Iranian Journal of Fuzzy Systems
Volume 17, Number 3, (2020), pp. 151-161

Interpolating time series based on fuzzy cluster analysis problem

V. V. Tai [1] and L. D. Nghiep[2]

[1] *College of Natural Science, Can Tho University, Can Tho city, Vietnam.*
[2] *College of Basis Science, Nam Can Tho University, Can Tho city, Vietnam.*

vvtai@ctu.edu.vn, nghiep1808@gmail.com

Abstract

This study proposes the model for interpolating time series to use them to forecast effectively for future. This model is established based on the improved fuzzy clustering analysis problem, which is implemented by the Matlab procedure. The proposed model is illustrated by a data set and tested for many other datasets, especially for 3003 series in M3-Competition data. Comparing to the existing models, the proposed model always gives the best result. We also apply the proposed model in forecasting the salt peak for a coastal province of Vietnam. Examples and applications show the potential of the studied problem.

Keywords: Cluster analysis, forecast, fuzzy time series model, interpolating data.

1 Introduction

Forecasting is the prediction the results for future based on the past data, knowledge and experience of the related problems. It is the scientific basis for plans and development strategies in all areas. Therefore, the forecast always receives the attention of managers, mathematicians and statistics. However, it is still a problem that has not been solved fully yet [1, 3, 29, 30]. In statisticians, using time series and regression models are common methods used to forecast. When constructing a regression model, we must constrain the conditions for data that are difficult to satisfy in reality. Although there are many different regression models with many continuous improvements, it often receives limited results in forecasting [2, 3, 14, 22].

Time series is a popular data type, which is stored in many fields. At the moment, it has a huge forecasting demand and has become an important research direction of statistics, attracting many interested scientists. Non-fuzzy time series models (NFTS) such as autoregressive, autoregressive integrated moving average have been commonly used in forecasting. To have an effective NTFS model, the time series must stop and the error must be a white noise. In fact, many series do not satisfy this condition, so NFTS model is limited in many cases [33, 34]. NFTS models were built on the association of data by mathematical expressions that are not language level, and this is the main disadvantage. Fuzzy time series (FTS) model has been proposed to solve this drawback.

FTS model is developed in two main directions. The first one is to build the model from the original data and directly use this model to forecast. Abbasov and Manedova [1] and Tai [33] had important contributions to this direction. The second one is to interpolate data in order to get the relation between elements in series, then to use this fuzzy data to forecast by the known forecasting models. This research has been of great interest by many statisticians. Song and Chissom [31] were the pioneer in this direction with data on enrollment of the University of Alabama (EnrollmentUA). Quang [26] used the triangular fuzzy relation for performing. Ming, Chen and Hsu [7, 23] improved the model of Qiang and Brad [26] when taking notice of fuzzy level. Huarng, Own and Yu [17, 25] presented a heuristic model for FTS using heuristic knowledge to improve the forecast for EnrollmentUA. Based on neural network, the model of Aladag *et al.* [4] gave the interesting results in some cases. From the fuzzy model in accordance with different linguistic levels, many scientists such as [15, 21, 28, 35] have proposed the new models. Recently, Tai [34] proposed a data fuzzy model based on cluster analysis problem.

Corresponding Author: V. V. Tai
Received: June 2019; Revised: September 2019; Accepted: November 2019.

In the second direction, a FTS model usually consists of three stages: (i) determining universal set, dividing intervals for universal set and finding the number of elements for each interval, (ii) building the fuzzy relationships, and (iii) interpolating for data. For (i), many authors used the values min and max of original data to divide the interval for a universal set [6, 7]. In addition, Huarng, Huarng and Yu [17, 18] proposed two new techniques for finding intervals based on the mean of the distributions. Abbasov and Manedova [1] have built the universal set based on the change of data between consecutive periods of time or their percentage change. Many authors divided the number of fuzzy set based on testing in many cases to find the suitable number for each case. This means that it is not a common rule for all cases. The number of fuzzy sets and their elements were also determined by the $k-$ mean [38] and the genetic algorithm [13]. A common feature of these studies have used the whole series to be a background. Then the relationship of an element with the whole series will be the basis of fuzzy data. We use fuzzy data to forecast without using the original data. Actually, when forecasting for the future, we only base on previous data. For this reason, taking the whole series as a background for interpolation is usually to give good results. However, It is limited for the forecast stage in our opinion. For (ii), several important studies have been performed. For instance, Song and Chissom [31] used matrix operations, Chen [6] took the fuzzy logic relations. Moreover, many authors in [4, 9, 10, 12, 18] used artificial neural networks to determine fuzzy relations. In addition, the fuzzy relationship based on the triangle and trapezoid fuzzy number was also considered in [15]. The relationship based on the fuzzy cluster analysis problem was also established by [34]. For (iii), many studies had used either the centroid method [6, 17, 18] or the adaptive expectation method [3, 7, 34] to perform.

