ISSN 2462-2052 Journal of Science and Mathematics Letters
DOI: https://doi.org/10.37134/jsml.vol11.2.11.2023 Volume 11, Issue 2, 93-103, 2023

RESEARCH ARTICLE

Fuzzy Time Series Forecasting Accuracy Based on Hybrid Similarity Measure

Nazirah Ramli[1*], Siti Musleha Ab Mutalib[2], Daud Mohamad[3], Mahmod Othman[4], Asyura Abd Nassir[1]

[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Pahang, 26400 Bandar Tun Abdul Razak Jengka, Pahang, Malaysia
[2]School of Professional and Continuing Education (UTMSPACE), Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[3]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
[4]Fundamental and Applied Sciences Department, Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia
*Corresponding author: nazirahr@uitm.edu.my

Received: 6 June 2023; **Accepted:** 15 July 2023; **Published:** 4 August 2023

ABSTRACT

The majority of fuzzy time series forecasting (FTSF) algorithms assess forecasting accuracy using an error-based distance. The predicted value is defuzzified to a crisp number and the error-based distance will be computed. Defuzzification causes some information to be lost, which leads to its inability to comprehend the level of uncertainty that has been preserved during the forecasting process. This paper proposes an enhanced FTSF model with forecasting accuracy developed based on a new hybrid similarity measure combining the centre of gravity and area and height. Three properties of the hybrid similarity measure are presented. The FTSF model is implemented in the case of the Malaysian unemployment rate. The findings indicate that, on average more than 94% of the predicted value was identical to historical data. The forecasting accuracy is produced directly from the forecasting value without undergoing the defuzzification process, which can preserve some information from being lost.

Keywords: Hybrid Similarity Measure, Fuzzy Time Series Forecasting, Forecasting Accuracy

1. INTRODUCTION

Fuzzy time series forecasting (FTSF) model was proposed to overcome the limitation of the traditional time series forecasting model that cannot handle linguistic information. Song and Chissom (1993, 1994) presented the FTSF model based on discrete fuzzy sets to overcome the issue. Many improvements to increase the forecasting accuracy (FA) have been made, such as the modification to the types of partition interval method (Singh, 2007; Kuo et al., 2009, Chen & Phuong, 2017; Pal & Kar, 2019; Hanif et al., 2023), fuzzy logical relation (FLR) order (Chen, 2014; Bisht & Kumar, 2016; Tinh & Dieu, 2017; Cheng & Chen, 2018), defuzzification method (Ramli & Tap, 2009), and aggregation operator (Gupta & Kumar, 2019). However, the discrete fuzzy sets were employed in the aforementioned methods for representing the linguistic terms of the time series data in which the forecasted value (FV) for various levels of confidence cannot be produced.

The evolution of FTSF was continued by Liu (2007, 2009), where he proposed the fuzzy numbers to characterise the data's linguistic term. The FV is in the form of fuzzy numbers and therefore the forecasted range at various levels of confidence can be produced. On the other hand, the FV were defuzzified to crisp values for determining the FA using the mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE). As known, the process of defuzzification will cause some information that has been kept losing. However, the majority of the FTSF models typically used the MSE, RMSE, MAPE, and MAE, such as studies by Chen and Phuong (2017), Alam et al. (2021), Solikhin et al. (2022), Khatoon et al. (2023) and Gamayanti et al. (2023). Therefore, this paper proposes a hybrid similarity measure combining centre of gravity (COG), and area and height in determining the forecasting accuracy.

The remaining parts of the paper are organised as follows: the literature review and the basic concept of fuzzy numbers and FTSF will be presented in the next section. The new hybrid similarity measure with its properties and performance, and the proposed FTSF are presented in the following section. The implementation of the proposed FTSF in the data of unemployment rate in Malaysia is discussed. The final section contains the conclusion.

2. METHODOLOGY

2.1. Preliminaries

This section briefly discusses several concepts of trapezoidal fuzzy numbers (Wang, 1997) and fuzzy time series (FTS) (Song & Chissom, 1993, 1994).

Definition 2.1. (Wang, 1997)
The membership function of a trapezoidal fuzzy number (TrFN) $P = (p_1, p_2, p_3, p_4)$ is as follows:

$$\mu_P(x) = \begin{cases} 0, & x < p_1 \\ \dfrac{x - p_1}{p_2 - p_1}, & p_1 \le x \le p_2 \\ 1, & p_2 \le x \le p_3 \\ \dfrac{p_4 - x}{p_4 - p_3}, & p_3 \le x \le p_4 \\ 0, & x > p_4 \end{cases}$$

Definition 2.2. (Song & Chissom, 1993)
Let $Y(t)$ $(t = \ldots, 0,1,2,\ldots)$ be a subset of $\Re$ and $Y(t)$ be a universe of discourse described by fuzzy set $\mu_i(t)(i = 1,2,\ldots)$, then $A(t)$ is called as FTS on $Y(t)$ $(t = \ldots, 0,1,2,\ldots)$.

