Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 73(2) 2019 111–120

# A New Approach for Spatio-Temporal Modelling and Forecasting based on Fuzzy Techniques in conjunction with K-means clustering

**Amit Saha[1], K.N. Singh[1], Mrinmoy Ray[1], Sanjay Kumar[2] and Santosha Rathod[3]**
*[1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
*[2]Govind Ballabh Pant University of Agriculture and Technology*
*[3]ICAR- Indian Institute of Rice Research, Hyderabad*

*Received 01 January 2019; Revised 14 March 2019; Accepted 19 March 2019*

## SUMMARY

The importance of spatio-temporal modeling is increasing day by day in many areas with the growing accessibility of spatio-temporal data. The spatio-temporal time series data exhibits spatio-temporal relationships among themselves. Therefore, it is essential to recognize those relations and incorporate it to the models for getting better forecast. On the other hand, forecasting based on fuzzy techniques are very much useful for the imprecise data like rainfall, temperature etc. In the last three decades, various univariate methods have been developed for forecasting the time series based on fuzzy techniques. Since these methods are only valid for the single time series, it cannot utilize for spatio-temporal time series. To overcome this limitation, a new forecasting method has been developed using fuzzy techniques in conjunction with the k-means clustering. The proposed method has been empirically illustrated for forecasting the annual precipitation data of six districts of West Bengal. The results from this study, reveals the superiority of the proposed approach as compared to classical univariate Chen's method.

*Keywords:* Fuzzy technique, K-means clustering, Rainfall, Spatio-temporal time series.

## 1. INTRODUCTION

Indian agriculture is mostly dependent upon rainfall rather than the irrigation. So that, it is necessary to forecast the rainfall for proper future planning of cropping pattern through capturing the trend and variability of the data. Many methods have been applied to forecast the rainfall based on time series. In time series forecasting, time series data are taken as crisp values. However, data may not be precise and complete in all the cases, viz.,rainfall data, water level data of river, temperature data etc. Fuzzy techniques are appropriate in those cases when vagueness has been seen in the data. Fuzzy data can be found in artificial intelligence, quality control, biology, psychometry, agriculture, social economy, image recognition etc. Fuzzy time series model can improve the utilization of the data. Fuzzy techniques are applied in following conditions where-

1. People's decisions are involved.
2. Data are imprecise.
3. Assumptions of distribution are not satisfied.
4. Number of observations for time series models are less.
5. The data is linear or non-linear or combination of both.

Other than the above things, it also has some key features like simplicity, readability, manageability and scalability. Many time invariant and time variant fuzzy time series model have been used in literature. Song and Chissom (Song and Chissom, 1993a, Song and Chissom, 1993b and Song and Chissom, 1994) proposed many fuzzy based time series models. Chen (1996) presented a new model based on fuzzy time series for forecasting the enrollment of university of Alabama. Many methodological aspects related to fuzzy set theory have been discussed in series of paper

*Corresponding author:* Amit Saha
*E-mail address:* amitsahauni@gmail.com

(Zadeh, 1965; Klir *et al.*, 1988; Kaufmann and Gupta, 1988; Klir *et al.*, 1995; Kandel, 1996; Lee and Chou, 2004;Singh, 2007; Liu, 2007; Qu and Chen, 2012; Kuo *et al.*, 2015 and Lee *et al.*, 2017).

In recent times, spatio-temporal time series forecasting is gaining lot of importance, as those time series involves spatial dependencies and univariate time series data depicts only temporal dependencies of the data. Thespatio-temporal time series analysis incorporates more information as compared to time series analysis alone in many domains including agriculture, traffic, meteorology, economics, disease mapping, sociology, environmental sciences, ecological sciences etc. Spatial weight grid incorporates the information of different location on a particular location. For example, the weather of Delhi not only depends on Delhi but its' adjacent areas like Punjab, Haryana, Rajasthan, U.P. etc. Examples of spatio-temporal time series data includes river flow data which is recorded daily on different river basins, carbon or Sulphur di oxide emission data of many location, daily temperature data over the various place, and annual precipitation data of different regions.

In this study, an attempt has been made to forecast spatio-temporal series by employing fuzzy approach. There are many fuzzy time series method in literature for forecasting time series data as said in earlier. However, fuzzy approach has not been yet used in spatio-temporal series for forecasting. So, there was a need to develop fuzzy techniques in spatio-temporal time series data. Hence, a new fuzzy approach has been proposed for spatio-temporal forecasting based on well-known k-means clustering.

## 2. MATERIALS AND METHODS

### 2.1 Data Description

Annual precipitation data of six districts viz., South Dinajpur (SD), Malda (MAL), North Dinajpur (ND), Darjeeling (DAR), Jalpaiguri (JAL) and Cooch behar (COO) of northern part of West Bengal from 1989 to 2010 has been collected from the India water portal (http://www.indiawaterportal.org/) and dairy knowledge portal under National Dairy Development Board (http://www.dairyknowledge.in/). The first 20 observations (around 90%) were utilized for model building and the last 2 data (around 10%) was used for testing purpose.

### 2.2 Chen's fuzzy time series forecasting models

A very well-known and renowned method of fuzzy time series forecasting is the Chen's method (1996).

The steps of Chen's methodsare:-

Step-1: The universe of discourse need to be defined and divide it into equal parts with the equal length.

