Implementation of Revised Heuristic Knowledge in Average-based Interval for Fuzzy Time Series Forecasting of Tuberculosis Cases in Sabah

Suriana Lasaraiya[1], Suzelawati Zenian[2], Risman Mat Hasim[3], Azmirul Ashaari[4]
Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia[1]
Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia[1, 2]
Dept. of Mathematics and Statistics-Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia[3]
Azman Hashim International Business School, Universiti of Technology Malaysia, Johor Bahru, Johor, Malaysia[4]

***Abstract*—Fuzzy time series forecasting is one method used to forecast in certain reality problems. The research on fuzzy time series forecasting has been increased due to its capability in dealing with vagueness and uncertainty. In this paper, we are dealing with implementation of revised heuristic knowledge to basic average-based interval and showing that these models forecast better than the basic one. We suggest three different lengths of interval, size 5, size 10 and size 20 to be used in comparing these models of average-based interval, average-based interval with implementation of heuristic knowledge and, average-based interval with implementation of revised heuristic knowledge. These models applied to forecast the number of tuberculosis cases reported monthly in Sabah starting from January 2012 until December 2019. A few numerical examples are shown as well. The performances of evaluations are shown by comparison on the values obtained by Mean Square error (MSE) and Root Mean Square Error (RMSE).**

Keywords—Fuzzy time series; forecasting; length of interval; average-based interval; heuristic knowledge

I. Introduction

Time series analysis, at the present time has become a very important research object since it has been utilized in large number of fields such as finance, health care, education and environmental. In all these fields, obtaining an accurate picture of the future is an essential requirement and this can only be possible with an appropriate and competent prediction tool. Since the main concern of the time series analysis is to predict the future, it can be considered as an inference problem [1]. We identify that there are three main kinds of inference system (IS) which are utilized in time-series analysis as a prediction tool, i.e Statistical Inference System (SIS), Fuzzy Inference System (FIS) and computational inference system (CIS).

Fuzzy time series (FTS) methods are one of the prediction tool groups based on fuzzy set theory designed for time-series and they take into consideration the dependency structure of time series. In FTS model, time series with crisp observations, by using various fuzzification techniques, is transformed into time series with fuzzy observations called fuzzy time series.

The concept of linguistics variables and their application reasoning was initially introduced by Zadeh [2], [3], [4], [5]. Song and Chissom [6], [7] were the first researchers who implemented and developed models to forecast in fuzzy time series. Their model is initially used to forecast the number of students enrolled in the University of Alabama from the year 1971 until year 1992. At the same time, they forecast using linear regression technique and compared the results obtained from time series technique. However, their methods require a large amount of computational time since their proposed methods is based on complex matrix operations. Due to limitation on time constraints, Chens [8] simplified the Song and Chissom methods by using simpler arithmetic operations and proved that his methods forecast better than the previous. Chens [9] continues his research, applying the same methods in [8], but he is focusing on high order fuzzy time series, which is more than one fuzzy logical relationship.

Huarng [10] has been concerned on dealing with the suitable length of interval. A key point in choosing effective length of intervals is that they should not be too large as there will be no fluctuations in the fuzzy time series or too small as the meaning of fuzzy time series will be diminished. In his research, he proposed distribution- and average-based length in determining the effective length of intervals and proved that average-based length increases forecast accuracy compared to distribution-based length. Meanwhile, Li and Chen [11] proposed a concept of 3-4-5 rules of natural partitions to determine the length of intervals to be used, in their forecasting models.

In 2005, Huarng and Yu [12] proposed a type-2 fuzzy time series model for stock index forecasting. They proved that type-2 model is much better for smooth defuzzification process and forecast consistently. Similar work on fuzzy time series can be found in Lee, et al. (2001) [13] and Tsai and Wu (1999) [14]. The other researchers such as Liu (2007) [15], Ozge, et al. (2020) [16] and Lee and Chou (2007) [17] does not specify on how they find the suitable the length of interval in their proposed methods. More research on defining the suitable interval also found in Chen (2014) [18], Wang et.al (2015) [19] as well as Lu et al (2015) [20] in their proposed model.

