Menemui Matematik (Discovering Mathematics)
Vol. 44, No. 2: 127-138 (2022)

Approximating The Effective Length of Interval to Forecast in Fuzzy Time Series

Suriana L.[1] , Suzelawati, Z.[2] , Risman, M. H.[3] and Amirul, A.[4]
[1]*Preparatory Centre for Science and Technology, University Malaysia Sabah, 88400 UMS Kota Kinabalu Sabah*
[2]*Faculty of Science and Natural Resources, University Malaysia Sabah, 88400 UMS Kota Kinabalu Sabah*
[3]*Department of Mathematics and Statistics, Faculty of Science, University of Putra Malaysia, 43400 UPM Serdang, Selangor*
[4]*Azman Hashim International Business School, University of Technology Malaysia, 81310 UTM Johor Bahru, Johor.*
[1]suriana@ums.edu.my, [2]suzela@ums.edu.my, [3]risman@upm.edu.my, [4]azmirul@utm.my

ABSTRACT

The research on forecasting in fuzzy time series has increased due to its capability in dealing with uncertainty and vagueness. However, in this research the effectiveness of the forecasting is hugely depending on the first step in every forecasting model being applied, that is the determination of the size of the intervals. However, previous study did not mention on the best length of interval to be used in the model. In this study, we suggested a few different lengths of interval to be used, to look for the best size of interval in fuzzy time series. The aim is to increase the accuracy of forecasting. This method is applied to the selected data of tuberculosis cases reported monthly in Sabah starting from January 2012 until May 2020. The data is collected from the Queen Elizabeth Hospital in Kota Kinabalu, Sabah. The performance of evaluations is showed by comparison on the values obtained for MSE and RMSE. One numerical data set from the whole tuberculosis data were used to illustrate the chosen methods.

Keywords: Fuzzy Time Series, Length of Interval, Forecasting, Average Based Interval, Tuberculosis

INTRODUCTION

It is obvious that forecasting plays an important role in our life. It is proved that the statistical approaches in time series may forecast the problems arising from new trends, but somehow fail to forecast data obtained from the linguistic facts.

In early 1980, Song and Chissoms introduced the concept of fuzzy time series and used this concept to forecast the enrolment of students in University of Alabama. The conventional linear regression technique is applied, and the forecasted number obtained is compared to the forecasted number obtained by fuzzy time series techniques. Later, the same researchers, Song and Chissoms (1993a, 1993b) proposed a new time series forecasting model called fuzzy time series whereby this proposed method is capable in dealing with uncertainty and do not require any assumptions. However, their methods require large amount of computational time since their proposed methods is based on complex matrix operations and there are no exact details on how to determine the length of interval being used.

Chens (1996) simplify Song and Chissoms (1993a, 1993b) methods by using simpler arithmetic operations instead of complex max-min operations method. Chens (2002) methods proved to make good forecast on the number of students enrollment in University of Alabama. Huarng (2001) who has been concerned on the effectiveness of length of Interval chosen proved that average-based intervals give more accurate forecast compared to distribution-based length. In 2002, Chens continues his study on the same methods, but he is focusing on high order fuzzy time series. Li and Chen (2004) proposed a concept of 3-4-5 rules of natural partitions which is then be applied by Ramli and Mohamad (2017) in their forecasting models in forecasting the unemployment rate under different degree of confidence.

Meanwhile Huarng and Yu (2006) proposed a type-2 fuzzy time series model for stock index forecasting. They proved that type-2 model is much better for smooth defuzzified forecast and forecast consistently. Similar work on fuzzy time series can be found in Lee, et. al. (2001) and Tsai and Wu (1999). The other researchers such as Liu (2007), Ozge, et. al. (2020) and Lee and Chou (2007) does not specify on how they find the suitable the length of interval in their proposed methods. Recently, Susilo, et. at. (2022) predict covid-19 in Central Java by applying average based fuzzy time series with some modification to frequency density partition. They conclude that their method could increase the accuracy in determining the number of positive cases which can be seen from the MAPE value obtained.

The rest of the paper is organized as follows. Section 2 will briefly explain on the concept of fuzzy time series forecasting. The next section is then explaining on the proposed methods in general starting from Step 1 until Step 5. Then, we have the empirical analysis with three different lengths of intervals being focused. The final step is the discussion on forecasting validation based on values of Mean Square Error (MSE) and Root Mean Square Error (RMSE), followed by conclusion.

