

Hindawi Publishing Corporation
Advances in Fuzzy Systems
Volume 2013, Article ID 494239, 10 pages
<http://dx.doi.org/10.1155/2013/494239>



Research Article

A New Method for Short Multivariate Fuzzy Time Series Based on Genetic Algorithm and Fuzzy Clustering

Kamal S. Selim¹ and Gihan A. Elanany²

¹ Faculty of Economics and Political Science, Cairo University, Egypt

² Sadat Academy for Management Sciences, Cairo, Egypt

Correspondence should be addressed to Gihan A. Elanany; gigi_aman@yahoo.com

Received 16 September 2013; Accepted 19 October 2013

Academic Editor: Katsuhiko Honda

Copyright © 2013 K. S. Selim and G. A. Elanany. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forecasting activities play an important role in our daily life. In recent years, fuzzy time series (FTS) methods were developed to deal with forecasting problems. FTS attracted researchers because of its ability to predict the future values in some critical situations where most standard forecasting models are doubtfully applicable or produce bad fittings. However, some critical issues in FTS are still open; these issues are often subjective and affect the accuracy of forecasting. In this paper, we focus on improving the accuracy of FTS forecasting methods. The new method integrates the fuzzy clustering and genetic algorithm with FTS to reduce subjectivity and improve its accuracy. In the new method, the genetic algorithm is responsible for selecting the proper model. Also, the fuzzy clustering algorithm is responsible for fuzzifying the historical data, based on its membership degrees to each cluster, and using these memberships to defuzzify the results. This method provides better forecasting accuracy when compared with other extant researches.

1. Introduction

Time series are widely observed in many aspects of our lives; therefore, the prediction of future values based on the past and present information is very useful. In practice, there are several emergent domains that require dealing with short multivariate time series. As a consequence, the prediction of such time series arises in many situations. There are many existing techniques that are well proven in forecasting with multivariate time series data, but they put constraints on the minimum number of observations and require distribution assumptions to be made regarding the observed time series.

Fuzzy time series (FTS) models have become increasingly popular in recent years because of their ability to deal with time series data without the need for validating any theoretical assumptions. However, how to select the proper model, how to partition the universe of discourse and determine effective lengths of intervals objectively to fuzzify the numerical data, and how to defuzzify the results are still open critical issues. These issues are very important and affect the model accuracy. The paper probes into these three questions in the modeling of FTS. The new method incorporates the fuzzy

clustering and genetic algorithms (GA) with FTS to reduce its subjectivity and improve its accuracy. More specifically, the new method uses the integer genetic algorithm to search for the optimal model that fits the available data. (In this paper, to solve integer optimization problems, we used the MI-LXPM algorithm which is a suitably modified and extended version of the real coded genetic algorithm, LXPM. In MI-LXPM, a tournament selection procedure, Laplace crossover, and power mutation are modified and extended for integer decision variables. Moreover, a special truncation procedure for satisfaction of integer restriction on decision variables and a “parameter free” penalty approach for constraint handling are used in MI-LXPM algorithm. More details of these operators are defined in [1].) In addition, fuzzy clustering is used to partition the universe of discourse objectively. Furthermore, the method employs clustering centers and the observations' fuzzy memberships to defuzzify the results, instead of the centers of each interval, which are used in numerous existing models. The empirical results show that the new model is able to forecast with high accuracy measures than the counterpart of existing models.

GA was used before in the literature of fuzzy time series. For example, Chen and Chung [2] use GA to tune up the length of each interval in the universe of discourse for one factor (variable), Lee et al. [3] use genetic algorithms to adjust the length of each interval in the universe of discourse for two factors, Kang [4] uses GA to obtain the optimal fuzzy membership function, while Egrioglu [5] uses genetic algorithm for finding the elements of fuzzy relation matrix.

Many authors applied fuzzy clustering in the fuzzification process. For example, Cheng et al. [6], Li et al. [7], Egrioglu [5], and Yolcu [8] used fuzzy c-means for the fuzzification of time series. While Lie et al. and Egrioglu applied their methods in univariate time series, Cheng et al. applied their method in both univariate and multivariate time series.

The remainder of this paper is organized as follows. Section 2 describes in brief the concept of FTS and its models. Section 3 presents the new FTS method; it begins with a method overview. Then the details of each step of the new FTS forecasting model are described. In Section 4, the method is evaluated by comparing its forecasts with those derived from other related FTS methods based on an example. Finally, Section 5 summarizes the main conclusions.

2. The Basic Concepts of Fuzzy Time Series and Its Models

Fuzzy set theory provides a powerful framework to cope with vague or ambiguous problems and can express linguistic values and human subjective judgments of natural language. In 1993, Song and Chissom [9, 10] successfully employed the concept of fuzzy sets presented by Zadeh [11] and the application of fuzzy logic to approximate reasoning presented by Mamdani [12] to develop the foundation of FTS forecasting. Initially, FTS was proposed to deal with forecasting problems where the historical data are linguistic values. But recently FTS models have become increasingly popular because of their ability to deal with quantitative time series data with limited or even no theoretical assumptions in contrast with conventional time series models. The main difference between a traditional time series and FTS is the values of observations. In a traditional time series, the observations are represented by crisp numerical values. However, in a FTS, the values of observations are represented by fuzzy sets. The basic definitions of FTS could be presented as follows [10].

