

A Hybrid Fuzzy Time Series Forecasting Model with 4253HT Smoother

Nik Muhammad Farhan Hakim Nik Badrul Alam^{1*}, Nazirah Ramli¹, Adie Safian Ton Mohamed², Noor Izyan Mohamad Adnan¹

¹Mathematical Sciences Studies, College of Computing, Informatics and Media, Universiti Teknologi MARA (UiTM) Pahang Branch, Jengka Campus, Bandar Tun Abdul Razak Jengka, Pahang, Malaysia

²School of Mathematics, Actuarial and Quantitative Studies, Asia Pacific University of Technology and Innovation, Malaysia

* Corresponding author: farhanhakim@uitm.edu.my

Received: 28 September 2022; Accepted: 1 November 2022; Available online (in press): 14 November 2022

ABSTRACT

Forecasting time series data is crucial for predicting upcoming observations, especially in the market and business. Proper actions can be taken when there are some figures on future data, which are predicted based on the previous data. The fusion of fuzzy time series in forecasting has made forecasting using linguistic variables possible. However, the existence of extreme values in the time series data has led to inaccurate forecasting since the values are too large or too small. Hence, this paper proposes a hybrid fuzzy time series forecasting model with the 4253HT smoother to reduce the uncertainty of data. In this study, students' enrolment data at the University of Alabama are implemented to illustrate the proposed hybrid forecasting model. The results show that the proposed model improves the forecasting performance since the mean square, root mean square, and mean absolute errors have been reduced. In the future, the implementation of data smoothing using the 4253HT smoother can be used in other fuzzy time series and intuitionistic fuzzy time series forecasting models.

Keywords: Fuzzy time series, 4253HT smoother, students' enrolment, time series forecasting.

1 INTRODUCTION

In the business world, proper planning should be done in order to sustain sales, especially in this era of economic difficulty. Accordingly, predicting future observations based on the previous figures should be done in order to gain as much profit from the sales as possible. Various forecasting methods have been developed previously, such as ARIMA, SARIMA, and GARCH. However, traditional forecasting methods can only deal with crisp data.

In 1993, Song and Chissom [1] proposed the concept of fuzzy time series, making the forecasting of linguistic values possible. They illustrated the theory by implementing it in predicting students' enrolment at the University of Alabama [2-3]. Since then, the fuzzy time series has received attention from researchers all over the world. In fact, most of the existing forecasting models using fuzzy sets do not decompose the components of time series such as trend, seasonal, irregular, and cyclical.

Chen [4] simplified Song and Chissom's [2] model by using simpler rules for calculating the forecasted data instead of using the max-min composition operation. Lee and Chou [5] further used the supports of fuzzy numbers to define the universe of discourse for fuzzy time series forecasting. Furthermore, instead of using fuzzy sets, Liu [6] used trapezoidal fuzzy numbers to forecast the students' enrolment data. The fuzzy numbers provide more useful information instead of the single-point values, which are similar to the traditional forecasting methods [6].

Huang [7] used the average-based length for partitioning the universe of discourse to forecast the students' enrolment data at the University of Alabama. In addition, Tsaur and Kuo [8] employed the fuzzy time series forecasting model to predict the number of tourists in Taiwan. Meanwhile, Jilani and Burney [9] used the frequency-density-based method to partition the universe of discourse into several intervals to improve the forecasting of TAIEX data.

Notwithstanding the implications, the main problem of the time series data is that it is usually characterized by a high level of uncertainty and extreme values. Hence, Velleman and Hoaglin [10] used the 4253HT smoother to remove extreme values, which reduces the fluctuation of the time series data. The 4253HT smoother, as recognized by Jin and Xu [11], is among the best smoothing techniques. Azmi et al. [12] modified the 4253HT smoother to extract heavy noises.

Combining fuzzy time series with the smoothing technique, Cheng and Li [13] proposed a fuzzy smoothing method to enhance forecasting based on the hidden Markov model. Their proposed model managed to handle various uncertainties and is capable of solving the forecasting problem involving both the fuzzy time series and traditional time series in crisp values.