Ramli and Mohamad (2017) [21] applied the models from [11] in their forecasting models. They used 3-4-5 natural partition rules to find the best length of intervals in their models in forecasting the unemployment rate under different degrees of confidence. The study [22] shows a comparison of the fuzzy time series methods of Chen, Cheng and Markov Chain model in predicting rainfall in Medan. In their study, Chen's model gives MAPE=8.002%, Markov chain's model gives MAPE=30.12% and Cheng's model gives MAPE=34.5%. According to [23] a model has a very good performance if the MAPE value is below 10% and has a good performance if the MAPE value is between 10% and 20%. Thus, Chen's model forecasting performances was very Good.

Norhayati et al. (2019) [24] proposed that fuzzy time series method is way more accurate compared to geometric brownian motion in forecasting stock prices in Bursa Malaysia. They apply average-based length focusing on trapezoidal membership members in their model Meanwhile, Susilo et al. (2022) [25] applied average-based fuzzy time series markov chain with some modification on partitioning using frequency density in predicting Covid-19 in Central Java, the ideal interval. The average based FTS Markov Chain approach with adjustments to the frequency density partition achieves an accuracy rate of 89.3%, according to the findings of this study. Lasaraiya et. al. (2022) [27] shows that determination of the length of interval by average-based interval gives nearest forecasting value to the actual values compared to forecasting value obtained from determination of length of interval by natural partitions as it shows the smallest percentage error.

This paper will merge Chen's model which is more on determination of suitable length of interval and Huarng's model which is more on choosing the suitable fuzzy logical relationship group (FLRG) in the model. Then, the comparison between the results of Chen's model with basic average-based interval, Huarng's model with average-based interval with implementation of heuristic knowledge and the proposed average-based interval model with implementation of revised heuristic knowledge will be shown. Section II will briefly explain the fuzzy time series background. Section III will be on the general step in the proposed methods. The Section IV presents results and discussion followed by future work and conclusions.

II. Fuzzy Time Series

Fuzzy time series concept was proposed by Song and Chissom [6], [7]. It can be used as prediction tools in a real-life problem or cases where historical data are formed in linguistic value. This means that the data in fuzzy time series is linguistic data while the actual data is real number. The definition of fuzzy time series is briefly reviewed below.

Let U be the universe of discourse, where $U = u_1, u_2, \dots, u_n$. A fuzzy set of A_i of U is defined by

$$A_i = \frac{f_{A_i}(u_1)}{u_1} + \frac{f_{A_i}(u_2)}{u_2} + \cdots + \frac{f_{A_i}(u_n)}{u_n} \qquad (1)$$

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

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

According to Definition 1, $F(t)$ can be understood as a linguistics variable and $f_i(t)(i = 1,2, \dots)$ as the possible linguistics values of $F(t)$.

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

$$F(t-1) \rightarrow F(t) \qquad (2)$$

Definition 3: Suppose $F(t-1) = A_i$ and $F(t) = A_j$, a fuzzy logical relationship (FLR) is defined as $A_i \rightarrow A_j$, 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 [8], the repeated fuzzy sets will be removed in the FLRG. For example,

$$\begin{matrix} A_i \rightarrow A_{j1} \\ A_i \rightarrow A_{j2} \\ \dots \end{matrix} \quad \rightarrow \quad A_i \rightarrow A_{j1}, A_{j2}, \dots \qquad (3)$$

Heuristic knowledge is introduced by [26] and this is used to guide the selection of proper fuzzy set for fuzzy logical relationship group (FLRG) before forecasting step. The definition is as shown below.

Definition 5: In the heuristic models, heuristic functions take fuzzy logical relationship groups and relevant variables as parameters. From these fuzzy logical relationship groups, the heuristic functions use the variables to select proper fuzzy set to establish heuristic fuzzy logical relationship groups.