Definition 2.3. (Song & Chissom, 1994)
Let $A(t)$ must be an FTS. $A(t)$ is produced from $A(t-1)$ if there exists a fuzzy relationship $B(t-1, t)$ such that $A(t) = A(t-1) \otimes B(t-1, t)$ whereby $\otimes$ denotes as a fuzzy operator. $A(t-1) \rightarrow A(t)$ denotes the relationship.

Definition 2.4. (Song & Chissom, 1994)
Given $A(t-1) = C_i$ and $A(t) = C_j$. The fuzzy logical relationship (FLR) can be described as $C_i \rightarrow C_j$ where C_i and C_j are the left and right side of FLR, respectively. The FLR can be further categorised into the same FLR group, if the FLR on the left side has the same fuzzy set.

2.2 Fuzzy Time Series Forecasting Based On Hybrid Similarity Measure

The procedure of the proposed model of FTSF based on hybrid similarity measure is presented as the following algorithm:

Step 1: Define the universe of discourse *(UD*) of the historical data. The universe of discourse is defined as $UD = [M_{min} - c_1, M_{max} + c_2]$ whereby $M_{\min}$ and $M_{\max}$ are the minimum and maximum value, respectively, and c_1 and c_2 are positive real numbers.
Step 2: Partition the *UD* with the same length in *m* intervals $u_1, u_2, u_3, ..., u_m$.
Step 3: Calculate the data frequency in the interval u_i. Based on their frequency level, categorise and subpartition intervals. The interval with the most frequency is categorised as Class 1 and is divided into four sub-intervals. Other detailed classification of intervals is shown in Table 1. List all new sub-intervals as $v_1, v_2, v_3, ..., v_n$ with $v_1 = [d_1, d_2]$, $v_2 = [d_2, d_3]$, …, $v_n = [d_n, d_{n+1}]$.

Table 1. Category of Sub-interval

Frequency Level of the Interval	Class	Number of Sub-interval with equal length
Largest	1	4
Second largest	2	3
Third largest	3	2
Fourth largest and above	4	1

Step 4: Define the TrFNs $W_1, W_2, \ldots, W_{n-1}, W_n$ as follows:

$W_1 = (d_0, d_1, d_2, d_3)$, $W_2 = (d_1, d_2, d_3, d_4)$,
$\vdots$
$W_{n-1} = (d_{n-2}, d_{n-1}, d_n, d_{n+1})$, $W_n = (d_{n-1}, d_n, d_{n+1}, d_{n+2})$.

Step 5: The historical data, M_t is fuzzified. The historical data M_t belongs to TrFN W_k if the value of M_t falls in the sub-interval of v_k.
Step 6: Develop the FLR based on Definition 2.4: "If the fuzzy production of time *t*-1 is W_j, then the fuzzy production of time *t* is W_k", thus, the FLR is labeled as $W_j \rightarrow W_k$.
Step 7: Develop the FLR groups. The FLR is arranged based on the same TrFN on the left-hand side of the FLR. The following are the FLR groups.

Group 1: $W_i \rightarrow W_{j1}, W_i \rightarrow W_{j2}, \cdots, W_i \rightarrow W_{jp}$
Group 2: $W_j \rightarrow W_{k1}, W_j \rightarrow W_{k2}, \cdots, W_j \rightarrow W_{kp}$
Group *n*: $W_n \rightarrow W_{l1}, W_n \rightarrow W_{l2}, \cdots, W_n \rightarrow W_{lp}$

Step 8: Compute the forecasted value, FV_t in the form of TrFNs based on heuristic rules (Cheng et al., 2008) as follows:

- If the FLR group of W_i is $W_i \rightarrow \phi$, then $FV_t = W_i$.
- If the FLR group of W_i is $W_i \rightarrow W_j$, then $FV_t = W_j$.
- If the FLR group of W_i is $W_i \rightarrow W_{j1}$, $W_i \rightarrow W_{j2}$, $\cdots$, $W_i \rightarrow W_{jp}$, then $FV_t = \dfrac{W_{j1} + W_{j2} + \ldots + W_{jp}}{p}$.

Step 9: Calculate the similarity of FV_t and W_t ($S(FV_t,\ W_t)$) by using the proposed hybrid similarity measure in Definition 3.1, with W_t is the fuzzy value of M_t.

3. RESULTS AND DISCUSSION

3.1. Hybrid Similarity Measure

A hybrid similarity measure based on similarity measure combining of COG (Xu et al., 2010), and area and height (Patra & Mondal, 2015) is presented in this section.