By adopting the strength of the smoothing, this paper aims to propose a hybrid fuzzy time series forecasting model with the 4253HT smoother. The data are first smoothed using the 4253HT smoother before conducting the fuzzy time series forecasting procedure. The purpose of smoothing is to remove extreme values from the time series data so as to reduce uncertainties.

2 HYBRID FUZZY TIME SERIES AND 4253HT SMOOTHER

The 4253HT smoother was proposed by Velleman and Hoaglin [10], in which the concept of moving median is used to remove extreme values. The first value is obtained by finding the 4-moving median and followed by its 2-moving median. Next, the 5-moving median is calculated, followed by the 3-moving median. The values are then smoothed through Hanning using several coefficients. Hanning, also known as running weighted average is to re-smooth the data by replacing each data with the average of its neighborhoods [10]. The natural of mean or average makes it effected by outliers so the Hanning process should be applied after removing away the outliers by one of the running smoothing techniques. Tukey's Hanning coefficients, {0.25,0.5,0.25} are used in this study [14]. In this section, the hybrid fuzzy time series forecasting model with the 4253HT smoother is proposed. Firstly, the data are smoothed using the 4253HT smoother, as follows:

Step 1: Find the 4-moving median and 2-moving median.

$$M_4(d_k) = \text{median}\{d_{k-2}, d_{k-1}, d_k, d_{k+1}\} \quad (1)$$

$$M_{42}(d_k) = \text{median}\{M_4(d_k), M_4(d_{k+1})\} \quad (2)$$

Step 2: Find the 5-moving median and 3-moving median.

$$M_{425}(d_k) = \text{median}\{M_{42}(d_{k-2}), M_{42}(d_{k-1}), M_{42}(d_k), M_{42}(d_{k+1}), M_{42}(d_{k+2})\} \quad (3)$$

$$M_{4253}(d_k) = \text{median}\{M_{425}(d_{k-1}), M_{425}(d_k), M_{425}(d_{k+1})\} \quad (4)$$

Step 3: Find the Hanning function using the coefficients {0.25, 0.5, 0.25}.

$$M_{4253H}(d_k) = 0.25M_{4253}(d_{k-1}) + 0.5M_{4253}(d_k) + 0.25M_{4253}(d_{k+1}) \quad (5)$$

Step 4: Smooth the residual and add it back to the smoothed value.

$$e_k = d_k - M_{4253H}(d_k) \quad (6)$$

$$M_{4253HT}(d_k) = M_{4253H}(d_k) + M_{4253H}(e_k) \quad (7)$$

Once we have obtained the smoothed data, we can subsequently perform the fuzzy time series forecasting on the data using the following steps:

Step 1: Define the universe of discourse using the formula

$$U = [d_{\min} - \underline{d}, d_{\max} + \bar{d}] \quad (8)$$

where d_{\min} and d_{\max} are the smallest and largest observations, respectively, while \underline{d} and \bar{d} are two properly chosen integers.

Step 2: Partition U into several intervals using a randomly chosen length. For each interval $[r, s]$, define a triangular fuzzy number (r, s, t) .

Step 3: Establish the fuzzy sets using the triangular fuzzy numbers defined by

$$\mu_F(x) = \begin{cases} \frac{x-r}{s-r} & , r < x \leq s \\ \frac{t-x}{t-s} & , s < x \leq t \\ 0 & , \text{elsewhere} \end{cases} \quad (9)$$

Step 4: The data are then fuzzified based on the membership values obtained in Step 3. The fuzzy set with a higher membership value is chosen to fuzzify the data.

Step 5: Form the fuzzy logical relationships (FLRs). In this case, the FLR that appears to occur more than one is counted only once [2].