Step-2: The fuzzy variable on the universe of discourse will be defined.

Step-3: Fuzzification of the data.

Step-4: Identification of the logical relationships among the fuzzy set has to be made.

Step-5: Setting up the fuzzy logical relational groups.

Step-6: Finally, defuzzification of forecasted value.

### 2.3 The Proposed spatio-temporal approach

The step by step procedure of the proposed algorithm has been described in the following. Steps are-

Step-1: Choose the spatio-temporal time series datasets.

Step-2: Formation of the weight matrices.

- Check for the spatio-temporal time series correlation using weight matrices through space-time series Box-Pierce test (Pfeifer and Deutsch, 1980 and Pfeifer and Deutsch, 1981).
- If, there is significant spatio-temporal correlation, then go for further step, otherwise not.

Step-3: Divide the time series datasets of each region along with its' neighboring areas by K-means clustering algorithm. Now, for each cluster; the following step will be followed.

Step-4: Define Universe of discourse.

Step-5: Make division of the universe of discourse into intervals of equal length.

Step-6: Define the fuzzy sets for the intervals using the triangular membership function.

Step-7: Fuzzification of dataset using triangular membership function to get the fuzzified input.

Step-8: Identification of fuzzy relations among the fuzzified input set.

Step-9: Computation of the fuzzy logical relationships groups.

Step-10: Various relations $(R_1, R_2, \ldots, R_i)$ are to be made based on fuzzy logical relational groups.

Step-11: Getting fuzzified forecasted output using the following model-

$$F_i = F_{i-1} \; o \; R \tag{1}$$

Where, $R = \bigcup_{i=1}^{n} R_i$, $F_i$ is the fuzzified forecasted row matrix at the $i^{th}$ unit and $F_{i-1}$ is the fuzzified row matrix of the $(i-1)^{th}$ unit and $o$ is the max-min composition operator.

Step-12: Defuzzification of the output using the centroid method.

After going through the above procedure for all the cluster of each region along with its neighbor; the value of forecast of that particular region will be made. In the same way, the forecasted value of other region will be prepared. Schematic representation of proposed algorithm has been given in Fig. 1 for a single cluster.

## 2.4 K-means Clustering

K-means Clustering method is a nonhierarchical clustering technique which groups items on the basis of data information (Johnson and Wichern, 1996). The steps are-

1. Pick k initial clusters.
2. Assign items to the nearest cluster center.
3. Recalculate the cluster center after the new assignment.
4. Repeat 2 and 3 until cluster center don't change.

# 3. RESULTS AND DISCUSSION

## 3.1 Results of Chen's method

An illustration is demonstrated in the following to provide the step by step procedure of Chen model:

