Industrial Management & Data Systems

Genetic algorithm based fuzzy time series tourism demand forecast model

Sumit Sakhuja Vipul Jain Sameer Kumar Charu Chandra Sarit K Ghildayal

Genetic algorithm based fuzzy time series tourism demand forecast model

Sumit Sakhuja
Mechanical Engineering Department, Indian Institute of Technology Delhi, New Delhi, India
Vipul Jain
Department of Industrial Engineering and Management, University of Sharjah, Sharjah, United Arab Emirates
Sameer Kumar
Department of Operations and Supply Chain Management, University of St Thomas, Opus College of Business, Minneapolis, Minnesota, USA
Charu Chandra
Dearborn College of Business Administration, University of Michigan – Dearborn, Dearborn, Michigan, USA, and
Sarit K. Ghildayal
Department of Computer Science, University of Minnesota, Minneapolis, Minnesota, USA

Received 1 May 2015
Revised 21 July 2015
20 September 2015
Accepted 24 September 2015

Abstract

Purpose – Many studies have proposed variant fuzzy time series models for uncertain and vague data. The purpose of this paper is to adapt a fuzzy time series combined with genetic algorithm (GA) to forecast tourist arrivals in Taiwan.

Design/methodology/approach – Different cases are studied to understand the effect of variation of fuzzy time series order, number of intervals and population size on the fitness function which decreases with increase in fuzzy time series order and number of fuzzy intervals, but do not have marginal effect due to change in population size.

Findings – Results based on an example of forecasting Taiwan's tourism demand was used to verify the efficacy of proposed model and confirmed its superiority to existing models providing solutions for different orders of fuzzy time series, number of intervals and population size with a smaller forecasting error as measured by root mean square error.

Originality/value – This study provides a viable forecasting methodology, adapting a fuzzy time series combined with an evolutionary GA. The proposed hybridized framework of fuzzy time series and GA, where GA is used to calibrate fuzzy interval length, is flexible and replicable to many industrial situations.

Keywords Decision support system, Evolutionary algorithm, GA, Adaptive fuzzy time series forecasting model, Interval calibration

Paper type Research paper

1. Introduction

In the past few decades, tourism has clearly become one of the most prominent economic trends for many countries. The tourism industry requires accurate forecasts in order to build and manage its service supply chains efficiently. The need for more

Industrial Management & Data Systems
Vol. 116 No. 3, 2016
pp. 483-507

0263-5577
DOI 10.1108/IMDS-05-2015-0165

accurate forecasts of tourism demand is driven by the desire to reduce risk and uncertainty. This need, according to Frechtling (2001), is further stressed by the perishable nature of tourism products, the simultaneous purchase-production process, and the role of complementary services in shaping consumer satisfaction and, thereby, its subsequent demand for future tourism services. Government bodies need accurate tourism demand forecasts to plan required tourism infrastructures, such as accommodation, site planning and transportation development. Tourism forecasting is classified into three main groups according to the methods and techniques adopted – an econometric-based approach, time series techniques and artificial intelligence (AI)-based methods. Before the 1990s, traditional regression approaches dominated the tourism forecasting and modeling literature. After pioneering up-to-date developments in econometric methodologies in recent years, the reputation of econometric forecasting models for improved accuracy has grown (Song and Li, 2008). Time series models such as ARIMA and GARCH (Alleyne, 2006; Gil-Alana *et al.*, 2004; Lee *et al.*, 2008; Lim and McAleer, 2002) and econometric models such as error correction model (ECM) and the vector autoregressive (VAR) models (Song and Witt, 2006; Wong *et al.*, 2007) are the most commonly used tourism demand forecasting techniques. The single exponential smoothing (SES) model is used to forecast a time series when there is no trend or seasonal pattern. According to Chen *et al.* (2008), SES is more suitable for a time series with seasonality removed. Other studies advocate that the ARIMA and SARIMA (Seasonal ARIMA) approaches are favored in tourism demand forecasting when the time series does not demonstrate structural breaks (Chu, 2008; Gustavsson and Nordström, 2001). Preez and Witt (2003) show that the ARIMA approach performs best in terms of forecasting accuracy and goodness of fit. Guo (2007) employs the gravity model to analyze inbound tourism demand to China, and Khadaroo and Seetanah (2008) use it to examine the effect of transportation infrastructure on tourism flows. According to Wang (2004), AI forecasting methods, including neural networks, rough sets theory, fuzzy time series theory, grey theory, genetic algorithms (GAs), and expert systems, tend to perform better than traditional forecasting methods. In traditional forecasting, piecewise linear function are the basic elements of the prediction model and users need to specify the functional form of the problem. Obtaining proper and valid models are done through experimentation with possible function relations and algorithms, and such experiments take comparatively longer time. Some popular techniques to solve the complex engineering and optimization problems (Konar, 2005) are AI techniques such as artificial neural networks (ANNs), fuzzy logic, and GAs. In conventional forecasting techniques a large amount of sample data and long-term historical data is required. Artificial models can be used to estimate the non-linear relationship, without the limits of traditional time series and econometric models. The new elements that are present now are the methods and techniques of soft computing and machine learning for decision making and forecasting, in particular neural networks and GAs, and the new science of the combination of those tools (Leigh *et al.*, 2002). Fuzzy time series combined with soft computing techniques can overcome these limitations and make appropriate short-term forecasting.

The objective of this paper is to put forward method of fuzzy time series combined with evolutionary algorithm (GA) to calibrate the length of fuzzy intervals of the fuzzy time series. The GA will aid in searching for the best fuzzy interval sizes corresponding to which the forecasting accuracy or mean square error is optimum. The study also aims to analyze the effect of different fuzzy time series order, number of intervals and population size.

This paper is organized as follows: a brief review of research on fuzzy time series and use of GA in fuzzy time series is discussed in Section 2. Overview of the procedure of forecasting using the fuzzy time series is described in Section 3. Section 4 describes the GA and also discusses the details of the new proposed forecast model. Section 5 discusses the experimental results obtained by the new proposed forecast model. Section 6 offers contribution of the study to knowledge and practice. Finally, Section 7 summarizes the conclusions and recommendations for future work.

2. Brief review of fuzzy time series and GA research

Short-time series data with limited past observations can be analyzed using fuzzy time series methods. Fuzzy time series method was first defined by Song and Chissom (1993). The definitions of fuzzy time series was proposed by them and also methods to model fuzzy relationships among observations were developed. Time variant and time invariant models were discussed and differentiated (Song and Chissom, 1994; Ju *et al.*, 1997; Chen and Hwang, 2000; Tsaur *et al.*, 2005; Cheikhrouhou *et al.*, 2011). Max-min composition operations in forecasting algorithm was presented by Song and Chissom (1993) and was followed by simplified arithmetic operations by Chen (1996) and Hwang *et al.* (1998). Chen (2002) suggested using higher order forecasting to deal with ambiguity. The effect of interval lengths on forecasting accuracy in fuzzy time series model was highlighted by Huarng (2001a, b), and Huarng proposed a method with distribution-based length and average-based length to reconcile this problem. Yu (2005) refined length of intervals. Huarng and Yu (2005) proposed type 2 fuzzy time series model and Huarng and Yu (2006) applied back propagation neural network for forecasting. Wang (2004) presented grey theory and fuzzy time series which are suitable for short-term time series data given a shortage of sample data. Singh (2007) proposed a new method of fuzzy time series forecasting based on difference parameters. There are also some prominent research studies that use higher order fuzzy time series. Eğrioğlu *et al.* (2010) proposed the approach which uses an optimization technique with a single-variable constraint is proposed to determine an optimal interval length in high-order fuzzy time series models. Gangwar and Kumar (2012) presented a computational method of forecasting based on multiple partitioning and higher order fuzzy time series. In a recent paper, Chan *et al.* (2015) test the effectiveness of fuzzy time series forecasting system in a supply chain with disruptions and demonstrate superiority of their model compared to other forecasting systems. They also show higher order fuzzy time series is more effective for upper tiers of supply chain in order to match autoregressive integrated moving average.