FUZZY TIME SERIES

Fuzzy logic and fuzzy set theory was introduced to deal with the vagueness and uncertainty in most real-world problems. This fuzzy set theory was initially introduced by Zadeh in (1965) It has been applied in many fields, for example, health care, finance, environment, energy, image processing, modelling, forecasting, optimization. In 1975, Zadeh (1971, 1973, 1975) introduced fuzzy arithmetics theory and its application. Meanwhile, Zenian et al. (2020) implemented advanced fuzzy set to enhance the image of Flat Electroencephalography by comparing the performance of intuitionistic fuzzy and type-2 fuzzy approaches.

The definition of fuzzy time series was proposed by Song and Chissoms (1993a, 1993b). Let U be the universe of discourse, where $U = u_1, u_2, \ldots, u_n$. A fuzzy set of A_i of U is defined by

$$A_i = \frac{\mathrm{f_{A_i}(u_1)}}{u_1} + \frac{\mathrm{f_{A_i}(u_2)}}{u_2} + \cdots + \frac{\mathrm{f_{A_i}}(\mathrm{u}_n)}{u_n} \tag{1}$$

where $\mathrm{f_{A_i}}$ is the membership function of the fuzzy set A_i, $\mathrm{f_{A_i}}: U \to [0,1]$. u_k is the element of fuzzy set A_i, and $\mathrm{f_{A_i}}(u_k)$ is the degree of belongingness of u_k to A_i for $\mathrm{f_{A_i}}(u_k) \in [0, 1]$ and $1 \leq k \leq n$.

Definition 1: Definition 1: $Y\,(t)(t\; = \ldots, 0, 1, 2, \ldots)$, is a subset of R. Let $Y(t)$ be the universe of discourse defined by fuzzy set $f_i(t)$. $F(t)$ is defined as a fuzzy time series on $Y\,(t)(t = \ldots, 0, 1, 2, \ldots)$

Definition 2: If there exist a fuzzy relationship $R(t\; - \;1, t)$, such that $F(t) = F(t - \;1) \; \times \; R(t\; - \;1, t)$, where $\times$ is an operator, then $F(t)$) is said to be caused by $F(t-1)$. The relationship between $F(t)$ and $F(t-1)$ can be denoted by

$$F\;(t-1) \to F\;(t). \tag{2}$$

Definition 3: Suppose $F(t-1) = A_i$ and $F(t) \;=\; A_j$, a fuzzy logical relationship (FLR) is defined as $A_i \to A_j$, where A_i and A_j is on the left- and right-hand side respectively, while the repeated fuzzy logical relationship is removed.

Definition 4: Fuzzy logical relationship can be group together into fuzzy logical relationship group (FLRG) according to the same left-hand sides FLR. According to Chen's model, the repeated fuzzy sets will be removed in the FLRG. For example,

$$\begin{matrix} A_i \rightarrow A_{j1} \\ A_i \rightarrow A_{j2} \\ \ldots \end{matrix} \quad \rightarrow \quad A_i \rightarrow A_{j1}, A_{j2}, \ldots \tag{3}$$

THE PROPOSED METHOD

In this section, we apply Chen's method (1996) with some modification in finding the length of intervals. The algorithm in average based interval by Huarng (2001) to set the length of interval is given below:

a. Determine the absolute difference between data n and $n+1$, and find their average.
b. Determine the value of half of their average value.
c. Determine the basis value of interval length according to basis mapping table below.

TABLE 1. Basis Mapping Table

Range	Basis
0, 1 – 1, 0	0, 1
1, 1 - 10	1
11 - 100	10
101 - 1000	100

The modification is as follows:

a. The length of interval obtained by using the algorithm above is 10.
b. The length of interval is then divided by 2, giving the length of interval is 5.
c. The length of interval is then multiplied by 2, giving the length of interval is 20.

The proposed method is presented as follow.
Step 1: Define the universe of discourse $U = [U_{min} - D_1, U_{max} + D_2]$ into n equal length of intervals $u_1, u_2, \ldots, U_n$, where U_{min} and U_{max} are minimum and maximum values in raw data and D_1, D_2 are two real numbers.

Step 2: Fuzzy sets A_i. The linguistics variable is the raw data, A_i as possible linguistics values of the raw data. Each is defined by the intervals $u_1, u_2, \ldots, u_n$.