Definition 3.1.
Let $P=\left(p_1,p_2,p_3,p_4;h_P\right)$ and $Q=\left(q_1,q_2,q_3,q_4;h_Q\right)$ be two generalized TrFNs. The degree of similarity between P and Q is defined as

$$S(P,Q)=\left[1-\frac{1}{8}\sum_{i=1}^{4}\left|p_i-q_i\right|-\frac{d(P,Q)}{2}\right]\left[\left(1-\frac{1}{4}\sum_{i=1}^{4}\left|p_i-q_i\right|\right)\times\left(1-\frac{1}{2}\left\{\left|ar(P)-ar(Q)\right|+\left|h_P-h_Q\right|\right\}\right)\right]$$

whereby $d(P,Q)=\dfrac{\sqrt{\left(x_P^*-x_Q^*\right)^2+\left(y_P^*-y_Q^*\right)^2}}{\sqrt{1.25}}$, $y_P^*=\begin{cases}\dfrac{h_P\left(\dfrac{p_3-p_2}{p_4-p_1}+2\right)}{6} & ,p_4\neq p_1\\ \dfrac{h_P}{2} & ,p_4=p_1\end{cases}$,

$y_Q^*=\begin{cases}\dfrac{h_Q\left(\dfrac{q_3-q_2}{q_4-q_1}+2\right)}{6} & ,q_4\neq q_1\\ \dfrac{h_Q}{2} & ,q_4=q_1\end{cases}$, $x_P^*=\begin{cases}\dfrac{y_P^*\left(p_3+p_2\right)+\left(p_4+p_1\right)\left(h_P-y_P^*\right)}{2h_P} & ,h_P\neq 0\\ \dfrac{p_4+p_1}{2} & ,h_P=0\end{cases}$

$x_Q^*=\begin{cases}\dfrac{y_Q^*\left(q_3+q_2\right)+\left(q_4+q_1\right)\left(h_Q-y_Q^*\right)}{2h_Q} & ,h_Q\neq 0\\ \dfrac{q_4+q_1}{2} & ,h_Q=0\end{cases}$, $ar(P)=\dfrac{\left(p_4+p_3-p_2-p_1\right)h_P}{2}$ and

$ar(Q)=\dfrac{\left(q_4+q_3-q_2-q_1\right)h_Q}{2}$.

The higher the $S(P,Q)$ value, the higher the likeness between TrFNs P and Q.

The proposed hybrid similarity measure has three properties as follows:

Property 3.1. $S(P,Q) = 1$ if and only if $P = Q$.

Proof: If $S(P,Q) = 1$, then by Definition 3.1

$$S(P,Q)=\left[1-\frac{1}{8}\sum_{i=1}^{4}\left|p_i-q_i\right|-\frac{d(P,Q)}{2}\right]\left[\left(1-\frac{1}{4}\sum_{i=1}^{4}\left|p_i-q_i\right|\right)\times\left(1-\frac{1}{2}\left\{\left|ar(P)-ar(Q)\right|+\left|h_P-h_Q\right|\right\}\right)\right]$$

ISSN 2462-2052 *Journal of Science and Mathematics Letters*
DOI: https://doi.org/10.37134/jsml.vol11.2.11.2023 *Volume 11, Issue 2, 93-103, 2023*

This implies $\frac{1}{8}\sum_{i=1}^{4}|p_i - q_i| = 0$, $d(P,Q) = 0$, $\frac{1}{4}\sum_{i=1}^{4}|p_i - q_i| = 0$ and $\left\{|ar(P) - ar(Q)| + |h_P - h_Q|\right\} = 0$ whereby $p_1 = q_1$, $p_2 = q_2$, $p_3 = q_3$, $p_4 = q_4$, $y_P^* = y_Q^*$, $x_P^* = x_Q^*$, $ar(P) = ar(Q)$ and $h_P = h_Q$. Thus, $P = Q$. On the other hand, if P and Q are identical, then, $p_1 = q_1$, $p_2 = q_2$, $p_3 = q_3$, $p_4 = q_4$ and $h_P = h_Q$. There are $y_P^* = y_Q^*$, $x_P^* = x_Q^*$, $d(P,Q) = 0$, $\frac{1}{8}\sum_{i=1}^{4}|p_i - q_i| = 0$, $\frac{1}{4}\sum_{i=1}^{4}|p_i - q_i| = 0$ and $ar(P) = ar(Q)$. This gives the degree of similarity between P and Q as $S(P,Q) = 1$.