Let

\mathcal{R} : the set of all real values

$Y(t)$: the universe of discourse, $t = \dots, 0, 1, 2, \dots$

$F(t)$: fuzzy time series, $t = \dots, 0, 1, 2, \dots$

$f_i(t)$: fuzzy set, $i = 1, 2, \dots$,

$R(t, t-1)$: the fuzzy relation between $F(t)$ and $F(t-1)$.

FTS. Let $Y(t)$, ($t = \dots, 0, 1, 2, \dots$), a subset of \mathcal{R} , be the universe of discourse on which fuzzy sets $f_i(t)$, ($i = 1, 2, \dots$) are defined and $F(t)$ is the collection of $f_i(t)$, ($i = 1, 2, \dots$). Then $F(t)$ is called a FTS on $Y(t)$, ($t = \dots, 0, 1, 2, \dots$).

$F(t)$ can be understood as a linguistic variable and $f_i(t)$ ($i = 1, 2, \dots$) as the possible linguistic values of $F(t)$. Because at different times, the values of $F(t)$ can be different, $F(t)$ is a function of time t . Also, since the universes of discourse can be different at different times, $Y(t)$ is used for the universe of discourse at time t .

The m th Order Model. Suppose $F(t)$ is caused by $F(t-1)$, $F(t-2)$, \dots , and $F(t-m)$ ($m \geq 1$) simultaneously. This relation can be expressed as the following fuzzy relation equation:

$$F(t) = (F(t-1) \times F(t-2) \times \dots \times F(t-m)) \circ R_m(t, t-m), \quad (1)$$

where \times is the Cartesian product and $R_m(t, t-m)$ is defined as the fuzzy relation between $F(t)$ and $F(t-1)$, $F(t-2)$, \dots , and $F(t-m)$. Then (1) is called the m th order model of $F(t)$. When $m = 1$, the model is called the first-order model.

Despite the large number of FTS methods developed since the introduction of the subject by Song and Chissom [9, 10, 16], the main steps in most of them are similar to the first method [9], which could be summarized as follows:

- (1) defining and partitioning the universe of discourse, defining fuzzy sets and fuzzifying time series,
- (2) establishing fuzzy logical rules,
- (3) grouping the fuzzy logical rules,
- (4) calculating the forecasted fuzzy sets,
- (5) defuzzifying the forecasted results.

Most of the work done in FTS employs one variable to build a fuzzy time-series model, that is, univariate model. Recently, FTS models have employed multiple variables in forecasting processes to deal with more complex forecasting problems, that is, multivariate model [17, 18]. It is important to note that the term "multivariate" in most FTS literature is limited to the prediction of one variable of interest, taking into consideration the multiple variables that could have effects on it, into the prediction process. To forecast some or all of the other variables, the same process has to be individually repeated for each variable. In this paper, we adopt the same understanding to deal with multivariate fuzzy time series data.

3. The New Method

In this section, we present the new forecasting method for fuzzy time series. Section 3.1 provides an overview of the new method. Section 3.2 presents the notations and definitions. Finally, the detailed description of the new method is given in Section 3.3.

3.1. Method Overview. Figure 1 summarizes the structure of the new method in the following steps.

- (1) Set the upper bound and lower bound for the model parameters, the probability of crossover (P_c), and probability of mutation (P_m). Also set the stopping

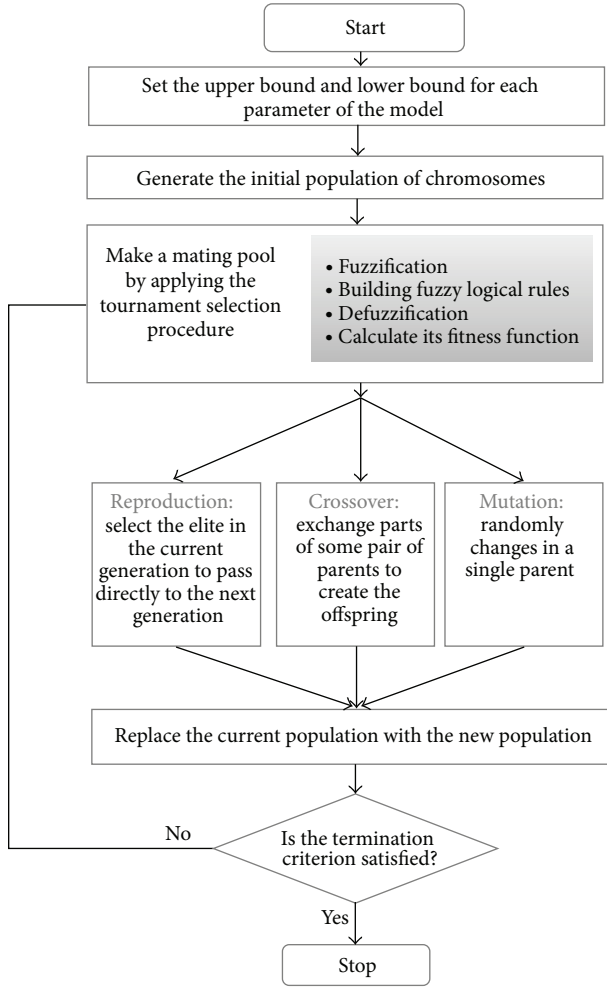


FIGURE 1: Flowchart for the suggested method.

criteria (the algorithm stops if the weighted average relative change in the best fitness function value over generations is less than or equal to tolerance (TolFun) or a prespecified maximum number of generations is reached [19]), maximum number of generations, and fitness function tolerance (TolFun).

