Zaria Journal of Electrical Engineering Technology, Department of Electrical Engineering, Ahmadu Bello University, Zaria – Nigeria. Vol. 8 No.2, September2019. ISSN: 0261 – 1570.

MULTIPLE MONTE CARLO SIMULATED HIDDEN MARKOV MODEL FOR FUZZY TIME SERIES FORECASTING

A. T. Salawudeen[1], A. Nurudeen[2], S. Garba[1],S. U. Hussein[3], M. L. Imam[4], B. Yahaya[1]
I. F. Egbujo[4]
[1]*Department of Electrical and Electronic Engineering, University of Jos, Nigeria.*
[2]*Department of Computer of Electrical Engineering, Ahmadu Bello University, Zaria, Nigeria.*
[3]*Department of Electrical and Electronic Engineering, Nile University, Abuja, Nigeria.*
[4]*Department of Electrical and Electronic Engineering, University of Jos, Nigeria.*
[*]tasalawudeen@abu.edu.ng

Keywords: –
FTS
HMM
Monte Carlo Simulation
GA
RMSE

Article History: –
Received: January, 2019.
Reviewed: April, 2019
Accepted: August, 2019
Published: September 2019

ABSTRACT

This paper presents a Monte Carlo based Hidden Markov Model (HMM) for fuzzy time series forecasting. To make the nature of conjecture and randomness of forecasting more realistic, the Monte Carlo method with different simulation size is adopted to estimate the forecasting outcome. To address the insufficiency in data associated with the HMM model, we adopted a method called smoothing. A number of simulations was performed using MATLAB simulation environment. The performance of the model was evaluated using the daily average temperature and cloud density of Taipei, Taiwan. In addition to improving forecasting accuracy, the proposed model adheres to the central limit theorem, and thus, the result statistically approximates to the real mean of the target value being forecasted. Results showed that the proposed model attain and MSE, RMSE, and AFEP of 0.8596, 2.4283, 0.9272 respectively.

1. INTRODUCTION

The forecasting problem of time series data plays an important role in various domains, such as air pollution, population growth, rainfall prediction, and load forecasting. It deals with forecasting future outcomes from a temporally ordered sequence of past-observed data points, whose values are usually real numbers. However, traditional time series analysis cannot handle the vagueness and uncertainty inherent in certain data due to inaccuracies in measurements, incomplete sets of observations, or difficulties in obtaining the measurements [1].

Time series is simply a collection of quantitative variables at regular intervals of time. Whether discreet or continuous, time series is always both nonlinear and non-stationary since they are sample functions realized from processes that are always stochastic [2]. Time series forecasting plays an important role in a great variety of applications, such as predicting university enrollments, stock prices, rainfall, blood pressure, and so on. Such forecasting usually uses a sequence of past data points which are typically measured successively for forecasting future outcomes [3]. Various techniques for time series forecasting have been evolved in recent decades. Compared with other models, Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA)-based models are prominent and highly useful. However, they cannot deal with time series vagueness and linguistic terms [4]. In addition, these statistical methods could not perform appropriately on time series with a small amount of data [5]. Furthermore, the necessary conditions for applying the conventional time series with probabilistic models which requires some assumptions such as number of observations, normal distribution, and linearity [6]. Thus, these approaches can lead to misleading forecasting results when these assumptions are not, satisfied. Therefore, non-probabilistic approaches have been put forward as an alternative to probabilistic time series forecasting models [7].

To deal with such deficiencies, fuzzy time series (FTS) have been developed and widely applied [8]. FTS was developed model based on fuzzy logic. In recent years, FTS models have attracted the attention of many researchers because of their advantages: better performance in some real forecasting problems terms [4], dealing with data in linguistic terms [9], and their ability to integrate with heuristic knowledge and models [10, 11].

One of the most important issues in FTS models is the determination of the relations [12].In the literature, many methods were used for determining fuzzy relations. These include fuzzy logic group relation tables, artificial neural networks, fuzzy relation matrices obtained from some fuzzy set operations, particle swarm optimization and genetic algorithms [12-14]. In the analysis of fuzzy time series, fuzzy logic group relationships tables have been generally preferred for determination of fuzzy logic relationships. The reason of this is that it doesn't need to perform complex matrix operations when these tables are used. On the other hand, when fuzzy logic group relationships tables are exploited, membership values of fuzzy sets are ignored. Thus, in defiance of fuzzy set theory, fuzzy sets' elements with the highest membership value are only considered [15]. This situation causes information loss and it may affect the forecasting performance, negatively. Since the fuzzy relationships can be nonlinear and complex, an intelligent method is needed to calculate these relationships.

To deal with such deficiencies, Hidden Markov Model (HMM) have been developed and applied in formulating the fuzzy relationship, where the model parameters were estimated using a local search technique, known as the Baum Welch algorithm [16].

Since parameter learning in Hidden Markov Model using the Baum-Welch algorithm is prone to be trapped in the local optima, it has become imperative that a technique for finding enhanced estimates of the fuzzy relations and also avoiding the local optima is required [17].

In this research work, a forecasting model is proposed to enhancing Sullivan and Woodall's Markov-based forecasting one to allow handling two-factor forecasting problems. This model is built on the basis of the hidden Markov model (HMM), a probabilistic model that is commonly applied to time series [18, 19]. Moreover, by applying the Monte Carlo method when estimating the forecasting outcome, the nature of conjecture and randomness of the forecasting are made more realistic [20, 21]. To test the effectiveness of the model, experiments is conduct in forecasting daily average temperature in Taipei, Taiwan and compare the results with those from other models.

The remainder of this paper is organized as follows. In Section 2, the basic concept of fuzzy time series is briefly introduced in the form of Materials, and in Section 3, the Monte-Carlo Simulation forecasting model based on HMM is presented. Section 4 shows the performance evaluation of the model and a comparison of the results. The last section describes our conclusions and directions for future work.

2. MATERIALS

In this section, the relevant information on the methods adopted for this paper is discoursed.

2.1 Fuzzy Set Theory

Fuzzy sets were introduced by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set. At the same time, Salii (1965) defined a more general kind of structure called an *L*-relation, which he studied in an abstract algebraic context. Fuzzy relations, which are used now in different areas, such as linguistics [22-24] and clustering[25], are special cases of *L*-relations when *L* is the unit interval [0, 1].

In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition, an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1.In fuzzy set theory, classical bivalent sets are usually called *crisp* sets. The fuzzy set theory can be used in a wide range of domains in which information is incomplete or imprecise, such as bioinformatics.