Step 6: Defuzzify the fuzzy sets obtained in Step 3 using the following formula:

$$F_{def} = \frac{\sum \mu_F(X) \cdot X}{\sum \mu_F(X)} \tag{10}$$

Step 7: Calculate the forecasted data using the following rules:

1. If the FLR from year k to $k+1$ is $F_i \rightarrow F_j$, then the forecasted data for year $k+1$ is $(F_j)_{def}$.
2. If the FLR from year k to $k+1$ is $F_i \rightarrow F_{j_1}, \dots, F_{j_m}$, then the forecasted data for year $k+1$ is $\text{average}\{F_{j_1}, \dots, F_{j_m}\}$.
3. If the FLR from year k to $k+1$ is $F_i \rightarrow \phi$, then the forecasted data for year $k+1$ is $(F_i)_{def}$.

The steps in data smoothing and fuzzy time series forecasting are summarized in Figure. 1.

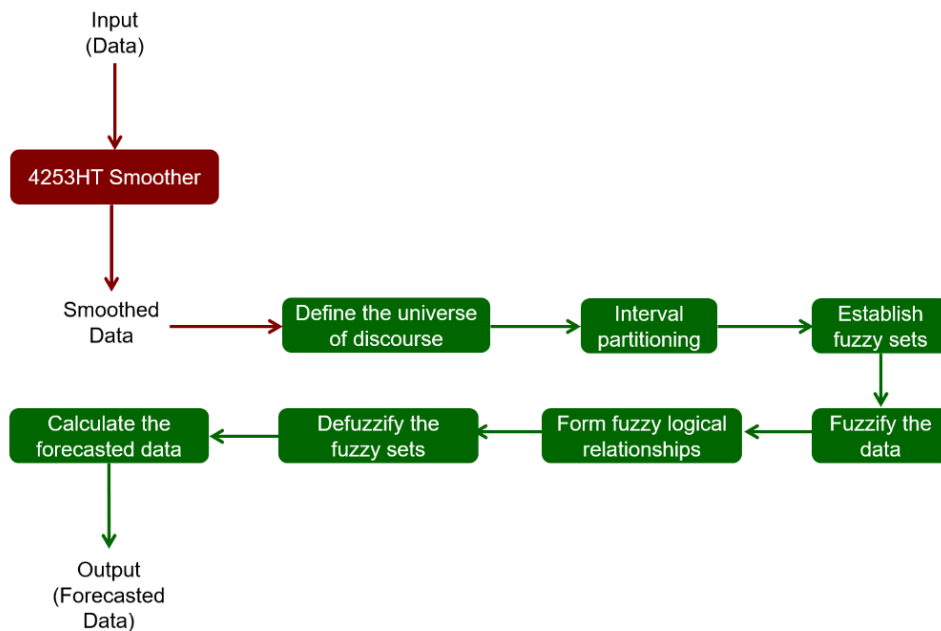


Figure 1 : Proposed hybrid fuzzy time series and 4253HT smoother

3 FORECASTING STUDENTS' ENROLMENT

In this section, we illustrate the proposed methodology presented in the previous section using the students' enrolment data at the University of Alabama obtained from [2]. This set of data has been extensively used by researchers in modelling fuzzy time series since 1993. Many time series forecasting models based on fuzzy sets, intuitionistic fuzzy sets, and intuitionistic fuzzy sets with de-fuzzification developed previously have used this set of data for illustration. Hence, the same set of data is used in this study to see the improvement throughout the proposed methodology.