GAs were proposed by Holland (1975) and developed by Goldberg (1985). A GA is a stochastic global search technique that solves problems by imitating processes observed during natural evolution. GA was used before in the literature of fuzzy time series. Eğrioğlu *et al.* (2010, 2011a) suggested considering the problem of finding intervals as an optimization problem. Chen and Chung (2006) used different interval lengths, instead of a fixed interval length, obtained by using a GA. Chen and Chung (2006) use GA to tune up the length of each interval in the universe of discourse for one factor (variable), Lee *et al.* (2007) use GA to adjust the length of each interval in the universe of discourse for two factors. Kang (2005) uses GA to obtain the optimal fuzzy membership function, while Eğrioğlu (2012) uses GA for finding the elements of fuzzy relation matrix.

3. Methodology

This section explains the fuzzy time series (Song and Chissom 1993; Chen, 1996), and forecasting with fuzzy time series.

3.1 Fuzzy time series

The main difference between fuzzy time series and traditional time series is that the values of fuzzy time series are represented by fuzzy sets rather than real values. Let U be the universe of discourse, where $U = \{u_1, u_2, u_3, \ldots, u_n\}$, A fuzzy set A defined in the universe of discourse U can be represented as:

$$A = \mu_A(u_1)/u_1 + \mu_A(u_2)/u_2 + \mu_A(u_3)/u_3 + \cdots + \mu_A(u_n)/u_n$$

where μ_A denotes the membership function of fuzzy set A, μ_A: $U \rightarrow [0,1]$, and $\mu_A(u_i)$ denotes the degree of membership of u_i belonging to the fuzzy set A, and $\mu_A(u_i) \in [0,1]$, and $1 \leqslant i \leqslant n$.

The "+" sign in the above notation indicates how many different crisp numbers of the time series are in the fuzzy set A.

3.2 Forecasting with fuzzy time series

Stepwise procedure proposed by Song and Chissom (1993, 1994) model is described as below:

Step 1. Define the universe of discourse U for the historical data. First, we find the minimum data D_{min} and the maximum data D_{max} individually in the historical time series data, and then we define the universal discourse U as $U = [D_{min} - D_1, D_{max} + D_2]$, where D_1 and D_2 are two proper positive numbers.

Step 2. Partition the universe of discourse into equal length of intervals: $u_1, u_2, \ldots, u_n$. The number of intervals will be in accordance with the number of fuzzy sets $A_1, A_2, \ldots, A_n$ to be considered.

Step 3. Define the fuzzy sets A_i on universe of discourse U in Step 2. If there are fuzzy sets $A_1, A_2, \ldots, A_n$, then the fuzzy sets A_i $\forall_i = 1, 2, 3, \ldots, n$ can be described as:

$$A_i = \mu_{A_i}(u_1)/u_1 + \mu_{A_i}(u_2)/u_2 + \mu_{A_i}(u_3)/u_3 + \ldots + \mu_{A_i}(u_n)/u_n$$

Step 4. Fuzzification is the process of identifying associations between the historical values in the data set and the fuzzy sets defined in the previous step. Each historical value is fuzzified according to its highest degree of membership. If the highest degree of belongingness of a certain historical time variable, say $F(t-1)$ occurs at fuzzy set A_k, then $F(t-1)$ is fuzzified as A_k.

Step 5. Determine fuzzy relation matrix R. Establishing the fuzzy logical relations of various orders as given below:

(1) If for year $n-1$ and n the fuzzified arrivals are A_i and A_j, then the first order fuzzy logical relationship is represented as $A_i \rightarrow A_j$.

(2) If for year $n-2$, $n-1$ and n, the fuzzified arrivals are A_{i1}, A_i and A_j, respectively, then the second order fuzzy logical relation is represented as $A_{i1}, A_i \rightarrow A_j$.

(3) If for year $n-3$, $n-2$, $n-1$ and n, the fuzzified arrivals are A_{i2}, A_{i1}, A_i and A_j, respectively, then the third order fuzzy logical relation is represented as $A_{i2}, A_{i1}, A_i \rightarrow A_j$.

In a similar way we can find the various fifth, sixth, seventh, eighth and other higher order fuzzy logical relations.

Step 6. Based on the jth order fuzzy logical relationships, where $j \geqslant 2$ forecast arrivals using the following principles:

(1) If the jth order fuzzified historical arrivals for year i are A_{ij}, $A_{i(j-1)}$, …, and A_{i1}, where:

$j \geqslant 2$, and if there is the following fuzzy logical relationship in which the current state is "A_{ij}, $A_{i(j-1)}$, …, A_{i1}," shown as follows:

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \rightarrow A_j$$

where A_{ij}, $A_{i(j-1)}$, …, and A_{i1} are fuzzy sets, the maximum membership value A_k occurs at interval u_k, and the midpoint of u_k is m_k, then the forecasted arrivals of the ith year is m_k.

(2) If the jth order fuzzified historical arrivals for year i are A_{ij}, $A_{i(j-1)}$, …, and A_{i1}, where:

$j \geqslant 2$, and if there are the following fuzzy logical relationship in which the current state is "A_{ij}, $A_{i(j-1)}$, …, A_{i1}," shown as follows:

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \rightarrow A_{j1}$$

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \rightarrow A_{j2}$$

$$A_{ij}, A_{i(j-1)}, \ldots, A_{i1} \rightarrow A_{jp}$$

where A_{ij}, $A_{i(j-1)}$, …, A_{i1}, A_{j1}, A_{j2}, A_{jp} are fuzzy sets, the maximum membership values of A_{j1}, A_{j2}, …, and A_{jp} occur at intervals u_1, u_2, …, and u_p, respectively, and the midpoints of the interval u_1, u_2, …, and u_p are m_1, m_2, …, and m_p, respectively, then the forecasted arrivals the ith year is $\left(m_1 + m_2 + \ldots + m_p\right)/p$.

4. GA-based fuzzy time series forecasting

This section explains GA in brief and also proposes hybridization of fuzzy time series with GA.

4.1 GA

The GA is based on the process of natural selection and evolution and is applied in searching for the global optimum for many applications.

4.1.1 Outline of the basic GA

[Start] Generate random population of n chromosomes (suitable solutions for the problem)

[Fitness] Evaluate the fitness *f(x)* of each chromosome x in the population

[New population] Create a new population by repeating the following steps until the new population is complete

[Selection] Select two parent chromosomes from a population according to their fitness (the better the fitness, the greater the chance to be selected)

[Crossover] With a crossover probability, crossover the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.

[Mutation] With a mutation probability mutate new offspring at each locus (position in chromosome).

[Accepting] Place new offspring in a new population

[Replace] Use newly generated population for a further run of algorithm

[Test] If the end condition is satisfied, stop, and return the best solution in current population
[Loop] Go to **[Fitness]**

In this section a hybridized approach to fuzzy time series is explained. The defined universe of discourse is partitioned into N unequal intervals and the size of the intervals is calibrated with the help of GA. The objective here is to partition the universe of discourse in the best possible way, corresponding to which the root mean square error (RMSE) is minimum. Here, RMSE is the fitness function, which is to be minimized. Each interval in the universe is represented as a gene in a chromosome. Hence, each chromosome consists of a number of genes equal to the (number of intervals–1). Once the representation is chosen, the process of evolution is started by applying different steps of GA as discussed in Section 3. Figure 1 shows the flowchart of hybridization of fuzzy time series and GA, where GA is used to calibrate fuzzy interval length.

The algorithm for the proposed methodology is coded in JAVA language. The program was run on Pentium dual core processor with 4 GB RAM. For different cases undertaken in this work, each simulation run was carried out for 1000 generations.

5. Empirical analysis

The adopted methodology is used to forecast tourist arrivals based on Taiwan's tourism demand (Tsaur and Kuo, 2011). Table I shows the previous years' tourist arrivals in Taiwan. Initially the forecasting with fuzzy time series is demonstrated with third order fuzzy time series. Later the fuzzy intervals are tuned using GA:

Step 1. Let D_{min} and D_{max} be the minimum and maximum number of tourist arrivals in a given period. Based on the given D_{min} and D_{max}, define the universe of discourse U as:

$[D_{min}–D_1, D_{max}+D_2]$, where D_1 and D_2 are positive integers. From Table II we can see that $D_{min}=40{,}256$, $D_{max}=298{,}282$, so taking $D_1=256$ and $D_2=1{,}718$, universe of discourse $U=[40000\ 300000]$.

