BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17Issue 3 Page 1245–1256
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1245-1256

FUZZY TIME SERIES BASED ON THE HYBRID OF FCM WITH CMBO OPTIMIZATION TECHNIQUE FOR HIGH WATER PREDICTION

Nursyiva Irsalinda [1*], Dera Kurnia Laely[2], Sugiyarto Surono[3]

[1,2,3]*Mathematics Department, Faculty of Applied Science and Technology, Universitas Ahmad Dahlan*
Jend. Ahmad Yani Street, Banguntapan, Yogyakarta, 55166, Indonesia

*Corresponding author's e-mail: * **nursyiva.irsalinda@math.uad.ac.id***

Article History:

Received: 7th December 2022
Revised: 2nd July 2023
Accepted: 10th July 2023

Keywords:

FTS;
FCM;
CMBO;
Forecasting;
Optimization.

ABSTRACT

Time series data represents measurements taken over a specific period and is often employed for forecasting purposes. The typical approach in forecasting involves the analysis of relationships among estimated variables. In this study, we apply Fuzzy Time Series (FTS) to water level data collected every 10 minutes at the Irish Achill Island Observation Station. The FTS, which is based on Fuzzy C-Means (FCM), is hybridized with the Cat and Mouse Based Optimizer (CMBO). This hybridization of FCM with the CMBO optimizer aims to address weaknesses inherent in FTS, particularly concerning the determination of interval lengths, with the ultimate goal of enhancing prediction accuracy. Before conducting forecasts, we execute the FCM-CMBO process to determine the optimal centroid used for defining interval lengths within the FTS framework. Our study utilizes a dataset comprising 52,562 data points obtained from the official Kaggle website. Subsequently, we assess forecasting accuracy using the Mean Absolute Percentage Error (MAPE), where a smaller percentage indicates superior performance. Our proposed methodology effectively mitigates the limitations associated with interval length determination and significantly improves forecasting accuracy. Specifically, the MAPE percentage for FTS-FCM before the optimization is 20.180%, while that of FCM-CMBO is notably lower at 18.265%. These results highlight the superior performance of the FCM-CMBO hybrid approach, which achieves a forecasting accuracy of 81.735% compared to actual data.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

N. Irsalinda, D. K. Laely and S. Surono,"FUZZY TIME SERIES BASED ON THE HYBRID OF FCM WITH CMBO OPTIMIZATION TECHNIQUE FOR HIGH WATERPREDICITON" *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1245-1256, September, 2023.

Copyright © 2023 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article • Open Access

1. INTRODUCTION

Forecasting is a process of estimating a future event by using present or previous data [1]. It usually uses time series data to observe the development of an object. To obtain good results, which are characterized by high accuracy and minimum error value, there is a need to select the correct method used.

A method used in forecasting is the Fuzzy Time Series (FTS), first proposed by Song and Chissom in 1993 and applied to a fuzzy logic concept to predict a scenario where historical data was formed into linguistic values [2]. However, the FTS method is very dependent on the interval length, thereby causing it to affect the results. To address this problem, the interval length formation can be optimized by using clustering techniques that were developed to minimize errors in the prediction process. An example of the clustering technique is Fuzzy C-Means (FCM) using the fuzzy logic concept introduced by Bezdek, in which the existence of each data point in a cluster is determined by the membership degree [3].

Several studies on prediction with FTS have been conducted. An example is a study performed by [4] to provide the latest development overview in the FTS field and identify opportunities that can be implemented in the future. The obstacle with the FTS method is determining and increasing the interval length. To overcome this problem, two techniques were proposed. One of them is the soft computing technique, which is classified into two categories, namely methods based on optimization and clustering [4].

In Cheng's study [5], the limitations associated with Fuzzy Time Series (FTS) were effectively addressed by employing Particle Swarm Optimization (PSO) to attain optimal interval partitioning within the matrix U. Additionally, the K-Means clustering technique was utilized to determine centroids, with the primary objective of enhancing forecasting accuracy. Similarly, in Nghiem Van Tinh's research conducted in 2020, the FTS method was combined with Fuzzy C-Means (FCM) and PSO techniques to further augment forecasting precision. The findings demonstrated that the FCM-PSO hybrid approach significantly improved FTS prediction accuracy. Furthermore, it suggested the potential for combining this method with other more potent optimization techniques to tackle more intricate challenges.

In different branches of science, several optimization problems need to be solved using the right techniques. One of the most important and practical techniques for solving optimization problems is population-based algorithms, which include 1) Genetic Algorithm (GA) based on Darwin's theory [6], 2) PSO by imitating the behavior of birds while fishing [7], and 3) Gravitational Search Algorithm (GSA) based on gravitational force relationships with Newton's laws of motion [8]. Meanwhile, some weaknesses were found in these techniques. For example, the GA requires quite a long time to reach the optimum [9], PSO only provides solutions quickly and locally but is unable to discover a wider solution space [10], and GSA still has poor convergence results [8].