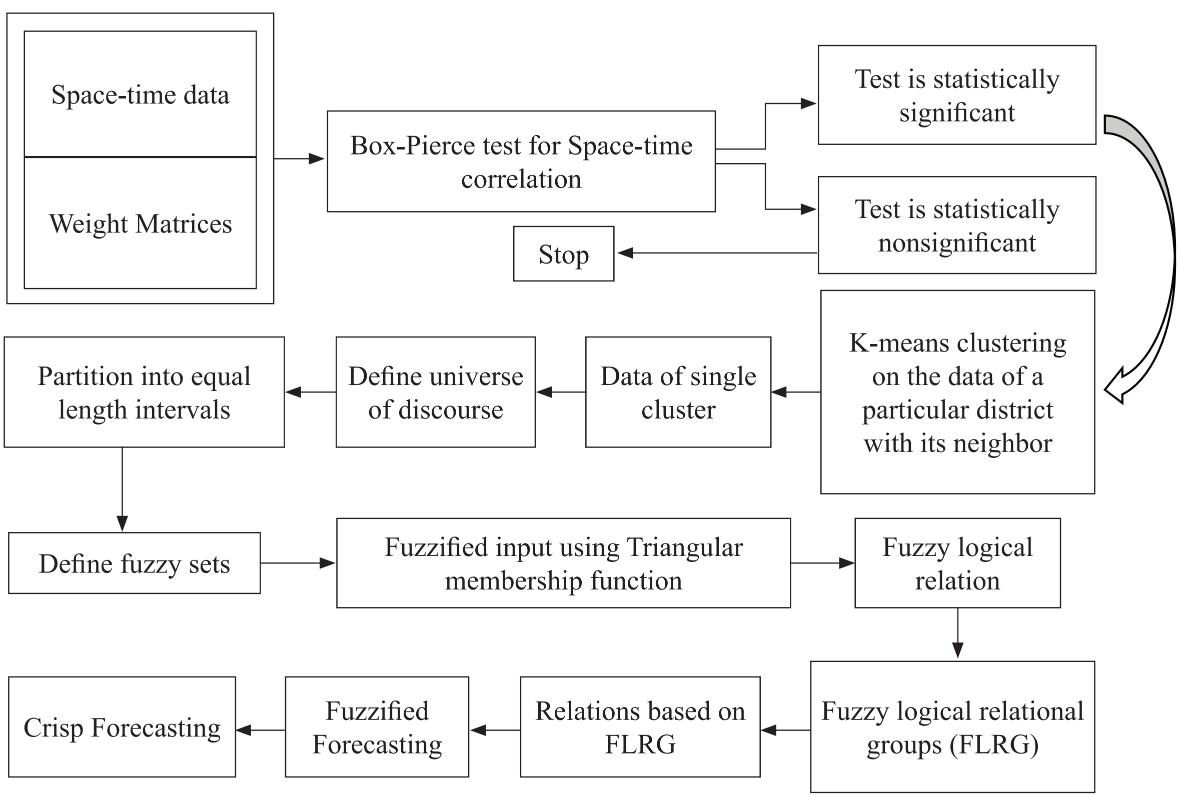


**Fig. 1.** Schematic representation of proposed forecasting procedure for a single cluster.

**Step-1:** Fixing the universe of discourse which is defined as-$U = [U_{min}$ -$U_1, U_{max}$ -$U_2]$ , where $U_{min}$ and $U_{max}$ are minimum and maximum value of the data and $U_1$ and $U_2$ are any two positive values which are selected by the modeler properly. In our case,

$U = [1000, 2000]$ where, $U_{min}$ =1133.1, $U_{max}$ =1905.523, $U_1$ =133.1, $U_2$=94.477

The universe of discourse is divided into five equal lengthy intervals- $I_1$= [1000, 1200], $I_2$= [1200, 1400], $I_3$= [1400,1600], $I_4$= [1600,1800], $I_5$= [1800,2000].

**Step-2:** Define the fuzzy sets as:

$$A_1 = 1/I_1 + 0/I_2 + 0/I_3 + 0/I_4 + 0/I_5$$

$$A_2 = 0/I_1 + 1/I_2 + 0/I_3 + 0/I_4 + 0/I_5$$

$$A_3 = 0/I_1 + 0/I_2 + 1/I_3 + 0/I_4 + 0/I_5$$

$$A_4 = 0/I_1 + 0/I_2 + 0/I_3 + 1/I_4 + 0/I_5$$

$$A_5 = 0/I_1 + 0/I_2 + 0/I_3 + 0/I_4 + 1/I_5$$

**Step-3:** Fuzzify the data-

See Table 1.

**Step-4:** Make the fuzzy logical relationship (FLR) in Table-2.

**Step-5:** Prepare the fuzzy logical relational groups in Table-3.

**Table 2.** Fuzzy logical relationships

| |
|---|
| $A_3 \rightarrow A_5, A_5 \rightarrow A_4, A_4 \rightarrow A_2, A_2 \rightarrow A_4,$ |
| $A_4 \rightarrow A_3, A_3 \rightarrow A_3, A_3 \rightarrow A_2, A_2 \rightarrow A_2,$ |
| $A_2 \rightarrow A_3, A_3 \rightarrow A_3, A_3 \rightarrow A_2, A_2 \rightarrow A_2, A_2 \rightarrow A_2,$ |
| $A_2 \rightarrow A_3, A_3 \rightarrow A_1, A_1 \rightarrow A_5, A_5 \rightarrow A_1,$ |
| $A_1 \rightarrow A_3, A_3 \rightarrow A_2$ |

**Table 3.** Fuzzy logical relational groups

| |
|---|
| Group 1: $A_1 \rightarrow A_3, A_1 \rightarrow A_5$. |
| Group 2: $A_2 \rightarrow A_2, A_2 \rightarrow A_3, A_2 \rightarrow A_4$. |
| Group 3: $A_3 \rightarrow A_1, A_3 \rightarrow A_2, A_3 \rightarrow A_3, A_3 \rightarrow A_5$. |
| Group 4: $A_4 \rightarrow A_2, A_4 \rightarrow A_3$. |
| Group 5: $A_5 \rightarrow A_1, A_5 \rightarrow A_4$. |

**Step-6:** Forecasting by using the following rules.

Rule 1: In One-to-One relationship, If, $A_i \rightarrow A_j$, then, forecasted value will be the midpoint of $I_j$.

Rule 2: If $A_i \rightarrow \emptyset$, then, the output is equal to the middle value of $I_i$.

Rule 3: In One-to-Many relationships, suppose $A_i \rightarrow A_1, A_2, \ldots, A_n$, then the forecasted value will be the average value of the midpoints of the $I_1, I_2, \ldots, I_n$.

**Table 1.** Fuzzified the rainfall data of South Dinajpur

| Year | Annual Rainfall | Fuzzified data | Forecast | Year | Annual Rainfall | Fuzzified data | Forecast |
|---|---|---|---|---|---|---|---|
| 1989 | 1473.312 | $A_3$ | | 1999 | 1503.701 | $A_3$ | 1450 |
| 1990 | 1905.523 | $A_5$ | 1450 | 2000 | 1223.909 | $A_2$ | 1500 |
| 1991 | 1631.869 | $A_4$ | 1400 | 2001 | 1254.143 | $A_2$ | 1500 |
| 1992 | 1302.148 | $A_2$ | 1400 | 2002 | 1229.561 | $A_2$ | 1500 |
| 1993 | 1700.469 | $A_4$ | 1500 | 2003 | 1493.899 | $A_3$ | 1500 |
| 1994 | 1431.333 | $A_3$ | 1400 | 2004 | 1162 | $A_1$ | 1450 |
| 1995 | 1544.136 | $A_3$ | 1450 | 2005 | 1823.8 | $A_5$ | 1700 |
| 1996 | 1249.317 | $A_2$ | 1450 | 2006 | 1113.1 | $A_1$ | 1400 |
| 1997 | 1375.237 | $A_2$ | 1500 | 2007 | 1582.2 | $A_3$ | 1700 |
| 1998 | 1437.133 | $A_3$ | 1500 | 2008 | 1296.8 | $A_2$ | 1450 |

### 3.2 Results of proposed method

The proposed approach has been empirically illustrated step by step for a single cluster. Those steps are discussed as follows:

**Step 1:** Annual rainfall data of six districts of north part of West Bengal, viz., South Dinajpur, Malda, Uttar Dinajpur, Darjeeling, Jalpaiguri and Cooch Behar has been taken.