$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_n$
$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_n$
$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_n$
$\ldots$
$A_n = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_n$

Step 3: Obtained fuzzy logical relationship (FLR) according to Definition 2. Removed repeated FLR as in Definition 3.

Step 4: Derived fuzzy logical relationship group (FLRG) as in Definition 4.

Step 5: Calculated the forecasted output. The calculation is based on three rules in Chens methods, 1996.

- Rule 1: If the fuzzified number of cases of day i is A_i, and there is only one FLR in the FLRG obtained in STEP 4, that is $A_i \rightarrow A_k$, whereby A_k occurs in the interval u_k and the midpoint of u_k is m_k, then the forecast number of cases of day $i+1$ is m_k.
- Rule 2: If the fuzzified number of cases of day A_i and there is more than one FLR in the FLRG obtained in STEP 4, that is, $A_i \rightarrow A_{k1},\ A_i \rightarrow A_{k2},\ A_i \rightarrow A_{k3}, \dots, A_i \rightarrow A_{kq}$ whereby $A_{k1}, A_{k2}, A_{k3}, \dots, A_{kq}$ occurs in the intervals $u_{k1}, u_{k2}, u_{k3}, \dots, u_{kq}$ and their midpoint are $m_{k1}, m_{k2}, m_{k3}, \dots, m_{kq}$, the the forecast number of cases of the day $i+1$ is $\frac{m_{k1}+m_{k2}+ m_{k3}+\cdots+m_{kq}}{q}$.
- If the fuzzified number of cases of day i is A_i , and there is empty FLR in the FLRG obtained in STEP 4, that is $A_i \rightarrow \emptyset$, whereby A_i occurs in the intervals u_i and the midpoint of u_i is m_i , then the forecast number of cases of the day $i+1$ is m_i.

RESULTS AND DISCUSSION

Table 2 shows the tuberculosis cases obtained from Hospital Queen Elizabeth, Kota Kinabalu Sabah. These data are used to be applied in the proposed method.

TABLE 2. The number of tuberculosis cases reported monthly in Sabah from January 2012 until May 2020.

Month Year	Cases	Month Year	Cases	Month Year	Cases	Month Year	Cases	Month Year	Cases
Jan 2012	199	Oct 2013	389	July 2015	332	Apr 2017	366	Jan 2019	382
Feb 2012	487	Nov 2013	359	Aug 2015	301	May 2017	437	Feb 2019	396
Mar 2012	356	Dec 2013	370	Sep 2015	397	Jun 2017	428	Mar 2019	375
Apr 2012	364	Jan 2014	394	Oct 2015	385	July 2017	404	Apr 2019	495
May 2012	419	Feb 2014	353	Nov 2015	393	Aug 2017	472	May 2019	376
Jun 2012	349	Mar 2014	341	Dec 2015	542	Sep 2017	347	Jun 2019	331
July 2012	341	Apr 2014	495	Jan 2016	338	Oct 2017	394	July 2019	590
Aug 2012	422	May 2014	356	Feb 2016	372	Nov 2017	532	Aug 2019	431
Sep 2012	365	Jun 2014	395	Mar 2016	505	Dec 2017	590	Sep 2019	500

Oct 2012	380	July 2014	389	Apr 2016	394	Jan 2018	392	Oct 2019	442
Nov 2012	351	Aug 2014	406	May 2016	401	Feb 2018	393	Nov 2019	440
Dec 2012	353	Sep 2014	332	Jun 2016	401	Mar 2018	371	Dec 2019	524
Jan 2013	380	Oct 2014	428	July 2016	305	Apr 2018	482	Jan 2020	441
Feb 2013	385	Nov 2014	376	Aug 2016	482	May 2018	307	Feb 2020	413
Mar 2013	379	Dec 2014	482	Sep 2016	410	Jun 2018	381	Mar 2020	440
Apr 2013	376	Jan 2015	296	Oct 2016	311	July 2018	515	Apr 2020	217
May 2013	402	Feb 2015	345	Nov 2016	601	Aug 2018	335	May 2020	332
Jun 2013	342	Mar 2015	330	Dec 2016	433	Sep 2018	372		
July 2013	432	Apr 2015	418	Jan 2017	299	Oct 2018	573		
Aug 2013	324	May 2015	321	Feb 2017	390	Nov 2018	351		
Sep 2013	388	Jun 2015	404	Mar 2017	446	Dec 2018	572		

Step 1: Define the course of universe. According to Table 2, $U_{min} = 199$ and $U_{max} = 601$ respectively. We choose $D_1 = 9$ and $D_2 = 9$ respectively. Thus, $U = [199 - 9, 601 + 9] = [190,610]$. From U, we apply average based algorithm and modification, considering three different lengths of interval, which is size 5, size 10, and size 20. The number of intervals and list of intervals are as shown in Table 3 below.