Subsequently, Dehghani examined the problems in the previous optimizers and discovered a new population-based optimization technique, the Cat and Mouse Based Optimizer (CMBO). The results showed the CMBO's performance analysis is more competitive than the other nine algorithms by providing a quasi-optimal solution closer to the global optimum [9].

Based on the previous studies, the FCM was hybridized with CMBO optimization techniques in order to obtain better forecasting accuracy results. Furthermore, this method was tested using a water level dataset at an observation station on Achill Island, Ireland.

2. RESEARCH METHODS

The FTS forecasting method based on the FCM hybrid with the CMBO optimization technique is built in the following stages:

2.1 Manhattan Distance

Manhattan distance is a calculation method for measuring the absolute distance between two points [11]. This distance is formulated as follows [12]:

$$Manhattan\,(x, v) = \sum_{i=1}^{n}\sum_{j=1}^{c}\left|x_i - v_j\right| \tag{1}$$

where,
$Manhattan\ (x, v)$ represents the manhattan distance between x and v,
x_i is the data i,
v_j denotes the j cluster center,
n indicates lots of data,
c represents lots of clusters.

2.2 Fuzzy C-Means

The FCM was developed by Dunn in 1973 and then enhanced by Bezdek in 1981 with the principle that the clustering process is based on splitting a dataset into multiple clusters having similarities [13]. Furthermore, the FCM is a development of the K-Means method, which adds the degree of membership when a data value's degree of membership is between $[0, 1]$ [14]. It is important to note that the clustering method allows data to have two or more clusters, and one of them is the FCM. In this method, the objective function is considered the sum of the square error. This is because the goal of FCM is to minimize the following objective functions [3]:

$$P_t = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^m D(x_i, v_j)^2 \tag{2}$$

where,
P_t represents the value of the t iteration objective function,
$D(x_i, v_j)$ indicates the value of the distance between the i data to the j centroid,
μ_{ij} is the degree of membership of the matrix U row i column j,
m denotes the fuzzy $m \geq 2$ rank.

The membership degree of the matrix U was formed using the following equation [15]:

$$U = (\mu_{ij}): \sum_{j=1}^{c}\mu_{ij} = 1, 1 \leq i \leq n \tag{3}$$

$$(\mu_{ij}) \in [0, 1], 1 \leq i \leq n, 1 \leq j \leq c \tag{4}$$

In each iteration, the membership matrix U was updated with the equation below [15]:

$$\mu_{ij} = \left[\left(\frac{D(x_i, v_j)}{\sum_{j=1}^{c} D(x_i, v_j)}\right)^{\frac{2}{m-1}}\right]^{-1} \tag{5}$$

After forming the membership matrix U, the cluster center can be calculated based on the following equation [15]:

$$v_j = \frac{\sum_{i=1}^{n}(\mu_{ij})^m x_i}{\sum_{i=1}^{n}(\mu_{ij})^m} \tag{6}$$

2.3 Fuzzy Time Series

The FTS method is a forecasting system for determining the patterns from historical data, which is projected to obtain new data using the fuzzy logic concept [16]. The stages in the FTS process include defining the universe of U conversations, forming interval lengths, fuzzification, forming Fuzzy Logic Relationships (FLR) as well as Fuzzy Logic Relationship Groups (FLRG), and defuzzifying forecasting values.

Definition 2.3.1 [5]: Given $U = \{u_1, u_2, u_3, \dots, u_n\}$ is a universe of speech in which $u_n(i = 1, \dots, n)$ is a possible deep linguistic value in U. The fuzzy set A_i is defined as follows:

$$A_i = \frac{\mu_{A_i}(u_1)}{u_1} + \frac{\mu_{A_i}(u_2)}{u_2} + \cdots + \frac{\mu_{A_i}(u_n)}{u_n} \quad (7)$$

Where $\mu_{A_i}: U \to [0, 1]$ represents the membership function of the fuzzy set A_i, $\mu_{A_i}(u_i)$ indicates the degree of membership u_i into a fuzzy set A_i, $\mu_A(u_i) \in [0,1]$ and $1 \le i \le c$.

Definition 2.3.2 [5]: The relationship between $F(t)$ and $F(t-1)$ can be presented as $F(t-1) \to F(t)$. Given $F(t) = A_j$ and $F(t-1) = A_i$, then the relationship between $F(t)$ and $F(t-1)$ is represented with FLR $A_i \to A_j$, in which A_i and A_j are on the left and right side of FTS.