Property 3.2. $S(P,Q) = S(Q,P)$

Proof: By Definition 3.1,

$$S(P,Q) = \left[1 - \frac{1}{8}\sum_{i=1}^{4}|p_i - q_i| - \frac{d(P,Q)}{2}\right]\left[\left(1 - \frac{1}{4}\sum_{i=1}^{4}|p_i - q_i|\right) \times \left(1 - \frac{1}{2}\left\{|ar(P) - ar(Q)| + |h_P - h_Q|\right\}\right)\right]$$

and

$$S(Q,P) = \left[1 - \frac{1}{8}\sum_{i=1}^{4}|q_i - p_i| - \frac{d(Q,P)}{2}\right]\left[\left(1 - \frac{1}{4}\sum_{i=1}^{4}|q_i - p_i|\right) \times \left(1 - \frac{1}{2}\left\{|ar(Q) - ar(P)| + |h_Q - h_P|\right\}\right)\right]$$

Since, $\frac{1}{8}\sum_{i=1}^{4}|p_i - q_i| = \frac{1}{8}\sum_{i=1}^{4}|q_i - p_i|$, $\frac{d(P,Q)}{2} = \frac{d(Q,P)}{2}$, $\frac{1}{4}\sum_{i=1}^{4}|p_i - q_i| = \frac{1}{4}\sum_{i=1}^{4}|q_i - p_i|$, $|ar(P) - ar(Q)| = |ar(Q) - ar(P)|$ and $|h_P - h_Q| = |h_Q - h_P|$. Hence, $S(P,Q) = S(Q,P)$.

Property 3.3. If $P = (p, p, p, p; 1)$ and $Q = (q, q, q, q; 1)$, then

$$S(P,Q) = (1 - k|p - q|) \times (1 - |p - q|) \text{ whereby } k = \frac{5 + 2\sqrt{5}}{10} .$$

Proof: For a real number, $y_P^* - y_Q^* = 0$, $d(P,Q) = \sqrt{\frac{(p-q)^2}{1.25}}$, $ar(P) = ar(Q)$ and $h_P = h_Q$. Then, Definition 3.1 calculates the similarity of P and Q as:

$$S(P,Q) = \left[1 - \frac{1}{2}|p - q| - \frac{|p - q|}{\sqrt{5}}\right] \times \left[1 - |p - q|\right] \times (1) = (1 - k|p - q|) \times (1 - |p - q|)$$

whereby $k = \frac{5 + 2\sqrt{5}}{10}$.

3.2. Performance of the Hybrid Similarity Measure

The performance of the hybrid similarity measure is compared in this section with those of Hsieh and Chen (1999), Chen and Chen (2001), Xu et al. (2010), Hejazi et al. (2011) and Patra and Mondal (2015) approaches. For comparison, ten sets of TrFNs P and Q from Xu et al. (2010) and Patra and Mondal (2015) are used. The TrFNs are as follows:

Set 1: P = (0.10, 0.20, 0.30,0.40; 1), Q = (0.10, 0.25, 0.25, 0.40; 1)
Set 2: P = (0.10, 0.20, 0.30,0.40; 1), Q = (0.10, 0.25, 0.25, 0.40; 0.2)
Set 3: P = (0.10, 0.20, 0.30,0.40; 1), Q = (0.10, 0.35, 0.35, 0.50; 0)
Set 4: P = (0.10, 0.25, 0.25,0.40; 0.01), Q = (0.10, 0.25, 0.25, 0.40; 0)
Set 5: P = (0.10, 0.20, 0.30,0.40; 1), Q = (0.10, 0.20, 0.30,0.40; 1)
Set 6: P = (0.10, 0.20, 0.30,0.40; 0), Q = (0.10, 0.20, 0.30,0.40; 0)
Set 7: P = (0.10, 0.20, 0.30,0.40; 1), Q = (0.30, 0.45, 0.45,0.60; 1)
Set 8: P = (0.10, 0.20, 0.30,0.40; 0), Q = (0.30, 0.40, 0.50,0.60; 1)
Set 9: P = (0.10, 0.20, 0.30,0.40; 0.8), Q = (0.20, 0.30, 0.40,0.50; 0.4)
Set 10: P = (0.10, 0.20, 0.30,0.40; 0.8), Q = (0.30, 0.45, 0.54,0.60; 0.6)

Table 2 shows the performance of the hybrid method compared to others. For Sets 1 and 2, the degree of similarity (DoS) by Hsieh and Chen (1999) method is equal to one, which means that the two TrFNs are equal. However, the membership functions of P and Q are not equal, and the DoS will not be equal to one. The proposed method produces the value of DoS not equal to one, which is consistent with the membership function.

Table 2. Comparison of Performance of the Proposed Method

TrFNs	Hsieh & Chen (1999)	Chen & Chen (2001)	Xu et al. (2010)	Hejazi et al. (2011)	Patra & Mondal (2015)	Proposed Hybrid Method
Set 1	1	0.8357	0.9627	0.9004	0.9506	0.9151
Set 2	1	0.1671	0.8434	0.0644	0.5021	0.4235
Set 3	0.9231	-	0.7704	0	0.36	0.2773
Set 4	1	-	0.9985	0	0.9943	0.9928
Set 5	1	1	1	1	1	1
Set 6	1	-	1	-	1	1
Set 7	0.8571	0.5486	0.8072	0.7407	0.78	0.6296
Set 8	0.8571	0.64	0.8126	0.8	0.8	0.6501
Set 9	0.9231	0.6075	0.8933	0.2624	0.684	0.6110
Set 10	0.8571	0.48	0.8057	0.4783	0.708	0.5704