TABLE 3. The length of intervals, the number of intervals and list of intervals.

Length of intervals	**Number of intervals**	**List of intervals**

Size 5	84	$u_1 = [190 - 195], u_2 = [195 - 200], \dots, u_{84} = [605 - 610]$
Size 10	42	$u_1 = [190 - 200], u_2 = [200 - 210], \dots, u_{42} = [600 - 610]$
Size 20	21	$u_1 = [190 - 210], u_2 = [210 - 230], \dots, u_{21} = [590 - 610]$

Step 2: Fuzzy sets A_i. The linguistics variable is the number of tuberculosis cases reported monthly, while A_i as possible linguistics values of the raw data. Each is defined by the intervals $u_1, u_2, \dots, u_n$ as in Table 4.

TABLE 4. The fuzzy sets, A_i according to the different length of intervals.

Length of intervals	**Number of intervals**	**Fuzzy sets, A_i**
Size 5	84	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{84}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{84}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_{84}$ … $A_{84} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_{84}$
Size 10	42	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{42}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{42}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_{42}$ … $A_{42} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_{42}$
Size 20	21	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ … $A_{21} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_{21}$

Step 3: Create fuzzy logical relationship (FLR) and fuzzy logical relationship group (FLRG) based on Definition 3 and Definition 4 as shown in Table 5 and Table 6 respectively.

TABLE 5. The fuzzy set, A_i and the fuzzy logical relationship (FLR).

Month / Year	**Cases**	**Fuzzy set (A_i)**			**Fuzzy logical relationship (FLR)**		
		Size 5	**Size 10**	**Size 20**	**Size 5**	**Size 10**	**Size 20**

…	…	…	…	…	…	…	…
Jan 2014	394	A_{41}	A_{21}	A_{11}	$A_{36} \rightarrow A_{41}$	$A_{18} \rightarrow A_{21}$	$A_{9} \rightarrow A_{11}$
Feb 2014	353	A_{33}	A_{17}	A_{9}	$A_{41} \rightarrow A_{33}$	$A_{21} \rightarrow A_{17}$	$A_{11} \rightarrow A_{9}$
Mar 2014	341	A_{31}	A_{16}	A_{8}	$A_{33} \rightarrow A_{31}$	$A_{17} \rightarrow A_{16}$	$A_{9} \rightarrow A_{8}$
Apr 2014	495	A_{61}	A_{31}	A_{16}	$A_{31} \rightarrow A_{61}$	$A_{16} \rightarrow A_{31}$	$A_{8} \rightarrow A_{16}$
May 2014	356	A_{34}	A_{17}	A_{9}	$A_{61} \rightarrow A_{34}$	$A_{31} \rightarrow A_{17}$	$A_{16} \rightarrow A_{9}$
Jun 2014	395	A_{41}	A_{21}	A_{11}	$A_{34} \rightarrow A_{41}$	$A_{17} \rightarrow A_{21}$	$A_{9} \rightarrow A_{11}$
July 2014	389	A_{40}	A_{20}	A_{10}	$A_{41} \rightarrow A_{40}$	$A_{21} \rightarrow A_{20}$	$A_{11} \rightarrow A_{10}$
Aug 2014	406	A_{44}	A_{22}	A_{11}	$A_{40} \rightarrow A_{44}$	$A_{20} \rightarrow A_{22}$	$A_{10} \rightarrow A_{11}$
Sep 2014	332	A_{29}	A_{15}	A_{8}	$A_{44} \rightarrow A_{29}$	$A_{22} \rightarrow A_{15}$	$A_{11} \rightarrow A_{8}$
Oct 2014	428	A_{48}	A_{24}	A_{12}	$A_{29} \rightarrow A_{48}$	$A_{15} \rightarrow A_{24}$	$A_{8} \rightarrow A_{12}$
Nov 2014	376	A_{38}	A_{19}	A_{10}	$A_{48} \rightarrow A_{38}$	$A_{24} \rightarrow A_{19}$	$A_{12} \rightarrow A_{10}$
Dec 2014	482	A_{59}	A_{30}	A_{15}	$A_{38} \rightarrow A_{59}$	$A_{19} \rightarrow A_{30}$	$A_{10} \rightarrow A_{15}$
…	…	…	…	…	…	…	…

TABLE 6. The fuzzy logical relationship group (FLRG).