This paper studies FTS model in the second direction, it makes contributions to all stages of (i), (ii) and (iii). For (i): to interpolate the mth element of the series, we use the previous $m-1$ elements to make the universe set without using the whole series. In our opinion, this will be more reasonable than existing models, then we use fuzzy data to forecast for the future. Efficiency is checked through a lot of well-know data sets. For (ii), we find the relationship of an element with others by the improved fuzzy cluster analysis problem. For (iii), based on the fuzzy relationship established from (ii), we suggested a new fuzzy solution method. Incorporating all these improvements, we have a new time series interpolating model (TSI) better than the existing ones through many well-know data sets. We also establish the Matlab procedure for the proposed model. This procedure can perform effectively the TSI model for real data. In addition, we also apply the proposed model to forecast salt peak for the rivers in the coast province in Vietnam.

The remainder of this paper is organized as follows. Section 2 introduces the related definitions and the proposed algorithms. Section 3 presents the numerical examples to illustrate the proposed model and compare it with others. Section 4 gives the application of the proposed model in forecasting the salt peaks of Cau Quan and Tra Vinh stations. Section 5 is the conclusion.

2 The proposed algorithm

2.1 Definitions

Definition 2.1. *Let U be universe set, $U=\{u_1,u_2,\ldots,u_n\}$. Fuzzy set A of U is defined as follows:*

$$A=\{\mu_A(u_1)/u_1,\mu_A(u_2)/u_2,....,\mu_A(u_n)/u_n\},$$

where $\mu_A(u_i)$ is the membership function, $\mu_A(u_i):U\to[0,1]$, $\mu_A(u_i)$ indicates the grade of membership of u_i in A, $\mu_A(u_i)\in[0,1]$, $1\le i\le n$.

Definition 2.2. *Let $X(t)$, $(t=1,2,\ldots)$ be universe set with fuzzy set $\mu_A(u_i),(i=1,2,\ldots)$ and $F(t)$ is a set of values $\mu_A(u_i)$, $(i=1,2,\ldots)$. Then $F(t)$ is called fuzzy time series (FTS) in $X(t)$.*

Definition 2.3. *Given a chain of historical data $\{X_i\}$ and the predictive value $\{\widehat{X}_i\}$, $i=1,2,\ldots,n$. Then, we have the following criteria for evaluating FTS models:*

Mean squared error:

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_i-X_i\right)^2. \quad (1)$$

Mean absolute error:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|\hat{X}_i-X_i\right|}{X_i}\right). \quad (2)$$

Mean absolute percentage error:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|\hat{X}_i - X_i\right|}{X_i}.100\right). \tag{3}$$

In forecasting, which model is better when the evaluation criteria are smaller.

2.2 The proposed algorithm

Let time series $x_i, i = 1, ..., n$. Based on the fuzzy cluster analysis problem, we propose the fuzzy time series model with the following steps:

Step 1. Considering $x_1, x_2, ..., x_{m-1}, 2 \leq m \leq n$ are the $m-1$ clusters and x_m is an element that needs to be inserted into the above clusters. At a time $t = 0$, establishing the original partition matrix $\mu^{(0)} = [\mu_{ij}]_{(m-1)\times m}$ with $\mu_{ij} = 1$ if the jth element belongs to the cluster x_i and $\mu_{ij} = 0$ for the opposite case, $j \neq m$. The last column of the matrix $\mu_{im}^{(0)} = \frac{1}{m-1}, i = 1, 2, ..., m-1$.

Step 2. Determining the representative element of each fuzzy set by the formula: $v_i = \frac{\sum_j^m {\mu_{ij}}^2 x_j}{\sum_j^m {\mu_{ij}}^2}$, where $1 \leq i \leq m-1$, μ_{ij} is probability of the jth element belonging to the cluster x_i.

Step 3. Updating the new partition matrix $\mu^{(t)}$ by the following rule:

$$\mu_{ij} = \begin{cases} \dfrac{1}{\sum\limits_{l=1}^{c} \left(d(v_i, v_j)/d(v_l, v_j)\right)^2}, & \text{if } d(v_j, v_i) > 0, \\ 0 & \text{if } d(v_j, v_i) = 0, \end{cases}$$

where $d(v_j, v_i)$ is the Euclidean distance from y_j to v_i.