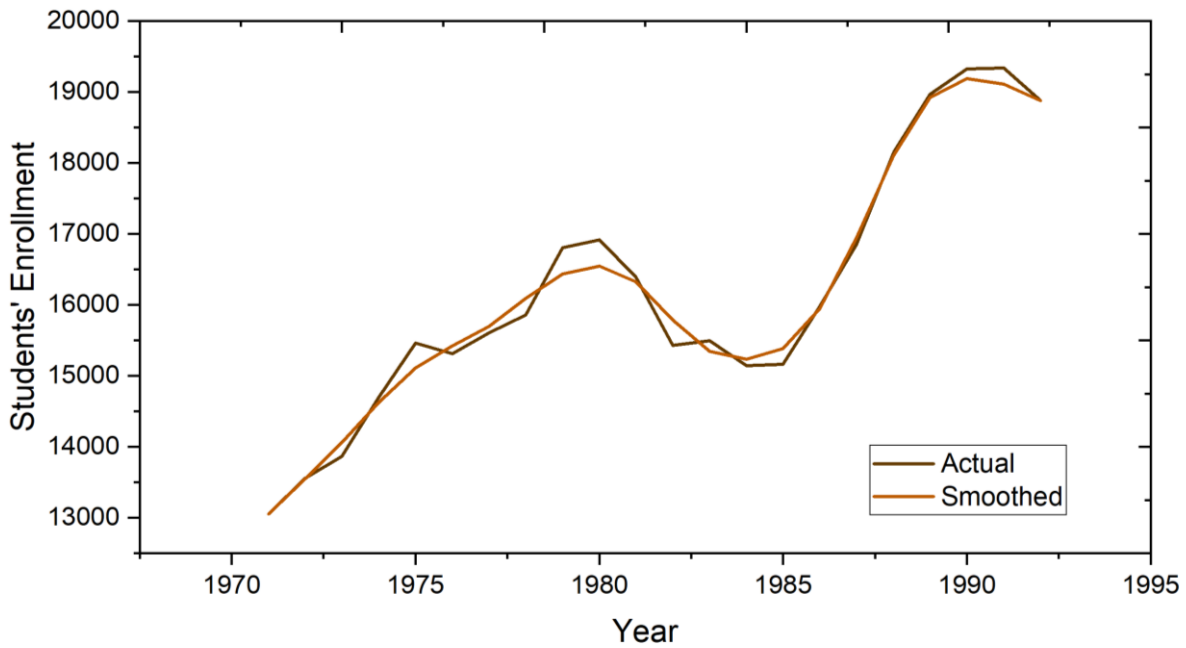


Figure 2 : Actual and smoothed students' enrolment

Firstly, the data are smoothed using the 4253HT smoother. The smoothed data are presented in Figure 2. Using the smoothed data, the fuzzy time series forecasting model is then implemented to forecast the data by converting the data into linguistic variables.

Step 1: The smallest and largest enrolments are 13055 and 19188, respectively. The universe of discourse is then defined as $[13000, 19200]$ using Equation (8).

Step 2: Using the interval length of 200, U is divided into 31 intervals as shown in Table 1.

Table 1 : Intervals with their corresponding triangular fuzzy numbers

Intervals	Fuzzy Numbers	Intervals	Fuzzy Numbers
[13000,13200]	(13000,13200,13400)	[16200,16400]	(16200,16400,16600)
[13200,13400]	(13200,13400,13600)	[16400,16600]	(16400,16600,16800)
[13400,13600]	(13400,13600,13800)	[16600,16800]	(16600,16800,17000)
[13600,13800]	(13600,13800,14000)	[16800,17000]	(16800,17000,17200)
[13800,14000]	(13800,14000,14200)	[17000,17200]	(17000,17200,17400)
[14000,14200]	(14000,14200,14400)	[17200,17400]	(17200,17400,17600)
[14200,14400]	(14200,14400,14600)	[17400,17600]	(17400,17600,17800)
[14400,14600]	(14400,14600,14800)	[17600,17800]	(17600,17800,18000)
[14600,14800]	(14600,14800,15000)	[17800,18000]	(17800,18000,18200)
[14800,15000]	(14800,15000,15200)	[18000,18200]	(18000,18200,18400)
[15000,15200]	(15000,15200,15400)	[18200,18400]	(18200,18400,18600)
[15200,15400]	(15200,15400,15600)	[18400,18600]	(18400,18600,18800)
[15400,15600]	(15400,15600,15800)	[18600,18800]	(18600,18800,19000)
[15600,15800]	(15600,15800,16000)	[18800,19000]	(18800,19000,19200)
[15800,16000]	(15800,16000,16200)	[19000,19200]	(19000,19200,19200)
[16000,16200]	(16000,16200,16400)		

Step 3: Using Equation (9), the fuzzy sets are formed from the historical data, as follows:

$$F_1 = 0.275 / 13055$$

$$F_2 = 0.232 / 13554$$

$$F_3 = 0.768 / 13554$$

⋮

$$F_{30} = 0.431 / 19114 + 0.062 / 19188$$

$$F_{31} = 0.569 / 19114 + 0.938 / 19188$$

Step 4: The historical data are fuzzified based on the membership values of the fuzzy sets. For example, observation 13554 has membership values of 0.232 and 0.768 in F_2 and F_3 , respectively; hence, it is fuzzified as F_3 . The rest of the observations are fuzzified analogously as presented in Table 2.