All fuzzy sets $A_1, A_2, \dots, A_n$ are well ordered. In other word, $A_k \geq A_l$ if $k > l$. Suppose $F(t-1) = A_j$ and the FLRG for A_j is given by $A_j \rightarrow A_q, A_r, \dots$. Proper fuzzy sets, $A_{p1}, A_{P2}, \dots, A_{pk}$, can be selected by heuristic function $h()$,

$$h(x_1, x_2, \dots; A_q, A_r, \dots) = A_{p1}, A_{P2}, \dots, A_{pk} \qquad (4)$$

where $x_1, x_2, \dots$ are heuristic variables; $A_{p1}, A_{P2}, \dots, A_{pk}$ are selected from $A_q, A_r, \dots$, by the heuristics function. A heuristic fuzzy logical relationship group is obtained below:

$$A_j \rightarrow A_{p1}, A_{P2}, \dots, A_{pk} \qquad (5)$$

The heuristic fuzzy logical relationship group is then used to forecast $F(t)$.

III. The Proposed Method

Average-based intervals and Average-based intervals with implementation of heuristic knowledge were introduced a few years ago in forecasting the number of students' enrollments in University of Alabama. These models were hugely applied in forecasting in other real-life problems. In this paper, we try to show the Average-based intervals with implementation of revised heuristic knowledge forecast better than the other two models. There are five steps involved in this proposed method. The data as shown in Table I, was obtained from the Hospital of Queen Elizabeth II.

TABLE I. THE NUMBER OF TUBERCULOSIS CASES REPORTED MONTHLY IN SABAH FROM JANUARY 2012 UNTIL DECEMBER 2019

Month/ Year	***Cases***	***Month/ Year***	***Cases***	***Month/ Year***	***Cases***	***Month/ Year***	***Cases***
Jan 2012	199	Jan 2014	394	Jan 2016	338	Jan 2018	392
Feb 2012	487	Feb 2014	353	Feb 2016	372	Feb 2018	393
Mar 2012	356	Mar 2014	341	Mar 2016	505	Mar 2018	371
Apr 2012	364	Apr 2014	495	Apr 2016	394	Apr 2018	482
May 2012	419	May 2014	356	May 2016	401	May 2018	307
Jun 2012	349	Jun 2014	395	Jun 2016	401	Jun 2018	381
July 2012	341	July 2014	389	July 2016	305	July 2018	515
Aug 2012	422	Aug 2014	406	Aug 2016	482	Aug 2018	335
Sep 2012	365	Sep 2014	332	Sep 2016	410	Sep 2018	372
Oct 2012	380	Oct 2014	428	Oct 2016	311	Oct 2018	573
Nov 2012	351	Nov 2014	376	Nov 2016	601	Nov 2018	351
Dec 2012	353	Dec 2014	482	Dec 2016	433	Dec 2018	572
Jan 2013	380	Jan 2015	296	Jan 2017	299	Jan 2019	382
Feb 2013	385	Feb 2015	345	Feb 2017	390	Feb 2019	396
Mar 2013	379	Mar 2015	330	Mar 2017	446	Mar 2019	375
Apr 2013	376	Apr 2015	418	Apr 2017	366	Apr 2019	495
May 2013	402	May 2015	321	May 2017	437	May 2019	376
Jun 2013	342	Jun 2015	404	Jun 2017	428	Jun 2019	331
July 2013	432	July 2015	332	July 2017	404	July 2019	590
Aug 2013	324	Aug 2015	301	Aug 2017	472	Aug 2019	431
Sep 2013	388	Sep 2015	397	Sep 2017	347	Sep 2019	500
Oct 2013	389	Oct 2015	385	Oct 2017	394	Oct 2019	442
Nov 2013	359	Nov 2015	393	Nov 2017	532	Nov 2019	440
Dec 2013	370	Dec 2015	542	Dec 2017	590	Dec 2019	524

[a] Data obtained from Queen Elizabeth II Hospital, Kota Kinabalu, Sabah.