Step 4. Computing the $S = \max_{ij}\left(\left|\mu_{ij}^{(t)} - \mu_{ij}^{(t-1)}\right|\right)$.

If $S < \varepsilon$, then the algorithm stops, opposite it repeats Step 2 and Step 3.

ε is an arbitrary small value that we choose $\varepsilon = 0.0001$ in this article.

Step 5. When the algorithm stops at Step 4, we have a partition matrix $\mu^{(t)} = [\mu_{ij}^{(t)}]_{(m-1)\times m}$. Then, the interpolation x_m is determined as follows:

$$\widehat{x}_m = \sum_{i=1}^{m-1} \mu_{im}^{(t)} x_i. \tag{4}$$

Continuing the interpolation for the values $x_{m+1}, x_{m+2}, \ldots, x_n$ then, we will get the whole series.

The steps of the proposed algorithm are illustrated by Figure 1.

3 The illustrative example and some comparisons

3.1 The illustrative example

In this section, we use the Foodgrain dataset to illustrate detail the steps of the proposed model. This dataset was used in many recent studies [15, 24, 27, 34]. Data are given in Table 1.

Step 1. Considering 7 initial elements $\{x_1, x_2, \ldots, x_7\}$ are 7 clusters and element x_8 is a new element that needs to be inserted into the above clusters. Establishing the original partition matrix $\mu^{(0)}$:

$$\mu^{(0)} = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,1/7 \\ 0\,1\,0\,0\,0\,0\,0\,1/7 \\ 0\,0\,1\,0\,0\,0\,0\,1/7 \\ 0\,0\,0\,1\,0\,0\,0\,1/7 \\ 0\,0\,0\,0\,1\,0\,0\,1/7 \\ 0\,0\,0\,0\,0\,1\,0\,1/7 \\ 0\,0\,0\,0\,0\,0\,1\,1/7 \end{bmatrix}$$

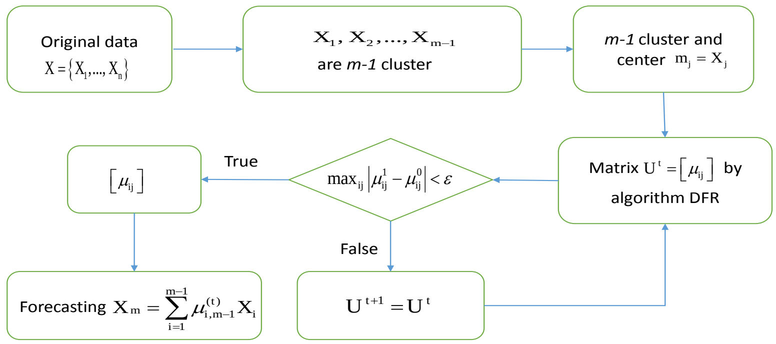

Figure 1: The diagram of the proposed algorithm

Table 1: Foodgrain dataset

Year	x_i	Year	x_i	Year	x_i
1966	74.2	1981	133.3	1996	199.4
1967	95.0	1982	129.5	1997	193.4
1968	94.0	1983	152.3	1998	203.6
1969	99.5	1984	145.5	1999	209.8
1970	108.4	1985	150.4	2000	196.8
1971	105.1	1986	143.4	2001	212.8
1972	97.0	1987	140.3	2002	174.7
1973	104.6	1988	169.9	2003	213.1
1974	99.8	1989	171.0	2004	198.3
1975	121.0	1990	176.3	2005	208.6
1976	111.1	1991	168.3	2006	217.2
1977	126.4	1992	179.4	2007	230.7
1978	131.9	1993	184.2	2008	234.4
1979	109.7	1994	191.5	2009	218.1
1980	129.5	1995	180.4	2010	244.7

Step 2. Determining the representative element of each fuzzy set $v_i = \{74.81; 95.20; 94.21; 99.60; 108.32; 105.09; 97.15\}$.
Step 3. Update the new partition matrix $\mu^{(t)}$ by (8). After 3 iterations, we have the matrix:

$$\mu^{(3)} = \begin{bmatrix} 0.99999 & 0.00000 \dots 0.00000 & 0.00007 \\ 0.00001 & 0.99999 \dots 0.00000 & 0.00067 \\ 0.00000 & 0.00001 \dots 0.00000 & 0.00055 \\ 0.00000 & 0.00000 \dots 0.00000 & 0.00238 \\ 0.00000 & 0.00000 \dots 0.00000 & 0.00428 \\ 0.00000 & 0.00000 \dots 0.00001 & 0.99098 \\ 0.00000 & 0.00000 \dots 0.99999 & 0.00107 \end{bmatrix}.$$

Step 4. From the results of $\mu^{(3)}$, we calculate $S = 0.000026 < \varepsilon = 0.0001$, so the algorithm stops.
Step 5. The 8th element is interpolated as follows:

$$X_8 = 0.00007 * 74.67 + 0.00067 * 95 + \ldots + 0.99098 * 105.09 + 0.00107 * 97.12 = 105.08.$$

Calculating similarly for the next years, we obtain Table 2 and it is shown by Figure. 2.