2.2 **Basic Definitions and Operations:**

The fuzzy time series definition used in this project was first suggested by Song and Chissom in 1994.

Definition 1: Let Y (t) (t =0, 1, 2, . . .), a subset of R, be the universe of discourse on which fuzzy sets fi(t) (i = 1, 2, . . .) are defined, and let F(t) be a collection of fi(t). Then, F(t) is called a fuzzy time series on Y (t) (t = 0, 1, 2, . . .). Song and Chissom employed a fuzzy relational equation to develop their forecasting model under the assumption that the observations at time t are dependent only upon the accumulated results of the observations at previous times, which is defined as follows.

Definition 2: If, for any fj(t) ∈ F(t), where j ∈ J, there exist an fi(t − 1) ∈ F(t − 1), where i ∈ I, and a fuzzy relation Rij(t, t − 1), such that fj(t) = fi(t − 1) ◦ R_{ij}(t, t − 1), let R(t, t − 1) = $\cup_{i,j}$

R_{ij}(t, − 1), where "∪" is the union operator and "◦" is the composition. R(t, t − 1) is called the fuzzy relation between F(t) and F(t − 1), which can be represented using the following fuzzy relational equation [3, 26]:

$$F(t) = F(t-1) \circ R(t, t-1). \quad (1)$$

Definition 3: If we suppose that F(t) is caused by F(t − 1), F(t − 2), . . ., or F(t − m) (m > 0), then the first-order model of F(t) can be expressed as

$$F(t) = F(t-1) \circ R(t, t-1) \quad (2)$$

Or

$$F(t) = (F(t-1) \cup F(t-2) \cup \cdots \cup F(t-m)) \circ Ro(t, t-m) \quad (3)$$

where "∪" is the union operator and "◦" is the composition. R(t, t − 1) is called the fuzzy relation between F(t) and F(t − 1), and Ro(t, t − k) is the fuzzy relation that joins F(t) with F(t − 1), F(t − 2), . . ., or F(t − k), where the subscript "o" denotes the relationship "or." In the literature, the fuzzy relation Rij(t, t − 1) is usually represented by a fuzzy logical relationship rule ("IF–THEN" rule), as in [9, 27]. In this project, the fuzzy relation is realized by an HMM, which will be discussed

2.3 Hidden Markov Model

The Markov model proposed in the early 1900s is often called an observable (or visible) Markov model [16-18], so as to be distinguished from the Hidden Markov Model. Since the HMM is derived from traditional Markov theory, this section begins with a brief description for a Markov Model. A Markov model is a statistical model, and can be viewed as a typical transitional diagram composed of the following components: a set of different states; transition between states; and transition probabilities, which are probabilities linked to transitions. Beginning from a start state, a transitional process continues until it reaches an end state. The outcome of this sequence of states (observations) within a Markov model is called Markov chain. Formally, a Markov chain can be represented by a series of random variables, $X^{(0)}, X^{(1)}, X^{(2)},..., X^{(m+1)}$ each of which takes on a value from the state space, $S = \{S_1, S_2,...., S_N\}$ [28]. The main characteristic of a Markov model is that it predicts the future based on the present rather than the past: this is termed Markov condition. When the probability of a sequence of random variables, $P(X^{(0)},..., X^{(m+1)})$ is calculated, all the previous random variables are considered according to [28, 29]. This is defined as:

$$P(X^{(0)},..., X^{(m+1)}) = P(X^{(0)})P(X^{(1)} / X^{(0)}...P(X^{(m)} / X^{(0)},..., X^{(m-1)})P(X^{(m+1)} / X^{(0)},..., X^{(m)}) \quad (4)$$

By assuming the Markov condition, however, Markov theory only takes into consideration the current random variable by [30]:

$$P(X^{(0)},..., X^{(m+1)}) = P(X^{(0)})P(X^{(1)} / X^{(0)}...P(X^{(m)} / X^{(m-1)})P(X^{(m+1)} / X^{(m)}) \quad (5)$$

To present the Markov model's features, a simple example was provided. Figure 1. shows an example of a simple Markov model for tossing a coin. Each state of the model corresponds to the outcome of an observation (also called an event), either Heads (H) or Tails (T) of a coin.

The numbers along the lines or curves specify transition probabilities linked to a directed path between the two states connected by the lines or curves. The arrows indicate the direction of the state transition. The S in figure 1 shows the special state serving as the starting point in the model. This model must start from the special state and can move to other states by adding the probabilities corresponding to the direction that it takes.

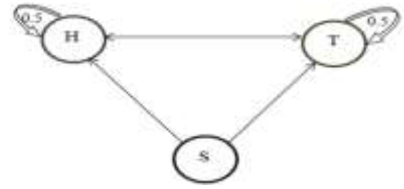

Figure 1: Markov Model for Tossing One Coin

3. METHODOLOGY

3.1 Monte Carlo Based HHMM FTS Model

Under realistic circumstances, there are usually multiple related factors that influence the behavior and outcome of any event. For example, when trying to predict today's temperature, we could easily look up and observe the clouds in the sky. If there are dense clouds, it can be intuitively inferred that the temperature will be low. However, temperature depends on not only cloud density but also temperature values in previous days. We thus might obtain a better forecast for today's temperature by combining knowledge about what happened in previous days with the observed cloud state. These kinds of problems are constantly encountered in the real world, which is why our paper focuses on targeting them. Of course, the state of the temperature is not merely controlled by both factors, as elements such as winds and air pressure are also likely to have an impact. However, in this paper, we limit ourselves to problems concerning two factors, in which both are probabilistically related. This can be formally represented as follows:

Given two fuzzy time series $F(t) = \{fi(t)|t = 1, 2, . . . , T, i = 1, 2, . . . , n\}cc$ and $G(t) = \{gi(t)|t = 1, 2, . . . , T, i = 1, 2, . . .,m\}$, where $fi(t)$ and $gi(t)$ are the respective states at time t, the fuzzy relation among $F(t)$,

$G(t)$, and $F(t - 1)$ can be formulated as a fuzzy relational equation [16]:

$F(t) = (F(t - 1),G(t)) \circ R(t, t - 1)$.

To solve the forecasting problem of $fi(t)$, which is dependent on $fi(t - 1)$ and $gi(t)$, the theory of HMM is applied, in which $F(t)$ and $G(t)$ are the hidden and observed state sequences, respectively.

