: https://doi.org/10.1016/j.techfore.2005.07.004

[36] R. Enayatifar, H. J. Sadaei, A. H. Abdullah, and A. Gani, "Imperialist competitive algorithm combined with refined high-order weighted fuzzy time series (RHWFTS-ICA) for short term load forecasting," *Energy Convers. Manag.*, vol. 76, pp. 1104–1116, 2013.

[37] H.-K. Yu, "Weighted fuzzy time series models for TAIEX forecasting," *Physica A; Statistical Mech. Appl.*, vol. 349, no. 3, pp. 609–624, 2005.
[38] H. J. Sadaei, R. Enayatifar, A. H. Abdullah, and A. Gani, "Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search," *Int. J. Elect. Power Energy Syst.*, vol. 62, pp. 118–129, 2014.
[39] M. H. Lee, H. J. Sadaei, and Suhartono,"Introducing polynomial fuzzy time series," *J. Intell. Fuzzy Syst.; Appl. Eng. Technol.*, vol. 25, no. 1, pp. 117–128, 2013.
[40] 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, pp. 2121–2135, 2015.
[41] C. A. Severiano, P. C. L. Silva, H. J. Sadaei, and F. G. Guimarães, "Very short-term solar forecasting using fuzzy time series," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Naples, Italy, 2017, pp. 1–6.
[42] M. Bose and K. Mali, "A novel data partitioning and rule selection technique for modeling high-order fuzzy time series," *Appl. Soft Comput.*, vol. 63, no. 17, pp. 87–96, 2017. [Online]. Available: http://doi.org/10.1016/j.asoc.2017.11.011
[43] L. Nguyen and V. Novák, "Forecasting seasonal time series based on fuzzy techniques," *Fuzzy Sets Syst.*, vol. 361, pp. 114–129, 2018.
[44] S. S. Zhang *et al.*, "A novel method based on FTS with both GA-FCM and multifactor BPNN for stock forecasting," *Soft Comput.*, vol. 23 no. 16, pp. 6979–6994, Aug. 2019.
[45] I. Perfilieva, "Fuzzy transforms: Theory and applications," *Fuzzy Sets Syst.*, vol. 157, pp. 993–1023, 2006. [Online]. Available: http://doi.org/10.1016/j.fss.2005.11.012
[46] M. Štěpnička, A. Dvořák, V. Pavliska, and L. Vavříčková, "A linguistic approach to time series modeling with the help of F-transform," *Fuzzy Sets Syst.*, vol. 180, pp. 164–184, 2011.
[47] V. Novák, "Linguistic characterization of time series," *Fuzzy Sets Syst.*, vol. 285, pp. 52–72, 2016.
[48] L. A. Zadeh, "Probability measures of fuzzy events," *J. Math. Anal. Appl.*, vol. 23, no. 2, pp. 421–427, 1968.
[49] L. A. Zadeh, "Fuzzy probabilities," *Inf. Process. Manag.*, vol. 20, no. 3, pp. 363–372, 1984.
[50] E. P. Klement, W. Schwyhla, and R. Lowen, "Fuzzy probability measures," *Fuzzy Sets Syst.*, vol. 5, pp. 21–30, 1981.
[51] D. Dubois and H. Prade, "Fuzzy sets, probability and measurement," *Eur. J. Oper. Res.*, vol. 40, pp. 135–154, 1989.
[52] Q. Song, R. P. Leland, and B. S. Chissom, "Fuzzy stochastic fuzzy time series and its models," *Fuzzy Sets Syst.*, vol. 88, no. 3, pp. 333–341, 1997.
[53] T. A. Jilani, A. S. Burney, and C. Ardil, "A new quantile based fuzzy time series forecasting model," *Int. J. Comput., Elect., Autom., Control Inf. Eng.*, vol. 2, no. 3, pp. 995–1001, 2008. [Online]. Available: http://waset.org/publications/14214/a-new-quantile-based-fuzzy-time-series-forecasting-model
[54] S. S. Gangwar and S. Kumar, "Probabilistic and intuitionistic fuzzy setsbased method for fuzzy time series forecasting," *Cybern. Syst.*, vol. 45, no. 4, pp. 349–361, 2014. [Online]. Available: https://doi.org/10.1080/01969722.2014.904135
[55] C. Chatfield, "Prediction intervals for time-series forecasting," in *Principles of Forecasting*, 1st ed., J. Armstrong, Ed. New York, NY, USA: Springer-Verlag, 2001, pp. 475–494.
[56] R. Koenker and Z. Xiao, "Quantile autoregression," *J. Amer. Statistical Assoc.*, vol. 101, no. 405, pp. 980–990, 2006. [Online]. Available: https://doi.org/10.1198/016214506000000672
[57] R. Koenker and K. F. Hallock, "Quantile regression," *J. Econ. Perspectives*, vol. 15, no. 4, pp. 143–156, 2001.
[58] P. C. L. Silva, H. J. Sadaei, and F. G. Guimarães, "Interval forecasting with fuzzy time series," in *Proc. IEEE Symp. Ser. Comput. Intell.*, Athens, Greece, 2016, pp. 1–8.
[59] R. T. Clemen, "Combining forecast: A review and annotated bibliography," *Int. J. Forecasting*, vol. 5, no. 4, pp. 559–583, 1989.
[60] T. Gneiting, "Editorial: Probabilistic forecasting," *J. Roy. Statistical Soc.*, vol. 171, no. 2, pp. 319–321, 2008. [Online]. Available: http://www.jstor.org/stable/30130759