Definition 2.3.3 [5][17]: The fuzzy logic relationships satisfying the left-hand side are grouped into FLRG. Suppose there is an FLR as follows $A_i \to A_{j1}, A_{j2}, \ldots, A_{jn}$.

Before performing the fuzzification process, which transforms all data into a linguistic variable, the formation of interval lengths is obtained by determining the upper and lower limits using the following equation:

$$IntervalUB_i = \frac{cluster\ center_i\ + cluster\ center_{i+1}}{2} \quad (8)$$

$$IntervalLB_{i+1} = IntervalUB_i \quad (9)$$

Where $i = 1,2, \ldots, c-1$, $IntervalUB_i$ represents the upper limit of the i, and the $IntervalLB_{i+1}$ denotes the lower limit of the $i + 1$ *interval* because there is no cluster center before the first and after the last cluster center. The lower limit of $IntervalLB_1$ and the upper limit for $IntervalUB_c$ was calculated using the following formula:

$$IntervalUB_c = claster\ center_c + |max_{data} - claster\ center_c| \quad (10)$$

$$IntervalLB_1 = claster\ center_1 - |claster\ center_1 - min_{data}| \quad (11)$$

Subsequently, the mean of each interval is obtained with the following equation:

$$Middlevalue_i = \frac{IntervalLB_i\ + IntervalUB_i}{2} \quad (12)$$

The mean value of each interval was used in the defuzzification process of forecasting values based on the following three rules [18]:

Rule 1: If A_i has no FLR in FLRG, then the predicted value at $n+1$ is the middle value of u_n, written as $Middlevalue_n$.
Rule 2: If A_i has one FLR in FLRG, which is $A_i \to A_{jn}$, then the estimated value at $n+1$ is the middle value of u_n, denoted as $Middlevalue_n$.
Rule 3: If A_i has a lot of FLR in FLRG, denoted as $A_i \to A_{j1}, A_{j2}, \ldots, A_{jn}$, then the forecasted value at $n+1$ can be calculated using the following equation:

$$forecasted(n+1) = \frac{Middlevalue_1 + Middlevalue_2 + \cdots + Middlevalue_n}{n} \quad (13)$$

2.4 Cat and Mouse-Based Optimizer

The CMBO method is inspired by the cat's natural behavior of chasing mouse, which also look for shelter. In the CMBO algorithm, the population is divided into the cat and the mouse. Also, there are two search phases when updating population members, which include calculating the cat-approaching mouse and the movements of mice looking for hiding places. Each population member is a vector whose value determines the problem variable. The initial population of CMBO is in the form of a matrix called the population matrix [9].

The first step is forming an initial population containing the solutions to the problem variables and calculating the objective function.

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,c} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N.1} & w_{N.2} & \cdots & w_{N.c} \end{bmatrix} \tag{14}$$

where W is the population matrix containing the solutions to the problem variables and $w_{i,j}$ $(i = 1,2,\dots,N\,; j = 1,2,\dots,c)$ denotes an element of the population matrix Z. Furthermore, the objective function of each population member was calculated.

The second step is to sort the population members based on the smallest to the largest objective function value. The third step entails dividing the sorted population matrix into 2 2, with 50% of the mice and cats having the smallest and largest objective functions, respectively. The mice and cat population matrices were written as follows:

$$Z = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots & w_{1,c} \\ w_{2,1} & w_{2,2} & \cdots & w_{2,c} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Nz,1} & w_{Nz,2} & \cdots & w_{Nz,c} \end{bmatrix}_{Nz\times c} \tag{15}$$

$$E = \begin{bmatrix} w_{Nz+1,1} & w_{1,2} & \cdots & w_{Nz+1,c} \\ w_{Nz+1,1} & w_{2,2} & \cdots & w_{Nz+2,c} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Nz+Ne,1} & w_{Nz+Ne,2} & \cdots & w_{Nz+Ne,c} \end{bmatrix}_{Ne\times c} \tag{16}$$

where Z represents the mouse population matrix, Nz denotes the number of mice population rows, E indicates the cat population matrix, Ne denotes the number of cat population rows, while Z_i is the i-th mouse agent and E_j is the j cat agent.

The fourth step is updating the search factor, in which there are two phases, namely changing the cat's position by moving towards the mouse and changing the mouse's position by moving towards the hiding place.

Phase 1: The change in the cat's position by moving towards the mouse is formulated as follows:

$$E_j^{new}: e_{j,d}^{new} = e_{j,d} + r \times (z_{k,d} - I \times e_{j,d}) \tag{17}$$

Given $j = 1,\dots,Ne,\ d = 1,2,\dots,c,\ k = 1,2,\dots,Nz, and\ I = round(1 + rand)$

$$E_j = \begin{cases} E_j^{new} & , \quad F_j^{e,new} < F_j^e \\ E_j & , \quad other \end{cases} \tag{18}$$

Where E_j^{new} represents the new j agent cat's position, $e_{j,d}^{new}$ denotes the positioning element of the cat agent in the new matrix, r is the random number in the interval [0,1], and $F_j^{e,new}$ indicates the objective function of the new cat position j.