Table 2: The result of the proposed model for Foodgrain data

Year	x_i	Proposed	Year	x_i	Proposed
1966	74.2	-	1988	169.9	152.30
1967	95.0	-	1989	171.0	169.75
1968	94.0	-	1990	176.3	171.00
1969	99.5	-	1991	168.3	169.90
1970	108.4	-	1992	179.4	175.43
1971	105.1	-	1993	184.2	179.40
1972	97.0	-	1994	191.5	184.20
1973	104.6	105.08	1995	180.4	179.30
1974	99.8	99.50	1996	199.4	191.41
1975	121.0	105.10	1997	193.4	191.10
1976	111.1	108.40	1998	203.6	198.33
1977	126.4	119.82	1999	209.8	203.60
1978	131.9	126.23	2000	196.8	196.72
1979	109.7	109.02	2001	212.8	208.57
1980	129.5	129.94	2002	174.7	175.93
1981	133.3	131.63	2003	213.1	212.78
1982	129.5	129.50	2004	198.3	198.76
1983	152.3	133.30	2005	208.6	209.52
1984	145.5	140.33	2006	217.2	213.10
1985	150.4	151.56	2007	230.7	213.10
1986	143.4	144.93	2008	234.4	228.72
1987	140.3	143.40	2009	218.1	216.99
			2010	244.70	234.40

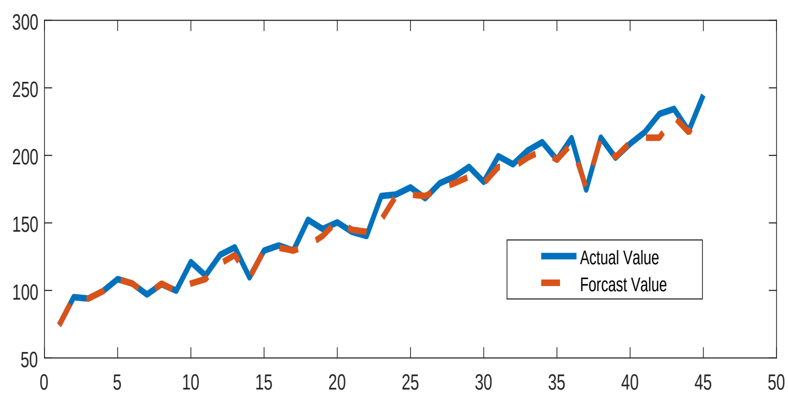

Figure 2: The original and forecasted values for Foodgrain data

Figure 2 shows that the actual and forecasted values of the proposed model almost coincident.

3.2 Comparing the interpolating result

a) *Comparing with the Foodgrain data set:*
In this section, we use the Foodgrain data set to make the comparison the effectiveness of the proposed model with existing fuzzy models such as L-C [21], Hua [17], A-M [1], Si [28], Gh [15], C-H [7], Y-H [36], Tai [33], C-K [8], B-R [5], Chen [6], Yus [37], Egr [11], Kha [19], Tai [34] based on the parameters MAE, MAPE and MSE. Their results are given in Table 3:

b) *Comparing with M3 - Competition data set:*
We use M3-Competition data to compare the results of models. This dataset is knowledge data in studying time series model that was set up by [32]. Entrants had to forecast 3003 time series. 3003 series of M3-Competition is selected

Table 3: Comparing the interpolating results of models for Foodgrain data

Criteria	L-C	Hua	A-M	Si	Gh	C-H	Y-H	Tai
MAE	47.76	58.64	89.6	8.69	8.17	107.71	67.23	60.35
MAPE	6.47	4.53	5.81	5.43	4.98	7.01	5.96	4.55
MSE	175.43	4772.00	10672.00	104.25	123.45	183.56	2987.15	6460.00