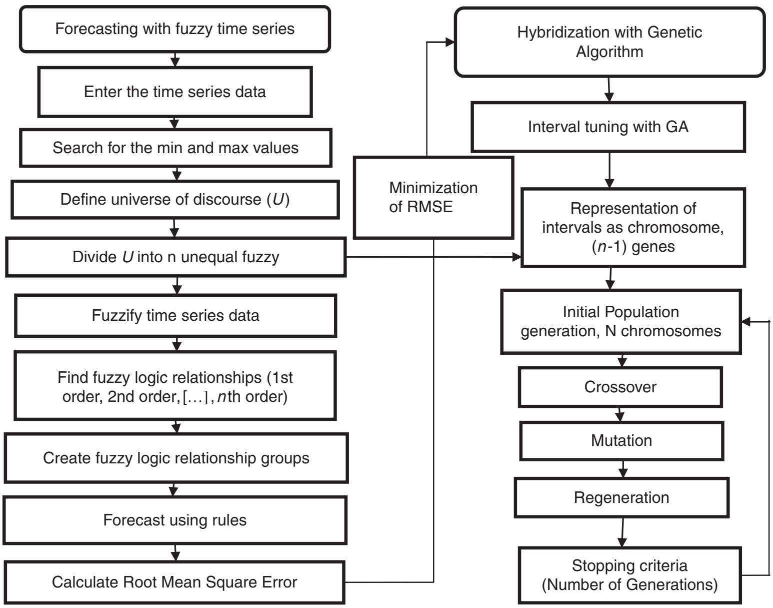

Figure 1.
Hybridization of fuzzy time series with GAs

Year/month	Arrivals	Year/month	Arrivals	Year/month	Arrivals
2000/05	216,692	2002/03	281,522	2004/01	212,854
2000/06	225,069	2002/04	245,759	2004/02	221,020
2000/07	217,302	2002/05	243,941	2004/03	239,575
2000/08	220,227	2002/06	241,378	2004/04	229,061
2000/09	221,504	2002/07	234,596	2004/05	232,293
2000/10	249,352	2002/08	246,079	2004/06	258,861
2000/11	232,810	2002/09	233,613	2004/07	243,396
2000/12	228,821	2002/10	258,360	2004/08	253,544
2001/01	199,800	2002/11	255,645	2004/09	245,915
2001/02	234,386	2002/12	285,303	2004/10	266,590
2001/03	251,111	2003/01	238,031	2004/11	270,553
2001/04	235,251	2003/02	259,966	2004/12	276,680
2001/05	227,021	2003/03	258,128	2005/01	244,252
2001/06	239,878	2003/04	110,640	2005/02	257,340
2001/07	218,673	2003/05	40,256	2005/03	298,282
2001/08	224,208	2003/06	57,131	2005/04	269,513
2001/09	193,254	2003/07	154,174	2005/05	284,049
2001/10	192,452	2003/08	200,614	2005/06	293,044
2001/11	190,500	2003/09	218,594	2005/07	268,269
2001/12	210,603	2003/10	223,552	2005/08	281,693
2002/01	217,600	2003/11	241,349	2005/09	270,700
2002/02	233,896	2003/12	245,682		

Source: Tsaur and Kuo (2011)

Table I. Previous year's tourist arrivals

Step 2. Divide the universe into n intervals, u_1, u_2, $u_3 \ldots u_n$, where $u_1 = [40000\, x_1]$, $u_2 = [x_1\, x_2] \ldots u_n = [x_{n\text{-}1} 300000]$ and $x_1 \leqslant x_2 \leqslant x_3 \leqslant \ldots \leqslant x_{n\text{-}1}$. Each chromosome is repritalic>−1 geitalic>−1 genes −1 genes as shown:

40,000	x_1	x_2	x_3	…	x_{n-2}	x_{n-1}	300,000

Let the universe of discourse be divided into nine intervals, that are $u_1 = [40000, 65419]$; $u_2 = [65419,\ 142914]$; $u_3 = [142914,\ 210164]$; $u_4 = [210164,\ 230853]$; $u_5 = [230853,\ 238009]$; $u_6 = [238009,\ 244466]$; $u_7 = [244466,\ 259300]$; $u_8 = [259300,\ 284659]$; $u_9 = [284659, 300000]$.

Step 3. Define the fuzzy sets A_i on universe of discourse U in Step 2. Figure 2 illustrates membership functions for fuzzy sets A_i.

Fuzzy sets A_i's are defined below:

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_9}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_9}$$

$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \cdots + \frac{0}{u_9}$$

$$\vdots$$

$$\vdots$$

$$A_8 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_6} + \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9}$$

$$A_9 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0.5}{u_8} + \frac{1}{u_9}$$

Step 4. The fuzzified results of tourist arrivals are shown in Table II.

Step 5. Generate fuzzy logical relationship based on the 3rd order fuzzy time series.

Step 6. Based on the 3rd order fuzzy logical relationship, calculate the forecasted arrivals using the fuzzy time series concepts discussed in Section 3. RMSE is calculated based on intermediate calculations shown in Table II:

$$\text{RMSE} = \sqrt{\frac{\sum (A_i - F_i)^2}{n}} = \sqrt{\frac{14{,}100{,}094{,}037.00}{65}} = 14{,}728.36$$

This result corresponds to the third order fuzzy time series by dividing the universe of discourse into nine unequal intervals. In order to find out the best combination of interval length corresponding to which the RMSE is minimum, genetic calibrating of intervals is done.

5.1 Genetic calibrating of intervals

To obtain the best possible set of fuzzy intervals corresponding to which RMSE is minimum, genetic calibration of intervals is performed. A set of randomly generated chromosomes is considered as the initial population for the evolution. The population is reproduced using GA operators. First, selection procedure is carried out in order to select the best chromosomes from the initial population. For this the tournament selection is done with selection pressure of 2. Two chromosomes are chosen randomly from the initial population and their fitness values (RMSE) are compared, the chromosome with the minimum fitness value is selected for the reproduction. A number of tournaments are carried out to get a number of chromosomes for reproduction. In the next step crossover operation is done. One-point crossover is done on the selected chromosomes. Crossover probability is generally kept close to one so as to evolve new population faster. In our study we have taken crossover probability $P_c = 0.9$. After crossover mutation is done, a gene is randomly taken from the chromosome and its value is changed arbitrarily within the range. Mutation probability is taken

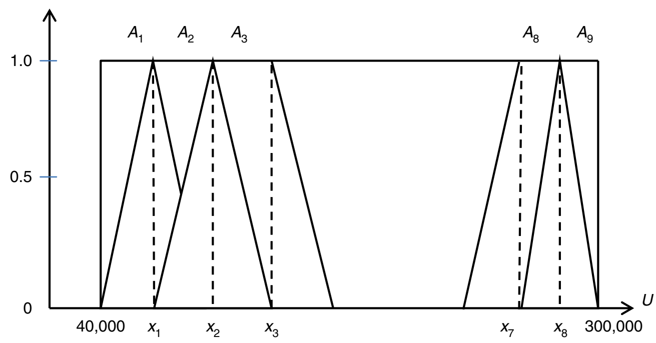

Figure 2. Membership functions constructed from the genes $x_1, x_2, x_3, \ldots, x_8$ of a chromosome