Table 2 : Fuzzification of students' enrolment

Enrolment	Fuzzy Set	Enrolment	Fuzzy Set	Enrolment	Fuzzy Set
13055	F_1	16435	F_{17}	16957	F_{20}
13554	F_3	16544	F_{18}	18108	F_{26}
14062	F_5	16329	F_{17}	18927	F_{30}
14625	F_8	15787	F_{14}	19188	F_{31}
15115	F_{11}	15348	F_{12}	19114	F_{31}
15421	F_{12}	15236	F_{11}	18876	F_{29}
15697	F_{13}	15384	F_{12}		
16094	F_{15}	15951	F_{15}		

Step 5: The fuzzy logical relationships (FLRs) are developed as follows: $F_1 \rightarrow F_3, F_3 \rightarrow F_5, F_5 \rightarrow F_8, \dots, F_{31} \rightarrow F_{29}, F_{29} \rightarrow \phi$. The FLRs are then grouped as shown in Table 3.

Table 3 : FLR groups

Group	FLR	Group	FLR
1	$F_1 \rightarrow F_3$	9	$F_{15} \rightarrow F_{17}, F_{20}$
2	$F_3 \rightarrow F_5$	10	$F_{17} \rightarrow F_{14}, F_{18}$
3	$F_5 \rightarrow F_8$	11	$F_{18} \rightarrow F_{17}$
4	$F_8 \rightarrow F_{11}$	12	$F_{20} \rightarrow F_{26}$
5	$F_{11} \rightarrow F_{12}$	13	$F_{26} \rightarrow F_{30}$
6	$F_{12} \rightarrow F_{11}, F_{13}, F_{15}$	14	$F_{30} \rightarrow F_{31}$
7	$F_{13} \rightarrow F_{15}$	15	$F_{31} \rightarrow F_{29}, F_{31}$
8	$F_{14} \rightarrow F_{12}$	16	$F_{29} \rightarrow \phi$

Step 6: Using Equation (10), the fuzzy sets are defuzzified and the values obtained are as follows:

$$(F_1)_{def} = 13055, (F_2)_{def} = 13554, (F_3)_{def} = 13554, \dots, (F_{30})_{def} = 18978, (F_{31})_{def} = 19160.$$

Step 7: Finally, the forecasted data are calculated using the proposed rules.