3.2 HMM Fuzzy Time Series Forecasting

The proposed forecasting model expands Sullivan and woodalls model by combining HMM and Monte Carlos simulation and consists of the following five steps. . In developing the HMM based FTS model, there are some very important pre-processing steps (sorting, fuzzification, relationship representation, optimization and defuzzification) that make up the model. In this research, the historical data was obtained from the work of Cheng & Sheng (2010). About 60% of the data was used for training the model and the rest was used to test the model.

3.2.1 Define the Universe of Discourse

In defining the universe of discourse, the minimum value $D_{\min}$ and the maximum value $D_{\max}$ of the historical data sets were first determined. Based on $D_{\min}$ and $D_{\max}$, the universe of discourse U is defined as follows:

$$U = \left[D_{\min} - D_1, D_{\max} + D_2\right] \tag{6}$$

where:

D_1 and D_2 are any two positive integers.

HMM model is a bivariate model and thus, it will then have two universes of discourse given as [31]:

$$U_s = \left[D_{s\min} - D_{s1}, D_{s\max} + D_{s2}\right] \tag{7}$$

$$U_o = \left[D_{o\min} - D_{o1}, D_{o\max} + D_{o2}\right] \tag{8}$$

where:

U_s and U_o are the universe of discourses for the hidden variables and observation variables.

$D_{s\min}, D_{s\max}$, $D_{o\min}$,and $D_{o\max}$ are the respective minimal and maximal values of thehistorical data of hidden and observation variables while D_{s1}, D_{s2}, D_{o1}, and D_{o2} are the corresponding proper positive numbers.

3.2.2 Partition the Universe of Discourse into Several Even Lengthy Intervals

In this section, the universe of discourse U is partitioned into n equal intervals using the equal interval method defined as:

$$I_1 = [D_{\min} \rightarrow (D_{\min} + l)] \tag{9}$$

$$I_2 = [(D_{\min} + l) \rightarrow (D_{\min} + 2l)] \quad (10)$$

$$I_n = [(D_{\min} + (n-1)l) \rightarrow (D_{\min} + nl)] \quad (11)$$

Where:

I_1, I_2 and I_n represents the first, second and nth partitions respectively, while l is the interval length defined as [5]:

$$l = \left(\frac{1}{n}\right) \times \left[(D_{\max} - D_2), (D_{\min} + D_1)\right] \quad (12)$$

where:

n is the number of intervals.

In this research, the hidden variables universe of discourse U_s is partitioned as:

$$u_1 = [D_{\min} \rightarrow (D_{\min} + l)] \quad (13)$$

$$u_2 = [(D_{\min} + l) \rightarrow (D_{\min} + 2l)] \quad (14)$$

$$u_n = [(D_{\min} + (n-1)l) \rightarrow (D_{\min} + nl)] \quad (15)$$

$$l_s = \left(\frac{1}{n}\right) \times \left[(D_{s\max} - D_{s2}), (D_{s\min} + D_{s1})\right] \quad (16)$$

Where: u_1, u_2 and u_n represents the first, second and nth partitions respectively, while l_s is its interval length.

Similarly, observation variables universe of discourse U_o is partitioned as:

$$v_1 = [D_{\min} \rightarrow (D_{\min} + l)] \quad (17)$$

$$v_2 = [(D_{\min} + l) \rightarrow (D_{\min} + 2l)] \quad (18)$$

$$v_n = [(D_{\min} + (n-1)l) \rightarrow (D_{\min} + nl)] \quad (19)$$

$$l_o = \left(\frac{1}{m}\right) \times \left[(D_{o\max} - D_{o2}), (D_{o\min} + D_{o1})\right] \quad (20)$$

Where: v_1, v_2 and v_n represents the first, second and nth partitions respectively, while l_o is the interval length.

3.2.3 Defining the Fuzzy Sets on the Universe of Discourse

Next, the fuzzy sets are defined for the intervals. There is no restriction on the number of the fuzzy sets defined. In this research n and m linguistic fuzzy setswere defined for the hidden and observation variables' intervals, respectively

3.2.4 Fuzzifying the Time Series Data

Given a traditional crisp time series, fuzzification procedure to obtain the corresponding fuzzy set is performed. With the fuzzy sets that are properly defined, the crisp time series are fuzzified using the method of general membership functions [32].

$$A_i = \sum_{j=1}^{n} \mu_{ij} / u_j \quad (21)$$

Where: μ_{ij} is the membership function of the fuzzy sets A_i, such that:

$$\mu_{A_i} = U \rightarrow [0,1] \quad (22)$$

Hence, a fuzzy value of some raw datum of time series was obtained.

In this research, for hidden states, an n fuzzy sets were defined on U_s using general membership functions expressed as follows:

$$S_i = \sum_{j=1}^{n} \mu_{ij} / u_j \quad (23)$$

where:

μ_{ij} is the membership degree of S_i belonging to u_j which is defined by [16]:

$$\mu_{ij} = \begin{cases} 1, & if\ j = 1 \\ 0.5, & if\ j = i-1\ or\ i+1 \\ 0, & if\ otherwise \end{cases} \quad (24)$$

For a given historical datum Y_t, its membership degree belonging to interval u_i is determined by the following heuristic rules.

Rule 1) If Y_t is located at u_1, the membership degrees are 1 for u_1, 0.5 for u_2, and 0 otherwise.

Rule 2) If Y_t belongs to u_i, $1 < i < n$, then the degrees are 1, 0.5, and 0.5 for u_i, $u_i - 1$, and $u_i + 1$, respectively, and 0 otherwise.

Rule 3) If Y_t is located at u_n, the membership degrees are 1 for u_n, 0.5 for $u_n - 1$, and 0 otherwise. Then, Y_t is fuzzified as S_j, where the membership degree in interval j is maximal.

For observable states, m fuzzy setscan be defined on U_o as expressed as follows [16]:

$$O_i = \sum_{j=1}^{m} \mu_{ij} / u_j \tag{25}$$

Where:

μ_{ij} is the membership degree of O_i belonging to v_j and is defined by

$$\mu_{ij} = \begin{cases} 1, & if\ j = 1 \\ 0.5, & if\ j = i - 1\ or\ i + 1 \\ 0, & if\ otherwise \end{cases} \tag{26}$$

The observation variables were fuzzified in the same way as the hidden variable.