Phase 2: The change in the mouse's position by moving it toward the hiding place is formulated below:

$$H_i: h_{i,d} = w_{l,d}, \qquad i = 1,2,\dots,Nz, d = 1,2,\dots,c,\ \ l = 1,2,\dots,N \tag{19}$$

$$Z_i^{new}: z_{i,d}^{new} = z_{i,d} + r \times (h_{i,d} - I \times z_{i,d}) \times sign(F_i^z - F_i^h) \tag{20}$$

Given $i = 1,\dots,Nz,$ and $d = 1,2,\dots,c,$

$$Z_i = \begin{cases} Z_i^{new} & , \quad F_i^{z,new} < F_i^z \\ Z_i & , \quad other \end{cases} \tag{21}$$

where H_i is the i-th mice's nest, F_i^h denotes the objective function of the i-th mice's nest, Z_i^{new} represents the new position (i) of the mice agent, $z_{i,d}^{new}$ indicates the new mice position in matrix Z_i^{new}, and $F_i^{z,new}$ is the objective function of the new mouse position (i).

The final step involves repeating the second to the fourth stage, until it satisfies the stopping criterion, namely the maximum iteration. The moment this condition is met, the optimal CMBO solution is obtained based on the population member with the lowest objective value.

2.5 Forecasting Evaluation

According to Lewis 1982, MAPE is effective and widely used in evaluating forecasting methods by comparing its accuracy with the actual data [19]. When the MAPE value is below 10%, the process has a very good performance. Furthermore, when it lies between 10% and 20%, it is considered to have good performance, and when it is between 20% and 50%, the method has acceptable or pretty good performance. Meanwhile, when it is above 50%, the performance is regarded as poor [20]. The values can be calculated using the following formula.

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Y_t - F_t}{F_t} \right|}{n} \times 100\% \tag{22}$$

Where,
Y_t : actual value on t data,
F_t : the predicted value for the t data,
n : lots of data.

2.6 Data Description

The dataset is time series data on water levels obtained from the official Kaggle website. Water level data in Achill Island, Ireland, are recorded every 10 minutes from January 1, 2019, to 2020. Additionally, the total data is 52,562 lines with 2 columns, namely the time column and the water level figure in meters. Table 1 shows a subset of the water level dataset.

Table 1. Achill Island Water Level Dataset, Ireland

Time	Water Level
2019/01/01 (00:00:00)	0.15486111
2019/01/01 (00:10:00)	0.15833333
2019/01/01 (00:20:00)	0.16180556
…	…
2019/12/31 (00:50:00)	0.13680556
2020/01/01 (00:00:00)	0.13055556
2020/01/01 (00:10:00)	0.12430556

2.7 Flowchart of FTS-FCMCMBO

The integration of FTS and FCM results in a powerful clustering algorithm that takes advantage of the strengths of both techniques. FTS provides a flexible and robust framework for modeling time series data, while FCM enables efficient and accurate clustering of the data. The integration of FTS and FCM results in a powerful clustering algorithm that takes advantage of the strengths of both techniques. FTS provides a flexible and robust framework for modeling time series data, while FCM enables efficient and accurate clustering of the data. This study combines the metaheuristic optimizer to improve the FTS-FCM to find the best centroid. The flowchart of FTS-FCMCMBO is as follows.