- (2) With the language of GA, the method begins with creating a random initial population of chromosomes. Each chromosome suggests one possible model and consists of number of genes that is equal to the model's parameters and the values which specify that model.
- (3) Make a mating pool by applying the tournament selection procedure on initial (old) population [1].
- (4) Applying FTS procedure which is conducted for each chromosome in the mating pool, this procedure can be briefed as follows.
 - (i) Fuzzification: applying fuzzy clustering to partition each time series according to the specific number of clusters (as specified in the chromosome), ranking cluster centers of each variable, and then fuzzifying the historical data.

- (ii) Building fuzzy rules: building fuzzy relations according to specific order and number of variables (as specified in the chromosome) in addition to assigning membership degree of each data point to each cluster.
- (iii) Defuzzification: defuzzifying the forecasted outputs using the membership degrees and the clusters centers.
- (iv) Calculate the penalty function which includes a term for infeasibility. If the member (chromosome) is feasible, the penalty function will be the fitness function; otherwise, it will be the maximum fitness function among feasible members of the population, plus a sum of the constraint violations of the (infeasible) point.

- (5) Apply Laplace crossover and power mutation to all individuals in mating pool, with P_c and P_m to make new population.
- (6) Replace the current population with the offsprings to form the next generation.
- (7) Check the stopping criteria. If satisfied stop; else go to (2) [19].

After this overview of the new method, we will go through the method in more detail.

3.2. Notations and Definitions. Assume that we have P time series (or variables) x_1, x_2, \dots, x_P ; and without loss of generality, assume that the first time series x_1 is the main time series that we want to forecast and the rest of P time series, that is, $x_p, p = 2, \dots, P$, are the auxiliary time series.

The following notations and definitions are used throughout the formulation of the new method: p denotes the variable's (time series) number; $p = 1, \dots, P$, where P is the total number of variables. t denotes the time point number; $t = 1, \dots, T$, where T is the total number of time points. x_{tp} denotes the actual value of the variable p at time t ; $t = 1, \dots, T$; $p = 1, \dots, P$. $x'_p = (x_{1p}, x_{2p}, \dots, x_{Tp})$; $p = 1, \dots, P$. \hat{x}_{tp} denotes the forecasted value of the variable p at time t ; $t = 1, \dots, T$; $p = 1, \dots, P$. $\hat{x}'_p = (\hat{x}_{1p}, \hat{x}_{2p}, \dots, \hat{x}_{Tp})$; $p = 1, \dots, P$. k denotes the cluster's number; $k = 1, \dots, K$, where K is the total number of clusters.

Generally speaking, the number of clusters K could take any value from 1 to T . However, in the new method a number of clusters lower than 3 is considered too small and a number of clusters higher than $(T/2)$ is considered too large. If the number of clusters is too small, there will be no fluctuations in FTS. On the other hand, if the number of clusters is too large, the meaning of FTS will be diminished and it creates computational complexities [20], although the higher number of intervals may contribute to reaching better forecasting accuracy.

In the new method we assume that K takes values ranging from 3 to $T/2$; that is,

$$3 \leq K \leq \frac{T}{2}. \quad (2)$$

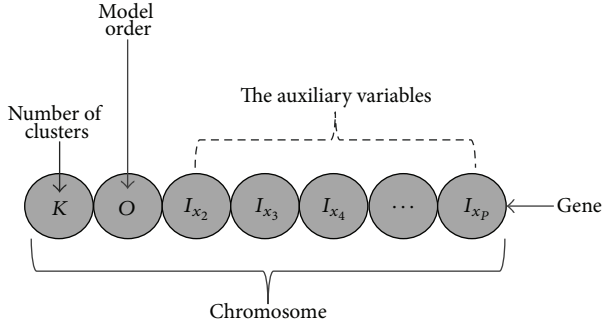


FIGURE 2: The component of chromosome in the proposed method.

O denotes the model order, which indicates how many previous time points we use to predict the present time. It takes values ranging from 1 to $(T - 1)$. Generally, lower and upper limits of the model order are optional. However, it is common in the literature that the model order takes values between 1 and 5. In the new method, we consider that the upper limit of the model order is $(T/3)$. That is,

$$1 \leq O \leq \frac{T}{3}. \quad (3)$$

$I_{x_p} \in [0, 1]$; $p = 2, \dots, P$: the indicator variables, each one takes the value 0 or 1 to indicate the absence or presence of the auxiliary variable in the model; that is,

$$I_{x_p} = \begin{cases} 0, & \text{if the auxiliary variable} \\ & x_p \text{ is excluded from the model} \\ 1, & \text{if the auxiliary variable} \\ & x_p \text{ is included in the model.} \end{cases} \quad p = 2, \dots, P \quad (4)$$

3.3. Description of the New Method. In this section, each step of the new method is further illustrated and detailed as follows.