Table II. Fuzzified and forecasted tourist arrivals

Year	Actual arrivals (A_i)	Fuzzified arrivals	Forecasted arrivals (F_i)	$A_i–F_i$	$(A_i–F_i)^2$	Year	Actual arrivals (A_i)	Fuzzified arrivals	Forecasted arrivals (F_i)	$A_i–F_i$	$(A_i–F_i)^2$
2000/05	216,692	A4	0			2003/02	259,966	A8	259,300	666	443,556
2000/06	225,069	A4	0			2003/03	258,128	A7	259,300	−1,172	137,3584
2000/07	217,302	A4	210,164	7,138	50,951,044	2003/04	110,640	A2	142,914	−32,274	1,041,611,076
2000/08	220,227	A4	237,976	−17,749	315,027,001	2003/05	40,256	A1	65,419	−25,163	633,176,569
2000/09	221,504	A4	237,976	−16,472	271,326,784	2003/06	57,131	A1	65,419	−8,288	68,690,944
2000/10	249,352	A7	237,976	11,376	129,413,376	2003/07	154,174	A3	142,914	11,260	126,787,600
2000/11	232,810	A5	230,853	1,957	3,829,849	2003/08	200,614	A3	210,164	−9,550	91,202,500
2000/12	228,821	A4	230,853	−2,032	4,129,024	2003/09	218,594	A4	210,164	8,430	71,064,900
2001/01	199,800	A3	22,7315	−27,515	757,075,225	2003/10	223,552	A4	220,508	3,044	9,265,936
2001/02	234,386	A5	230,853	3,533	12,482,089	2003/11	241,349	A6	237,976	3,373	11,377,129
2001/03	251,111	A7	259,300	−8,189	67,059,721	2003/12	245,682	A7	244,466	1,216	1,478,656
2001/04	235,251	A5	230,853	4,398	19,342,404	2004/01	212,854	A4	210,164	2,690	7,236,100
2001/05	227,021	A4	230,853	−3,832	14,684,224	2004/02	221,020	A4	210,164	10,856	117,852,736
2001/06	239,878	A6	227,315	12,563	157,828,969	2004/03	239,575	A6	190,461	49,114	2,412,184,996
2001/07	218,673	A4	210,164	8,509	72,403,081	2004/04	229,061	A4	230,853	−1,792	3,211,264
2001/08	224,208	A4	210,164	14,044	197,233,936	2004/05	232,293	A5	230,853	1,440	2,073,600
2001/09	193,254	A3	190,461	2,793	7,800,849	2004/06	258,861	A7	259,300	−439	192,721
2001/10	192,452	A3	142,914	49,538	2,454,013,444	2004/07	243,396	A6	244,466	−1,070	1,144,900
2001/11	190,500	A3	142,914	47,586	2,264,427,396	2004/08	253,544	A7	259,300	−5,756	33,131,536
2001/12	210,603	A4	210,164	439	19,2721	2004/09	245,915	A7	264,562	−18,647	347,710,609
2002/01	217,600	A4	210,164	7,436	55,294,096	2004/10	266,590	A8	259,300	7,290	53,144,100
2002/02	233,896	A5	220,508	13,388	179,238,544	2004/11	270,553	A8	264,562	5,991	35,892,081
2002/03	281,522	A8	284,659	−3,137	9,840,769	2004/12	276,680	A8	284,659	−7,979	63,664,441
2002/04	245,759	A7	244,466	1,293	167,1849	2005/01	244,252	A6	264,562	−20,310	412,496,100
2002/05	243,941	A6	244,466	−525	275,625	2005/02	257,340	A7	259,300	−1,960	3,841,600
2002/06	241,378	A6	244,466	−3,088	9,535,744	2005/03	298,282	A9	284,659	13,623	185,586,129
2002/07	234,596	A5	230,853	3,743	14,010,049	2005/04	269,513	A8	284,659	−15,146	229,401,316
2002/08	246,079	A7	244,466	1,613	260,1769	2005/05	284,049	A8	264,562	19,487	379,743,169
2002/09	233,613	A5	230,853	2,760	761,7600	2005/06	293,044	A9	271,979	21,065	443,734,225
2002/10	258,360	A7	259,300	−940	883,600	2005/07	268,269	A8	271,979	−3,710	13,764,100
2002/11	255,645	A7	259,300	−3,655	13,359,025	2005/08	281,693	A8	284,659	−2,966	8,797,156
2002/12	285,303	A9	284,659	644	414,736	2005/09	270,700	A8	284,659	−13,959	194,853,681
2003/01	238,031	A6	238,009	22	484						

as $P_m = 0.05$. This reproduction process is carried out for several generations in order to get the chromosomes with minimum RMSE. The stopping criterion is number of generations which are fixed at 1000 generations.

Let chromosomes 1 and 2 are randomly chosen and go for the tournament selection process. It can be seen that, RMSE of chromosome 1 is 14,716 and that of chromosome 2 is 20,373, so chromosome 1 (minimum of two RMSE values) will be selected and will go to the population:

Chromosome 1	40,000	62,338	147,698	207,715	236,543	238,627	251,439	281,915	300,000	14,716.70075
Chromosome 2	40,000	91,719	100,009	171,189	204,860	232,058	256,366	292,755	300,000	20,373.54374

Let two chromosomes are randomly selected for the one-point crossover process. In this, a point is randomly chosen in the two selected chromosomes, and the genes are exchange between the two chromosomes after the selected point of crossover. For example let the crossover point is 5, and then genes after point 5 will be exchanged between the two chromosomes:

Crossover Point

Chromosome 1	40,000	62,338	147,698	207,715	236,543	238,627	251,439	281,915	300,000
Chromosome 2	40,000	91,719	100,009	171,189	204,860	232,058	256,366	292,755	300,000

Before Crossover

Chromosome 1	40,000	62,338	147,698	207,715	236,543	232,058	256,366	292,755	300,000
Chromosome 2	40,000	91,719	100,009	171,189	204,860	238,627	251,439	281,915	300,000

After Crossover

The mutation is carried out by randomly selecting a point in a chromosome and then replacing the gene at that point by some other random number between the adjacent genes as shown below:

Mutation Point

Chromosome	40,000	62,338	147,698	207,715	236,543	238,627	251,439	281,915	300,000

Before Mutation

Chromosome	40,000	62,338	147,698	186,776	236,543	238,627	251,439	281,915	300,000

After Mutation

A set of 50 randomly generated chromosomes forms the initial population for the evolution shown in Table III. Tables IV-VI show the various operations of GA performed on the initial population.

In order to analyze the effect of different fuzzy time series parameters, the forecasting problem is solved with different combinations of parameters. In this study the following cases are undertaken.

Case I. Effect of change of order (fuzzy time series parameter): in this case population size and number of intervals are kept constant and the orders of fuzzy time series are varied from 1 to 10. Population size is taken as 200, and number of intervals is 8. The minimum RMSE obtained is 5,851 for order 10.

Table III. Initial population (randomly generated, population size: 50)

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 1	40000	62338	147698	207715	236543	238627	251439	281915	300000	14,716.70075
Chromosome 2	40000	91719	100009	171189	204860	232058	256366	292755	300000	20,373.54374
Chromosome 3	40000	59690	98767	176002	214995	225675	269748	276997	300000	22,720.26607
Chromosome 4	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.16305
Chromosome 5	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 6	40000	63417	130199	144067	172703	234844	251173	282149	300000	23,371.27791
Chromosome 7	40000	53217	76785	180454	196721	228281	268456	283346	300000	23,387.07329
Chromosome 8	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 9	40000	48463	125640	184273	196546	201561	225364	244015	300000	24,771.5018
Chromosome 10	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 11	40000	46014	66388	98694	99048	216526	242277	291426	300000	25,028.44017
Chromosome 12	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.79442
Chromosome 13	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.73144
Chromosome 14	40000	44920	45892	153881	194435	220624	270446	277174	300000	25,523.23659
Chromosome 15	40000	68080	118976	152603	186172	240334	250637	262292	300000	25,549.18707
Chromosome 16	40000	44670	85649	132836	172309	216130	228433	246661	300000	25,913.84605
Chromosome 17	40000	48210	154202	204156	214917	255458	273037	273767	300000	26,083.53283
Chromosome 18	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.02161
Chromosome 19	40000	53829	79780	109667	163116	163354	216950	239539	300000	26,321.47435
Chromosome 20	40000	82320	93909	100673	229873	231931	256378	281393	300000	26,346.46339
Chromosome 21	40000	45483	101395	126573	186571	228250	265131	293047	300000	26,812.51814
Chromosome 22	40000	53258	92611	128073	137692	176317	225493	239619	300000	26,963.97553
Chromosome 23	40000	72138	88004	98760	108717	162289	201109	250495	300000	27,098.31479
Chromosome 24	40000	85167	145335	238743	243992	273770	277751	284919	300000	27,109.51098
Chromosome 25	40000	56518	92361	100172	136989	163247	241799	255465	300000	27,307.80317
Chromosome 26	40000	59866	70689	143693	239624	240379	247569	261321	300000	27,372.23276
Chromosome 27	40000	43420	172497	182044	215194	228407	284478	286945	300000	27,503.95424
Chromosome 28	40000	98694	104210	149348	151334	233158	265445	296809	300000	27,774.94559
Chromosome 29	40000	77163	100961	104415	126069	149643	259284	293481	300000	27,909.8545
Chromosome 30	40000	66268	112323	133178	147240	228669	237623	280460	300000	27,979.04956

(*continued*)

Table III.