Criteria	C-K	B-R	Chen	Yus	Egr	Kha	Tai	Proposed
MAE	7.45	7.95	16.181	109.15	6.98	5.98	4.99	4.61
MAPE	5.21	6.62	10.13	7.57	5.09	4.87	3.93	2.72
MSE	2345.21	124.07	440.26	256.57	123.08	98.28	60.10	47.31

based on basic criteria including many types of time series data (Very small, industrial, very big,...) and different times between continuous observations (every year, quarterly,...). In order to ensure that enough data were available to develop an adequate forecasting model, it was decided to have a minimum number of observations for each type of data. This minimum was set as 14 observations for yearly series (the median length of the 645 year series is 19 observations), 16 observations for quarterly series (the median length of the 756 quarterly series is 44 observations), 48 observations for monthly series (the median length of the 1428 monthly series is 115 observations) and 60 for another series (the median length of the 174 other series is 63 observations). All the data are publicly available in the Mcomp package for R software. The considered important models are ForecastPro, ForecastX, BJ automatic, Autobox1, Autobox2, Autobox3, Hybrid, ETS and AutoARIMA (see https://robjhyndman.com/m3comparisons.R).

Using the proposed model and the above existing models, we perform for each series of the M3-Competition. For the proposed model, we use the established Matlab procedure to run and using the Rob's M3comp package to perform the existing models. The result is given by Table 4.

Table 4: Comparing the result of models for the M3-competition data set

Methods	MAPE
ForecastPro	18.00
ForecastX	17.35
BJ Automatic	19.13
AutoBox1	18.23
Autobox2	20.36
Autobox3	19.31
ETS	17.38
AutoARIMA	18.92
Hybrid	17.59
Proposed model	6.354

Table 4 shows that the advantage of proposed model is more than the existing models. With the large numbers of the considered time series and the different features of the M3-competition data set, this comparison is very meaningful to evaluate the advantages of the proposed model.

3.3 Comparing the forecasting result

In this section, we use 5 data sets: Foodgrain, Taifex, Outpatient, Lahi, and EnrollmentUA to compare the effectiveness of using the original data and the fuzzy data when applying the popular forecasting models. These are the popular data sets used to compare the effectiveness of time series models in many articles [29, 16, 20]. It is performed as follows:

Dividing each data into two parts: seventy five percent of them are used as the training set to build the models and about twenty five percent of the remaining data is used as the validation set. For the training set, the ARIMA, NFTS and Abbasov-Manenova (AM) models with original data (ARIMAR, NFTSR, AMR), ARIMA, NFTS, and AM models with fuzzy data of the proposed method (ARIMAP, NFTSP, AMP) are established. Using the established models from the training set to forecast for the validation set. The result is present in Table 5.

Table 5 shows that the results when using data from the proposed model for the test set are better than ones when using the original data with all cases.

In short, from theory and comparison the proposed with other ones, we have the following comments:

Table 5: Result of the forecasting models for validation set of five datasets

Data	Model	MAE	MAPE	MSE
Foodgrain	AMR	15.77	7.91	404.69
	ARIMAR	12.91	6.33	281.57
	AMP	13.36	6.66	292.77
	ARIMAP	20.41	9.16	595.70
	NFTSR	20.88	10.30	589.72
	NFTSP	14.06	7.06	329.91
Taifex	AMR	79.00	1.16	7117.50
	ARIMAR	79.61	1.17	7740.10
	AMP	11.50	0.17	162.44
	ARIMAP	42.40	0.62	2246.21
	NFTSR	88.50	1.30	8708.75
	NFTSP	39.78	0.58	1609.17
Outpatient	AMR	930.75	19.52	1303655.09
	ARIMAR	335.57	7.09	195066.15
	AMP	527.58	9.88	390820.60
	ARIMAP	410.73	8.46	244203.00
	NFTSR	939.77	19.68	1304506.00
	NFTSP	1012.31	21.10	1439333.00
Lahi	AMR	351.51	37.96	132401.90
	ARIMAR	187.04	19.81	42759.88
	AMP	284.10	31.80	104595.20
	ARIMAP	75.34	7.78	10469.55
	NFTSR	360.28	38.93	138643.30
	NFTSP	360.09	40.23	151581.10
Enrollments	AMR	1785.28	9.39	3326909.30
	ARIMAR	742.27	3.93	901655.37
	AMP	935.66	4.89	1107895.00
	ARIMAP	564.49	2.96	374186.90
	NFTSR	2082.71	10.95	4528381.00
	NFTSP	890.11	4.65	1021535.00

i) The fuzzy principle of an element in series is based on the relationship of this element to the previous elements, without using the behind elements. While many existing algorithm, rely on both the previous elements and the behind elements.

ii) Besides the idea of using the fuzzy cluster analysis for time series data, it is important to identify the specific parameter for each time series. In the article, we investigated many common series (more than 3000 series) with different number of elements, fields, and characteristics to choose the appropriate parameter m of Step 1. The parameter is chosen so that MSE, MAE, and $MAPE$ obtain the optimal values.

iii) The interpolating principle (Step 5) of the proposed model is simple but it is effective: The fuzzy value is the weighted average of past values, in which weights are the probabilities determined from Step 1, Step 2, Step 3, and Step 4.

iv) The proposed algorithm is implemented quickly and efficiently by the established Matlab procedure.