4 RESULTS AND DISCUSSION

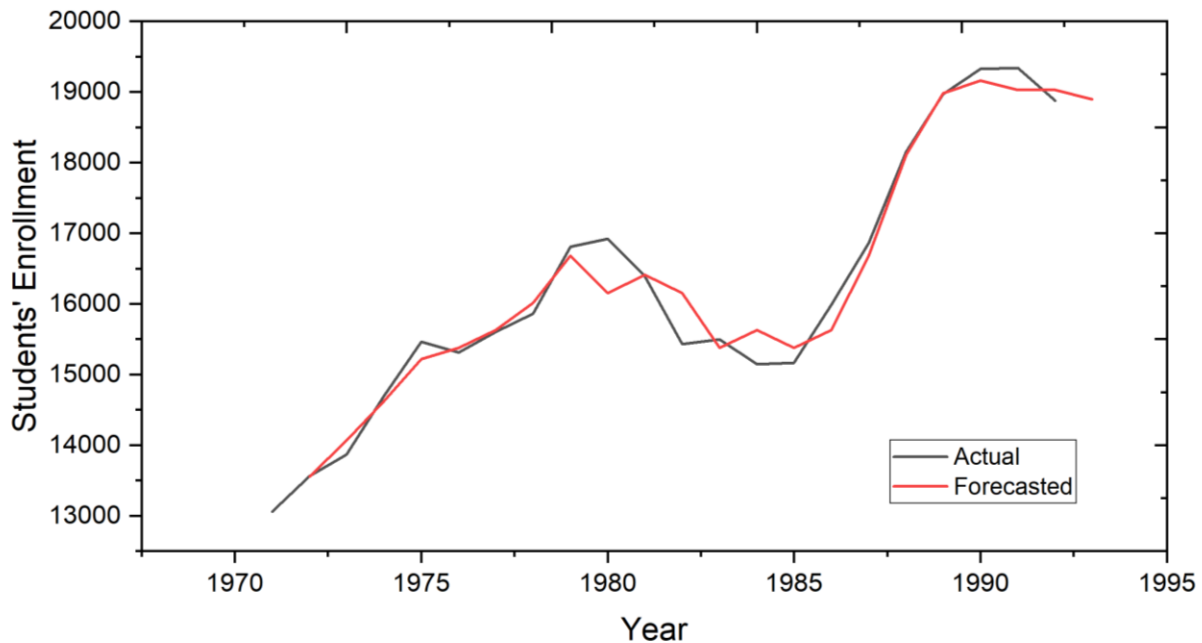


Figure 3 : Actual and forecasted students' enrolment

The results obtained using the proposed hybrid forecasting model are shown in Figure 3. When comparing the forecasted enrolment to the actual enrolment, the hybrid model has graphically shown a good forecasting result. As can be seen in Figure 3, the forecasted enrolment data (represented using the red line) is almost near to the actual enrolment data (represented using the black line). Hence, the obtained predicted results are almost accurate. Furthermore, a calculation is also carried out to measure the deviation of the forecasted enrolment from the actual enrolment.

In this study, the mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE) are used to measure the forecasting accuracy of the proposed model. These error values are then compared with the existing model to highlight the strength of the proposed model.

Table 4 : Comparison with the existing models

Model	MSE	RMSE	MAE
Fuzzy time series [4]	407507	638.363	498.810
Intuitionistic fuzzy time series [15]	145949	382.033	272.048
Intuitionistic fuzzy time series via de-i-fuzzification [16]	130282	360.945	254.697
Hybrid fuzzy time series and 4253HT smoother	88036	296.709	210.387

In reference to Table 4, the proposed model has shown its strength in reducing forecasting inaccuracy. In the model based on fuzzy sets [4], the universe of discourse was partitioned into seven sub-intervals of equal length. This method of partitioning the intervals has led to high forecasting inaccuracy since many interval partitioning methods were improved afterwards and the forecasting accuracy was increased. In the intuitionistic fuzzy time series forecasting model proposed by Abhishekh et al. [15], the intuitionistic fuzzy sets were developed to represent the historical data. However, the mid-points of intervals were used in the defuzzification, which did not present the strength of intuitionistic fuzzy sets in reducing uncertainties. Hence, the model was further improved by Alam and Ramli [16] by using the equal distribution of hesitancy de-i-fuzzification as an extension of the intuitionistic fuzzy time series model.

In the previously mentioned models, no smoothing method was adopted to remove the extreme values. As such, the proposed model entails a hybrid of fuzzy time series and the 4253HT smoother, in which the historical data were firstly smoothed prior to forecasting using fuzzy time series. Using the 4253HT smoother, the extreme values are better handled; hence, uncertainties in the time series data can be slightly reduced. In the students' enrolment data used in this study, the extreme values are not obvious. However, the use of the 4253HT smoother could smooth the data before the fuzzy time series forecasting procedure could be conducted. The smoothing process has reduced the uncertainties in the set of data. Hence, implementing the proposed hybrid model in other sets of data can remove extreme values and reduce uncertainties to obtain better forecasting results.