Step 1 (initialization). The first step is to create an initial population of possible models or chromosomes. For each model, three main factors are of concern:

- (1) the number of clusters or fuzzy sets which will be used to fuzzify the historical data;
- (2) the order of the model, whether first order or high order;
- (3) the selected auxiliary variables to be entered in the forecasting model.

These factors are considered as the parameters of the model, so each chromosome the GA created represents one possible model. Accordingly, the number of genes in each chromosome is set equal to the number of auxiliary variables $(P - 1)$ plus 2. That is, each chromosome consists of $(P + 1)$ genes. Figure 2 illustrates the model parameter as a chromosome.

The population of chromosomes is generally chosen at random. There are no hard rules for determining the size of

the population. Generally, the size is defined to range between 20 and 100. Large populations guarantee greater diversity and may produce more robust solutions. Following this, the initial population is used to make a mating pool by applying the tournament selection procedure [19]; for more details, see [1].

Step 2. Applying the fuzzy time series procedure for each chromosome in the mating pool as follows.

(1) Each time series x_p , $p = 1, 2, \dots, P$, in the possible model at hand is partitioned using fuzzy c -means into k clusters (for more details about fuzzy c -means, see [21]). When fuzzy c -means algorithm is applied, the clusters' centers for each time series, denoted by c_{kp} , $k = 1, \dots, K$, and $p = 1, \dots, P$, are calculated iteratively. After the clusters' centers are fixed, the memberships of each time point with respect to these centers, denoted by μ_{tkp} , $t = 1, \dots, T$, $k = 1, \dots, K$, and $p = 1, \dots, P$, are computed. Then, the clusters of each variable are ranked ascendingly according to their centers' values, and these ranks are used as linguistic values (fuzzy sets) A_{kp} , $k = 1, \dots, K$, and $p = 1, \dots, P$ such that $A_{1p} < A_{2p} < \dots < A_{Kp}$ and $p = 1, \dots, P$.

(2) Fuzzify each time series by replacing each value with the equivalent linguistic value to obtain the FTS, $F(t)$ that is, each crisp value is mapped into a fuzzy set where its membership degree has maximum value.

(3) Given that the model order O and the indicator variables I_x 's are as specified in the randomly selected chromosome at hand, the fuzzy rule is established as follows:

$$\begin{aligned} & f_1(t - o), \dots, f_1(t - 1), f_2(t - o) \times I_{x_2}, \dots, f_2(t - 1) \\ & \times I_{x_2}, \dots, f_p(t - o) \times I_{x_p}, \dots, f_p(t - 1) \\ & \times I_{x_p} \longrightarrow [\mu_{f_1}(t)], \end{aligned} \quad (5)$$

where $f_p(t)$ is the value of the FTS F_p at time t , $p = 1, 2, \dots, P$, and $\mu_{f_1}(t) = (\mu_{t11} \ \mu_{t21} \ \dots \ \mu_{tK1})$ is the row vector of membership degrees of the main variable at time t to the k clusters; $t = 1, 2, \dots, T$.

For example, assume that we have two fuzzy time series F_1 (the main variable) and F_2 (the auxiliary variable) both clustered and fuzzified into four fuzzy sets (i.e., $K = 4$) and assume also that the model is of order 2 (i.e., $O = 2$). Let the current states of the main variable F_1 be $f_1(t) = A_{31}$ and $f_1(t + 1) = A_{41}$, let its next state be $f_2(t + 2) = A_{12}$, and let the next state vector of membership degrees to the four fuzzy sets (i.e., $\mu_{f_1(t+2)}$) be equal to [0.006 0.017 0.153 0.823].

Similarly, let the current states of the auxiliary variable F_2 be $f_2(t) = A_{42}$ and $f_2(t + 1) = A_{32}$. Then, under these assumptions, the rule can be induced as follows:

$$A_{31} \ A_{41}, \ A_{42} \ A_{32}, \longrightarrow [0.006 \ 0.017 \ 0.153 \ 0.823]$$

Now moving one time point ahead, let the current states of F_1 become $f_1(t + 1) = A_{41}$ and $f_1(t + 2) = A_{41}$, let the next state be $f_1(t + 3) = A_{41}$, and let its vector of membership degrees $\mu_{f_1(t+3)}$ be equal to [0.185 0.756 0.042 0.018].

And for the auxiliary variable F_2 let $f_2(t+1) = A_{32}$, $f_2(t+2) = A_{12}$. Then, the corresponding rule can be stated as

$$\boxed{A_{41} \ A_{41}, \ A_{32} \ A_{12}, \ \longrightarrow \ [0.185 \ 0.756 \ 0.042 \ 0.018]}$$

It is important to notice that, for each of the above two rules, the vector of membership degrees in the right-hand side is the main variable's next state row vector of memberships; that is, it has no relation with the vectors of memberships of all of the current states appearing in the left side of the rule whether they belong to the main variable or to the auxiliary variable.

(4) Group all generated fuzzy logical rules that have the same left-hand side. If this grouping process resulted in any group with more than one relation, then the average vector of all related membership vectors in this group is calculated. The resulting groups will be denoted by G_j ; $j = 1, 2, \dots, J$, $J \leq T - o$, where J is the total number of groups with at least one relation. The average membership vector μ_{G_j} ; $j_1 = 1, 2, \dots, J_1$ and $J_1 \leq J$ where J_1 is the total number of groups with at least two relations.