**Step 2:** In next step, weight matrices have been formed to include the temporal as well as spatial pattern. The zero order identity weight matrix (Table-4) and first order matrix (Table-5)is an asymmetric matrix based on the percentage of boundary sharing is considered (Fig. 2).

**Table 4.** Weight matrix of zero order to include the temporal pattern

| Sl. No. | District | SD | MAL | ND | DAR | JAL | COO |
|---|---|---|---|---|---|---|---|
| 1 | SD | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | MAL | 0 | 1 | 0 | 0 | 0 | 0 |
| 3 | ND | 0 | 0 | 1 | 0 | 0 | 0 |
| 4 | DAR | 0 | 0 | 0 | 1 | 0 | 0 |
| 5 | JAL | 0 | 0 | 0 | 0 | 1 | 0 |
| 6 | COO | 0 | 0 | 0 | 0 | 0 | 1 |

**Table 5.** Weight matrix of first order based on percentage of boundary sharing to include the spatial pattern

| Sl. No. | District | SD | MAL | ND | DAR | JAL | COO |
|---|---|---|---|---|---|---|---|
| 1 | SD | 0 | 0.49812 | 0.50188 | 0 | 0 | 0 |
| 2 | MAL | 0.51859 | 0 | 0.48141 | 0 | 0 | 0 |
| 3 | ND | 0.45562 | 0.41978 | 0 | 0.12461 | 0 | 0 |
| 4 | DAR | 0 | 0 | 0.10349 | 0 | 0.89651 | 0 |
| 5 | JAL | 0 | 0 | 0 | 0.40533 | 0 | 0.59467 |
| 6 | COO | 0 | 0 | 0 | 0 | 1 | 0 |

Box-pierce test has been implemented to test the correlation among the space-time series. The null hypothesis is that the space-time data is not correlated, whereas alternative hypothesis is presence of space-time correlation. The null hypothesis should be rejected if the p-value is less than critical values. The result shows that p-value is <0.0001. That means, the data has a significant spatio-temporal correlation because the null hypothesis is rejected at $1\%$ level of significance.

**Step 3:** Considering the case for South Dinajpur (1) , it has its two neighboring districts. *i.e.* Malda (2) and North Dinajpur (3). So, all the data has been taken in a single frame. After that, the K-means clustering has been performed to form two clusters.

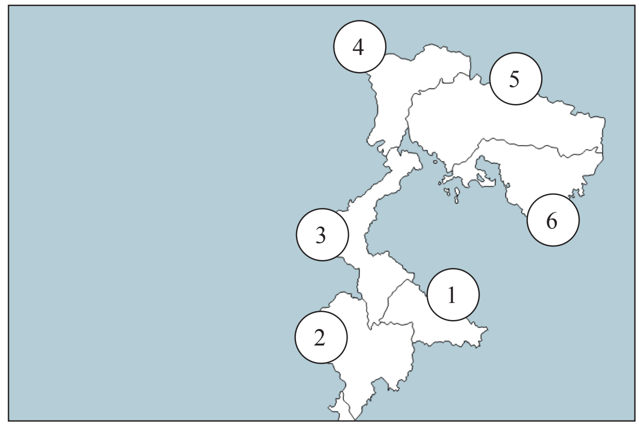


**Fig. 2.** Map of six districts of northern part of West Bengal

**Step 4:** Considering the first cluster, fixing the universe of discourse which is defined as $U = [U_{min} - U_1, U_{max} - U_2]$ , where $U_{min}$ and $U_{max}$ are minimum and maximum value of the data of that cluster and $U_1$ and $U_2$ are two any two positive values which are selected by the modeler properly. In this case,

$U = [900, 1400]$ where, $U_{min}$ =1011.8, $U_{max}$ =1389.511, $U_1$ =111.8, $U_2$=10.489

**Step-5:** The universe of discourse has been partitioned into five equal lengthy intervals- $I_1$=[900,1000], $I_2$=[1000, 1100], $I_3$=[1100, 1200], $I_4$=[1200, 1300], $I_5$=[1300, 1400].

**Step-6:** Fuzzy sets are defined for the intervals by the triangular membership functions (Table- 6 and Fig. 3). Zadeh (1965) first introduced the membership functions in his research paper "Fuzzy Sets".