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 31	40000	50618	50858	56784	85759	132356	241182	287267	300000	29,405.66035
Chromosome 32	40000	67476	81680	134876	165247	185295	251123	253194	300000	31,334.10926
Chromosome 33	40000	43102	54826	72433	110266	168350	193813	268151	300000	31,668.09085
Chromosome 34	40000	68832	70819	71266	173331	177468	225949	297227	300000	32,240.80683
Chromosome 35	40000	117991	149824	174649	200545	218259	227094	282338	300000	32,705.89428
Chromosome 36	40000	40244	103783	107886	117814	178858	213623	281933	300000	32,742.47525
Chromosome 37	40000	70302	71276	86928	92768	144243	180915	280832	300000	33,433.72428
Chromosome 38	40000	44761	48216	82197	109759	243217	262188	276770	300000	33,445.13211
Chromosome 39	40000	93213	103227	158195	199647	201638	218499	291796	300000	34,100.59337
Chromosome 40	40000	50720	98261	124881	181210	181363	273079	295092	300000	34,413.08514
Chromosome 41	40000	71108	88198	121751	154646	194192	261374	297649	300000	36,915.71579
Chromosome 42	40000	42740	49901	113847	115164	156407	232703	295907	300000	37,036.65842
Chromosome 43	40000	57248	76276	131424	177653	215707	226414	234323	300000	40,400.93222
Chromosome 44	40000	43029	53309	68752	79021	134514	143816	271146	300000	40,641.33631
Chromosome 45	40000	85096	116702	130125	214208	227987	233580	234510	300000	40,968.82875
Chromosome 46	40000	124940	125536	129380	129632	196548	208711	277975	300000	41,823.61428
Chromosome 47	40000	117111	121579	147776	190854	202695	213395	228328	300000	50,766.36682
Chromosome 48	40000	42617	74543	80930	174593	187177	198806	200402	300000	54,105.00766
Chromosome 49	40000	56276	73236	112435	135050	165226	192337	215343	300000	57,670.59126
Chromosome 50	40000	45566	68052	91704	106883	107951	132525	190753	300000	68,167.92898

Table IV. Selected population (tournament selection; selection pressure: 2)

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 1	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.73144
Chromosome 2	40000	59690	98767	176002	214995	225675	269748	276997	300000	22,720.26607
Chromosome 3	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 4	40000	44920	45892	153881	194435	220624	270446	277174	300000	25,523.23659
Chromosome 5	40000	68080	118976	152603	186172	240334	250637	262292	300000	25,549.18707
Chromosome 6	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 7	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 8	40000	63417	130199	144067	172703	234844	251173	282149	300000	23,371.27791
Chromosome 9	40000	59866	70689	143693	239624	240379	247569	261321	300000	27,372.23276
Chromosome 10	40000	63417	130199	144067	172703	234844	251173	282149	300000	23,371.27791
Chromosome 11	40000	91719	100009	171189	204860	232058	256366	292755	300000	20,373.54374
Chromosome 12	40000	53217	76785	180454	196721	228281	268456	283346	300000	23,387.07329
Chromosome 13	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 14	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 15	40000	117991	149824	174649	200545	218259	227094	282338	300000	32,705.89428
Chromosome 16	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 17	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 18	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 19	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 20	40000	53217	76785	180454	196721	228281	268456	283346	300000	23,387.07329
Chromosome 21	40000	48210	154202	204156	214917	255458	273037	273767	300000	26,083.53283
Chromosome 22	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 23	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.73144
Chromosome 24	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.16305
Chromosome 25	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 26	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.79442
Chromosome 27	40000	77163	100961	104415	126069	149643	259284	293481	300000	27,909.8545
Chromosome 28	40000	56518	92361	100172	136989	163247	241799	255465	300000	27,307.80317
Chromosome 29	40000	82320	93909	100673	229873	231931	256378	281393	300000	26,346.46339
Chromosome 30	40000	82320	93909	100673	229873	231931	256378	281393	300000	26,346.46339

(*continued*)

Table IV.

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 31	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 32	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.02161
Chromosome 33	40000	98694	104210	149348	151334	233158	265445	296809	300000	27,774.94559
Chromosome 34	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 35	40000	44920	45892	153881	194435	220624	270446	277174	300000	25,523.23659
Chromosome 36	40000	68080	118976	152603	186172	240334	250637	262292	300000	25,549.18707
Chromosome 37	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.79442
Chromosome 38	40000	53829	79780	109667	163116	163354	216950	239539	300000	26,321.47435
Chromosome 39	40000	56518	92361	100172	136989	163247	241799	255465	300000	27,307.80317
Chromosome 40	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 41	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.99431
Chromosome 42	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 43	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.16305
Chromosome 44	40000	62338	147698	207715	236543	238627	251439	281915	300000	14,716.70075
Chromosome 45	40000	45483	101395	126573	186571	228250	265131	293047	300000	26,812.51814
Chromosome 46	40000	68080	118976	152603	186172	240334	250637	262292	300000	25,549.18707
Chromosome 47	40000	68080	118976	152603	186172	240334	250637	262292	300000	25,549.18707
Chromosome 48	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.23139
Chromosome 49	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.3352
Chromosome 50	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.73144

Table V. Population after crossover (one-point crossover; crossover probability, $P_c = 0.9$)

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 1	40000	62481	80394	145226	145766	174794	229466	258520	300000	22,989.429
Chromosome 2	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 3	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 4	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 5	40000	48468	87079	99909	172135	211563	259058	281393	300000	18,156.784
Chromosome 6	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 7	40000	59630	134980	159907	189941	196234	241799	255465	300000	23,298.79
Chromosome 8	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794
Chromosome 9	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 10	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 11	40000	63417	130199	144067	172703	234844	250637	262292	300000	19,380.114
Chromosome 12	40000	63417	130199	144067	172703	234844	251173	282149	300000	23,371.278
Chromosome 13	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 14	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 15	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 16	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 17	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.731
Chromosome 18	40000	117991	149824	168080	174649	233225	251766	286212	300000	26,914.35
Chromosome 19	40000	89112	146349	208024	221517	262571	266113	278959	300000	23,720.785
Chromosome 20	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.231
Chromosome 21	40000	63417	104618	116965	145226	174794	229466	258520	300000	23,000.033
Chromosome 22	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 23	40000	57357	99909	109017	172135	211563	259058	294655	300000	22,910.459
Chromosome 24	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 25	40000	57357	109017	188423	199626	211078	238590	262571	300000	22,016.299
Chromosome 26	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.231
Chromosome 27	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.022
Chromosome 28	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.022
Chromosome 29	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 30	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335

(*continued*)

Table V.