Month / Year	**Cases**	**Fuzzy logical relationship group (FLRG)**		
		Size 5	**Size 10**	**Size 20**

…	…	…	…	…
Jan 2014	394	$A_{36} \rightarrow A_{37}, A_{41}, A_{50}$	$A_{18} \rightarrow A_{19}, A_{21}, A_{23}, A_{25}$	$A_{9} \rightarrow A_{8}, A_{9}, A_{10}, A_{11}, A_{12}$ A_{13}, A_{20}
Feb 2014	353	$A_{41} \rightarrow A_{33}, A_{37}, A_{40},$ $A_{41}, A_{43}, A_{69}, A_{71}$	$A_{21} \rightarrow A_{17}, A_{19}, A_{20},$ $A_{21}, A_{22}, A_{35}, A_{36}$	$A_{11} \rightarrow A_{6}, A_{7}, A_{8}, A_{9}, A_{10}$ A_{11}, A_{15}, A_{18}
Mar 2014	341	$A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$	$A_{17} \rightarrow A_{16}, A_{17}, A_{18}, A_{19},$ A_{21}, A_{39}	$A_{9} \rightarrow A_{8}, A_{9}, A_{10}, A_{11}, A_{12}$ A_{13}, A_{20}
Apr 2014	495	$A_{31} \rightarrow A_{28}, A_{47}, A_{49}, A_{61}$	$A_{16} \rightarrow A_{14}, A_{16}, A_{24}, A_{25}, A$	$A_{8} \rightarrow A_{6}, A_{7}, A_{8}, A_{9}, A_{10},$ $A_{11}, A_{12}, A_{13}, A_{16}, A_{20}$
May 2014	356	$A_{61} \rightarrow A_{34}, A_{38}$	$A_{31} \rightarrow A_{17}, A_{19}, A_{26}$	$A_{16} \rightarrow A_{9}, A_{10}, A_{11}, A_{13}$
Jun 2014	395	$A_{34} \rightarrow A_{35}, A_{36}, A_{41}$	$A_{17} \rightarrow A_{16}, A_{17}, A_{18}, A_{19},$ A_{21}, A_{39}	$A_{9} \rightarrow A_{8}, A_{9}, A_{10}, A_{11}, A_{12}$ A_{13}, A_{20}
July 2014	389	$A_{41} \rightarrow A_{33}, A_{37}, A_{40},$ $A_{41}, A_{43}, A_{69}, A_{71}$	$A_{21} \rightarrow A_{17}, A_{19}, A_{20}, A_{21},$ A_{22}, A_{35}, A_{36}	$A_{11} \rightarrow A_{6}, A_{7}, A_{8}, A_{9}, A_{10}$ A_{11}, A_{15}, A_{18}
Aug 2014	406	$A_{40} \rightarrow A_{34}, A_{40}, A_{44}, A_{52}$	$A_{20} \rightarrow A_{17}, A_{18}, A_{19}, A_{20},$ $A_{21}, A_{22}, A_{26}, A_{33}$	$A_{10} \rightarrow A_{8}, A_{9}, A_{10}, A_{11}, A_{13}$ $A_{15}, A_{16}, A_{17}, A_{20}$
Sep 2014	332	$A_{44} \rightarrow A_{25}, A_{29}$	$A_{22} \rightarrow A_{12}, A_{13}, A_{15}, A_{16}$ A_{22}, A_{29}	$A_{11} \rightarrow A_{6}, A_{7}, A_{8}, A_{9}, A_{10}$ A_{11}, A_{15}, A_{18}
Oct 2014	428	$A_{29} \rightarrow A_{23}, A_{37}, A_{48}, A_{80}$	$A_{15} \rightarrow A_{12}, A_{19}, A_{24}, A_{40}$	$A_{8} \rightarrow A_{6}, A_{7}, A_{8}, A_{9}, A_{10},$ $A_{11}, A_{12}, A_{13}, A_{16}, A_{20}$