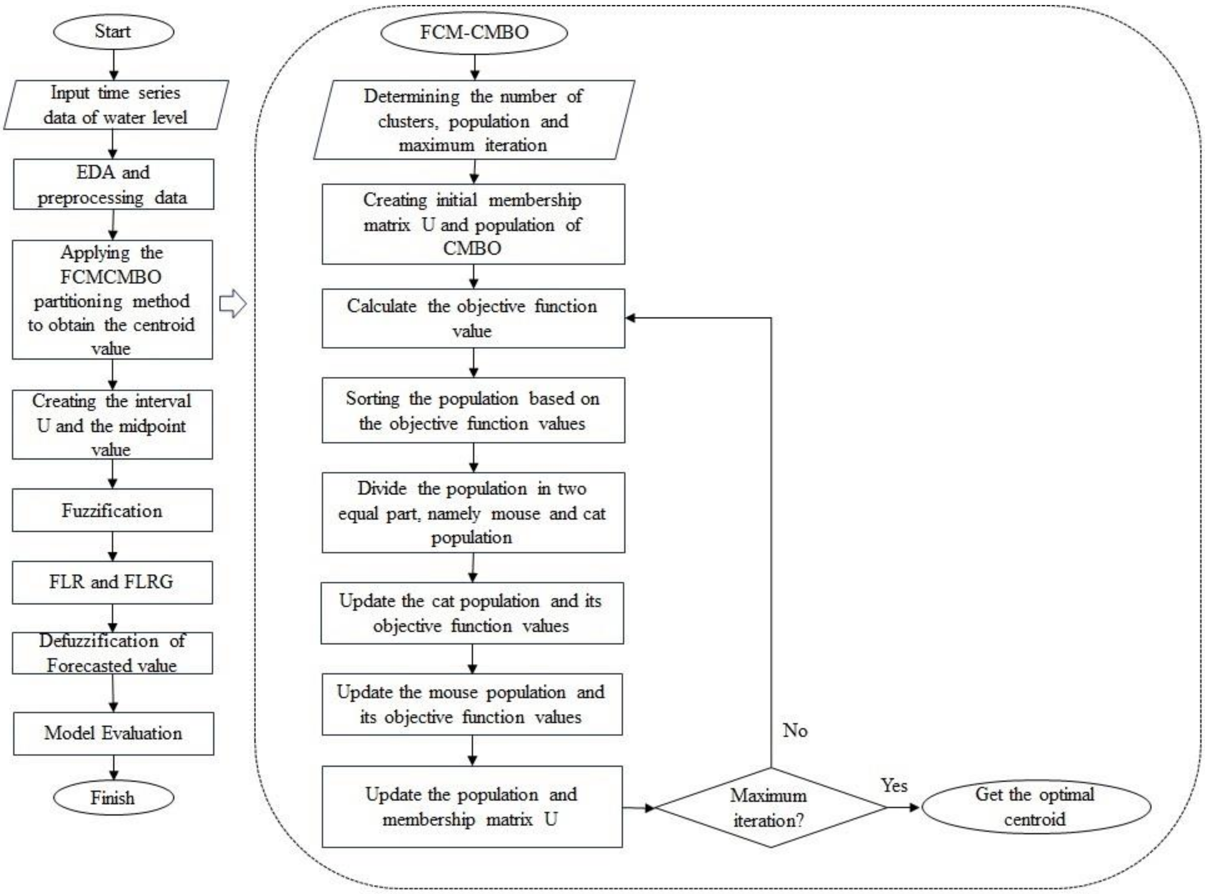

Figure 1. The Flowchart of FTS-FCMCMBO

3. RESULTS AND DISCUSSION

The process begins with the implementation of FCM with a CMBO optimizer in the water level dataset in order to obtain optimal centroid values used in forming intervals in the FTS method. Parameters used include eight clusters obtained with the elbow method, 10 search agents in the initial population, and 100 maximum iterations. Subsequently, the optimal centroid was determined with the FCCMBO method by forming an initial membership matrix and an initial CMBO population representing the centroid. The membership matrix U was formed based on **Equation (3)** and **Equation (4)** with a population size of $n \times c$. After this process, the objective function was calculated based on **Equation (2)** for each member as a search agent and membership matrix U that was formed. Furthermore, the objective functions obtained were sorted from the smallest to the largest values. **Table 2** shows the obtained population that has been sorted.

Table 2. Population Sorted

Search Agent	V_1	V_2	⋯	V_7	V_8	F_i
Agent 1	0.16687242	0.05859649	⋯	0.18206422	0.12865993	634.41105530
Agent 2	0.19879847	0.20055275	⋯	0.15148163	0.03641175	713.59658300
Agent 3	0.05118775	0.06441630	⋯	0.06920046	0.00971002	746.92885810
Agent 4	0.04324392	0.12965057	⋯	0.05952803	0.12838872	659.46512633
Agent 5	0.19552723	0.20632797	⋯	0.12143456	0.00479636	851.25602911
Agent 6	0.16774768	0.05928686	⋯	0.17580506	0.12473774	622.16079726
Agent 7	0.19719283	0.19162708	⋯	0.14529979	0.03480495	693.33393370
Agent 8	0.18169847	0.03053861	⋯	0.20795178	0.05608429	901.49197888
Agent 9	0.08655698	0.01617068	⋯	0.19653417	0.01830853	905.47692633
Agent 10	0.18514643	0.22465482	⋯	0.02608442	0.04395530	978.83992085

From the table above, the sorted population was divided into two parts, namely agents 1 to 5 and 6 to 10 representing mice and cats, respectively. It is important to mention that the two populations were updated based on the two phases of the search factor. Specifically, the phase of the cat's movement toward the mouse was updated based on Equation (17), and the change in its new position was based on Equation (18). Meanwhile, the mouse's movement toward the hiding place in phase 2 was updated using Equation (19), and Equation (20), and the change in its new position is based on Equation (21). Table 3 and Table 4 show the updated mouse and cat population results.