Step 1: Define the universe of discourse, $U = [U_{min} - D_1, U_{max} + D_2]$, where U_{min} and U_{max} are minimum and maximum values in raw data, D_1, D_2 are two real numbers. U will be divided into n equal length of intervals $u_1, u_2, \dots, u_n$. We applied Chen's method [8] with some modifications in finding the length of intervals. The algorithm in average based interval by [10] to set the suitable length of interval is according to Table II.

TABLE II. BASIS MAPPING TABLE

Range	**Basis**
0.1 – 1.0	0.1
1.1 - 10	1
11 - 100	10
101 - 1000	100

[b] Basis Mapping Table by [10].

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 $u_1, u_2, \dots, u_n$.

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

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

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

$$\dots$$

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

Step 3: Defined fuzzy logical relationship FLR. Repeated FLR will be removed.

Step 4: Group FLR in fuzzy logical relationship group FLRG.

Step 5: Calculated the forecasted output. There are 3 rules involved in Chen's Rule.

- 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.

IV. Results and Discussion

According to Section III, there are 5 steps counted. The data used in this study is the number of tuberculosis cases reported monthly in Sabah, period of January 2012 to December 2019 as you can see from Fig. 1. The time series graphic as in Fig. 1 demonstrates that there is no discernible pattern in the number of tuberculosis cases reported. This can be caused by any unexpected factors, such as gender, family background, knowledge, etc., and did not become observation in this study.

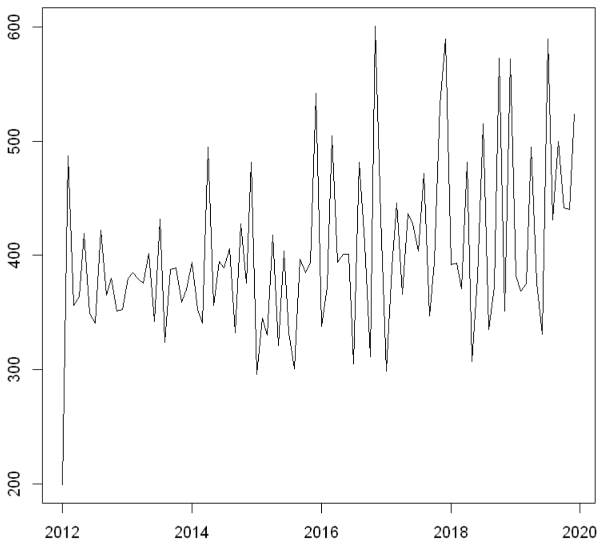

Fig. 1. Plot actual number of tuberculosis cases reported in sabah period January 2012 – December 2019.

A. Length of Interval

The first modification was done in determining the length of intervals. According to the data in Table I, $U_{min} = 199$ and $U_{max} = 601$ respectively. We choose $D_1 = 9$ and $D_2 = 9$ respectively. Thus, $U = [199 - 9\,, 601 + 9] = 190{,}610]$. The modifications (b) and (c) are as follows:

- Based on Table II, the length of interval obtained is 10.
- The length of interval is then divided by 2, giving the length of interval is 5.
- The length of interval is then multiplied by 2, giving the length of interval is 20.

Then, the lengths of intervals considered are size 5, size 10 and size 20, and hence list of intervals is shown below. All three sizes applied to the average-based interval model, average-based interval model with implementation of heuristic knowledge and average-based interval model with implementation of revised heuristic knowledge.