[61] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly Weather Rev.*, vol. 133, no. 5, pp. 1155–1174, 2005.
[62] T. Gneiting *et al.*, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation," *Monthly Weather Rev.*, vol. 133, no. 5, pp. 1098–1118, 2005.
[63] C. Fraley, A. E. Raftery, T. Gneiting, and J. M. Sloughter, "ensembleBMA: An R package for probabilistic forecasting using ensembles and bayesian model," Dept. of Statistics, University of Washington, Seattle, WA, USA, Tech. Rep. 516, 2013.
[64] J. Xie and T. Hong, "GEFCom2014 probabilistic electric load forecasting: An integrated solution with forecast combination and residual simulation," *Int. J. Forecasting*, vol. 32, pp. 1012–1016, 2016.
[65] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, vol. 26. Boca Raton, FL, USA: CRC Press, 1986.
[66] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *J. Roy. Statistical Soc.; Ser. B (Statistical Methodol.)*, vol. 53, no. 3, pp. 683–690, 1991.
[67] F. Olivas, F. Valdez, O. Castillo, C. I. Gonzalez, G. Martinez, and P. Melin, "Ant colony optimization with dynamic parameter adaptation based on interval type-2 fuzzy logic systems," *Appl. Soft Comput. J.*, vol. 53, pp. 74–87, 4 2017.
[68] A. García, A. Luviano-Juárez, I. Chairez, A. Poznyak, and T. Poznyak, "Projectional dynamic neural network identifier for chaotic systems: Application to Chua's circuit," *Int. J. Artif. Intell.*, vol. 6, no. S11, pp. 1–18, 2011.
[69] C. Pozna, R. E. Precup, J. K. Tar, I. Škrjanc, and S. Preitl, "New results in modelling derived from Bayesian filtering," *Knowl.-Based Syst.*, vol. 23, no. 2, pp. 182–194, 2010.
[70] L. Maciel and R. Ballini, "Interval fuzzy rule-based modeling approach for financial time series forecasting," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2017, pp. 1–6.
[71] M. A. Alves, P. C. D. L. Silva, C. A. J. Severiano, G. L. Vieira, F. G. Guimaraes, and H. J. Sadaei, "An extension of nonstationary fuzzy sets to heteroskedastic fuzzy time series," in *Proc. 26th Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, Bruges, Belgium, vol. 1, 2018, pp. 591–596.
[72] J. Luo and S. M. Bridges, "Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection," *Int. J. Intell. Syst.*, vol. 15, no. 8, pp. 687–703, 2000. [Online]. Available: https://doi.org/10.1002/1098-111X(200008)15:83C687::AID-INT13E3.0.CO;2-X
[73] C.-T. Chen, C.-T. Lin, S.-F. Huang, L. Da, and K.-C. Li, "A fuzzy approach for supplier evaluation and selection in supply chain management," *Int. J. Prod. Econ.*, vol. 102, pp. 289–301, 2006.
[74] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: An analysis and review," *Int. J. Forecasting*, vol. 16, no. 4, pp. 437–450, 2000. [Online]. Available: http://doi.org/10.1016/S0169-2070(00)00065-0
[75] S. M. Chen and C. D. Chen, "TAIEX forecasting based on fuzzy time series and fuzzy variation groups," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 1, pp. 1–12, Feb. 2011.
[76] N. S. Bajestani and A. Zare, "Forecasting TAIEX using improved type 2 fuzzy time series," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5816–5821, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2010.10.049
[77] 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 Trans. Cybernet.*, vol. 45, no. 3, pp. 391–403, Mar. 2015.
[78] H. J. Sadaei and M. H. Lee, "Multilayer stock forecasting model using fuzzy time series," *Sci. World J.*, vol. 2014, 2014, Art. no. 610594.
[79] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Inf. Sci.*, vol. 180, pp. 2044–2064, 2010.
[80] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm Evol. Comput.*, vol. 1, pp. 3–18, 2011.
[81] B. Trawiński, M. Smetek, Z. Telec, and T. Lasota, "Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms," *Int. J. Appl. Math. Comput. Sci*, vol. 22, no. 4, pp. 867–881, 2012. [Online]. Available: https://doi.org/10.2478/v10006-012-0064-z
[82] J. L. Hodges and E. L. Lehmann, "Rank methods for combination of independent experiments in analysis of variance," *Ann. Math. Statist.*, vol. 33, no. 2, pp. 482–497, 1962. [Online]. Available: https://doi.org/10.1214/aoms/1177704575
[83] H. Finner, "On a monotonicity problem in step-down multiple test procedures," *J. Amer. Statistical Assoc.*, vol. 88, no. 423, pp. 920–923, 1993. [Online]. Available: http//doi.org/10.1080/01621459.1993.10476358