Nov 2014	376	$A_{48} \rightarrow A_{38}, A_{43}$	$A_{24} \rightarrow A_{18}, A_{19}, A_{22}$	$A_{12} \rightarrow A_7, A_8, A_9, A_{10}, A_{11}, A_{13}$
Dec 2014	482	$A_{38} \rightarrow A_{29}, A_{33}, A_{38}, A_{39}, A_{43}, A_{59}$	$A_{19} \rightarrow A_{15}, A_{17}, A_{19}, A_{20} A_{22}, A_{30}, A_{31}, A_{32}, A_{35}$	$A_{10} \rightarrow A_8, A_9, A_{10}, A_{11}, A_{13} A_{15}, A_{16}, A_{17}, A_{20}$
…	…	…	…	…

Step 5: Calculate the forecasted outputs. The numerical example of Jan 2014 chosen is shown below.

[Size 5]: According to Table 5., the FLRG of Jan 2014 is $A_{36} \rightarrow A_{37}, A_{41}, A_{50}$. Thus, according to Chens methods in Rule 2, the forecasted values will the average of intervals midpoints for A_{37}, A_{41}, A_{50}.

$$\frac{m_{37} + m_{41} + m_{50}}{3} = 400.83 \tag{4}$$

[Size 10]: According to Table 5., the FLRG of Jan 2014 is $A_{18} \rightarrow A_{19}, A_{21}, A_{23}, A_{25}$. Thus, according to Chens methods in Rule 2, the forecasted values will the average of intervals midpoints for $A_{19}, A_{21}, A_{23}, A_{25}$.

$$\frac{m_{19} + m_{21} + m_{23} + m_{25}}{4} = 405.50 \tag{5}$$

[Size 20]: According to Table 5., the FLRG of Jan 2014 is $A_9 \rightarrow A_8, A_9, A_{10}, A_{11}, A_{12}, A_{13}, A_{20}$. Thus, according to Chens methods in Rule 2, the forecasted values will the average of intervals midpoints for $A_8, A_9, A_{10}, A_{11}, A_{12}, A_{13}, A_{20}$.

$$\frac{m_8 + m_9 + m_{10} + m_{11} + m_{12} + m_{13} + m_{20}}{7} = 417.14 \tag{6}$$

The forecasted tuberculosis cases with respect to length of interval size 5, size 10 and size 20 are as shown in Table 7.

TABLE 7. Comparison on forecasted tuberculosis cases with respect to different length of interval.

Month/ Year	Actual cases	Forecast cases according to different length of interval			Month/ Year	Actual cases	Forecast cases according to different length of interval		
		Size 5	Size 10	Size 20			Size 5	Size 10	Size 20
Jan-12	199				Apr-16	394	392.50	395.50	395.00
Feb-12	487	487.50	485.50	480.00	May-16	401	426.07	428.36	390.00
Mar-12	356	357.50	340.50	350.00	Jun-16	401	370.50	363.83	390.00
Apr-12	364	374.17	402.17	417.14	Jul-16	305	370.50	363.83	390.00