TABLE III. The Length of Intervals and List of Intervals

Length of intervals	***Number of intervals***	***List of intervals***
Size 5	84	$u_1 = [190 - 195], u_2 = [195 - 200], \ldots, u_{84} = [605 - 610]$
Size 10	42	$u_1 = [190 - 200], u_2 = [200 - 210], \ldots, u_{42} = [600 - 610]$
Size 20	21	$u_1 = [190 - 210], u_2 = [210 - 230], \ldots, u_{21} = [590 - 610]$

c. Jupyter Notebook tools under R kernel

B. Fuzzy Logical Relationship (FLR)

Next, is to find FLR. The number of tuberculosis cases reported monthly is to be defined according to intervals as shown in Table III and IV. The same step is repeated for all three models: average-based interval, average-based interval with implementation of heuristic knowledge and proposed average-based interval with implementation of revised heuristic knowledge.

TABLE IV. The Fuzzy Sets, A_i

Length of intervals	***Number of intervals***	***List of intervals***
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	$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ $A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ $A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \cdots + 0/u_{n-1} + 0/u_{21}$ … $A_{21} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \cdots + 0.5/u_{n-1} + 1/u_{21}$

d. Jupyter Notebook tools under R kernel

C. Fuzzy Logical Relationship Group (FLRG)

Group of FLR is defined according to same fuzzy set in left hand side of the FLR. As for example, numerical FLR in the model for size 5 is shown. FLRG for fuzzy set A_{33} is chosen.

1) Average-based interval: FLRG will be grouped according to Definition 3. Refer Fig. 2.

FLRG for item (12) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (13) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (84) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

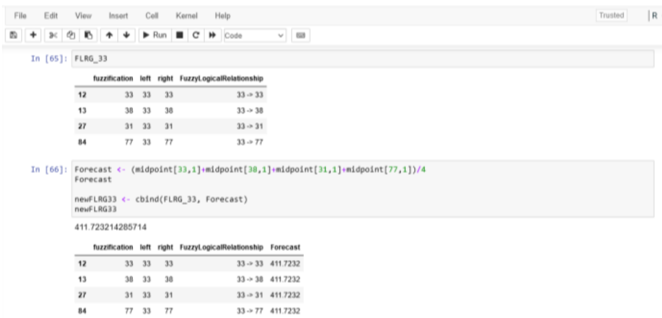

Fig. 2. The FLRG of A_{33} in Average-Based Intervals for Size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

2) Average-based intervals with heuristic knowledge according to Definition 5. Let the month $t-1$ be $F(t-1)$ and the FLRG for $F(t-1)$ be $A_j \rightarrow A_q, A_r, \ldots$ This study assumes that there is heuristic knowledge showing the increment or decrement in the number of cases reported. The heuristic function is set as $h(x; A_q, A_r, \ldots)$ where x is an indicator for number of cases forecasting. Fig. 3 shows the example on fuzzy sets of A_{33}. The heuristic FLRG is shown below.

FLRG for item (12) is $A_{33} \rightarrow A_{31}$

FLRG for item (13) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}$

FLRG for item (84) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

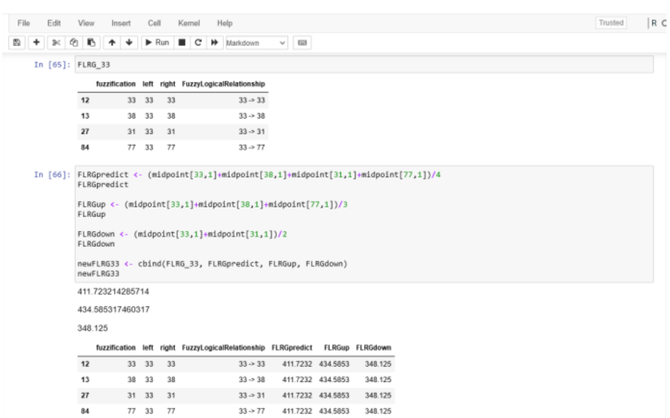

Fig. 3. The FLRG of A_{33} in average-based intervals with heuristic knowledge for size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

3) Average-based intervals with revised heuristic knowledge. The modification is that, for the FLR who have the same value on left and right side, the whole FLRG will be chosen as the heuristic FLRG instead of looking back at the raw data to look whether the value increase or decrease. Fig. 4 shows the example on fuzzy sets of A_{33}. The heuristic FLRG is shown below.