3.2.5 Building the HMM Model to Estimate the Fuzzy Relations

The HMM was characterized into the following three matrices.

$$A = \{a_{ij}\}, \tag{27}$$

where:

$$a_{ij} = P_r\left(S_{j,} t / S_i, t-1\right) \tag{28}$$

$$\pi = \{\pi_i\}, \tag{29}$$

where:

$$\pi_i = P_r(S_i, 1) \tag{30}$$

$$B = \{b_{ij}\} \tag{31}$$

Where:

$$b_{ij} = P_r\{O_j = t / S_i = t-1\} \tag{32}$$

The parameter set were estimated using the following expressions:

a) The initial state vector π is set to be a $1 \times n$ matrix and is defined as:

$$\pi_i = P_r(S_i, 1) = \frac{n(S_i 1)}{N_1} \tag{33}$$

Where; $n(S_i 1)$ is the number of the initial state S_i in the data set and

$$N_1 = \sum_{i=1}^{n} n(S_i, 1) \tag{34}$$

b) The state-transition matrix $A = \{a_{ij}\}$, is an $n \times n$ matrix defined as[16]:

$$a_{ij} = P_r\left(S_{j,} t / S_i, t-1\right) = \frac{n(S_{j,} t - S_i, t-1)}{n(S_i, t-1)} \tag{35}$$

Where:

$n(S_{j,} t - S_i, t-1)$ is the number of transitions from state i at time $t-1$ to state j at time t, $n(S_i, t-1)$ is the total number of transitions to state i at time $t-1$, $\forall a_{ij} \geq 0$ and $\sum_{j=1}^{n} a_{ij} = 1, 1 \leq i \leq n$

c) The confusion matrix $B = \{b_{ij}\}$ is an $n \times m$ matrix represented as[16]:

$$b_{ij} = P_r\{O_j = t / S_i = t-1\} = \frac{n(O_j, t / S_i, t)}{n(S_i, t)} \tag{36}$$

where:

$n(O_j, t / S_i, t)$ is the number of observing the observable symbol j and state i at the same time t,

$n(S_i, t)$ total number of transitions to the S_i at time t,

$\forall b_{ij} \geq 0$ and $\sum_{j=1}^{m} b_{ij} = 1, 1 \leq i \leq n$.

3.2.6 Smoothing HMM Model parameters.

The traditional smoothing methods reviewed in the literature were commonly applied in the field of speech recognition, in which the states are crisp in nature, and thus, they are inappropriate for the case of fuzzy time series. Thus, a smoothing technique that can deal with the fuzziness existing in the HMM based FTS will therefore be employed. Thus, the following smoothing techniques were adopted in this research [16]:

$$A^s = A + A^L \qquad (37)$$

where:

A^s is the smoothed transition matrix, A is the original estimated transition matrix and A^L is the Adjusted transition matrix defined as:

$$A^L(:,[i]) = \frac{1}{2}k \times A^P(:,[i-1]) - k \times A^P(:,[i]) + \frac{1}{2}k \times A^P(:,[i+1]), i = 2,3,4,...,n-1 \qquad (38)$$

where; k is the smoothing factor that represents the degree of smoothness and A^P is the peak matrix of A determined as[16, 19]:

$$a_y^p = \begin{cases} a_y \ if \ a_y \ is \ a \ peak \ in \ row, \\ 0, \ if \ Otherwise \end{cases}$$

The confusion matrix B was smoothed similarly using the expression:

$$B^s = B + B^L \qquad (40)$$

Where; B^s is the smoothed transition matrix, B is the original estimated transition matrix and B^L is the Adjusted transition matrix defined as [33-35]:

$$B^L(:,[i]) = \frac{1}{2}k \times B^P(:,[i-1]) - k \times B^P(:,[i]) + \frac{1}{2}k \times B^P(:,[i+1]), i = 2,3,4,...,n-1 \qquad (41)$$

where; B^P is the peak matrix of B determined as:

$$a_y^p = \begin{cases} a_y, \quad if \ a_y \ is \ a \ peak \ in \ row, \\ 0, \ if \ Otherwise \quad , \end{cases}$$

The procedural program for smoothing the model parameters can be found.

3.2.7 Calculating Forecast Outputs

In this research, the probabilistic forecast method was adopted at computing the forecast output. The forecasting problem was represented as:

Given two fuzzy time series,

$$F(t) = \{f(t) = (S_i, t)/t = 1,2,...,T, i = 1,2,...,n\} \qquad (43)$$

$$G(t) = \{g(t) = (O_i, t)/t = 1,2,...,T, j = 1,2,...,m\} \qquad (44)$$

The fuzzy relation among $F(t)$, $G(t)$, and $F(t-1)$ can be formulated as a fuzzy relational equation:

$$F(t) = (F(t-1), G(t)) \circ R(t, t-1) \qquad (45)$$

where; $F(t)$ and $G(t)$ are the hidden and observed state sequences, respectively. To solve the forecasting problem of $f_i(t)$, which is dependent on $f_i(t-1)$ and $g_i(t)$, the theory of HMM is applied. Thus, the probability of $S_i t$ is determined by $P_r(S_i, t / S_x, t-1)$ and $P_r(O_y, t / S_i, t)$.

The probabilities of all possible hidden states occurring at time $t(t \geq 2)$ with the transition influence of the previous hidden state $S_x, t-1$ and the observation state O_k, t are then computed.

This relation is represented by the function $\alpha_t(S_x, t-1, O_y, t)$ which is defined as follows:

$$\alpha_t(S_x, t-1, O_y, t) = B^S(:,[y])^T .* A^S([x],:) \qquad (46)$$

where; $A^S([x],:)$ is the xth row of state-transition matrix A^S and $B^S(:,[y])$ is the yth column of confusion matrix B^S. The symbol of operator ".*" is an array multiplication, and thus, $A \cdot * B$ means the element-by-element vector multiplication of A^S and B^S.

For the special case when $t = 1$, where the observation state is the only information available, the previous hidden state does not exist, and therefore the probability of the hidden state S_i at $t=1$ was determined by a function $\beta_1((O_y,1))$, which is defined as:

$$\beta_1((O_y,1)) = \pi \cdot * B^S(:,[y])^T \tag{47}$$

where; ".*" is defined as earlier, $\alpha_t(S_x,t-1,O_y,t)$ and $\beta_1((O_y,1))$ result in a probability matrix presenting the probabilities of hidden states that may occur.