Table 3. Cat Population Update Results

Cat Agent	V_1	V_2	⋯	V_7	V_8	F_i
Cat 1	0.22318714	0.19659342	⋯	0.1244774	0.04081481	789.09895075
Cat 2	0.03658590	0.19780677	⋯	0.20744585	0.09504631	794.40850167
Cat 3	0.05654029	0.13659368	⋯	0.00688936	−0.00152005	749.17857613
Cat 4	0.04324392	0.12965057	⋯	0.05952803	0.12838872	659.46512633
Cat 5	0.19552723	0.20632797	⋯	0.12143456	0.00479636	851.25602911

Table 4. Results of Mice Population Updates

Mice Agent	V_1	V_2	⋯	V_7	V_8	F_i
Mice 1	0.16774768	0.05928686	⋯	0.17580506	0.12473774	622.16079726
Mice 2	0.19719283	0.19162708	⋯	0.14529979	0.03480495	693.33393370
Mice 3	0.05118775	0.06441630	⋯	0.06920046	0.00971002	746.92885810
Mice 4	0.03178021	0.17619944	⋯	0.01907969	0.16015119	731.17456894
Mice 5	0.09589207	0.21584720	⋯	0.21247943	0.19163788	779.27020824

In the tables above, the updated mice and cat populations were merged into a new population. Furthermore, the membership matrix U was also updated for each new population member based on Equation (5). Iterations were performed, beginning with the step of sorting the objective function value through the process of updating the membership matrix U and the new occupation until the maximum iteration of 100 is reached. This implies the repetition stops at the 100th iteration by selecting the optimal centroid of the search agent that has the smallest objective function value in the new population. Table 5 shows the optimal centroid results.

Table 5. The Optimal Centroid of The FCCMBO method

Clustering	Centroid
Cluster 1	0.039547
Cluster 2	0.059197
Cluster 3	0.067612
Cluster 4	0.102132
Cluster 5	0.129127
Cluster 6	0.139819
Cluster 7	0.167110
Cluster 8	0.184457

The optimal centroid obtained was used to form intervals on the FTS based on Equations (8)-(11), and the mean was calculated based on Equation (12). Table 6 shows the results of forming interval lengths and mean values.

Table 6. Intervals and Mean Values

Interval	Mean Values
$u_1 = [0.002083, 0.049372)$	0.025728
$u_2 = [0.049372, 0.063404)$	0.056388
$u_3 = [0.063404, 0.084872)$	0.074138
$u_4 = [0.084872, 0.115630)$	0.100251
$u_5 = [0.115630, 0.134473)$	0.125052
$u_6 = [0.134473, 0.153465)$	0.143969
$u_7 = [0.153465, 0.175784)$	0.164624
$u_8 = [0.175784, 0.226389)$	0.201086

After obtaining the interval and mean values, the fuzzification process was performed by changing the actual data included in the interval u_i into a fuzzy set A_i. Subsequently, the FLR and FLRG were formed based on definitions 2.2.2 and 2.2.3, respectively. **Table 7** and **Table 8** show the subset of fuzzification process results as well as the FLR and FLRG formation.

Table 7. Fuzzification and FLR Results

Date	AQI	Fuzzification	FLR
2019/01/01 (00:00:00)	0.15486111	A_7	–
2019/01/01(00:10:00)	0.15833333	A_7	$A_7 \rightarrow A_7$
2019/01/01 (00:20:00)	0.16180556	A_7	$A_7 \rightarrow A_7$
⋮	⋮	⋮	⋮
2019/12/31 (00:50:00)	0,13680556	A_6	$A_6 \rightarrow A_6$
2020/01/01 (00:00:00)	0.13055556	A_5	$A_6 \rightarrow A_5$
2020/01/01 (00:10:00)	0.12430556	A_5	$A_5 \rightarrow A_5$

Table 8. FLRG results

Group	FLRG
1	$A_1 \rightarrow A_1, A_2, A_3, A_4$
2	$A_2 \rightarrow A_1, A_2, A_3$
3	$A_3 \rightarrow A_1, A_2, A_3, A_4$
4	$A_4 \rightarrow A_1, A_3, A_4, A_5, A_6$
5	$A_5 \rightarrow A_4, A_5, A_6, A_7, A_8$
6	$A_6 \rightarrow A_4, A_5, A_6, A_7$
7	$A_7 \rightarrow A_5, A_6, A_7, A_8$
8	$A_8 \rightarrow A_5, A_7, A_8$

The final stage of predicting the FTS value was the defuzzification process. **Table 9** shows a portion of the entire data defuzzification results of the FCCMBO hybrid FTS forecasting value.