[84] I. Rodríguez-Fdez, A. Canosa, M. Mucientes, and A. Bugarín, "STAC: A web platform for the comparison of algorithms using statistical tests," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Istanbul, Turkey, 2015, pp. 1–8. [Online]. Available: https://github.com/citiususc/stac

[85] P. C. d. L. Silva, P. O. Lucas, H. J. Sadaei, and F. G. Guimarães, "pyFTS: Fuzzy time series for Python," Belo Horizonte, 2018. [Online]. Available: https://doi.org/10.5281/zenodo.597359

[86] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Am. Statistical Assoc.*, vol. 102, no. 477, pp. 359–378, 2007.

[87] P. C. de Lima e Silva, M. A. Alves, C. A. Severiano Jr, G. L. Vieira, F. G. Guimarães, and H. J. Sadaei, "Probabilistic forecasting with seasonal ensemble fuzzy time-series," presented at 13th Brazilian Congr. Comput. Intell., Rio de Janeiro, Brazil, 2017, pp. 1–12.

[88] Y. Zhang and J. Wang, "GEFCom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2015, pp. 1–5.

[89] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1074–1080, 2016. [Online]. Available: https://doi.org/10.1016/j.ijforecast.2015.11.006

[90] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *J. Roy. Statistical Soc. Series B (Statistical Methodol.)*, vol. 69, no. 2, pp. 243–268, 2007. [Online]. Available: http://www.jstor.org/stable/4623266

[91] R. L. Winkler, "A decision-theoretic approach to interval estimation," *J. Am. Statistical Assoc.*, vol. 67, no. 337, pp. 187–191, 1972.

[92] J.-R. Hwang, S.-M. Chen, C.-H. Lee, J.-R. Hwang, S.-M. Chen, and C.-H. Lee, "Handling forecasting problems using fuzzy time series," *Fuzzy Sets Syst.*, vol. 100, no. 1, pp. 217–228, 1998.