The forecasting result is located at the state with the highest probability. However, according to the probability theorem, an event with a higher probability only has a greater chance of occurring, but may not necessarily occur. Due to the nature of conjecture and randomness, a stochastic simulation, based on Monte Carlo method is adopted to estimate the true outcome of each forecast. The Monte Carlo method provides approximate solutions by stochastic sampling experiments and solves problems based on random numbers and probability statistics. The forecasting process consists of two subsequent tasks: normalization and Monte Carlo simulation. For the model to obey the row stochasticity constraint, the normalization was performed for the probability vectors of functions $\alpha_t(S_x,t-1,O_y,t)$ and $\beta_1((O_y,1))$, which are expressed, respectively, as [33]:

$$N\alpha_t(S_x,t-1,O_y,t) = [\alpha_t^{S_1},\alpha_t^{S_2},...,\alpha_t^{S_n}] \tag{48}$$

$$N\beta_1((O_y,1)) = [\beta_1^{S_1},\beta_1^{S_2},...,\beta_1^{S_n}] \tag{49}$$

Where:

$$\alpha_t^{S_i} = \frac{P_r(O_y,t/S_i,t)P_r(S_i,t/S_x,t-1)}{\sum_{i=1}^{n} P_r(O_y,t/S_i,t)P_r(S_i,t/S_x,t-1)} \tag{50}$$

$$\beta_1^{S_i} = \alpha_t^{S_i} = \frac{P_r(O_y,1/S_i,1)}{\sum_{i=1}^{n} P_r(O_y,1/S_i,1)} \tag{51}$$

A stochastic experiment consisting of l Monte Carlo simulations is conducted to determine the forecasting result S_i from $\alpha_t(S_x,t-1,O_y,t)$ or $\beta_1((O_y,1))$. These forecasting results were represented with a vector C as:

$$C = [c_1,c_2,...,c_n] \tag{52}$$

Where; c_i is the number of forecasting hidden states belonging to s_i and

$$l = \sum_{i=1}^{n} c_i \tag{53}$$

The procedural program for carrying out these processes can be found in appendix X.

3.2.8 Defuzzifying the Forecasting Outputs

For simplicity, the defuzzification method based on center of gravity method was employed. This is expressed as follows:

$$\frac{\sum_{i=1}^{n} c_i \times t_i}{\sum_{i=1}^{n} c_i} \tag{54}$$

where:

$$t_{i=} = \begin{cases} \frac{1}{1.5}\ (m_1 + 0.5 \times m_2 & i = 1 \\ \frac{1}{2}(0.5 \times m_{i-1} + m_i + 0.5 \times m_{i+1}) & i = 2,3,\dots n-1 \\ \frac{1}{1.5}(0.5 \times m_{n-1} + m_n) & i = n \end{cases}$$

and m_i is the middle point of interval u_i

4. EXPERIMENT AND EVALUATION

In order to demonstrate the effectiveness of the proposed forecasting model, we conducted one experiments using real world data. The experiment consisted of forecasting temperature with observable cloud density in Taiwan.

4.1 Experiments Forecasting Temperature with Cloud Density Data

The time series data considered in this project is the historical data of the daily average temperature and cloud density in Taipei, Taiwan, obtained from the work of Cheng & Sheng (2010).

TheHMM FTS model was applied to forecast temperature with cloud density data. The time span of the data used is from June to September for 1993, 1994, 1995, and 1996. The data were divided into two parts, i.e., that for 1993, 1994, and 1995 was used as the training set to establish the parameter of λ, and that for 1996 was used as the testing set to evaluate forecasting performance

4.2 Defining the Universe of Discourse of the Historical Data

From the data, the following values for the hidden variables and observation variables were obtained. For the hidden variable (Temperature data)

$D_{s\max}$ =32.2, $D_{s\min}$ =21.6, $U_s = (21,32)$

Where; D_{s1} and D_{s2} were assumed to be 2.7 and 1.4 respectively. These were adequately selected to ensure the smoothness of boundaries of the interval.

For the observation variable (Cloud density data)

$D_{s\max}$ =31.6, $D_{s\min}$ = 23.7 $U_o = (0,100)$

Where; D_{o1} and D_{o2} were assumed to have 0 values respectively.

Using equation 16 and 20 the length of interval of both hidden and observable were calculated and the partition was obtained.

$$= (\tfrac{1}{6}) \times [(31.6 + 1.4) - (23.7 - 2.7)]$$

$$= 2$$

4.3 Partitioning the Obtained Universe of Discourse

For the hidden variables (Temperature data), the universe of discourse $U_s = (21,32)$, was partitioned into 6 temperature intervals using an interval length of 2 as follows:

$$u_1 = [21,23], u_2 = [23,25], u_3 = [25,27]$$
$$u_4 = [27,29], u_5 = [29,31], u_6 = [31,33]$$

Similarly, the observation variable's universe of discourse $U_o = (0,100)$ was partitioned into 5 cloud density intervals expressed as:

$v_1 = [0,20]$, $v_2 = [20,40]$, $v_3 = [40,60]$ $v_4 = [60,80]$ and $v_5 = [80,100]$ using the interval length of 20.

4.4 Fuzzifying the Time Series Data

For the linguistic variable 'Temperature', 6 fuzzy sets were defined which corresponds to the number to which its universe of discourse was partitioned. The used linguistic values are:

S_1 = (Freezing), S_2 = (Cold), S_3 = (Cool)

S_4 = (Mild), S_5 = (Warm) and S_6 = (Hot)

Similarly, for the linguistic variable 'cloud density', 5 fuzzy sets were defined expressed as:

O_1 = (Very low), O_2 = (Low), O_3 = (Medium)

O_4 = (High) and O_5 = (Very high)

Figure 2 shows the flow chart adopted for the implementation of the proposed model.

5 RESULTS AND DISCUSSION

Different number of Monte Carlo simulation was carried using 1000 (M1000), 500 (M500), 200 (M200) and 100 (M100) as number of Monte Carlo samples. The Monte Carlo simulation was performed using the equation given as:

$$y(t) = \sum_{i=1}^{m} \frac{x(t)}{m} \quad (55)$$

where *y*(t) is the output, *x*(t) is the input and *m* are the number of Monte Carlo simulation.

5.1 Results of the Defuzzified Forecasting Outputs:

The technique discussed in section 5 was employed with different values of Monte Carlo simulation and the forecasted values are shown in table VII.