Table 9. Fuzzification and FLR results

Time	Water Level	Fuzzification	FLR	F_t
2019/01/01 (00:00:00)	0.15486111	A_7	–	–
2019/01/01 (00:10:00)	0.15833333	A_7	$A_7 \rightarrow A_7$	0.15868279
2019/01/01 (00:20:00)	0.16180556	A_7	$A_7 \rightarrow A_7$	0.15868279
⋮	⋮	⋮	⋮	⋮
2019/12/31 (00:50:00)	0.13680556	A_6	$A_6 \rightarrow A_6$	0.13347394
2020/01/01 (00:00:00)	0.13055556	A_5	$A_6 \rightarrow A_5$	0.13347394
2020/01/01 (00:10:00)	0.12430556	A_5	$A_5 \rightarrow A_5$	0.14699641

The FTS results of the hybrid FCCMBO forecasting were evaluated based on MAPE with a percentage of 18.265%. The graph for obtaining the MAPE percentage value can be seen in **Figure 2**.

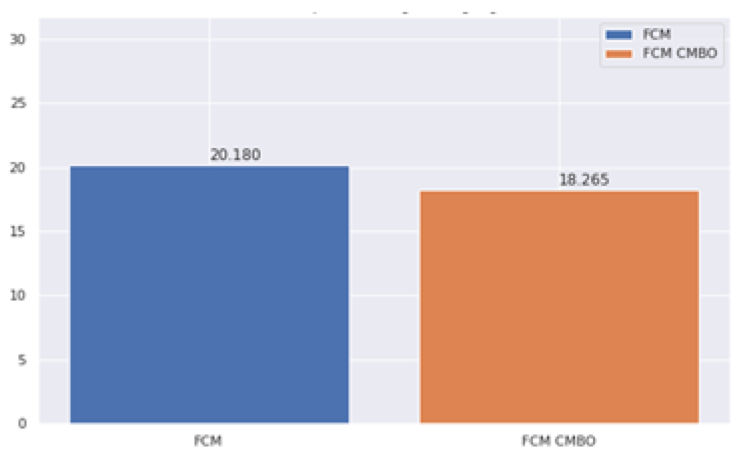

Figure 2. MAPE percentage chart

Figure 3 and **Figure 4** show the comparison graphs of actual data with the forecasting results of the hybrid FTS-FCMMBO and FTS-FCM.

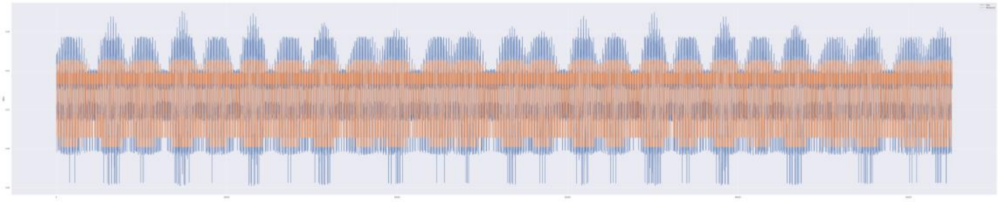

Figure 3. Graph of actual and forecasted FTS *hybrid* FCCMBO data

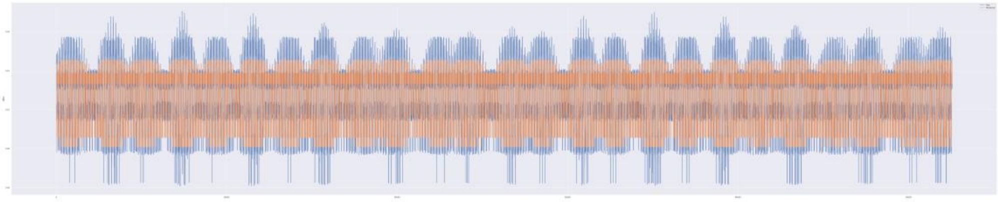

Figure 4. Graph of FTS-FCM actual and forecast data

According to **Figure 2** and **Figure 3**, the forecasting results of the hybrid FCM-CMBO method were closer to the actual data compared to that of FTS-FCM. Also, the MAPE percentage of the hybrid FCM-CMBO shows that the method used falls into the good performance criteria as its value was between 10% and 20%.