FLRG for item (12) is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$

FLRG for item (13) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

FLRG for item (27) is $A_{33} \rightarrow A_{31}$

FLRG for item (84) is $A_{33} \rightarrow A_{33}, A_{38}, A_{77}$

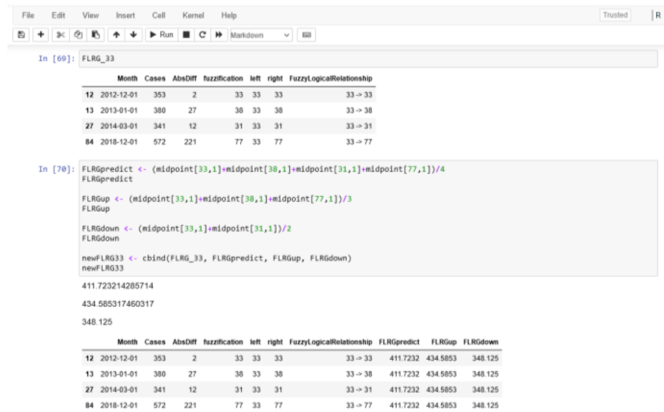

Fig. 4. The FLRG of A_{33} in average-based intervals with revised heuristic knowledge for size 5 is $A_{33} \rightarrow A_{31}, A_{33}, A_{38}, A_{77}$.

D. Forecasting

The output will be defined according to Step 5. The forecast number of tuberculosis cases with respect to length of intervals size 5, size 10 and size 20 for the three models; average based intervals, average-based intervals with implementation of heuristic knowledge and average-based intervals with implementation of revised heuristic knowledge is shown in Table V, Table VI, and Table VII.

TABLE V. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL

Month/ Year	***Cases***	***Forecast cases (according to respected size of intervals)***		
		Size 5	***Size 10***	***Size 20***
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	358.1012	4340.6429	350.6190
Apr 2012	364	374.7282	402.1627	417.6020
May 2012	419	398.0060	405.4881	417.6020
Jun 2012	349	335.6548	368.9087	373.8968
July 2012	341	368.0774	403.8254	409.3677
…	…	…	…	…
July 2019	590	422.9464	422.9464	409.3677
Aug 2019	431	412.9702	415.4643	4395.5119
Sep 2019	500	373.0655	380.5476	385.5357
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	385.5357
Dec 2019	524	389.6925	380.5476	385.5357

e. Jupyter notebook tools under R kernel

TABLE VI. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL WITH IMPLEMENTATION OF HEURISTIC KNOWLEDGE

Month/ Year	*Cases*	*Forecast cases (according to respected size of intervals)*		
		Size 5	*Size 10*	*Size 20*
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	385.1012	340.6429	350.6190
Apr 2012	364	374.7282	413.4690	430.4286
May 2012	419	398.0060	405.4881	430.4286
Jun 2012	349	335.6548	335.6548	360.5952
July 2012	341	343.1369	335.6548	320.6905
…	…	…	…	…
July 2019	590	462.8512	462.0198	437.5544
Aug 2019	431	412.9702	4415.4643	395.5119
Sep 2019	500	497.7679	510.2381	440.4048
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	360.5952
Dec 2019	524	522.7083	510.2381	440.4048