The experimental results are plotted in graphs for the month of June, 1996 as shown in Figures 1 to 4and the one with best trend is selected

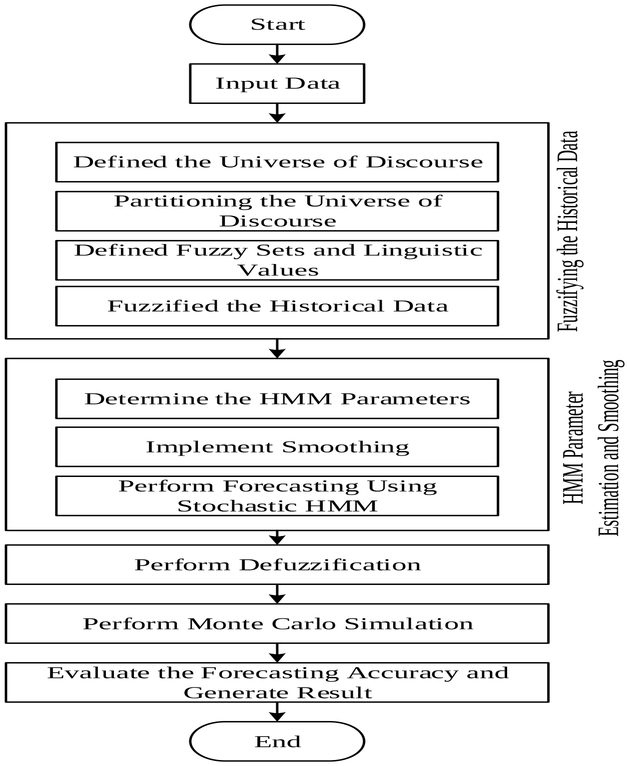

Figure 2: The flow chart of the proposed model

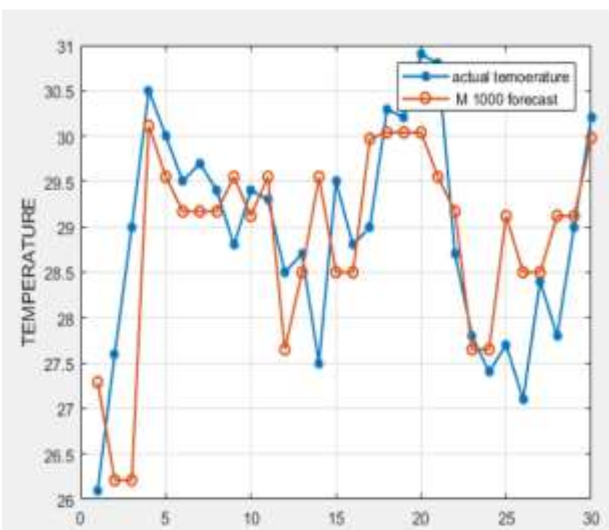

Figure 3 Comparison between Actual and Forecasted Values for June, 1996 using 1000 Monte Carlo simulation

From the Figure 3, it shows that for 1000 Monte Carlo simulation, the predicted values are in trend with the actual values. In figure 3 the blue line shows the actual values and the red line shows the predicted values.

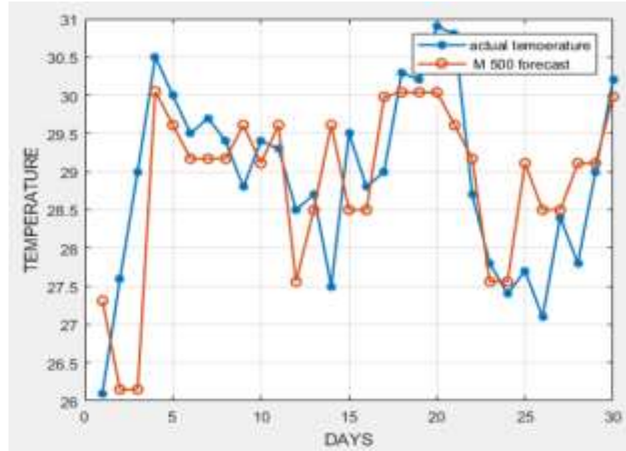

Figure 4. Comparison between Actual and Forecasted Values for June, 1996 using 500 Monte Carlo Simulation

From the figure above (Figure 4) it shows that for 500 Monte Carlo simulations, the predicted values are in trend with the actual values and it indicate best trend compare to 1000 Monte Carlo simulation. In Figure 4 the blue line shows the actual values and the red line shows the predicted values.

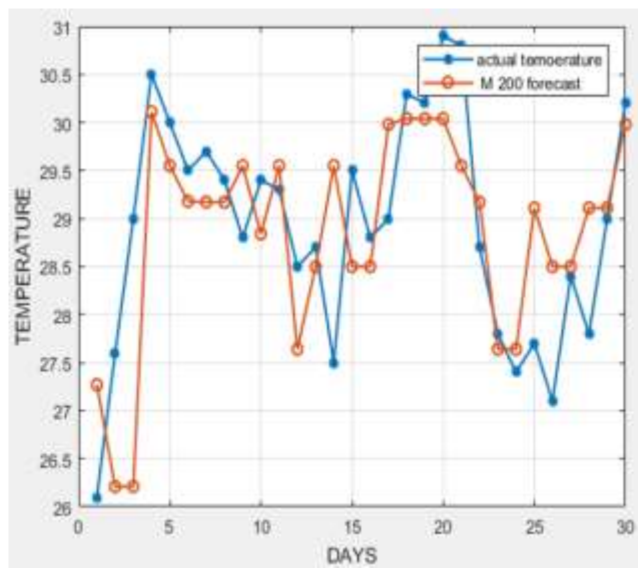

Figure 5 Comparison between Actual and Forecasted Values for June, 1996 using 200 Monte Carlo Simulation

From the figure above (Figure 5) it shows that for 200 Monte Carlo simulations the predicted values are in trend with the actual values and it indicate best trend compare to 1000 Monte Carlo simulation but not as best as 500 Monte Carlo simulations. In figure 5 the blue line shows the actual values and the red line shows the predicted values.

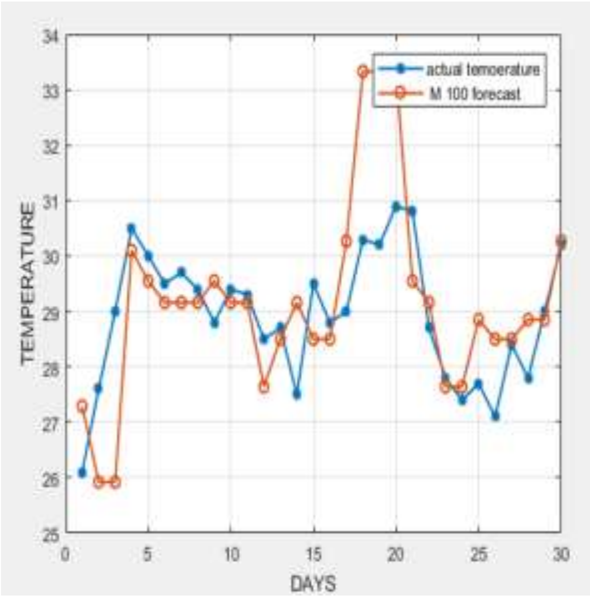

Figure 6 Comparison between Actual and Forecasted Values for June, 1996 using 100 Monte Carlo Simulation

From the figure above (Figure 6) it shows that for 100 Monte Carlo simulation the predicted values are in trend with the actual values and it indicate best trend compare to 1000 Monte Carlo simulation but not as best as 500 Monte Carlo simulations.