**Table 6.** Fuzzy sets are defined using the triangular membership function in First Cluster

| Fuzzy Set | Parameters |
|---|---|
| $A_1$ | [900, 1000, 1100] |
| $A_2$ | [1000, 1100, 1200] |
| $A_3$ | [1100, 1200,1300] |
| $A_4$ | [1200,1300,1400] |
| $A_5$ | [1300,1400,1400] |

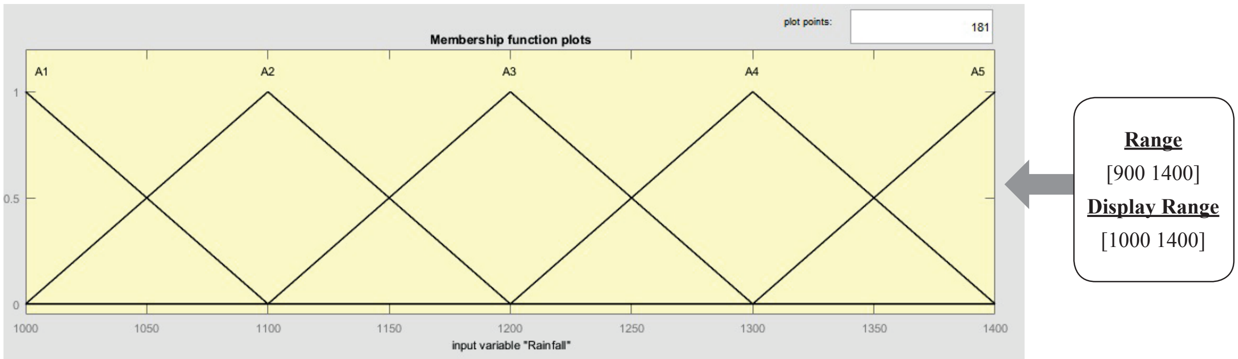


**Fig. 3.** Plotting of triangular membership function

**Table 7.** Fuzzifying the annual rainfall data based on membership grade in first cluster

| Sl. No. | Data | A1 | A2 | A3 | A4 | A5 | Fuzzified Input |
|---|---|---|---|---|---|---|---|
| 1 | 1347.312 | 0 | 0 | 0 | 0.52688 | 0.47312 | A4 |
| 2 | 1302.148 | 0 | 0 | 0 | 0.97852 | 0.02148 | A4 |
| 3 | 1088.095 | 0.11905 | 0.88095 | 0 | 0 | 0 | A2 |
| 4 | 1296.674 | 0 | 0 | 0.03326 | 0.96674 | 0 | A4 |
| 5 | 1255.248 | 0 | 0 | 0.44752 | 0.55248 | 0 | A4 |
| 6 | 1369.492 | 0 | 0 | 0 | 0.30508 | 0.69492 | A5 |
| 7 | 1326.425 | 0 | 0 | 0 | 0.73575 | 0.26425 | A4 |
| 8 | 1249.317 | 0 | 0 | 0.50683 | 0.49317 | 0 | A3 |
| 9 | 1113.425 | 0 | 0.86575 | 0.13425 | 0 | 0 | A2 |
| 10 | 1265.135 | 0 | 0 | 0.34865 | 0.65135 | 0 | A4 |
| 11 | 1375.237 | 0 | 0 | 0 | 0.24763 | 0.75237 | A5 |
| 12 | 1251.713 | 0 | 0 | 0.48287 | 0.51713 | 0 | A4 |
| 13 | 1389.511 | 0 | 0 | 0 | 0.10489 | 0.89511 | A5 |
| 14 | 1333.367 | 0 | 0 | 0 | 0.66633 | 0.33367 | A4 |
| 15 | 1326.381 | 0 | 0 | 0 | 0.73619 | 0.26381 | A4 |
| 16 | 1223.909 | 0 | 0 | 0.76091 | 0.23909 | 0 | A3 |
| 17 | 1063.865 | 0.36105 | 0.63895 | 0 | 0 | 0 | A2 |
| 18 | 1223.275 | 0 | 0 | 0.76725 | 0.23275 | 0 | A3 |
| 19 | 1254.143 | 0 | 0 | 0.45857 | 0.54143 | 0 | A4 |
| 20 | 1113.945 | 0 | 0.86055 | 0.13945 | 0 | 0 | A2 |
| 21 | 1255.398 | 0 | 0 | 0.44602 | 0.55398 | 0 | A4 |
| 22 | 1229.561 | 0 | 0 | 0.70439 | 0.29561 | 0 | A3 |
| 23 | 1101.809 | 0.98191 | 0.01809 | 0 | 0 | 0 | A2 |
| 24 | 1260.409 | 0 | 0 | 0.39591 | 0.60409 | 0 | A4 |
| 25 | 1379.9 | 0 | 0 | 0 | 0.201 | 0.799 | A5 |
| 26 | 1162 | 0 | 0.38 | 0.62 | 0 | 0 | A3 |
| 27 | 1266.7 | 0 | 0 | 0.333 | 0.667 | 0 | A4 |
| 28 | 1133.1 | 0 | 0.669 | 0.331 | 0 | 0 | A2 |
| 29 | 1221.3 | 0 | 0 | 0.787 | 0.213 | 0 | A3 |
| 30 | 1011.8 | 0.882 | 0.118 | 0 | 0 | 0 | A1 |
| 31 | 1296.8 | 0 | 0 | 0.032 | 0.968 | 0 | A4 |
| 32 | 1161.3 | 0 | 0.387 | 0.613 | 0 | 0 | A3 |

The triangular membership function is defined by three parameters which is written as:

$$w_F = \begin{cases} \dfrac{x-a}{b-a}, & a \le x \le b \\ \dfrac{c-x}{c-b}, & b \le x \le c \\ 0, & x > c \end{cases} \qquad (2)$$

**Step-7**: Fuzzification through the triangular membership function is depicted in Table 7. Based on fuzzified input and membership grade, fuzzy sets on the universe of discourse have been defined (Table 8).