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 31	40000	117991	149824	174649	200545	218259	227094	258520	300000	22,503.202
Chromosome 32	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 33	40000	53829	79780	109667	163116	174794	229466	258520	300000	23,056.341
Chromosome 34	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 35	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 36	40000	89112	116965	145226	146349	174794	229466	258520	300000	23,218.46
Chromosome 37	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 38	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 39	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 40	40000	57357	109017	172135	188423	211563	259058	294655	300000	22,918.309
Chromosome 41	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 42	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 43	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 44	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 45	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 46	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 47	40000	57357	109017	180454	196721	228281	268456	283346	300000	22,999.226
Chromosome 48	40000	53217	76785	180454	196721	228281	268456	283346	300000	23,387.073
Chromosome 49	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794
Chromosome 50	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794

Table VI. Population after mutation (random mutation; mutation probability, $P_m = 0.05$)

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 1	40000	62481	80394	145226	145766	174794	229466	258520	300000	22,989.429
Chromosome 2	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 3	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 4	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 5	40000	48468	87079	99909	172135	211563	259058	281393	300000	18,156.784
Chromosome 6	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 7	40000	59630	134980	159907	189941	196234	241799	255465	300000	23,298.79
Chromosome 8	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794
Chromosome 9	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 10	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 11	40000	63417	130199	144067	172703	234844	250637	262292	300000	19,380.114
Chromosome 12	40000	63417	130199	144067	172703	234844	251173	282149	300000	23,371.278
Chromosome 13	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 14	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 15	40000	77911	104618	116965	120440	174794	229466	258520	300000	23,069.409
Chromosome 16	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 17	40000	62481	80394	145766	168080	233225	251766	286212	300000	25,359.731
Chromosome 18	40000	117991	149824	168080	174649	233225	251766	286212	300000	26,914.35
Chromosome 19	40000	89112	146349	208024	221517	262571	266113	278959	300000	23,720.785
Chromosome 20	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.231
Chromosome 21	40000	63417	104618	116965	145226	174794	229466	258520	300000	23,000.033
Chromosome 22	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 23	40000	57357	99909	109017	172135	211563	259058	294655	300000	22,910.459
Chromosome 24	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 25	40000	57357	109017	188423	199626	211078	238590	262571	300000	22,016.299
Chromosome 26	40000	57357	109017	188423	199626	262571	266113	278959	300000	23,988.231
Chromosome 27	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.022
Chromosome 28	40000	108612	132996	147344	183972	208439	211078	238590	300000	26,252.022
Chromosome 29	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 30	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335

(*continued*)

Table VI.

Chromosomes	Gene1	Gene2	Gene3	Gene4	Gene5	Gene6	Gene7	Gene8	Gene9	RMSE
Chromosome 31	40000	117991	149824	174649	200545	218259	227094	258520	300000	22,503.202
Chromosome 32	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 33	40000	53829	79780	109667	163116	174794	229466	258520	300000	23,056.341
Chromosome 34	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 35	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 36	40000	89112	116965	145226	146349	174794	229466	258520	300000	23,218.46
Chromosome 37	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 38	40000	89112	146349	208024	221517	234754	235463	279603	300000	24,779.994
Chromosome 39	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 40	40000	57357	109017	172135	188423	211563	259058	294655	300000	22,918.309
Chromosome 41	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 42	40000	48468	87079	99909	172135	211563	259058	294655	300000	22,918.163
Chromosome 43	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 44	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 45	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 46	40000	77911	104618	116965	145226	174794	229466	258520	300000	23,066.335
Chromosome 47	40000	57357	109017	180454	196721	228281	268456	283346	300000	22,999.226
Chromosome 48	40000	53217	76785	180454	196721	228281	268456	283346	300000	23,387.073
Chromosome 49	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794
Chromosome 50	40000	59630	134980	159907	189941	196234	211107	241124	300000	25,142.794

Case II. Effect of change of number of intervals (fuzzy time series parameter): in this case population size and fuzzy time series order are kept constant and four sets (number of intervals = 6, 8, 10, 12) of interval size are analyzed. In all cases population size is taken as 200 and is solved for 3rd order fuzzy time series. The minimum RMSE obtained is 3,757 for 12 intervals.

Case III. Effect of change of population size (GA parameter): in this case effect of change of population size is analyzed. For 3rd order fuzzy time series with eight intervals, results are analyzed for population size of 100, 200 and 300, respectively. No marginal difference was noted for change in population size.

For each setting three runs are carried out and are plotted on the number of generations vs RMSE curve. In all the cases a convergence curve is obtained showing that GA is searching the best possible chromosome from the given population in each run. The convergence curves for the three cases studied are shown in Figures 3-5.

From the results obtained for different settings, a simulation run was carried out for a best combination of parameter, that was population size = 200, number of intervals = 12 and fuzzy time series order = 9.

The minimum RMSE obtained was 3,140 and the corresponding fuzzy intervals are:

u_1 = [40000 40361]; u_2 = [40361 79514]; u_3 = [79514 135505]; u_4 = [135505 179054]; u_5 = [179054 210507]; u_6 = [210507 222361]; u_7 = [222361 230649]; u_8 = [230649 238073]; u_9 = [238073 248935]; u_{10} = [248935 264279]; u_{11} = [264279 276717]; u_{12} = [276717 300000].

The convergence curve for this case is shown in Figure 6.

Table VII shows the summary of RMSE for different parameter settings and a plot is made for each setting, Figure 7(a-c) representing Case I, Case II and Case III, respectively, described above from which conclusions can be drawn.

The results obtained from the adopted methodology are compared with the already existing and traditional forecasting methods and this comparison of results is shown in Table VIII. The adopted methodology gives the least value of RMSE thereby having the maximum forecast accuracy.

6. Contribution to knowledge and practice

This study provides a viable methodology for determining tourist arrival forecast in a country as an example, adapting a fuzzy time series combined with an evolutionary GA. The proposed hybridized framework of fuzzy time series and GA, where GA is used to calibrate fuzzy interval length is flexible and replicable to many industrial situations. The GA facilitates in searching for the best interval sizes corresponding to which the forecasting accuracy or mean square error is optimum. In order to analyze the effect of different fuzzy time series parameters, the forecasting problem is solved with different combinations of parameters such as the effect of different fuzzy time series order, number of intervals and population size. The research focusses on an important application of decision support system for forecasting for a significant service industry. The proposed model will provide decision makers with improved estimation and decreased error in complex and uncertain environment.

7. Conclusions and scope for future work

In this study a hybrid fuzzy time series is adopted which combines fuzzy time series with GA and is used to forecast tourist arrivals. The proposed model will provide

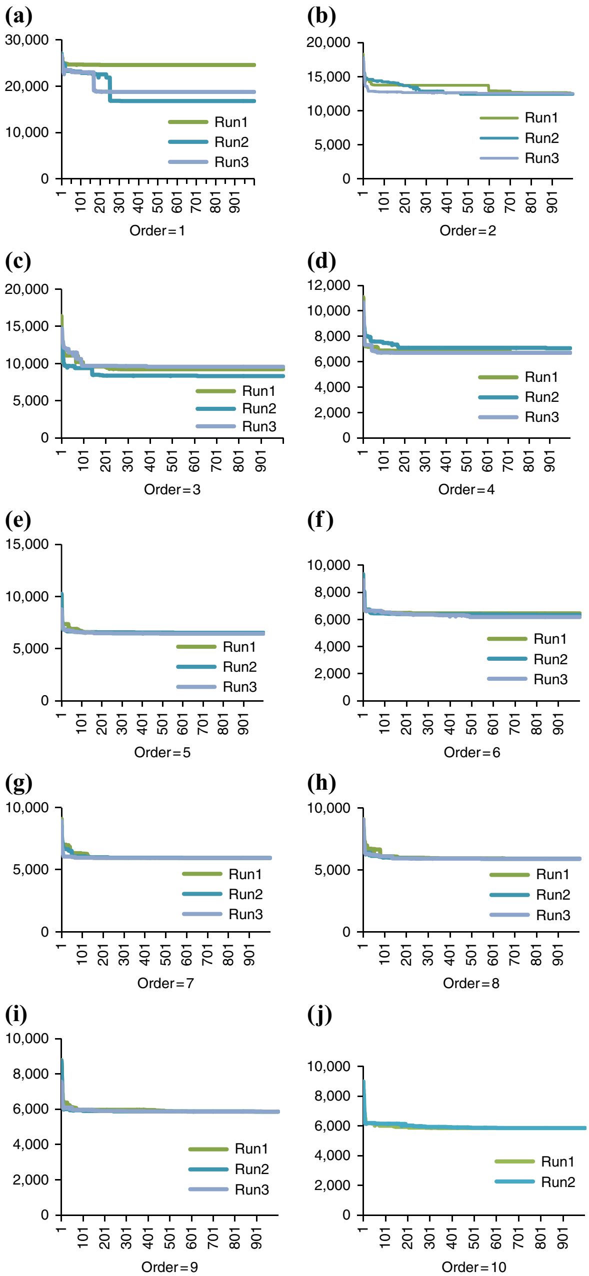

Figure 3.
CASE I change of order (population size = 200; number of intervals = 8)