[93] C.-H. Cheng, T.-L. Chen, and C.-H. Chiang, "Trend-weighted fuzzy time-series model for TAIEX forecasting," in *Neural Information Processing*, I. King, J. Wang, L.-W. Chan, and D. Wang, Eds., Berlin, Germany: Springer-Verlag, 2006, pp. 469–477. [Online]. Available: https://doi.org/10.1007/11893295_52

Petrônio Candido de Lima Silva was born in Janauba, Brazil, in 1982. He received the bachelor's degree in information systems from the Santo Agostinho Faculty of Exact and Technological Sciences, Montes Claros, Brazil, in 2005, and the M.Sc. degree in informatics from the Pontifical Catholic University of Minas Gerais, Belo Horizonte, Brazil, in 2010. He is currently working toward the Ph.D. degree in electrical engineering at the Machine Intelligence and Data Science Laboratory, Federal Institute of Northern Minas Gerais (IFNMG), Januaria, Brazil.

In 2013, he joined the IFNMG as a Lecturer in undergraduate courses in software development, databases, data-warehousing, and machine intelligence. He is a member with the Research Group on Data Science and Computational Intelligence, Instituto Federal do Norte de Minas Gerais, Brazil. His current research interests include data science methods for spatio–temporal data, computational intelligence, and machine learning, including scalable soft computing methods for probabilistic forecasting of dynamic and complex systems.

Hossein Javedani Sadaei received the B.Sc. degree in applied mathematics from the University of Sistan and Baluchestan, Zahedan, Iran, in 2001, the M.Sc. degree in pure mathematics from Zanjan University, Zanjan, Iran, in 2004, and the Ph.D. degree in statistics from Universiti Teknologi Malaysia, Johor Bahru, Malaysia, in 2013.

Starting from 2014, he was a Postdoctoral Fellow with Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil. He is currently a Machine Learning Practice Lead in Avenue Code, San Francisco, CA, USA, and also an active member of MINDS Lab, UFMG. He has more than ten years of experience in developing advanced forecasting methods and has authored more than ten scientific publications and a few patents. Most of his efforts are about combining fuzzy time series methods with novel concepts and deep learning to improve the performance of forecasting. He is conducting a new concept in the fuzzy time series, namely, polynomial fuzzy time series. Following the Ph.D. degree in statistics, he took up a position as the Leader Analytics of the big data team in the research and development center of Telecom Malaysia. His research interests include fuzzy time series, forecasting, learning systems, optimization algorithm, big data, voice, and text mining.

Rosangela Ballini received the B.Sc. degree in applied mathematics from the Federal University of São Carlos, São Carlos, Brazil, in 1994, the M.Sc. degree in mathematics and computation science from the University of São Paulo, São Carlos, Brazil, in 1996, and the Ph.D. degree in electrical engineering from the School of Electrical and Computer Engineering, University of Campinas, Campinas, Brazil, in 2000.

She has been a Professor with the Department of Economic Theory, Institute of Economics, University of Campinas, since 2002. Her research interests include time-series models, neural computation, fuzzy systems, modeling and optimization, forecasting, decision making, and applications.

Dr. Ballini is on the Editorial Boards of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, *Applied Soft Computing*, and *Evolving Systems*. She serves on the IEEE Task Forces on Adaptive Fuzzy Systems, Computational Finance, and Economics Technical Committee of the IEEE Computational Intelligence Society.

Frederico Gadelha Guimarães received the B.Eng. and M.Sc. degrees in electrical engineering from the Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, in 2003 and 2004, respectively, and the Ph.D. degree in electrical engineering from the UFMG, in 2008, with a one-year visiting student scholarship at McGill University, Montreal, QC, Canada (2006–2007).

He was a Postdoctoral Fellow (2017–2018) with the Laboratoire Images, Signaux et Systèmes Intelligents, Vitry-sur-Seine, France, linked to the Université Paris-Est Créteil, Paris, France. In 2010, he joined the Department of Electrical Engineering, UFMG, where in 2018, he became an Associate Professor. He is responsible for the Machine Intelligence and Data Science Laboratory, UFMG, for computational intelligence research. He has authored or coauthored more than 200 papers in journals, congresses, and chapters of national and international books. He has experience in electrical engineering and computer engineering, with emphasis on optimization, computational intelligence, genetic algorithms, and evolutionary computation.

Prof. Guimarães is a member of the IEEE Computational Intelligence Society and the IEEE Systems, Man, and Cybernetics Society.