**Table 8: Fuzzy sets are defined on the universe of discourse**

$$A_1 = 0.882/\mathrm{I}_1 + 0.118/\mathrm{I}_2 + 0/\mathrm{I}_3 + 0/\mathrm{I}_4 + 0/\mathrm{I}_5$$
$$A_2 = 0.361/\mathrm{I}_1 + 0.981/\mathrm{I}_2 + 0.331/\mathrm{I}_3 + 0/\mathrm{I}_4 + 0/\mathrm{I}_5$$
$$A_3 = 0/\mathrm{I}_1 + 0.387/\mathrm{I}_2 + 0.787/\mathrm{I}_3 + 0.493/\mathrm{I}_4 + 0/\mathrm{I}_5$$
$$A_4 = 0/\mathrm{I}_1 + 0/\mathrm{I}_2 + 0.482/\mathrm{I}_3 + 0.978/\mathrm{I}_4 + 0.473/\mathrm{I}_5$$
$$A_5 = 0/\mathrm{I}_1 + 0/\mathrm{I}_2 + 0/\mathrm{I}_3 + 0.305/\mathrm{I}_4 + 0.895/\mathrm{I}_5$$

**Step-8:** In this step, fuzzy logical relationship has been made which is shown in Table-9.

**Table 9.** Fuzzy logical relationship (FLR)

$$A_4 \to A_4, A_4 \to A_2, A_2 \to A_4, A_4 \to A_4, A_4 \to A_5, A_5 \to A_4, A_4 \to A_3, A_3 \to A_2,$$
$$A_2 \to A_4, A_4 \to A_5, A_5 \to A_4, A_4 \to A_5, A_5 \to A_4, A_4 \to A_4, A_4 \to A_3, A_3 \to A_2, A_2 \to A_3,$$
$$A_3 \to A_4, A_4 \to A_2, A_2 \to A_4, A_4 \to A_3, A_3 \to A_2, A_2 \to A_4, A_4 \to A_5, A_5 \to A_3, A_3 \to A_4,$$
$$A_4 \to A_2, A_2 \to A_3, A_3 \to A_1, A_1 \to A_4, A_4 \to A_3$$

**Step-9:**Fuzzy logical relational groups has been formed in Table-10**.**

**Table 10.** Fuzzy logical Relational Groups (FLRG)

Group 1: $A_1 \to A_4$.
Group 2: $A_2 \to A_3, A_2 \to A_4$.
Group 3: $A_3 \to A_1, A_3 \to A_2, A_3 \to A_4$.
Group 4: $A_4 \to A_2, A_4 \to A_3, A_4 \to A_4, A_4 \to A_5$.
Group 5: $A_5 \to A_3, A_5 \to A_4$.

**Step 10 to Step 12:**

Total twelve relationships are to be made based on the fuzzy logical relational groups. If, fuzzy logical relation is $A_j \to A_k$; then relation will be: $R_{jk} = A_j^T \times A_k$. The elements of $R_{jk}$ are to be computed as $e_{jk} = \min\left(A_j^T \times A_k\right)$; $j, k = 1,2,\ldots n$.

In the above data, the 12 relations are depicted in Table-11. For $A_2$, $A_3$, $A_4$, $A_5$; the union of their corresponding relations are defined as above relation (Table-11).

**Table 11: Fuzzy logical Relational Groups (FLRG)**

| FLR | Relations | FLR | Relations |
|---|---|---|---|
| $A_1 \to A_4$ | $R_{14} = A_1^T \times A_4$ | $A_4 \to A_2$ | $R_{42} = A_4^T \times A_2$ |
| $A_2 \to A_3$ | $R_{23} = A_2^T \times A_3$ | $A_4 \to A_3$ | $R_{43} = A_4^T \times A_3$ |
| $A_2 \to A_4$ | $R_{24} = A_2^T \times A_4$ | $A_4 \to A_4$ | $R_{44} = A_4^T \times A_4$ |
| $A_3 \to A_1$ | $R_{31} = A_3^T \times A_1$ | $A_4 \to A_5$ | $R_{45} = A_4^T \times A_5$ |
| $A_3 \to A_2$ | $R_{32} = A_3^T \times A_2$ | $A_5 \to A_3$ | $R_{53} = A_5^T \times A_3$ |
| $A_3 \to A_4$ | $R_{32} = A_3^T \times A_4$ | $A_5 \to A_4$ | $R_{54} = A_5^T \times A_4$ |