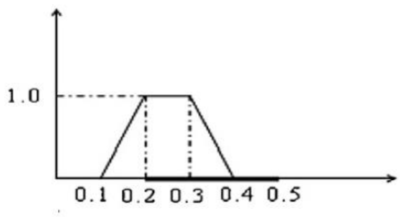

Figure 1. Set 3 TrFNS

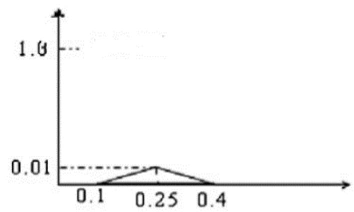

Figure 2. Set 4 TrFNS

Chen and Chen (2001) cannot provide the DoS for Sets 3 and 4. This is due to the weaknesses of the method that was unable to calculate the TrFN with height zero (TrFN Q has height zero for Sets 3 and 4). Hejazi et al. (2011) method produced DoS equal to zero for Sets 3 and 4, while Hshieh and Chen (1999) gave equal to one for Set 4. The values of DoS are not consistent with the graphical representation as in Figures 1 and 2. The proposed method gives better results that are consistent with the graphical representation. For Set 5, all methods in Table 1 produced DoS equal to one and this is true since the two TrFNs are equal. However, although the two TrFNs are equal for Set 6, Chen and Chen (2001) and Hejazi et al. (2011) failed to calculate the similarities. Chen and Chen's (2011) method was unable to calculate the cases if any one of the TrFN has height zero and Hejazi et al.'s (2011) method was unable to calculate the TrFNs with both height zeros. The proposed method produces consistent results since the TrFNs in sets 5 and 6 are equal. For all sets of TrFNs (except Sets 5 and 6), the DoS of the hybrid method is lower than the methods by Xu et al. (2010) and Patra and Mondal (2015). The similarity measure of the hybrid method is the combination of both methods (Xu et al., 2010; Patra & Mondal, 2015) through a multiplication operation. Thus, if one of the methods or both methods have DoS less than one, then the DoS of the hybrid method will be less than the DoS of each method. However, the DoS of the hybrid method will be equal to one if both methods have DoS equal to one.

3.3. Numerical Example

The proposed FTSF model with hybrid similarity measure is implemented in Malaysia's unemployment rate data. Figure 3 shows Malaysia's unemployment rate from 1982 to 2013 (Department of Statistic Malaysia, 2014).

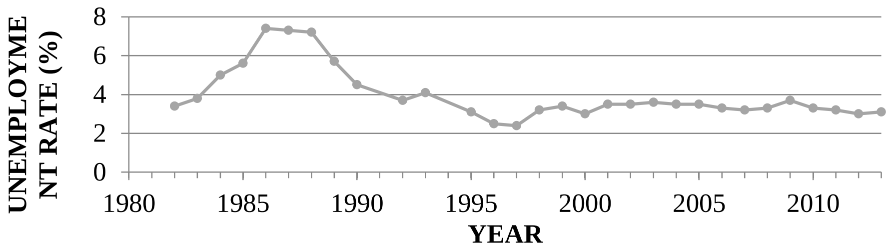

Figure 3. The Malaysian unemployment rate data for 1982 to 2013

Step 1: From the unemployment rate data, M_{min} = 2.4% and M_{max} = 7.4%. By assigning two numbers c_1 = 0.4 and c_2 = 0.6, $UD = [2.0, 8.0]$.

Step 2: UD is divided into eight equal lengths, u_1=[2.00,2.75], u_2=[2.75,3.50], u_3=[3.50,4.25], u_4=[4.25,5.00], u_5=[5.00,5.75], u_6=[5.75,6.50], u_7=[6.50,7.25], u_8=[7.25,8.00].

Step 3: The frequency, classification, and number of sub-intervals are shown in Table 3.

Table 3. The Number of Sub-Interval for Each Interval for Unemployment Rate

Interval	Frequency	Class	Number of Sub-interval
u_1=[2.00,2.75]	2	3	2
u_2=[2.75,3.50]	16	1	4
u_3=[3.50,4.25]	7	2	3
u_4=[4.25,5.00]	2	3	2
u_5=[5.00,5.75]	2	3	2
u_6=[5.75,6.50]	0	4	1
u_7=[6.50,7.25]	1	4	1
u_8=[7.25,8.00]	2	3	2