When the graphs were analyzed, results obtained show that increasing the Monte Carlo simulation affects the general data trend of the time series in varying degrees; decreasing the ratio of positive-to-negative trends. Also, results show that increasing the Monte Carlo simulation does not necessarily increase the forecast accuracy regardless of the structure of the time series. The error is increasing in most cases probably because of the increase in the uncertainty of the data trend accessioned by increase in Monte Carlo simulation.

5.2 Performance Evaluation

In order to test the superiority of the proposed model, the performance of the proposed fuzzy time series forecasting model was evaluated using the performance measures of Mean Square Error (MSE) and the results are compared with those obtained in the work of Cheng & Sheng (2012). To make the comparison fair, the same intervals and factors were used. The performance was further validated using the Average Forecasting Error percentage (AFEP). The MSE and AFEP results are shown in Table 2.

It can be seen from Table 2 that the result of the proposed methodhas the smallest MSE and AFEP values of 0.8596 and 2.4283respectively, when compared with the values obtained for both models in Cheng & Sheng (2012).

Furthermore, the performance of the model was also compared with the methods proposed by Cheng & Cheng (2010), Cheng & Sheng (2006) and Lee (2006) in terms of the MSE, RMSE and AFEP presented in Table 2.

The comparative analyses in TableVIII signify that the proposed model exhibits higher forecasting accuracy than those of considered competing models in terms of MSE, RMSE and AFEP for the bivariate daily average temperature data.

6. CONCLUSION

This paper has implemented a stochastic forecasting model for fuzzy time series by extending Sullivan and Woodall's Markov-based model. It was built upon an HMM in which the fuzzy relationships were formulated as state transitions so that it can handle two-factor forecasting problems. To allow the model to more effectively reflect the real-world situation and randomness of forecasting, Monte Carlo simulation was applied to estimate the stochastic outcome. The computations involved in forecasting were simple matrix operations, and thus, the model is more efficient than other "IF–THEN"-based models. The model was used to forecasting problem of daily average temperature and cloud density in Taipei, Taiwan, as the benchmark and conducted performance comparisons with other models. The results demonstrated the superiority of our model in forecasting. Moreover, the implemented probabilistic forecasting model adheres to the central limit theorem, proved by an experiment of sensitiveness, and thus, the forecasting results statistically approximate to the real mean of the target value being forecast.

Table 1: Forecasting Results of the Proposed Method for Average Temperatures of the Month of June 1996 in Taipei, Taiwan

Days	Actual Value (AV)	Forecast Value (FV)	(FV-AV)	$(FV-AV)^2$	| FV-AV |
1	26.1	27.31	1.21	1.4641	1.21
2	27.6	26.15	-1.45	2.1025	1.45
3	29	26.15	-2.85	8.1225	2.85
4	30.5	30.11	-0.39	0.1521	0.39
5	30	29.61	-0.39	0.1521	0.39
6	29.5	29.17	-0.33	0.1089	0.33
7	29.7	29.18	-0.52	0.2704	0.52
8	29.4	29.17	-0.23	0.0529	0.23
9	28.8	29.61	-0.81	0.6561	0.81
10	29.4	29.11	-0.29	0.0841	0.29
11	29.3	29.63	0.33	0.1089	0.33
12	28.5	29.63	1.13	1.2769	1.13
13	28.7	27.6	-0.11	0.0121	0.11
14	27.5	28.5	1	1	1
15	29.5	28.5	-1	1	1
16	28.8	29.61	0.81	0.6561	0.81
17	29	28.5	-0.5	0.25	0.5
18	30.3	29.98	-0.32	0.1024	0.32
19	30.2	30.04	-0.16	0.0256	0.16
20	30.9	30.06	-0.86	0.7396	0.86
21	30.8	29.61	-1.19	1.4161	1.19
22	28.7	29.17	0.47	0.2209	0.47
23	27.8	27.6	-0.2	0.04	0.2
24	27.4	27.6	0.2	0.04	0.2
25	27.7	29.11	1.41	1.9881	1.41
26	27.1	28.5	1.4	1.96	1.4
27	28.4	28.5	0.1	0.01	0.1
28	27.8	29.11	1.31	1.7161	1.31
29	29	29.11	0.11	0.0121	0.11
30	30.2	29.98	-0.22	0.0484	0.22
			SUM	25.789	

Table 2: Performance Comparison

Performance Measures	Cheng & Sheng (2006)	Lee et al., (2006)	Cheng & Sheng (2010)	Proposed Method
MSE	1.6948	1.034	0.995	**0.8596**
AFEP	3.7409	2.97	2.8071	**2.4283**
RMSE	1.1799	1.0361	1.0481	**0.9272**

REFERENCES

[1] F. G. Schmitt and Y. Huang, *Stochastic analysis of scaling time series: from turbulence theory to applications*. Cambridge University Press, 2016.

[2] M. H. Amini, A. Kargarian, and O. J. E. P. S. R. Karabasoglu, "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation," vol. 140, pp. 378-390, 2016.

[3] S.-T. Li, S.-C. Kuo, Y.-C. Cheng, and C.-C. Chen, "Deterministic vector long-term forecasting for fuzzy time series," *Fuzzy Sets and Systems*, vol. 161, no. 13, pp. 1852-1870, 2010/07/01/ 2010.

[4] Q. Song, B. S. J. F. s. Chissom, and systems, "Forecasting enrollments with fuzzy time series—part I," vol. 54, no. 1, pp. 1-9, 1993.

[5] R.-C. Tsaur, J.-C. O. Yang, and H.-F. Wang, "Fuzzy relation analysis in fuzzy time series model," *Journal of Computers Mathematics with Applications*, vol. 49, no. 4, pp. 539-548, 2005.

[6] E. Egrioglu, Y. Aslan, and C. H. J. T. J. o. F. S. Aladag, "A new fuzzy time series method based on Artificial Bee Colony Algorithm," vol. 5, no. 1, pp. 59-77, 2014.