May-12	419	397.50	405.50	417.14	Aug-16	482	440.00	422.17	400.00
Jun-12	349	335.00	368.83	373.33	Sep-16	410	337.50	340.50	350.00
Jul-12	341	367.50	405.50	408.89	Oct-16	311	322.50	363.83	390.00
Aug-12	422	418.75	405.50	408.89	Nov-16	601	602.50	605.50	450.00
Sep-12	365	362.50	382.17	373.33	Dec-16	433	432.50	435.50	440.00
Oct-12	380	397.50	405.50	417.14	Jan-17	299	372.50	380.50	397.78
Nov-12	351	388.33	431.06	444.44	Feb-17	390	365.00	365.50	400.00
Dec-12	353	411.25	402.17	417.14	Mar-17	446	400.00	405.50	444.44
Jan-13	380	411.25	402.17	417.14	Apr-17	366	367.50	403.00	397.78
Feb-13	385	388.33	431.06	444.44	May-17	437	400.83	405.50	417.14
Mar-13	379	412.50	405.50	444.44	Jun-17	428	389.17	380.50	397.78
Apr-13	376	388.33	431.06	444.44	Jul-17	404	390.00	382.17	373.33
May-13	402	388.33	431.06	444.44	Aug-17	472	370.50	363.83	390.00
Jun-13	342	370.50	363.83	390.00	Sep-17	347	347.50	445.50	350.00
Jul-13	432	418.75	405.50	408.89	Oct-17	394	367.50	403.00	408.89
Aug-13	324	372.50	380.50	397.78	Nov-17	532	426.07	428.36	390.00
Sep-13	388	395.00	402.17	450.00	Dec-17	590	587.50	470.50	455.00
Oct-13	389	400.00	405.50	444.44	Jan-18	392	412.50	415.50	395.00
Nov-13	359	400.00	405.50	444.44	Feb-18	393	426.07	428.36	390.00
Dec-13	370	374.17	402.17	417.14	Mar-18	371	426.07	428.36	390.00
Jan-14	394	400.83	405.50	417.14	Apr-18	482	512.50	431.06	444.44
Feb-14	353	426.07	428.36	390.00	May-18	307	337.50	340.50	350.00
Mar-14	341	411.25	402.17	417.14	Jun-18	381	382.50	422.17	400.00
Apr-14	495	418.75	405.50	408.89	Jul-18	515	412.50	405.50	444.44
May-14	356	367.50	392.17	395.00	Aug-18	335	332.50	335.50	390.00
Jun-14	395	374.17	402.17	417.14	Sep-18	372	422.50	423.00	408.89
Jul-14	389	426.07	428.36	390.00	Oct-18	573	512.50	431.06	444.44
Aug-14	406	400.00	405.50	444.44	Nov-18	351	367.50	470.50	395.00
Sep-14	332	322.50	363.83	390.00	Dec-18	572	411.25	402.17	417.14
Oct-14	428	422.50	423.00	408.89	Jan-19	382	367.50	385.50	395.00
Nov-14	376	390.00	382.17	373.33	Feb-19	369	412.50	405.50	444.44
Dec-14	482	388.33	431.06	444.44	Mar-19	375	400.83	405.50	417.14
Jan-15	296	337.50	340.50	350.00	Apr-19	495	512.50	431.06	444.44
Feb-15	345	365.00	365.50	400.00	May-19	376	367.50	392.17	395.00
Mar-15	330	418.75	405.50	408.89	Jun-19	331	388.33	431.06	444.44
Apr-15	418	417.50	402.17	450.00	Jul-19	590	422.50	423.00	408.89
May-15	321	335.00	368.83	373.33	Aug-19	431	412.50	415.50	395.00
Jun-15	404	395.00	402.17	450.00	Sep-19	500	372.50	380.50	397.78
Jul-15	332	370.50	363.83	390.00	Oct-19	442	442.50	392.17	395.00
Aug-15	301	422.50	423.00	408.89	Nov-19	440	425.00	403.00	397.78
Sep-15	397	440.00	422.17	400.00	Dec-19	524	389.17	380.50	397.78
Oct-15	385	382.50	428.36	390.00	Jan-20	441	442.50	445.50	390.00
Nov-15	393	412.50	405.50	444.44	Feb-20	413	425.00	403.00	397.78
Dec-15	542	426.07	428.36	390.00	Mar-20	440	437.50	368.83	373.33
Jan-16	338	337.50	335.50	455.00	Apr-20	217	389.17	380.50	397.78
Feb-16	372	372.50	423.00	408.89	May-20	332	332.50	335.50	340.00
Mar-16	505	512.50	431.06	444.44					

The Forecasting Accuracy Validation

In this section, we compare a performance of proposed method with different length of interval as mentioned in the previous section. As we all know, the aim of forecasting is to be accurate as possible. For this purpose, we consider a performance measure providing forecasting error, which is the difference between the actual data case and its forecasted data case value. The Mean Square Error (MSE) and Root Mean Square Error (RMSE) are used to measure the forecasting accuracy:

$$MSE = \frac{\sum(Forecast - Actual)^2}{n} \quad (7)$$

$$RMSE = \sqrt{\frac{\sum(Forecast - Actual)^2}{n}} \quad (8)$$

TABLE 8. Comparison on the value of The Mean Square Error (MSE) and Root Mean Square Error (RMSE)

Average Based Interval	Mean Square Error (MSE)	Root Mean Square Error (RMSE)
Size 5	2786.16	52.78
Size 10	3871.43	62.22
Size 20	4533.55	67.33

Based on the MSE and RMSE values shown in Table 8, we may conclude that the smallest value of MSE contribute to more accurate forecasting values. We proposed that, the length of intervals with size 5 is better than size 10 and size 20.