4 Application in forecasting the salt peak

In this section, we forecast the salt peaks for two main measuring stations of Tra Vinh, a coastal province of Vietnam. In the past years, Tra Vinh province has been heavily affected by salt intrusion. Salt intrusion here takes place very complicated that scientists cannot forecast accurately. For this reason, we forecast salt peaks for the main two stations located on the two rivers of Tra Vinh. They are called Tra Vinh and Cau Quan stations. The forecasting results will be the scientific basis for solutions to limit the harmful effects of salt intrusion. Data are given in Table 6.

The purpose of this application is to find the most suitable model to forecast the salt peak for the main stations in Tra Vinh province. This problem is implemented by two phases:

Table 6: Salt peak at two main stations in Tra Vinh province in the period of 2002-2017

Year	Tra Vinh	Proposed	Cau Quan	Proposed
2002	7.9	-	6.3	-
2003	11.3	-	7.9	-
2004	8.3	-	10.6	-
2005	10.7	-	10.9	-
2006	9.0	-	9.7	-
2007	9.5	-	8.9	-
2008	9.9	-	10.0	-
2009	9.9	9.90	6.3	6.30
2010	10.8	10.70	11.8	10.90
2011	11.1	11.15	8.3	8.11
2012	9.1	9.02	9.10	8.93
2013	12.4	11.30	9.2	9.00
2014	10.0	9.90	5.9	6.39
2015	8.9	9.00	8.5	8.38
2016	10.7	10.70	10.4	10.53
2017	11.7	11.30	11.5	11.58

Phase 1. *Finding the best model:* Dividing the data into two parts: training and test set with 80% and 20%, respectively. Interpolating the training set with the proposed model. Using predictive models for the future, such as ARIMA, AM, NFTS to forecast for the test set from the original data, interpolated data received from the proposed model. The MAE, MAPE and MSE parameters are used to compare the effectiveness of the models.

Phase 2. *Forecasting of salt peak in future:* Using the best model received from Phase 1, forecasting the salt peak for two stations to 2022.

Performing Phase 1, we have Table 7:

Table 7 shows that using predictive models from fuzzy data will yield better results from the original data. Because

Table 7: Comparing the result of models for test sets of two stations

Data	Model	MAE	MAPE	MSE
Tra Vinh	AMR	2.39	25.84	8.14
	ARIMAR	3.11	31.66	12.06
	AMP	2.55	22.97	9.12
	ARIMAP	1.96	17.86	5.01
	NFTSR	11.16	105.10	133.40
	NFTSP	2.97	27.24	11.38
Cau Quan	AMR	3.07	42.9	12.26
	ARIMAR	1.91	17.53	5.17
	AMP	0.47	4.97	0.27
	ARIMAP	1.95	17.79	5.74
	NFTSR	2.79	27.08	8.34
	NFTSP	0.56	6.34	0.62

ARIMAP model gives the best results for the Tra Vinh station and AMP model gives the best result for Cau Quan station, we use them to forecast the future. For Phase 2, the results of forecasting are presented in Table 8 and shown by Figure 3 and Figure 4.

Figure 3 and Figure 4 show the actual value and forecast for the period 2002-2017 almost matching. In the future, the saline peak at both stations are high, but there is no much fluctuation.

5 Conclusion

Based on establishment the fuzzy relationship between an element in series at a time with previous elements, this article has proposed the fuzzy time series model. The proposed model is illustrated in detail by the steps from the numerical

Table 8: The result of forecasting the salt peak for two stations to 2022

Year	Cau Quan	Tra Vinh
2018	9.33	9.82
2019	9.10	10.18
2020	9.26	10.24
2021	9.14	10.25
2022	9.23	10.25