For example, in $A_2$, consider the union of $R_{23}$ and $R_{24}$ which is computed by taking the maximum value of these two relations element wise and it is denoted by, $R_{234} = R_{23} \cup R_{24}$ and the elements of $R_{234}$ is computed by $e_{234} = \max(e_{23}, e_{24}) = \max\left[\min(A_2^T \times A_3), \min(A_2^T \times A_4)\right]$. For illustration,

$$R_{23} = A_2^T \times A_3 = \min\left(A_2^T, A_3\right)$$

$$= \min\left(\begin{bmatrix} 0.36105 \\ 0.98191 \\ 0.331 \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0.387 & 0.787 & 0.49317 & 0 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 0 & 0.36105 & 0.361050.36105 & 0 \\ 0 & 0.38700 & 0.787000.49317 & 0 \\ 0 & 0.33100 & 0.331000.33100 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Similarly,

$$R_{24} = \begin{bmatrix} 0 & 0.36105 & 0.361050.36105 & 0.36105 \\ 0 & 0 & 0.482870.97852 & 0.47312 \\ 0 & 0 & 0.331000.33100 & 0.33100 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$R_{234} = R_{23} \cup R_{24} = \max\left[\min(A_2^T \times A_3), \min(A_2^T \times A_4)\right]$$

$$= \begin{bmatrix} 0 & 0.36105 & 0.36105 & 0.36105 & 0.36105 \\ 0 & 0.38700 & 0.78700 & 0.97852 & 0.47312 \\ 0 & 0.33100 & 0.33100 & 0.33100 & 0.33100 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For example, the forecasted value for the data with Sl. No. 4,

$$F_4 = F_3 \ o \ R = A_2 \ oR_{234} \quad (3)$$

$$= [0.36105 \quad 0.98191 \quad 0.33100 \quad 0 \quad 0]$$

$$o \begin{bmatrix} 0 & 0.36105 & 0.36105 & 0.36105 & 0.36105 \\ 0 & 0.38700 & 0.78700 & 0.97852 & 0.47312 \\ 0 & 0.33100 & 0.33100 & 0.33100 & 0.33100 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= [0 \quad 0.38700 \quad 0.78700 \quad 0.97852 \quad 0.47312]$$

The above output fuzzified. Now, it has to be converted into the crisp value by centroid method, it is a method of weighted average in which it determines the centroid value of the sets. It is also known as center of area method. The mathematical formula of centroid method is-

$$Y = \frac{\sum_{i=1}^{n} \omega_F(x_i) x_i}{\sum_{i=1}^{n} \omega_F(x_i)}$$

Where, Y is the crisp output, $\omega_F(x_i)$ is the fuzzy output value or value of the membership function of $x_i$ and $x_i$ is the middle value of $I_i$.

Now,

$$F_4 = \frac{[(0\times950)+(0.38700\times1050)+(0.78700\times1150)+(0.97852\times1250)+(0.47312\times1350)]}{[0.38700+0.78700+0.97852+0.47312]}$$

= 1208.567 (Forecasted value for the data with Sl. No. 4, whereas the actual value was 1296.674)

By using the above procedure, forecasted value for other values have been worked out.

Table -12 shows the comparative results of Chen and Proposed method for both training and testing dataset in terms of Mean Absolute Percentage Error (MAPE). It is observed that the proposed method has a lower MAPE compared to the Chen method in all the districts in both training and testing dataset. In addition, Table 13 reveals the overall performance of proposed approach is better as compare to conventional Chen approach in both cases in-sample as well as out-of-sample forecast. Therefore, it can be inferred that the proposed method is evincing better than Chen method.

**Table 12.** MAPE of Chen and Proposed method

| District | Training | | Testing | |
|---|---|---|---|---|
| | Chen | Proposed | Chen | Proposed |
| SD | 12.29 | 7.39 | 30.92 | 19.25 |
| MAL | 13.10 | 7.50 | 18.80 | 8.31 |
| ND | 7.77 | 5.84 | 12.78 | 6.59 |
| DAR | 12.61 | 6.97 | 8.57 | 7.87 |
| JAL | 20.00 | 5.22 | 13.17 | 13.02 |
| COO | 9.74 | 8.99 | 31.19 | 22.36 |

**Table 13.** Forecast value of two in sample and two out of sample data

| District | Year | Actual | In Sample | | Year | Actual | Out of Sample | |
|---|---|---|---|---|---|---|---|---|
| | | | Chen | Proposed | | | Chen | Proposed |
| SD | 2007 | 1582.2 | 1700 | 1626.6 | 2009 | 980.5 | 1550 | 1284.2 |
| | 2008 | 1296.8 | 1450 | 1249.4 | 2010 | 1393 | 1450 | 1284.2 |
| MAL | 2007 | 1554.1 | 1500 | 1550.8 | 2009 | 1325 | 1300 | 1358 |
| | 2008 | 1503.4 | 1300 | 1612.1 | 2010 | 1189.8 | 1500 | 1358 |
| ND | 2007 | 1547.5 | 1583.3 | 1546.6 | 2009 | 1310.9 | 1583.3 | 1350 |
| | 2008 | 1161.3 | 1250 | 1160.9 | 2010 | 1502.9 | 1250 | 1349.7 |
| DAR | 2007 | 3285.3 | 2700 | 3367.6 | 2009 | 3009.1 | 3450 | 3166.7 |
| | 2008 | 3316.1 | 3450 | 3092.2 | 2010 | 3538.6 | 3450 | 3166.7 |
| JAL | 2007 | 3507.8 | 3000 | 3501.7 | 2009 | 3137.8 | 3400 | 3136.7 |
| | 2008 | 3559.63 | 3400 | 3124.9 | 2010 | 4146.2 | 3400 | 3067.7 |
| COO | 2007 | 2420.3 | 2200 | 2111.9 | 2009 | 2316.1 | 1800 | 3200 |
| | 2008 | 2830.6 | 2600 | 3400 | 2010 | 3338.9 | 2000 | 3119.9 |

BDS test for checking the non-linearity has also be done for all the districts. It has been found from Table-14 that most of the data was a mixture of more non-linearity and less of linearity; whereas data of North Dinajpur district was purely non-linear. So, proposed model can cope up the linearity and nonlinearity in a well manner and it can also furnish better result in the less number of data.