[7] E. Bas, U. Yolcu, E. Egrioglu, and C. H. J. A. J. o. I. S. Aladag, "A fuzzy time series forecasting method based on operation of union and feed forward artificial neural network," vol. 5, no. 3, pp. 81-91, 2015.

[8] F. Radmehr and N. J. S. A. J. o. I. E. Gharneh, "Forecasting method based on high order fuzzy time series and simulated annealing technique," vol. 23, no. 2, pp. 176-190, 2012.

[9] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series—part II," *Journal of Fuzzy Sets and System*, vol. 62, no. 1, pp. 1-8, 1994.

[10] E. Bai, W. K. Wong, W. C. Chu, M. Xia, and F. Pan, "A heuristic time-invariant model for fuzzy time series forecasting," *Expert Systems with Applications,* vol. 38, no. 3, pp. 2701-2707, 2011/03/01/ 2011.

[11] S.-L. Lu, "Integrating heuristic time series with modified grey forecasting for renewable energy in Taiwan," *Renewable Energy,* 2018/09/07/ 2018.

[12] E. Egrioglu, C. H. Aladag, and U. J. E. S. w. A. Yolcu, "Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks," vol. 40, no. 3, pp. 854-857, 2013.

[13] C. H. Aladag, U. Yolcu, E. Egrioglu, and E. Bas, "Fuzzy lagged variable selection in fuzzy time series with genetic algorithms," *Applied Soft Computing,* vol. 22, pp. 465-473, 2014/09/01/ 2014.

[14] E. Bas, C. Grosan, E. Egrioglu, and U. Yolcu, "High order fuzzy time series method based on pi-sigma neural network," *Engineering Applications of Artificial Intelligence,* vol. 72, pp. 350-356, 2018/06/01/ 2018.

[15] C. H. Aladag, U. Yolcu, E. Egrioglu, and A. Z. Dalar, "A new time invariant fuzzy time series forecasting method based on particle swarm optimization," *Applied Soft Computing,* vol. 12, no. 10, pp. 3291-3299, 2012/10/01/ 2012.

[16] S.-T. Li, Y.-C. J. I. T. o. S. Cheng, Man,, and P. B. Cybernetics, "A stochastic HMM-based forecasting model for fuzzy time series," vol. 40, no. 5, pp. 1255-1266, 2010.

[17] A. Babakura, M. N. Sulaiman, N. Mustapha, and T. J. I. J. o. S. H. Perumal, "Hmm-based decision model for smart home environment," vol. 8, no. 1, pp. 129-138, 2014.

[18] M. R. Hassan, B. Nath, M. Kirley, and J. J. N. Kamruzzaman, "A hybrid of multiobjective Evolutionary Algorithm and HMM-Fuzzy model for time series prediction," vol. 81, pp. 1-11, 2012.

[19] R. Hassan, B. Nath, and M. Kirley, "HMM based fuzzy model for time series prediction," in *Fuzzy Systems, 2006 IEEE International Conference on*, 2006, pp. 2120-2126: IEEE.

[20] S.-T. Li, S.-C. Kuo, Y.-C. Cheng, and C.-C. Chen, "A vector forecasting model for fuzzy time series," *Applied Soft Computing,* vol. 11, no. 3, pp. 3125-3134, 2011/04/01/ 2011.

[21] L. Maciel and R. Ballini, "A fuzzy inference system modeling approach for interval-valued symbolic data forecasting," *Knowledge-Based Systems,* vol. 164, pp. 139-149, 2019/01/15/ 2019.

[22] E. Bas, C. Grosan, E. Egrioglu, and U. J. E. A. o. A. I. Yolcu, "High order fuzzy time series method based on pi-sigma neural network," vol. 72, pp. 350-356, 2018.

[23] M.-Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," *Future Generation Computer Systems,* vol. 37, pp. 461-467, 2014/07/01/ 2014.

[24] J. A. J. J. o. m. a. Goguen and applications, "L-fuzzy sets," vol. 18, no. 1, pp. 145-174, 1967.

[25] L. Kuncheva, J. C. Bezdek, and M. A. Sutton, "On combining multiple classifiers by fuzzy templates," in *Fuzzy Information Processing Society-NAFIPS, 1998 Conference of the North American*, 1998, pp. 193-197: IEEE.

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

[27] A. Rubio, J. D. Bermúdez, and E. Vercher, "Forecasting portfolio returns using weighted fuzzy time series methods," *International Journal of Approximate Reasoning,* vol. 75, pp. 1-12, 2016/08/01/ 2016.

[28] P. Singh and B. Borah, "Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization," *International Journal of Approximate Reasoning,* vol. 55, no. 3, pp. 812-833, 2014/03/01/ 2014.

[29] U. M. Umer, T. Sevil, G. J. T. J. o. F. Sevil, and D. Science, "Forecasting performance of smooth transition autoregressive (STAR) model on travel and leisure stock index," 2018.

[30] O. Cagcag Yolcu, U. Yolcu, E. Egrioglu, and C. H. Aladag, "High order fuzzy time series forecasting method based on an intersection operation," *Applied Mathematical Modelling,* vol. 40, no. 19, pp. 8750-8765, 2016/10/01/ 2016.

[31] P. Singh and G. Dhiman, "A hybrid fuzzy time series forecasting model based on granular computing and bio-inspired optimization approaches," *Journal of Computational Science,* vol. 27, pp. 370-385, 2018/07/01/ 2018.

[32] H. J. Teoh, T.-L. Chen, C.-H. Cheng, and H.-H. Chu, "A hybrid multi-order fuzzy time series for forecasting stock markets," *Expert Systems with Applications,* vol. 36, no. 4, pp. 7888-7897, 2009/05/01/ 2009.

[33] Y.-C. Cheng and S.-T. J. I. T. o. F. S. Li, "Fuzzy time series forecasting with a probabilistic smoothing hidden Markov model," vol. 20, no. 2, pp. 291-304, 2012.

[34] Y. Syu, J.-Y. Kuo, and Y.-Y. Fanjiang, "Time series forecasting for dynamic quality of web services: An empirical study," *Journal of Systems and Software,* vol. 134, pp. 279-303, 2017/12/01/ 2017.

[35] F. Ye, L. Zhang, D. Zhang, H. Fujita, and Z. Gong, "A novel forecasting method based on multi-order fuzzy time series and technical analysis," *Information Sciences,* vol. 367-368, pp. 41-57, 2016/11/01/ 2016.