f. Jupyter notebook tools under R kernel

TABLE VII. THE FORECAST NUMBER OF CASES FOR AVERAGE-BASED INTERVAL WITH IMPLEMENTATION OF REVISED HEURISTIC KNOWLEDGE

Month/ Year	*Cases*	*Forecast cases (according to respected size of intervals)*		
		Size 5	*Size 10*	*Size 20*
Feb 2012	487	487.7917	485.2976	480.3095
Mar 2012	356	358.1012	340.6429	350.6190
Apr 2012	364	374.7282	413.4690	417.6020
May 2012	419	398.0060	405.4881	430.4286
Jun 2012	349	335.6548	335.6548	360.5952
July 2012	341	343.1369	403.8254	409.3677
…	…	…	…	…
July 2019	590	462.8512	462.0198	437.5544
Aug 2019	431	412.9702	415.4643	395.5119
Sep 2019	500	497.7679	510.2381	486.9603
Oct 2019	442	442.8988	392.1865	395.5119
Nov 2019	440	425.4405	405.4881	409.0510
Dec 2019	524	522.7083	510.2381	486.9603

g. Jupyter notebook tools under R kernel

E. Accuracy Validation

The performance for the proposed method is compared based on their value of Mean Square Error (MSE) and Root Mean Square Error (RMSE). The smaller the MSE value, the better the forecast. Meanwhile, RMSE is known to be a good measurement of how accurate the model is in prediction. Lower values of RMSE indicate better fit. Based on Table VIII, the average-based intervals with implementation of revised heuristic knowledge for size 5 give the smallest $MSE = 1309.11$ and $RMSE = 36.18168$ respectively compared to average-based interval and average-based interval with implementation of heuristic knowledge. Based on this MSE and RMSE value, it is concluded that the average-based intervals with implementation of revised heuristic knowledge model increase the accuracy of the forecasting value.

TABLE VIII. THE COMPARISON ON MSE AND RMSE OF EACH SIZE

Models	*Length of intervals*	*MSE*	*RMSE*
Average-based interval	Size 5	2618.065	51.15726
	Size 10	3337.299	57.76936
	Size 20	4392.890	66.2788
Average-based intervals with implementation of heuristic knowledge	Size 5	1436.139	37.89642
	Size 10	1926.855	43.89595
	Size 20	2624.269	51.22762
Average-based intervals with implementation of revised heuristic knowledge	Size 5	1309.114	36.18168
	Size 10	1887.522	43.44562
	Size 20	2289.208	47.84564

h. Jupyter Notebook tools under R kernel

V. FUTURE WORK

Overall, this paper shows only the results on training phase of the comparing model and length of interval chosen size 5, size 10 and size 20. Training phase is done for the period of January 2012 until December 2019. Testing phase on the next year 2020 (January-December) can be done as a future work for length of size 5, 10 and 20, limited to list of FLRG obtained from the training phase. The forecasting results will be compared with the actual number of tuberculosis cases.

ACKNOWLEDGMENT

The author would like to thank the 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), Hospital Queen Elizabeth (HQE) as well as Dr. Roddy Teo from Communicable Disease Unit, Sabah State Health Department for their help throughout this research.

REFERENCES

[1] Ozge, C. Y., Eren, B., Erol, E. and Ufuk, Y. "A New Intuitionistic Fuzzy Functions Approach Based On Hesitation Margin For Time-Series Prediction". Soft Computing 2020, 24 pp. 8211-8222.

[2] Zadeh, L.A. Fuzzy Sets, *Information and Control* 1965 **8**, pp. 338-353.

[3] Zadeh, L.A. "The concept of Linguistic variable and its application to approximate reasoning, Part 1-3" Information Science 1975 **8**, pp. 199-249, 301-357; 43-80.

[4] Zadeh, L.A. "Outline of a new approach to the analysis of complex system and decision process", in IEEE Trans. System Man Cybernet 1973 **3**, pp. 28-44.

[5] Zadeh, L.A. *"Similarity relations and fuzzy orderings",* Information Science 1971 **3**, pp. 177-200.

[6] Song, Q. and Chissom, B. S. Fuzzy forecasting enrollments with fuzzy time series-Part 1, *Fuzzy Sets and Systems* **54**, 1993a, pp. 1-9.

[7] Song, Q. and Chissom, B. S. Fuzzy time series and its models. *Fuzzy Sets and Systems* **3(54)**, 1993b, pp. 269–277.