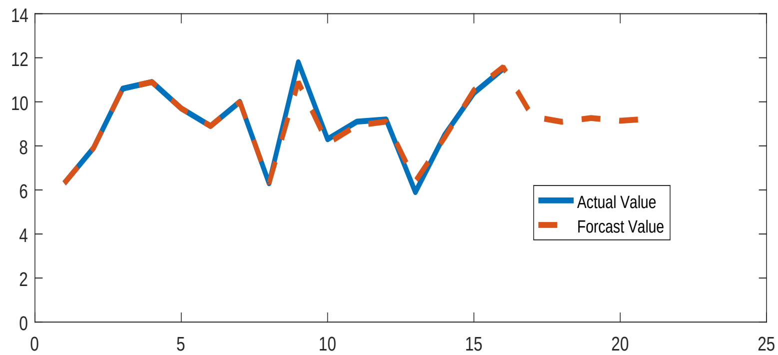

Figure 3: The graph of actual and forecasted value for the Cau Quan station

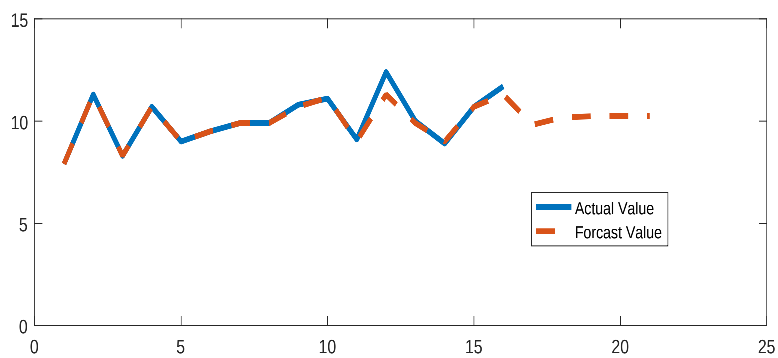

Figure 4: The graph of actual and forecasted value for the Tra Vinh station

example and can be quickly implemented by the established Matlab procedure. Performing for a lot of series with the different numbers and characteristics, the proposed model has shown advantages in comparison with the existing ones. The practical application of the article may apply similarly to many other real issues. It also shows the potential of the researched problems. In the future, we will apply the proposed model to forecast many urgent real problems in climate changes. This model does not distinguish seasonal data or non-seasonal data in performing. This is an advantages of the proposed model because it can apply for all data. However, it is also disadvantage because model may not be suitable for seasonal data in some cases. Thus, our further studies will focus on this problem in the next time.

References

[1] A. Abbasov, M. Mamedova, *Application of fuzzy time series to population forecasting*, Vienna University of Technology, **12** (2003), 545-552.

[2] P. H. Abreu, D. C. Silva, J. Mendes Moreira, L. P. Reis, J. Garganta, *Using multivariate adaptive regression splines in the construction of simulated soccer team's behavior models*, International Journal of Computational Intelligence, **6**(5) (2013), 893-910.

[3] S. Aladag, C. H. Aladag, T. Mentes, E. Egrioglu, *A new seasonal fuzzy time series method based on the multiplicative neuron model and SARIMA*, Hacettepe Journal of Mathematics and Statistics, **41**(3) (2012), 337-345.

[4] C. H. Aladag, M. A. Basaran, E. Egrioglu, U. Yolcu, V. R. Uslu, *Forecasting in high order fuzzy times series by using neural networks to define fuzzy relations*, Expert Systems with Applications, **36**(3) (2009), 4228-4231.

[5] G. Bindu, G. Rohit, *Enhanced accuracy of fuzzy time series model using ordered weighted aggregation*, Applied Soft Computing, **48** (2016), 265-280.

[6] S. M. Chen, *Forecasting enrollments based on fuzzy time series*, Fuzzy Sets and Systems, **81**(3) (1996), 311-319.

[7] S. M. Chen, C. Hsu, *A new method to forecast enrollments using fuzzy time series*, International Journal of Applied Science and Engineering, **2** (2004), 3234-3244.

[8] S. M. Chen, P. Y. Kao, *TAIEX forecasting based on fuzzy time series particle swarm optimization techniques and support vector machines*, Information Sciences, **247** (2013), 62-71.

[9] E. Egrioglu, C. Aladag., U. Yolcu, M. Basaran, V. Uslu, *A new hybrid approach based on SARIMA and partial high order bivariate fuzzy time series forecasting model*, Expert Systems with Applications, **36**(4) (2009a), 7424-7434.

[10] E. Egrioglu, C. Aladag, U. Yolcu, V. Uslu, M. A. Basaran, *A new approach based on artificial neural networks for high order multivariate fuzzy time series*, Expert Systems with Applications, **36**(7) (2009b), 10589-10594.

[11] S. Egrioglu, E. Bas, C. H. Aladag, U. Yolcu, *Probabilistic fuzzy time series method based on artificial neural network*, American Journal of Intelligent Systems, **62** (2016), 42-47.

[12] E. Egrioglu, V. Uslu, U. Yolcu, M. Basaran, C. Aladag, *A new approach based on artificial neural networks for high order bivariate fuzzy time series*, Applications of Soft Computing, (2009c), 265-273.