ACKNOWLEDGMENTS

The author would like to thank to main supervisor, Dr. Suzelawati Zenian and co supervisor for their continuous support in this research. Besides, million thanks to Mdm. Nadirah Sulaiman from Clinical Research Centre (CRC), HQE as well as Dr. Roddy Teo from Communicable Disease Unit, Sabah State Health Department for their help throughout this research.

REFERENCES

Chen, S. M. (1996), Fuzzy Forecasting with DNA Computing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS*., 324-336.

Chen, S. M. (2002), Forecasting Enrollments Based on High Order Fuzzy Time Series. *Cybernetics and Systems*., **(33)**: 1-16.

Huarng, K. (2001), Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. *Fuzzy Sets and Systems*., **3(123)**: 387–394.

Huarng, K. and Yu, T. F., (2006), The Application of Neural Networks to Forecast Fuzzy Time Series. *Physica A: Statistical Mechanics and Its Applications*., **363(2)**: 481-491.

Li, S. T. and Chen, Y. P. (2004), Natural Partition-Based Forecasting Model for Fuzzy Time Series. *Fuzzy-IEEE, Budapest*., 1355-1359.

Liu, H. T. (2007), An Improved Fuzzy Time Series Forecasting Method Using Trapezoidal Fuzzy Numbers. *Fuzzy Optimisation Decision Making*., **(6)**: 63-80.

Lee, T. S., Chiu, C. C. and Lin, F.C. (2001), Prediction of The Unemployment Rate Using Fuzzy Time Series with Box-Jenkins Methodology. *International Journal of Fuzzy Systems*., **3(4)**: 577-585.

Lee, H. S. & Chou, M. T. (2007), Fuzzy Forecasting on Fuzzy Time Series. *International Journal of Computer Mathematics*., **81(7**): 781-789.

Ozge, C. Y., Eren, B., Erol, E. and Ufuk, Y. (2020), A New Intuitionistic Fuzzy Functions Approach Based on Hesitation Margin for Time-Series Prediction. *Soft Computing*., **(24)**: 8211-8222.

Ramli, N., M. S. M. A. and Mohamad, D. (2017), Fuzzy Time Series Forecasting Model With Natural Partitioning Length Approach for Predicting the Unemployment Rate Under Different Degree Of Confidence. *AIP Conference Proceedings*.

Song, Q. and Chissom, B. S. (1993a), Fuzzy Forecasting Enrollments with Fuzzy Time Series-Part 1. *Fuzzy Sets and Systems*., (54): 1-9.

Song, Q. and Chissom, B. S. (1993b), Fuzzy Time Series and Its Models. *Fuzzy Sets and Systems*., **3(54)**: 269-277.

Susilo, H., Zaenurrohman, and Titi, U. (2022), Average Based-FTS Markov Chain with Modifications to the Frequency Density Partition to Predict COVID-19 in Central Java. *Jurnal Matematika Murni dan Aplikasi*., **7(2)**: 231-239.

Tsai, C. C. and Wu, S. J. (1999), A Study for Second Order Modelling of Fuzzy Time Series. *Proceedings of the IEEE International Conference on Fuzzy Systems*., **(2)**: 719-725.

Zadeh, L. A. (1965), Fuzzy Sets. *Information and Control*., **(8)**: 338-353.

Zadeh, L.A. (1971), Similarity Relations and Fuzzy Orderings. *Information Science*., **(3)**: 177-200.

Zadeh, L.A. (1973), Outline of A New Approach to The Analysis of Complex System and Decision Process. *IEEE Trans. System and Cybernet*., **(3)**: 28-44.

Zadeh, L.A. (1975), The Concept of Linguistic Variable and Its Application to Approximate Reasoning, Part 1-3. *Information Science*., **(8)**: 199-249; 301-357; 43-80.

Zenian, S., Ahmad, T., and Idris, A. (2020), Advanced Fuzzy Set: An Application to Flat Electroencephalography Image. *Computational Science and Technology*. *Lecture Notes in Electrical Engineering*., **(603)**: 649–657.