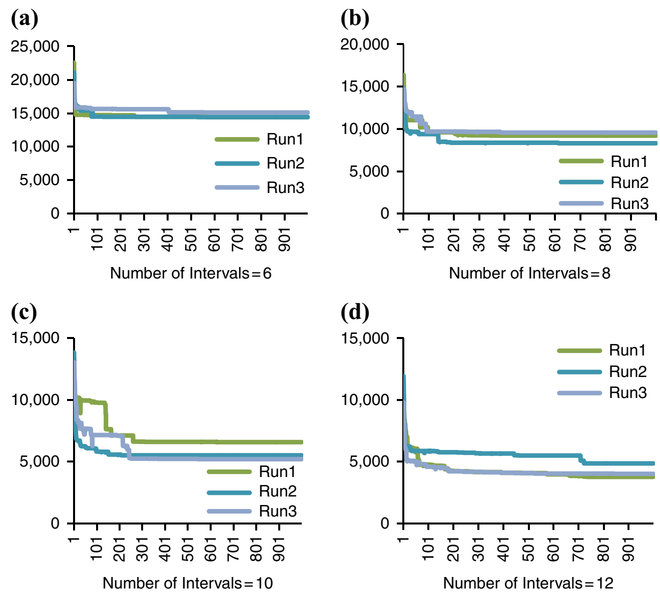

Figure 4.
CASE II change of intervals (population size = 200; order = 3)

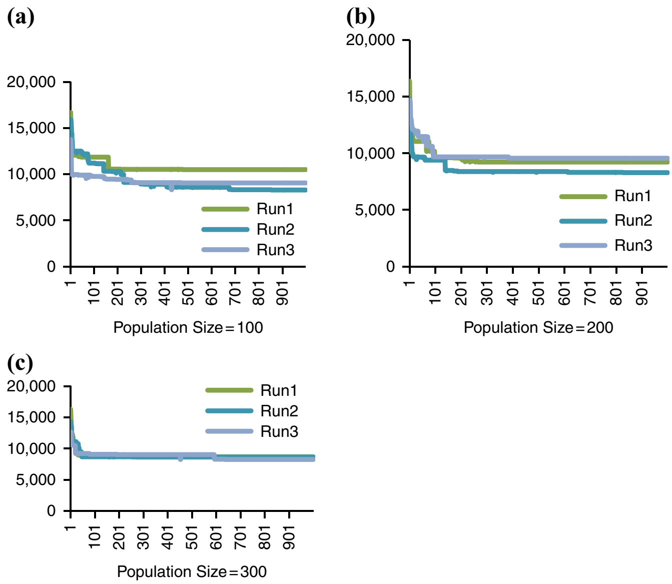

Figure 5.
CASE III change of population size (number of intervals = 8; order = 3)

decision makers with improved estimation and decreased error in complex and uncertain environment such as, the historical time series tourism data for Taiwan used in the present study. GA is used to calibrate the length of intervals of the universe of discourse. The problem is tested for different combinations of parameter values, such

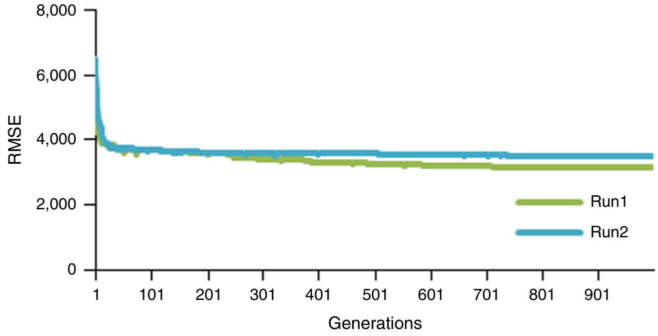

Figure 6.
Convergence curve for case (12 intervals; order 9)

Table VII.
Summary of results

Pop size – number of intervals – order	RMSE
200-8-1	16,800
200-8-2	12,394
200-8-3	8,299
200-8-4	6,696
200-8-5	6,433
200-8-6	6,175
200-8-7	5,936
200-8-8	5,907
200-8-9	5,863
200-8-10	5,851
200-6-3	14,425
200-8-3	8,299
200-10-3	5,196
200-12-3	*3,757*
100-8-3	8,301
200-8-3	8,299
300-8-3	8,272

as fuzzy time series order, number of fuzzy intervals and population size. The following inferences can be made from the results. RMSE decreases with the increase in order of fuzzy time series (Figure 7(a)). RMSE also decreases with the increase in number of intervals (Figure 7(b)). Population size does not affect the RMSE (Figure 7(c)). Finally, the results are compared with other methods in the literature and it shows that the proposed model offers better forecast than existing models, and it can get better quality solutions for different orders of fuzzy time series.

The present study can be extended by studying the affect of GA parameters such a crossover and mutation probability on forecasting accuracy. The proposed model can be extended by incorporating other evolutionary algorithms or AI tools such as neural networks and particle swarm optimization together with fuzzy time series and their results can be compared. Different forecasting problems such as weather forecast, enrollment forecast, stock forecast can be solved with the help of these methods.

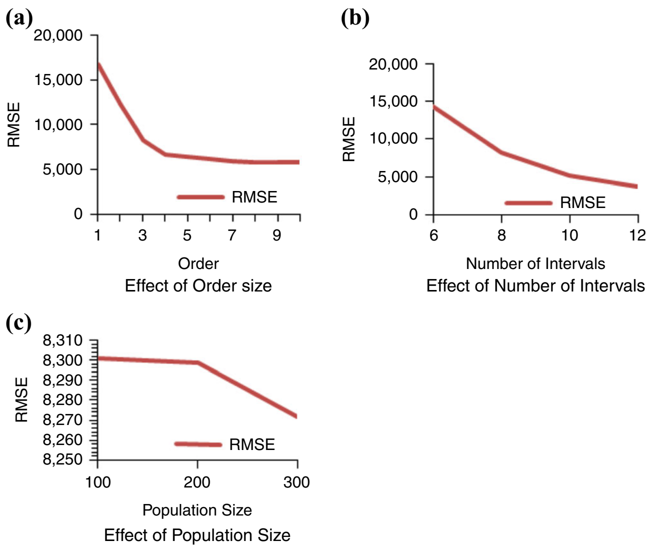

Figure 7. Effect of parameters on RMSE

	Chen (1996)	Huarng *et al.* (2007)	Huarng and Yu (2012)	FTS (order-3, number of intervals 9)	FTSGA (200-12-9)	Double exponential smoothing	Winters additive
RMSE	32,759	30,939	30,789	*14,728*	*3,140*	3,839	24,491
MAPE	13.38	12.5	12.54	*8.4*		12	10

Table VIII. Comparison of forecasts

References

Alleyne, D. (2006), "Can seasonal unit root testing improve the forecasting accuracy of tourist arrivals?", *Tourism Economics*, Vol. 12 No. 1, pp. 45-64.

Chan, F.T.S., Samvedi, A. and Chung, S.H. (2015), "Fuzzy time series forecasting for supply chain disruptions", *Industrial Management and Data Systems*, Vol. 115 No. 3, pp. 419-435.

Cheikhrouhou, N., Marmier, F., Ayadi, O. and Wieser, P. (2011), "A collaborative demand forecasting process with event-based fuzzy models: interest rate forecasting problem", *Computers & Industrial Engineering*, Vol. 61 No. 2, pp. 409-421.