For example, assume that we have the following 3 rules attributed to 3 different points of time:

A_{31}	A_{13}	A_{12}	A_{12}	\longrightarrow	0.039	0.242	0.641	0.078
A_{31}	A_{13}	A_{12}	A_{12}	\longrightarrow	0.001	0.004	0.988	0.006
A_{31}	A_{13}	A_{12}	A_{12}	\longrightarrow	0.005	0.986	0.006	0.002

These rules have similar left-hand sides (same fuzzified classes of the current states) but different vectors of membership degrees of the 3 next states (different right-hand sides). Hence, by calculating the average vector of membership degrees, we get the following final rule representing this group:

$$\boxed{A_{31} \ A_{31}, \ A_{12} \ A_{12} \ \longrightarrow \ [0.015 \ 0.411 \ 0.545 \ 0.029]}$$

(5) To calculate the forecasted output for each time point t , that is, to forecast $f_1(t+1)$, $t = 1, 2, \dots, T$, we have to match the current state of time point t with the left-hand sides of the grouped constructed fuzzy rules. Three possibilities could be faced.

- (1) If the current state exists in one of the groups, say G_j , then the forecasted output of year $t+1$, that is, $f_1(t+1)$, is calculated as

$$\mu_{G_j} * C' = [\mu_{t11} \ \mu_{t21} \ \dots \ \mu_{tK1}] \begin{bmatrix} C_{11} \\ C_{21} \\ \vdots \\ C_{K1} \end{bmatrix}, \quad (6)$$

where $C = [C_{11} \ C_{21} \ \dots \ C_{K1}]$ is the row vector of clusters' centers of the main variable.

- (2) If there are no fuzzy relations with the same left-hand side in the grouped constructed fuzzy rules whose current state is A_k which is the case if one or more of the time series points is missing or for the first future

time point in the time series, then we try to reduce the order of the model gradually to reach the same left-hand side; then we apply case (1).

- (3) If there are no fuzzy relations in the grouped constructed fuzzy logical rules with any order whose current state is A_k , then the forecasted output of time point $t + 1$ will be taken as the center of the cluster k .

(6) Calculate the fitness value that reflects the prediction accuracy of the main variable under the specified model. Theoretically speaking, the search is for the proper model that

- (1) minimizes a certain measure of solution error;
- (2) maximizes the separation between the clusters centers of the main variable.

Hence, a fitness function composed of two measures is proposed in this method to compare the possible models performance or prediction accuracy: the symmetric mean absolute percentage error (sMAPE) for measuring the solution error and the minimum absolute distance for measuring the cluster separation.

sMAPE can be defined as follows [22]:

$$sMAPE = \frac{1}{T} \sum_{t=1}^T \frac{|x_t - \hat{x}_t|}{(|x_t| + |\hat{x}_t|) / 2} \times 100, \quad (7)$$

where x_t is the actual value at time point t , \hat{x}_t is the forecasted value at time point t , and T is the number of time points used in the calculation. The formula above provides a result between 0% and 200%, which is considered an advantage to this measure. The closer the value of sMAPE to 0 is, the better the forecasting accuracy is.

sMAPE is well known in FTS literature as a measurement of solution error as it has an independent scale and it is less influenced by extreme values. It also corrects the computation asymmetry of the percentage error [22]. Hence, the fitness function in the new method depends mainly on the sMAPE as a measure of solution error; to improve the accuracy of the results, sMAPE has to be minimized.

The minimum absolute distance between the centers of clusters is chosen to measure the overlapping between clusters. The minimum absolute distance can be defined as follows:

$$d_{\min} = \min_{\substack{k,m=1,\dots,K \\ k \neq m}} |C_{k1} - C_{m1}|, \quad (8)$$

where k and m are any two clusters for the main variable. Conventionally, d_{\min} can take any positive value.

Now, we have a two-objective optimization problem. Since the parameters of the model take only integer values, it is not easy to find software that can solve multiobjective integer genetic algorithm. To overcome this deficiency, the weighted sum of objectives method [23] is used. This method is the simplest approach and is probably the most widely used one.

TABLE 1: Yearly car accident casualties in Belgium from 1974 to 2004.

Year	Killed	Mortally wounded	Died within 30 days	Severely wounded	Light casualties
1974	1,574	819	2,393	16,506	44,640
1975	1,460	701	2,161	15,792	42,423
1976	1,536	728	2,264	16,057	44,227
1977	1,597	701	2,298	15,830	44,995
1978	1,644	728	2,372	16,645	44,797
1979	1,572	544	2,116	15,750	42,346
1980	1,616	557	2,173	15,915	42,670
1981	1,564	454	2,018	15,091	41,915
1982	1,464	406	1,870	14,601	40,936
1983	1,479	412	1,891	14,864	42,023
1984	1,369	363	1,732	14,471	42,456
1985	1,308	352	1,660	13,287	39,879
1986	1,456	330	1,786	13,764	42,965
1987	1,390	380	1,770	13,809	44,090
1988	1,432	339	1,771	14,029	45,956
1989	1,488	312	1,800	14,515	46,667
1990	1,574	190	1,764	13,864	46,818
1991	1,471	209	1,680	12,965	43,578
1992	1,380	173	1,553	12,113	41,772
1993	1,346	171	1,517	11,680	41,736
1994	1,415	149	1,564	11,160	40,294
1995	1,228	109	1,337	10,267	39,140
1996	1,122	115	1,237	9,123	38,390
1997	1,150	105	1,255	9,229	39,594
1998	1,224	121	1,345	8,784	41,038
1999	1,173	126	1,299	8,461	41,841
2000	1,253	103	1,356	7,990	39,719
2001	1,288	90	1,378	7,319	38,747
2002	1,145	118	1,263	6,834	39,522
2003	1,035	101	1,136	6,898	42,445
2004	953	141	1,094	5,949	41,627

Source: [13].