[13] B. Eren, R. Vedide, E. Erol, *A modified genetic algorithm for forecasting fuzzy time series*, Applied Intelligence, **41**(2) (2014), 453-463.

[14] J. H. Friedman, *Multivariate adaptive regression splines*, The Annals of Statistics, (1991), 1-67.

[15] H. Ghosh, S. Chowdhury, S. Prajneshu, *An improved fuzzy time-series method of forecasting based on L–R fuzzy sets and its application*, Journal of Applied Statistics, **43**(6) (2016), 1128-1139.

[16] H. Goossens, M. Ferech, R.Vander Stichele, M. Elseviers, ESAC Project Group, *Outpatient antibiotic use in Europe and association with resistance: A cross-national database study*, The Lancet, **365**(9459) (2005), 579-587.

[17] K. Huarng, *Heuristic models of fuzzy time series for forecasting*, Fuzzy Sets and Systems, **123**(3) (2001), 369-386.

[18] K. Huarng, T. H. K. Yu, *Ratio-based lengths of intervals to improve fuzzy time series forecasting*, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), **36**(2) (2006), 328-340.

[19] M. Khashei, M. Bijari, C. S. R. Hejazi, *An extended fuzzy artificial neural networks model for time series forecasting*, Iranian Journal of Fuzzy Systems, **3** (2011), 45-66.

[20] K. I. Hong, et al. *Forecasting TAIFEX based on fuzzy time series and particle swarm optimization*, Expert Systems with Applications, **37**(2) (2010), 1494-1502 .

[21] H. S. Lee, M. T. Chou, *Fuzzy forecasting based on fuzzy time series*, International Journal of Computer Mathematics, **81**(7) (2004), 781-789.

[22] P. A. Lewis, J. G. Stevens, *Nonlinear modeling of time series using multivariate adaptive regression splines (mars)*, Journal of the American Statistical Association, **86**(416) (1991), 864-877.

[23] C. S. Ming, *Forecasting enrollments based on high-order fuzzy time series*, Fuzzy Sets and Systems, **33**(1) (2002), 1-16.

[24] S. Narayanan, *Food security in India: The imperative and its challenges*, Asia, the Pacific Policy Studies, **2**(1) (2015), 197-209.

[25] C. M. Own, P. T. Yu, *Forecasting fuzzy time series on a heuristic high-order model*, Cybernetics and Systems: An International Journal, **62**(1) (2005), 1-8.

[26] S. Qiang, C. Brad, *Forecasting enrollments with fuzzy time series-part II* , Fuzzy Sets and Systems, **62**(1) (1994), 1-8.

[27] R. Selvaraju, *Impact of El Niosouthern oscillation on Indian foodgrain production*, International Journal of Climatology, **23**(2) (2003), 187-206.

[28] S. Singh, *A simple method of forecasting based on fuzzy time series*, Applied Mathematics and Computation, **186**(1) (2007), 330-339.

[29] S. R. Singh, *A computational method of forecasting based on fuzzy time series*, Mathematics and Computers in Simulation, **79**(3) (2008), 539-554.

[30] P. Singh, *Rainfall and financial forecasting using fuzzy time series and neural networks based model*, International Journal of Machine Learning and Cybernetics, **9**(3) (2018), 491-506.

[31] Q. Song, B. S. Chissom, *Forecasting enrollments with fuzzy time series-Part I*, Fuzzy Sets and Systems, **54**(3) (1993), 269-277.

[32] M. Spyros, H. Michle, *The M_3-competition: Results, conclusions and implications*, International Journal of Forecasting, **16**(4) (2000), 451-476.

[33] V. V. Tai, *An improved fuzzy time series forecasting model using variations of data*, Fuzzy Optimization and Decision Making, (2018), 1-23.

[34] V. V. Tai, L. D. Nghiep, *A new fuzzy time series model based on cluster analysis problem*, International Journal of Fuzzy Systems, **21**(3) (2019), 852-864.

[35] H. J. Teoh, C. H. Cheng, H. H. Chu, J. S. Chen, *Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets*, Data and Knowledge Engineering, **67**(1) (2008), 103-117.

[36] H. K. Yu, K. Huarng, *A neural network-based fuzzy time series model to improve forecasting*, Expert Systems with Application, **37** (2010), 3366-3372.

[37] S. M. Yusuf, A. Mohammad, A. A. Hamisu, *A novel two-factor high order fuzzy time series with applications to temperature and futures exchange forecasting*, Nigerian Journal of Technology, **36**(4) (2017), 1124-1134.

[38] Z. Zhiqiang, Z. Qiong, *Fuzzy time series forecasting based on k-means Clustering*, Open Journal of Applied Sciences, **2**(1) (2012), 100-103.