4. CONCLUSIONS

The hybrid FCM-CMBO forecasting method was applied to the water level dataset to search for optimal centroids. It is important to note that centroid values were used in determining interval lengths and improving forecasting accuracy in order to overcome the FTS weaknesses. The results were evaluated by comparing actual data with forecasted values using MAPE and graphs. The MAPE percentage and the comparison chart obtained showed that the method has good performance with a forecasting error rate of 18.265% and a percentage of 81.735% according to actual data. In further studies, stopping criteria need to be used in the iteration of the CMBO optimization method. This can be achieved by using the difference in the objective function value and trying multivariable cases.

ACKNOWLEDGEMENT

This work was supported by Lembaga Penelitian dan Pengabdian kepada Masyarakat Universitas Ahmad Dahlan under the grant number PD-074/SP3/LPPM-UAD/VII/2022.

REFERENCES

[1] M. Makridakis, Wheelwright., *Metode dan Aplikasi Peramalan*. Jakarta: Binarupa Aksara, 1999.

[2] Y. Safitri, S. Wahyuningsih, and R. Goejantoro, “Peramalan dengan Metode Fuzzy Time Series Markov Chain (Studi Kasus : Harga Penutupan Saham PT . Radiant Utama Interinsco Tbk Periode Januari 2011 – Maret 2017) Forecasting with Fuzzy Time Series Markov Chain Method (Case Study : Closing Stock Price of,” *J. Eksponensial*, vol. 9, no. 1, pp. 51–58, 2018.

[3] J. . Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[4] M. Bose and K. Mali, "Designing Fuzzy Time Series Forecasting Models: A Survey," *Int. J. Approx. Reason.*, pp. 78–79, 2019.

[5] N. Van Tinh, “Enhanced Forecasting Accuracy of Fuzzy Time Series Model Based on Combined Fuzzy C-Mean Clustering with Particle Swam Optimization,” *Int. J. Comput. Intell. Appl.*, vol. 19, no. 2, 2020.

[6] W. Zhang, S. Zhang, S. Zhang, D. Yu, and N. N. Huang, “A novel method based on FTS with both GA-FCM and multifactor BPNN for stock forecasting,” *Soft Comput.*, vol. 23, no. 16, pp. 6979–6994, 2019.

[7] N. Van Tinh and N. C. Dieu, “a New Hybrid Fuzzy Time Series Forecasting Model Based on Combining Fuzzy C-Means Clustering and Particle Swam Optimization,” *J. Comput. Sci. Cybern.*, vol. 35, no. 3, pp. 267–292, 2019.

[8] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, “GSA: A Gravitational Search Algorithm,” *Inf. Sci. (Ny).*, vol. 179, no. 13, pp. 2232–2248, 2009.

[9] M. Dehghani, Š. Hubálovský, and P. Trojovský, "Cat and mouse-based optimizer: A new nature-inspired optimization algorithm," *Sensors*, vol. 21, no. 15, pp. 1–30, 2021.

[10] J. Kennedy and R. Eberhart, "In Particle Swarm Optimization," *Proc. ICNN'95-International Conf. Neural Networks*, vol. 4, 1995.

[11] P. Bhatia, *Data Mining and Data Warehousing Principles and Practical Techniques*, vol. 47. Cambridge University Press, 2019.

[12] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, 2019.

[13] C. Bai, D. Dhavale, and J. Sarkis, "Complex investment decisions using rough set and fuzzy c-means: An example of investment in green supply chains," *Eur. J. Oper. Res.*, vol. 248, no. 2, pp. 507–521, 2016.

[14] B. Haviluddin, A. Fanany, and O. Gafar, "Proceedings of the Eleventh International Conference on Management Science and Engineering Management," *Proc. Elev. Int. Conf. Manag. Sci. Eng. Manag.*, vol. 2, 2018.

[15] S. Miyamoto, H. Ichihashi, and K. Honda, *Algorithms for Fuzzy Clustering Methods in C-Means Clustering with Applications*. Springer-Verlag Berlin Heidelberg, 2008.

[16] A. S. Wicaksana, B. D. Setiawan, and B. Rahayudi, "Algoritma Genetika Untuk Optimasi Fuzzy Time Series Dalam Memprediksi Kepadatan Lalu Lintas di Jalan Tol," vol. 2, no. 3, pp. 1063–1071, 2018.

[17] A. Konar and D. Bhattacharya, *An introduction to time-series prediction*, vol. 127. 2017.

[18] E. Bulut, O. Duru, and S. Yoshida, "A Fuzzy Time Series Forecasting Model for Multi-Variate Forecasting Analysis with Fuzzy C-Means Clustering," *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, vol. 6, no. 3, pp. 671–677, 2012.

[19] C. D. Lewis, *Industrial, and Business Forecasting Methods*. London: Butterworths, 1982.

[20] E. Ostertagová and O. Ostertag, "Forecasting using simple exponential smoothing method," *Acta Electrotech. Inform.*, vol. 12, no. 3, pp. 62–66, 2013.