From Table 3, interval u_2 has the highest frequency followed by u_3. Thus, u_2 and u_3 are categorised as Class 1 and Class 2 with 4 and 3 sub-intervals respectively. The sub-intervals for u_2 = [2.75, 3.50] are [2.75, 2.938], [2.938, 3.125], [3.125, 3.313] and [3.313, 3.5], and the sub-intervals for u_3 = [3.50, 4.25] are [3.5, 3.75], [3.75, 4.00] and [4.00, 4.25]. Intervals u_1, u_4, u_5 and u_8 have the third highest frequency and categorised as Class 3 with 2 sub-intervals. The sub-intervals for u_1 = [2.00, 2.75] are [2.00, 2.375] and [2.375, 2.75], sub-intervals for u_4 = [4.25, 5.00] are [4.25, 4.625] and [4.625, 5.00], sub-intervals for u_5 = [5.00, 5.75] are [5.00, 5.375] and [5.375, 5.75], and sub-intervals for u_8 = [7.25, 8.00] are [7.25, 7.625] and [7.625, 8.00]. u_7 and u_6 are categorised as Class 4 and the sub-interval remains the same as [6.5,7.25] and [5.75,6.5] respectively. Thus, in total, there are 17 new sub-intervals obtained, and by arranging them, the sub-intervals are given as follows:

v_1=[2.00,2.375]	v_7=[3.5,3.75]	v_{13}=[5.375,5.75]
v_2=[2.375,2.75]	v_8=[3.75,4.00]	v_{14}=[5.75,6.5]
v_3=[2.75,2.938]	v_9=[4.00,4.25]	v_{15}=[6.5,7.25]
v_4=[2.938,3.125]	v_{10}=[4.25,4.625]	v_{16}=[7.25,7.625]
v_5=[3.125,3.313]	v_{11}=[4.625,5.00]	v_{17}=[7.625,8.00]
v_6= [3.313,3.5]	v_{12}=[5.00,5.375]	

Step 4: The unemployment rates in the form of TrFNs are as follows:

W_1 = (1.625, 2.00, 2.375, 2.75), W_2 = (2.00, 2.375, 2.75, 2.938),
W_3 = (2.375, 2.75, 2.938, 3.125), W_4 = (2.75, 2.938, 3.125, 3.313),
W_5 = (2.938, 3.125, 3.313, 3.5), W_6 = (3.125, 3.313, 3.5, 3.75)
W_7 = (3.313, 3.5, 3.75, 4.00), W_8 = (3.5, 3.75, 4.00, 4.25)
W_{16} = (6.50, 7.25, 7.625, 8.00), W_{17} = (7.25, 7.625, 8.00, 8.375)

Step 5: Table 4 shows the years 2005 to 2013 fuzzified unemployment rate.

Table 4. Fuzzified Rate of Unemployment in TrFNs Form for the years 2005 to 2013

Year	Unemployment rate	TrFNs	Year	Unemployment rate	TrFNs
2005	3.5	W_6	**2010**	3.3	W_5
2006	3.3	W_5	**2011**	3.2	W_5
2007	3.2	W_5	**2012**	3.0	W_4
2008	3.3	W_5	**2013**	3.1	W_4
2009	3.7	W_7			

The unemployment rate for the year 2005 is 3.5%, thus, it falls under sub-interval v_6 and it belongs to TrFN W_6. Similarly, the unemployment rates for the years 2006, 2007, 2008, 2010, and 2011 are 3.3%, 3.2%, 3.3%, 3.3% and 3.2% respectively, thus, they fall under sub-interval v_6 and belong to TrFN W_6.

Steps 6-7: Table 5 shows the unemployment rate FLR group.

Table 5. Unemployment Rate FLR Group

Group	FLR	Group	FLR
1	$W_2 \to W_2$, $W_2 \to W_5$	**8**	$W_{10} \to W_7$, $W_{10} \to W_9$
2	$W_4 \to W_2$, $W_4 \to W_2$, $W_4 \to W_2$	**9**	$W_{11} \to W_{13}$
3	$W_5 \to W_4$, $W_5 \to W_5$, $W_5 \to W_6$, $W_5 \to W_7$	**10**	$W_{13} \to W_{10}$, $W_{13} \to W_{16}$
4	$W_6 \to W_4$, $W_6 \to W_5$, $W_6 \to W_6$, $W_6 \to W_7$, $W_6 \to W_8$	**11**	$W_{15} \to W_{13}$
5	$W_7 \to W_4$, $W_7 \to W_5$, $W_7 \to W_6$, $W_7 \to W_9$	**12**	$W_{16} \to W_{15}$, $W_{16} \to W_{16}$
6	$W_8 \to W_{11}$	**13**	$W_4 \to \phi$
7	$W_9 \to W_7$		

Step 8: The forecast value FV_t is calculated on the basis of the heuristic rules. Table 6 shows the FV_t values for the years 1983 to 2013.