The weighted sum method [23] scalarizes a set of objectives into a single objective by premultiplying each objective by a user-supplied weight. There is one critical question that arises here: what values of the weights must one use? The answer depends on the importance of each objective and the scaling of them.

The scale of the first objective (sMAPE) ranges from 0 to 2, while the second objective can take any nonnegative value. To guarantee that both objectives have the same scale, we redefine the second objective as follows:

$$\text{separation measure} = \frac{d_{\min}}{(|C_{k1}| + |C_{m1}|)/2}. \quad (9)$$

Equation (9) is considered as a proportional absolute distance between the centers of two clusters C_{k1} and C_{m1} relative to their average.

Accordingly, the overall fitness function can be defined as a linear combination of the two objectives (7) and (9) as follows:

fitness function

$$= \min \left[\lambda \left(\frac{1}{T} \sum_{t=1}^T \frac{|x_t - \hat{x}_t|}{(|x_t| + |\hat{x}_t|/2)} \right) - (1 - \lambda) \frac{d_{\min}}{(|C_{k1}| + |C_{m1}|/2)} \right]; \quad 0 < \lambda < 1, \quad (10)$$

where λ is the weight (a simulation study is conducted to determine suitable weights for the two objectives. The simulation result shows that the best value for the weight λ is 0.9 because it gives lower sMAPE) and it takes any arbitrary

TABLE 2: The comparison of the forecasting results between the new method and other methods for yearly car accident casualties in Belgium from 1974 to 2004.

Years	Actual killed	Jilani and Burney (2008) method 1	Jilani and Burney (2008) method 2	Lee et al. (2007)	Egrioglu et al. (2009)	The new method
1974	1574	—	—	—	—	—
1975	1460	—	—	—	—	—
1976	1536	—	—	—	—	—
1977	1597	1497	1497	1500	1600	1593
1978	1644	1497	1497	1500	1600	1616
1979	1572	1497	1497	1500	1600	1567
1980	1616	1598	1497	1500	1600	1621
1981	1564	1598	1497	1500	1600	1567
1982	1464	1498	1497	1500	1500	1458
1983	1479	1498	1497	1500	1500	1462
1984	1369	1398	1497	1500	1400	1353
1985	1308	1298	1396	1400	1300	1335
1986	1456	1498	1296	1300	1500	1458
1987	1390	1398	1497	1500	1400	1389
1988	1432	1398	1396	1400	1400	1452
1989	1488	1498	1396	1400	1500	1470
1990	1574	1598	1497	1500	1600	1567
1991	1471	1498	1497	1500	1500	1459
1992	1380	1398	1497	1500	1400	1370
1993	1346	1298	1396	1400	1300	1337
1994	1415	1398	1296	1300	1400	1435
1995	1228	1198	1396	1400	1200	1222
1996	1122	1098	1095	1100	1100	1169
1997	1150	1198	1196	1200	1200	1173
1998	1224	1198	1196	1200	1200	1215
1999	1173	1198	1196	1200	1200	1171
2000	1253	1298	1296	1300	1300	1273
2001	1288	1298	1296	1300	1300	1325
2002	1145	1098	1095	1100	1100	1173
2003	1035	997	995	1000	1000	1024
2004	953	997	995	1000	1000	1010
MSE	—	2188	7333	7285	978	445
MAPE	—	2.72	5.25	5.24	2.17	1.28

Source: except for the last column and the last two rows the table is replicated from [13].

value between 0 and 1. d_{\min} is the shortest distance between cluster centers.

The formula in (10) is considered as a tradeoff between the prediction accuracy and the reasonability of the clustering results.

Step 3 (selection and mutation for all chromosomes in the mating pool). To produce the chromosomes for the next generation, genetic algorithm creates three types of children for the next generation:

- (i) selection of elite children which are the individuals in the current generation with the best fitness values.

These individuals automatically survive to the next generation.

- (ii) Crossover children, created by combining the vectors of a pair of parents.
- (iii) Mutation children, created by introducing random changes, or mutations, to a single parent.

Step 4. Replace the current population with the offsprings to form the next generation.

Step 5. If the end conditions are not satisfied, go to Step 2. Otherwise, stop and retain the best solution in current population.

TABLE 3: Actual and forecasted values of the Alabama University enrollments.