Chen, R.J.C., Bloomfield, P. and Cubbage, F.W. (2008), "Comparing forecasting models in tourism", *Journal of Hospitality & Tourism Research*, Vol. 32 No. 1, pp. 3-21.

Chen, S.M. (1996), "Forecasting enrollments based on fuzzy time series", *Fuzzy Sets and Systems*, Vol. 81 No. 3, pp. 311-319.

Chen, S.M. (2002), "Forecasting enrollments based on high-order fuzzy time series", *Cybernetics and Systems*, Vol. 33 No. 1, pp. 1-16.

Chen, S.M. and Chung, N.Y. (2006), "Forecasting enrollments using high-order fuzzy time series and GAs", *International Journal of Intelligent Systems*, Vol. 21 No. 5, pp. 485-501.

Chen, S.M. and Hwang, J.R. (2000), "Temperature prediction using fuzzy time series", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 30 No. 2, pp. 263-275.

Chu, F. (2008), "A fractionally integrated autoregressive moving average approach to forecasting tourism demand", *Tourism Management*, Vol. 29 No. 1, pp. 79-88.

Eğrioğlu, E. (2012), "A new time-invariant fuzzy time series forecasting method based on genetic algorithm", *Advances in Fuzzy Systems*, pp. 1-6.

Eğrioğlu, E., Aladag, C.H., Basaran, M.A., Yolcu, U. and Uslu, V.R. (2011a), "A new approach based on the optimization of the length of intervals in fuzzy time series", *Journal of Intelligent and Fuzzy Systems*, Vol. 22 No. 1, pp. 15-19.

Eğrioğlu, E., Aladag, C.H., Yolcu, U., Uslu, V.R. and Basaran, M.A. (2010), "Finding an optimal interval length in high order fuzzy time series", *Expert Systems with Applications*, Vol. 37 No. 1, pp. 5052-5055.

Frechtling, D.C. (2001), *Forecasting Tourism Demand: Methods and Strategies*, Butterworth-Heinemann, Oxford.

Gangwar, S.S. and Kumar, S. (2012), "Partitions based computational method for high-order fuzzy time series forecasting", *Expert Systems with Applications*, Vol. 39 No. 1, pp. 12158-12164.

Gil-Alana, L.A., Gracia, F.P. and Cunado, J. (2004), "Seasonal fractional integration in the Spanish tourism quarterly time-series", *Journal of Travel Research*, Vol. 42 No. 1, pp. 408-414.

Goldberg, D.E. (1985), *Optimal Initial Population Size for Binary Coded Genetic Algorithms*, Department of Engineering Mechanics, University of Alabama, Tuscaloosa, AL.

Guo, W. (2007), "Inbound tourism, an empirical research based on gravity model of international trade", *Tourism Tribune*, Vol. 22 No. 3, pp. 30-34.

Gustavsson, P. and Nordström, J. (2001), "The impact of seasonal unit roots and vector ARMA modelling on forecasting monthly tourism flows", *Tourism Economics*, Vol. 7 No. 2, pp. 117-133.

Holland, J.H. (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.

Huarng, K.H. (2001a), "Effective lengths of intervals to improve forecasting in fuzzy time series", *Fuzzy Sets and Systems*, Vol. 123 No. 3, pp. 387-394.

Huarng, K.H. (2001b), "Heuristic models of fuzzy time series for forecasting", *Fuzzy Sets and Systems*, Vol. 123 No. 3, pp. 369-386.

Huarng, K.H. and Yu, T.H.K. (2005), "A type 2 fuzzy time series model for stock index forecasting", *Physica A: Statistical Mechanics and Its Applications*, Vol. 353 Nos 1-4, pp. 445-462.

Huarng, K.H. and Yu, T.H.K. (2006), "The application of neural networks to forecast fuzzy time series", *Physica A: Statistical Mechanics and Its Applications*, Vol. 363 No. 2, pp. 481-491.

Huarng, K.H. and Yu, T.H.K. (2012), "Modeling fuzzy time series with multiple observations", *International Journal of Innovative Computing Information and Control*, Vol. 8 No. 10B, pp. 7415-7426.

Huarng, K.H., Yu, T.H.K. and Hsu, Y.W. (2007), "A multivariate heuristic model for fuzzy time-series forecasting", *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, Vol. 37 No. 4, pp. 836-846.

Hwang, J.R., Chen, S.M. and Lee, C.H. (1998), "Handling forecasting problems using fuzzy time series", *Fuzzy Sets and Systems*, Vol. 100 Nos 1-3, pp. 217-228.

Ju, Y.J., Kum, C.E. and Shim, J.C. (1997), "Genetic-based fuzzy models: interest rate forecasting problem", *Computers & Industrial Engineering*, Vol. 33 Nos 3-4, pp. 561-564.

Kang, H.I. (2005), "A fuzzy time series prediction method using the evolutionary algorithm", *Advances in Intelligent Computing Springer Berlin Heidelberg*, Vol. 3645, pp. 530-537.

Khadaroo, J. and Seetanah, B. (2008), "The role of transport infrastructure in international tourism development: a gravity model approach", *Tourism Management*, Vol. 29 No. 5, pp. 831-840.

Konar, A. (2005), *Computational Intelligence: Principles, Techniques*, Springer, Berlin.

Lee, C.K., Song, H.J. and Mjelde, J.W. (2008), "The forecasting of international expo tourism using quantitative and qualitative techniques", *Tourism Management*, Vol. 29 No. 1, pp. 1084-1098.

Lee, L.W., Wang, L.H. and Chen, S.M. (2007), "Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms", *Expert Systems with Applications*, Vol. 33 No. 3, pp. 539-550.

Leigh, W., Purvis, R. and Ragusaa, J.M. (2002), "Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and GA: a case study in romantic decision support", *Decision Support Systems*, Vol. 32 No. 1, pp. 361-377.

Lim, C. and McAleer, M. (2002), "Time-series forecasts of international travel demand for Australia", *Tourism Management*, Vol. 23 No. 1, pp. 389-396.

Preez, J. and Witt, S.F. (2003), "Univariate versus multivariate time series forecasting: an application to international tourism demand", *International Journal of Forecasting*, Vol. 19 No. 3, pp. 435-451.

Singh, S.R. (2007), "A simple method of forecasting based on fuzzy time series", *Applied Mathematics and Computation*, Vol. 186 No. 1, pp. 330-339.

Song, H. and Li, G. (2008), "Tourism demand modelling and forecasting: a review of recent research", *Tourism Management*, Vol. 29 No. 2, pp. 203-220.

Song, H. and Witt, S.F. (2006), "Forecasting international tourist flows to Macau", *Tourism Management*, Vol. 27 No. 1, pp. 214-224.

Song, Q. and Chissom, B.S. (1993), "Forecasting enrollments with fuzzy time series – part I", *Fuzzy Sets and Systems*, Vol. 54 No. 1, pp. 1-9.

Song, Q. and Chissom, B.S. (1994), "Forecasting enrollments with fuzzy time series – part II", *Fuzzy Sets and Systems*, Vol. 62 No. 1, pp. 1-8.

Tsaur, R.C. and Kuo, T.C. (2011), "The adaptive fuzzy time series model with an application to Taiwan's tourism demand", *Expert Systems with Applications*, Vol. 38 No. 8, pp. 9164-9171.

Tsaur, R.C., Yang, J.C.O. and Wang, H.F. (2005), "Fuzzy relation analysis in fuzzy time series model", *Computers and Mathematics with Applications*, Vol. 49 No. 4, pp. 539-548.

Wang, C.H. (2004), "Predicting tourism demand using fuzzy time series and hybrid grey theory", *Tourism Management*, Vol. 25 No. 3, pp. 367-374.

Wong, K.K., Song, H., Witt, S.F. and Wu, D.C. (2007), "Tourism forecasting: to combine or not to combine?", *Tourism Management*, Vol. 28 No. 1, pp. 1068-1078.

Yu, H.K. (2005), "A refined fuzzy time-series model for forecasting", *Physica A: Statistical Mechanics and its Applications*, Vol. 346 Nos 3-4, pp. 657-681.

Corresponding author
Sameer Kumar can be contacted at: skumar@stthomas.edu