[8] Chen, S. M., "Forecasting enrollments based on fuzzy time series", *Fuzzy Sets and Systems* **81**, 1996, pp. 311-318.

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

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

[11] Li, S. T. and Chen, Y. P. "Natural Partitioning-based Forecasting Model for Fuzzy Time Series", 2004 Budapest, Hungary.

[12] Huarng, K. and Yu, T. F, *"The Application of neural networks to forecast fuzzy time series"* Physica A: Statistical Mechanics and Its Applications 2006, Vol. 363, Issue 2, pp 481-491.

[13] Lee, T. S., Chiu, C. C. and Lin, F.C., "*Prediction of the unemployment rate using fuzzy time series with Box-Jenkins methodology*" International Journal of Fuzzy Systems 2001, Vol.3. No. 4. pp. 577-585.

[14] Tsai, C. C. and Wu, S. J., *"A study for second order modelling of fuzzy time series"* Proceedings of the IEEE International Conference on Fuzzy Systems 1999, Vol. 2, pp. 719-725.

[15] Liu, H. T. An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimisation Decision Making* 2007, **6** pp. 63-80.

[16] Ozge, C. Y., Eren, B., Erol, E. and Ufuk, Y. "A new intuitionistic fuzzy functions approach based on hesitation margin for time-series prediction". Soft Computing 2020, **24** pp. 8211-8222.

[17] Lee, H. S. & Chou, M. T. "*Fuzzy forecasting on fuzzy time series"* International Journal of Computer Mathematics 2007 **81:7**, pp. 781-789.

[18] Chen, M.Y., "A high order fuzzy time series forecasting model for internet stock trading", Future Generation Computer System 37:461-467.

[19] Wang, W. Liu, X., "Fuzzy forecasting based on automatic clustering and axiomatic fuzzy set classification", Inf. Sci. (Ny) 294:78-94

[20] Lu. W., Chen, X. Pedrycz, W., " Using interval information granules to improve forecasting in fuzzy time series", Int. J. Approx Reason 57:1-18

[21] Ramli, N., M. S. M. A. and Mohamad, D. "Fuzzy time series forecasting model with natural partitioning length approach for predicting the unemployment rate under different degree of confidence" 2017, *AIP conference proceeding*.

[22] Arnita, N. Afnisah, and F. Marpaung. *"A Comparison of the Fuzzy Rime Series Methods of Chen, Cheng and Markov Chain in Predicting Rainfall in Medan."* Journal of Physics.: Conference. Series. 1462 (2020) 012044

[23] Raharja, A., Anggraeni, W., dan Vinarti, R. A., 2010. Penerapan Metode Exponential Smoothing untuk Peramalan Penggunaan Waktu Telepon di PT. Telkomsel DIVRE3 Surabaya. Jurnal Sistem Informasi. Surabaya: Institut Teknologi Sepuluh Nopember.

[24] Norhayati, S., Nur Ezzati, D.M.R., Rohana, A. and Nur Fatihah, F., *"Fuzzy Time Series and Geometric Brownian Motion in Forecasting Stock Prices in Bursa Malaysia"*, 2019, Jurnal Intelek UITM Perlis, 14:2, DOI: http://10.24191/ji.v14i2.241

[25] Susilo, H., Zaenurrohman, Titi, U. S., "Average Based-FTS Markov Chain with Modifications to the Frequency Density Partition to Predict Covid-19 in Central Java", CAUCHY –Jurnal Matematika Murni dan Aplikasi 2022, 7:2, 213-239

[26] Huarng, K., Heuristic models of fuzzy time series for forecasting. *Fuzzy Sets and Systems*., 2001, **123**: pp. 369-386.

[27] Lasaraiya, S., Gabda, D., Che Hussin, C. H. & Mandangan, A. (2022). " An effective length of intervals for forecasting in fuzzy time series". Journal of Islamic, Social, Economics and Development. (JISED), 7(47), 163-176. DOI: 10.55573/JISED.074717