Year	Enrollment	Kuo et al. (2009) [15]	Egrioglu (2012) [5]	Forecasted enrollment
1971	13055	—	—	—
1972	13563	—	13481	—
1973	13867	14000	13481	—
1974	14696	14000	14242	—
1975	15460	15500	15710	—
1976	15311	15500	15485	15401
1977	15603	16000	15936	15451
1978	15861	16000	15936	15756
1979	16807	16000	16838	16729
1980	16919	17500	17500	17448
1981	16388	16000	17500	16632
1982	15433	16000	16737	15397
1983	15497	15500	15485	15407
1984	15145	16000	15485	15417
1985	15163	16000	15710	15416
1986	15984	16000	15710	15989
1987	16859	16000	16838	16729
1988	18150	17500	17500	17448
1989	18970	18500	19144	19073
1990	19328	19500	19144	19092
1991	19337	19500	19144	19087
1992	18876	18500	19144	18964
MSE*	—	275539	265203	69492
MAPE*	—	2.65	2.24	1.16

*To have a valid comparison, MSE and MAPE are calculated for the same period (1976–1992).
Source: the 3 inner columns are taken from [15].

4. Experimental Results

In this section, we consider two examples, one for multivariate time series and the other for univariate case. Predictions that resulted from the new method are compared with the well-known results reported in the literature by other researchers. The forecast accuracy is compared by the mean absolute percentage error (MAPE) and mean square error (MSE). Suppose the actual value in time point t is y_t , the forecasted value in time point t is \hat{y}_t , T is the total number of time points used in calculation; then the MAPE and MSE are computed by the following equation, respectively:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t}, \quad \text{MSE} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2. \quad (11)$$

4.1. The Multivariate Case. Table 1 shows the total number of annual car accidents casualties in Belgium from 1974 to 2004 [13]. Here, the main time series is the number of killed persons and the auxiliary time series are mortally wounded and died within 30 days, severely wounded, and light casualties.

To implement the new method, The mixed integer genetic algorithm module in MATLAB package is used after making

slight modifications to cope with the nature of the FTS problem. The solution steps of the new method are carried out using a specially developed MATLAB code.

The best model selected by the new method, in the language of Figure 2, is represented by the chromosome (6 3 0 0 0 1). This means that the best model found is of order three (second gene) with one auxiliary variable (last gene); namely, the light casualties and a number of fuzzy sets equal 6 (first gene). The other three auxiliary variables are ruled out from the fitted model as their corresponding indicator variables take the value 0 in the chromosome. The trends of the actual values and the forecasted ones obtained by the new method are shown in Figure 3.

As seen in Table 2, the new method has the smallest values for MSE and MAPE: 445 and 1.28, respectively; that is, the new method outperforms the Jilani and Burney model [14] and Egrioglu et al. model [13] which are superior to other existing methods in the literature in terms of the mean square error and the average forecasting error rate.

4.2. The Univariate Case. New method is applied on the famous enrollment data in FTS literature (Table 3). The best model according to the results of the new method is (5 5). This means that the best model found is of order five and a



FIGURE 3: The actual and forecasted values of killed persons.

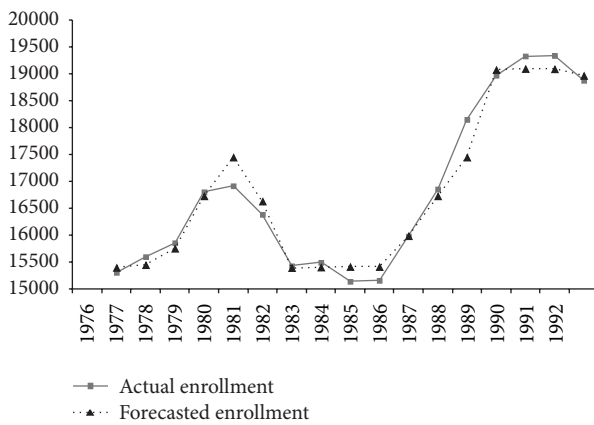


FIGURE 4: The actual and forecasted values of enrollments.

number of fuzzy sets equal 5. It should be noted that in the univariate time series the chromosome genes reduce to two as we have no auxiliary variables in this case. The trends of the actual values and the forecasted ones obtained by the new method are shown in Figure 4.

Comparison of the forecasting results is made between the new method and two other recently published ones, the Kuo et al. method [15] and the Egrioglu method [5]. Both mentioned methods considered the same enrollment data set and proved being superior to other existing methods in the literature in terms of the mean square error and the average forecasting error rate. From Table 3, it can be seen that the forecasting accuracy of the new method outperforms both mentioned methods in terms of MSE and MAPE.

5. Conclusion

When dealing with short multivariate time series, it is obvious that the assumptions of classical time series are not satisfactory. This paper introduces a new fuzzy time series method to deal with such cases. The new method drew on the ability of FTS approaches to deal with time series data without the need for checking any various theoretical assumptions. It also integrates the fuzzy clustering with genetic algorithm to