5 CONCLUSION

The 4253HT smoother has been hybridised with the fuzzy time series forecasting model in this paper. The data are first smoothed to remove extreme values using the moving median of the 4253HT smoother and the smoothed data are then forecasted using the fuzzy time series. The results show that the proposed hybrid model gives a better performance compared to the existing fuzzy time series forecasting models. The adoption of the 4253HT smoother has also reduced the uncertainties of the fuzzy time series since extreme values have been removed. For future works, it is recommended to combine the 4253HT smoother with other time series forecasting models based on both fuzzy sets and intuitionistic fuzzy sets. Besides, it is also suggested that future works use a modified 4253HT smoother as an advanced technique to obtain better forecasting accuracy.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Teknologi MARA for supporting this research under Geran Lestari Khas 600-TNCPI 5/3/DDN (06) (009/2020).

REFERENCES

- [1] Q. Song and B. S. Chissom, "Fuzzy Time Series and Its Models," *Fuzzy Sets Syst.*, vol. 54, pp. 269–277, 1993.
- [2] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part I," *Fuzzy Sets Syst.*, vol. 54, no. 1, pp. 1–9, 1993, doi: 10.1016/0165-0114(93)90355-L.
- [3] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - Part II," *Fuzzy Sets Syst.*, vol. 62, no. 1, pp. 1–8, 1994, doi: 10.1016/0165-0114(93)90355-L.
- [4] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets Syst.*, vol. 81, no. 3, pp. 311–319, Aug. 1996, doi: 10.1016/0165-0114(95)00220-0.
- [5] H. S. Lee and M. T. Chou, "Fuzzy forecasting based on fuzzy time series," *Int. J. Comput. Math.*, vol. 81, no. 7, pp. 781–789, 2004, doi: 10.1080/00207160410001712288.
- [6] H. T. Liu, "An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers," *Fuzzy Optim. Decis. Mak.*, vol. 6, no. 1, pp. 63–80, 2007, doi: 10.1007/s10700-006-0025-9.
- [7] K. Huarng, "Effective lengths of intervals to improve forecasting in fuzzy time series," *Fuzzy Sets Syst.*, vol. 123, no. 3, pp. 387–394, 2001, doi: 10.1016/S0165-0114(00)00057-9.
- [8] R. C. Tsauro and T. C. Kuo, "The adaptive fuzzy time series model with an application to Taiwan's tourism demand," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 9164–9171, 2011, doi: 10.1016/j.eswa.2011.01.059.
- [9] T. A. Jilani and S. M. A. Burney, "A refined fuzzy time series model for stock market forecasting," *Phys. A Stat. Mech. its Appl.*, vol. 387, no. 12, pp. 2857–2862, 2008, doi: 10.1016/j.physa.2008.01.099.
- [10] P. F. Velleman and D. C. Hoaglin, *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press, 1981.
- [11] Z. Jin and B. Xu, "A novel compound smoother - RMMEH to reconstruct MODIS NDVI time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 942–946, 2013, doi: 10.1109/LGRS.2013.2253760.
- [12] N. N. K. Azmi, M. B. Adam, and N. Ali, "Modified compound smoother in median algorithm of span size 42," *e-Academia*, vol. 8, pp. 78–84, 2019.
- [13] Y. C. Cheng and S. T. Li, "Fuzzy time series forecasting with a probabilistic smoothing hidden Markov model," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 291–304, 2012, doi: 10.1109/TFUZZ.2011.2173583.
- [14] J. W. Tukey, *Exploratory Data Analysis*. Massachusetts: Addison-Wesley Publishing Company Reading, 1977.

- [15] Abhishekh, S. S. Gautam, and S. R. Singh, “A new method of time series forecasting using intuitionistic fuzzy set based on average-length,” *J. Ind. Prod. Eng.*, vol. 37, no. 4, pp. 175–185, 2020, doi: 10.1080/21681015.2020.1768163.
- [16] N. M. F. H. N. B. Alam and N. Ramli, “Time Series Forecasting Model Based on Intuitionistic Fuzzy Set via Equal Distribution of Hesitancy De-I-Fuzzification,” *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 29, no. 06, pp. 1015–1029, 2021, doi: 10.1142/s0218488521500458.