Table 6. Fuzzy Historical and Fuzzy Forecasted for the Year 1983 to 2013

Year	Fuzzy historical data (W_t)	Fuzzy forecasted (FV_t)	Year	Fuzzy historical data (W_t)	Fuzzy forecasted (FV_t)
1983	(3.5, 3.75, 4, 4.25)	(3.125, 3.325, 3.538, 3.763)	**1999**	(3.125, 3.313, 3.5, 3.75)	(3.031, 3.219, 3.422, 3.641)
1984	(4.25, 4.625, 5,	(4.25, 4.625, 5,	**2000**	(2.75, 2.938, 3.125,	(3.125, 3.325, 3.538,

	5.375)	5.375)		3.313)	3.763)
1985	(5, 5.375, 5.75, 6.5)	(5, 5.375, 5.75, 6.5)	**2001**	(3.125, 3.313, 3.5, 3.75)	(2.625, 2.875, 3.125, 3.333)
1986	(6.5, 7.25, 7.625, 8)	(5.25, 5.75, 6.125, 6.5)	**2002**	(3.125, 3.313, 3.5, 3.75)	(3.125, 3.325, 3.538, 3.763)
1987	(6.5, 7.25, 7.625, 8)	(6.125, 6.875, 7.438, 7.813)	**2003**	(3.313, 3.5, 3.75, 4)	(3.125, 3.325, 3.538, 3.763)
1988	(5.75, 6.5, 7.25, 7.625)	(6.125, 6.875, 7.438, 7.813)	**2004**	(3.125, 3.313, 3.5, 3.75)	(3.141, 3.344, 3.547, 3.797)
1989	(5, 5.375, 5.75, 6.5)	(5, 5.375, 5.75, 6.5)	**2005**	(3.125, 3.313, 3.5, 3.75)	(3.125, 3.325, 3.538, 3.763)
1990	(4, 4.25, 4.625, 5)	(5.25, 5.75, 6.125, 6.5)	**2006**	(2.938, 3.125, 3.313, 3.5)	(3.125, 3.325, 3.538, 3.763)
1991	(3.75, 4, 4.25, 4.625)	(3.75, 4, 4.25, 4.625)	**2007**	(2.938, 3.125, 3.313, 3.5)	(3.031, 3.219, 3.422, 3.641)
1992	(3.313, 3.5, 3.75, 4)	(3.313, 3.5, 3.75, 4)	**2008**	(2.938, 3.125, 3.313, 3.5)	(3.031, 3.219, 3.422, 3.641)
1993	(3.75, 4, 4.25, 4.625)	(3.141, 3.344, 3.547, 3.797)	**2009**	(3.313, 3.5, 3.75, 4)	(3.031, 3.219, 3.422, 3.641)
1994	(3.313, 3.5, 3.75, 4)	(3.313, 3.5, 3.75, 4)	**2010**	(2.938, 3.125, 3.313, 3.5)	(3.141, 3.344, 3.547, 3.797)
1995	(2.75, 2.938, 3.125, 3.313)	(3.141, 3.344, 3.547, 3.797)	**2011**	(2.938, 3.125, 3.313, 3.5)	(3.031, 3.219, 3.422, 3.641)
1996	(2, 2.375, 2.75, 2.938)	(2.625, 2.875, 3.125, 3.333)	**2012**	(2.75, 2.938,3.125, 3.313)	(3.031, 3.219, 3.422, 3.641)
1997	(2, 2.375, 2.75, 2.938)	(2.469, 2.75, 3.031, 3.219)	**2013**	(2.75, 2.938,3.125, 3.313)	(2.625, 2.875, 3.125, 3.333)
1998	(2.938, 3.125, 3.313, 3.5)	(2.469, 2.75, 3.031, 3.219)			

Step 9: The hybrid similarity measure is computed based on the normalized W_t and FV_t as shown in Table 7. Table 7 shows that the DoS is equal to one for the years 1984, 1985, 1989, 1991, 1992, and 1994. It shows that 19.4% of the FV is similar to the actual values. 93.5% of the FV is more than 85% similar and on average the FV has 94.3% similar to the actual values. The results demonstrate that the FV is quite close to the actual data.

Table 7. The Hybrid Similarity Measure for Unemployment Rate

Year	Similarity	Year	Similarity	Year	Similarity
1983	0.913	**1994**	1	**2004**	0.992
1984	1	**1995**	0.916	**2005**	0.994
1985	1	**1996**	0.901	**2006**	0.955
1986	0.736	**1997**	0.926	**2007**	0.977
1987	0.936	**1998**	0.926	**2008**	0.977
1988	0.936	**1999**	0.981	**2009**	0.937
1989	1	**2000**	0.920	**2010**	0.951
1990	0.735	**2001**	0.914	**2011**	0.977
1991	1	**2002**	0.994	**2012**	0.941
1992	1	**2003**	0.958	**2013**	0.985
1993	0.863				
					Average: 0.943

4. CONCLUSION

This paper proposes a FTSF model based on hybrid similarity measure. The FA of the enhanced FTSF model is given in the form of DoS. Three properties of the hybrid similarity measure are presented. The hybrid similarity measure outperforms some of the previous studies in terms of consistency with the membership function and graphical presentation, and ability to calculate the similarity of TrFNs with height zero. The FTSF model with forecasting accuracy developed based on hybrid similarity is able to preserve some information at various levels of confidence.