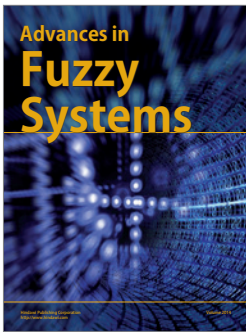
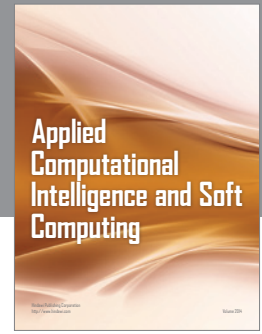
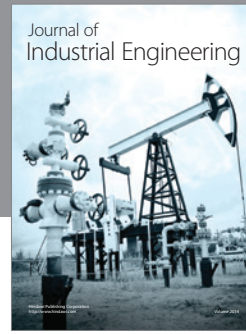
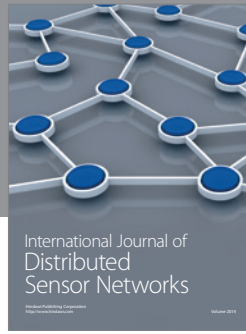
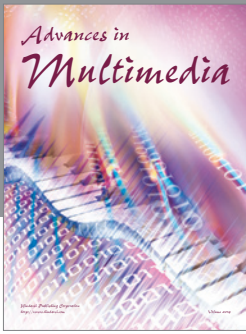
improve the prediction accuracy for solving the multivariate high order FTS.

In the new method, fuzzy clustering algorithm is responsible for fuzzifying the historical data based on its membership degrees to each cluster, then using these memberships to defuzzify the results. Also, genetic algorithm is responsible for selecting the proper model. The method is applied to time series of the yearly car accident casualties in Belgium from 1974 to 2004 and to the famous university enrollment data set. It provided forecasts with MSE and MAPE values that are smaller than those obtained from the other existing fuzzy time series methods.

References

- [1] K. Deep, K. P. Singh, M. L. Kansal, and C. Mohan, "A real coded genetic algorithm for solving integer and mixed integer optimization problems," *Applied Mathematics and Computation*, vol. 212, no. 2, pp. 505–518, 2009.
- [2] S.-M. Chen and N.-Y. Chung, "Forecasting enrollments using high-order fuzzy time series and genetic algorithms," *International Journal of Intelligent Systems*, vol. 21, no. 5, pp. 485–501, 2006.
- [3] L.-W. Lee, L.-H. Wang, and S.-M. Chen, "Temperature prediction and TAIEX forecasting based on fuzzy logical relationships and genetic algorithms," *Expert Systems with Applications*, vol. 33, no. 3, pp. 539–550, 2007.
- [4] H. I. Kang, "A fuzzy time series prediction method using the evolutionary algorithm," in *Advances in Intelligent Computing*, vol. 3645 of *Lecture Notes in Computer Science*, pp. 530–537, 2005.
- [5] E. Egrioglu, "A new time-invariant fuzzy time series forecasting method based on genetic algorithm," *Advances in Fuzzy Systems*, vol. 2012, Article ID 785709, 6 pages, 2012.
- [6] C.-H. Cheng, G.-W. Cheng, and J.-W. Wang, "Multi-attribute fuzzy time series method based on fuzzy clustering," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1235–1242, 2008.
- [7] S.-T. Li, Y.-C. Cheng, and S.-Y. Lin, "A FCM-based deterministic forecasting model for fuzzy time series," *Computers and Mathematics with Applications*, vol. 56, no. 12, pp. 3052–3063, 2008.
- [8] U. Yolcu, "The forecasting of Istanbul stock market with a high order multivariate fuzzy time series forecasting model," *An Official Journal of Turkish Fuzzy Systems Association*, vol. 3, no. 2, pp. 118–135, 2012.
- [9] 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.
- [10] Q. Song and B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 54, no. 3, pp. 269–277, 1993.
- [11] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [12] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synpaper," *IEEE Transactions on Computers*, vol. C-26, no. 12, pp. 1182–1191, 1977.
- [13] E. Egrioglu, C. H. Aladag, U. Yolcu, V. R. Uslu, and M. A. Basaran, "A new approach based on artificial neural networks for high order multivariate fuzzy time series," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10589–10594, 2009.
- [14] T. A. Jilani and S. M. A. Burney, "Multivariate stochastic fuzzy forecasting models," *Expert Systems with Applications*, vol. 35, no. 3, pp. 691–700, 2008.

- [15] I.-H. Kuo, S.-J. Horng, T.-W. Kao, T.-L. Lin, C.-L. Lee, and Y. Pan, "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6108–6117, 2009.
- [16] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—part II," *Fuzzy Sets and Systems*, vol. 62, no. 1, pp. 1–8, 1994.
- [17] S.-M. Chen and J.-R. Hwang, "Temperature prediction using fuzzy time series," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 30, no. 2, pp. 263–275, 2000.
- [18] T. H.-K. Yu and K.-H. Huarng, "A bivariate fuzzy time series model to forecast the TAIEX," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2945–2952, 2008.
- [19] M. A. T. LAB, *Global Optimization Toolbox User's Guide*, The Math Works, Natick, Mass, USA, 2011.
- [20] K. Huarng, "Heuristic models of fuzzy time series for forecasting," *Fuzzy Sets and Systems*, vol. 123, no. 3, pp. 369–386, 2001.
- [21] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Computers and Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984.
- [22] M. Hibon and S. Makridakis, *Evaluating Accuracy (or Error) Measures*, INSEAD, Fontainebleau, France, 1995.
- [23] K. Deb, *Multi-objective optimization using evolutionary algorithms*, John Wiley & Sons, New York, NY, USA, 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