Declaration of Interest
The authors declare that there is no conflict of interest.

Acknowledgement
The authors would like to thank Universiti Teknologi MARA for the support in making this research a success.

REFERENCES

Alam NMFHNA, Ramli N, Mohammed N. (2021). Fuzzy time series forecasting model based on intuitionistic fuzzy sets via delegation of hesitancy degree to the major grade de-i-fuzzification method. *Mathematics and Statistics*, 9(1), 46-53.
Bisht K, Kumar S. (2016). Fuzzy time series forecasting method based on hesitant fuzzy sets. *Expert Systems with Applications*, 64, 557-568.
Chen SM. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets Systems*, 81, 311-319.
Chen MY. (2014). A high-order fuzzy time series forecasting model for internet stock trading. *Future Generation Computer Systems*, 37, 461-467.
Chen SJ, Chen, SM. (2001). A new method to measure the similarity between fuzzy numbers. Proceedings of the 10th IEEE International Conference on Fuzzy Systems, p. 208-214.
Chen SM, Phuong BDH. (2017). Fuzzy time series forecasting based on optimal partition of intervals and optimal weighting vectors. *Knowledge-Based Systems*, 118, 204-216.
Cheng CH, Chen CH. (2018). Fuzzy time series model based on weighted association rule for financial market forecasting. *Expert Systems*, 35(4), e12271.
Department of Statistic Malaysia. Time series data of unemployment. Accessed January 13, 2014.
Gamayanti NH, Junaida J, Fitri F. (2023). Application of fuzzy time series to forecast COVID-19 cases in Central Sulawesi. *AIP Conference Proceedings*, 2719, 040001.
Gupta KK, Kumar S. (2019). Fuzzy time series forecasting method using probabilistic fuzzy sets. In Mandal, J., Bhattacharyya, D., Auluck, N. (Eds.). *Advanced Computing and Communication Technologies*, Springer, Singapore.
Hanif R, Mustafa S, Iqbal S, Piracha S. (2023). A study of time series forecasting enrollments using fuzzy interval partitioning method. *Journal of Computational and Cognitive Engineering*, 2(2), 143-149.
Hejazi SR, Doostparast A, Hosseini SM. (2011). An improved fuzzy risk analysis based on new similarity measure of generalized fuzzy numbers. *Expert Systems with Applications*, 38, 9179-9185.
Hsieh CH, Chen SH. (1999). Similarity of generalized fuzzy numbers with graded mean integration. Proceedings of the 8th International Fuzzy System Association World Congress, 2, p. 551-555.
Khatoon S, Ibraheem, Gupta P, Shahid M. (2023). Comparison of fuzzy time series, ANN, and wavelet techniques for short term load forecasting. *International Journal of Power Electronics and Drive Systems*, 14(2), 1260-1269.
Kuo IH, Horng SJ, Kao TW, Lin TL, Lee CL, Pan Y. (2009). An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Systems with Applications*, 36(3), 6108-6117.
Liu H. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optimization and Decision Making*, 6, 63-80.
Liu HT. (2009). An integrated fuzzy time series forecasting system. *Expert Systems with Applications*, 36(6), 10045-10053.
Pal SS, Kar S. (2019). Fuzzy time series model for unequal interval length using genetic algorithm. *Advances in Intelligent Systems and Computing*, 699, 205-216.
Patra K, Mondal SK. (2015). Fuzzy risk analysis using area and height based similarity measure on generalized trapezoidal fuzzy numbers and its application. *Applied Soft Computing*, 28, 276-284.
Ramli N, Tap AOM. (2009). Forecasting students' enrolment in fuzzy time series based on three classes of t-norm of subsethood defuzzification. *Gading Business and Management Journal*, 13(1), 1-14.
Singh SR. (2007). A simple method of forecasting based on fuzzy time series. *Applied Mathematics and Computation*, 186, 330-339.

Solikhin, Lutfi S, Purnomo, Hardiwinoto. A machine learning approach in Phyton is used to forecast the number of train passengers using a fuzzy time series model. *Bulletin of Electrical Engineering and Informatics*, 11(5), 2746-2755.

Song Q, Chissom BS. (1993). Forecasting enrollments with fuzzy time series-Part I. *Fuzzy Sets and Systems*, 54, 1-9.

Song Q, Chissom BS. (1994). Forecasting enrollments with fuzzy time series-Part II. *Fuzzy Sets and Systems*, 62, 1-8.

Tinh NV, Dieu NC. (2017). A new hybrid fuzzy time series forecasting model combined the time variant fuzzy logical relationship groups with particle swam optimization. *Computer Science and Engineering*, 7(2), 52-66.

Wang LX (1997). A course in fuzzy systems and control. Prentice-Hall.

Xu Z, Shang S, Qian W, Shu W. (2010). A method for fuzzy risk analysis based on the new similarity of trapezoidal fuzzy numbers. *Expert Systems with Applications*, 37(3), 1920-1927.