## 4. CONCLUSION

There are many fuzzy time series method for modelling and forecasting the time series data available in literature. However, fuzzy approach has not been yet used in spatio-temporal framework. Hence in this study, a new method has been proposed by utilizing the fuzzy techniques and k-means clustering algorithm

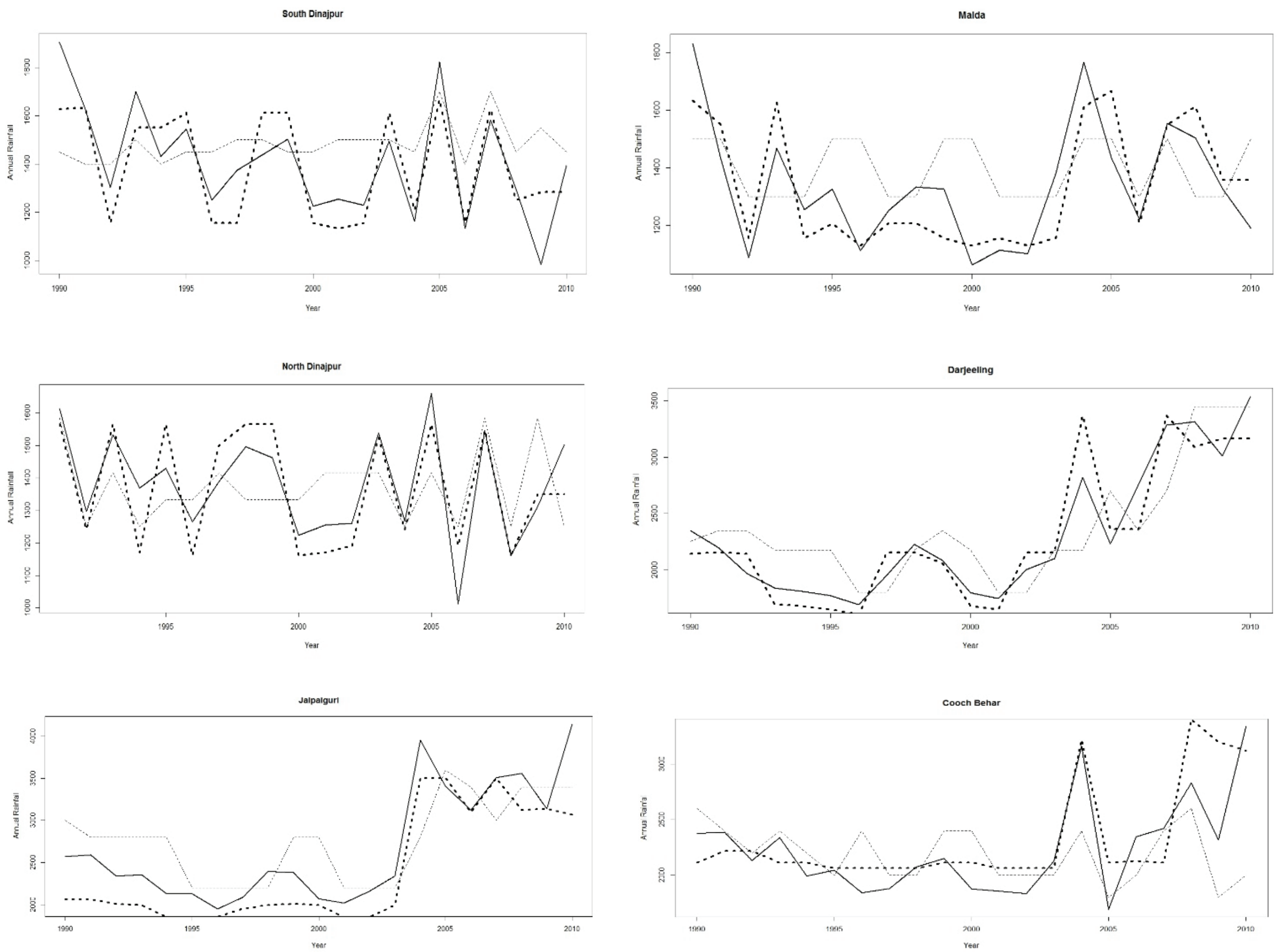


**Fig. 4.** Graphical comparison of proposed approach with Chen's method
(Solid and thin line- Chen's method, solid and thick-actual data and dashed- proposed approach)

**Table 14.** BDS test for non-linearity (the value in parenthesis represents $p$-value)

| Districts | eps(1) | | eps(2) | | eps(3) | | eps(4) | |
|---|---|---|---|---|---|---|---|---|
| | **Statistic at m=2** | **Statistic at m=3** | **Statistic at m=2** | **Statistic at m=3** | **Statistic at m=2** | **Statistic at m=3** | **Statistic at m=2** | **Statistic at m=3** |
| SD | -14.76 | -81.66 | 0.25 | 2.98 | 1.55 | 4.30 | -0.53 | 1.33 |
| | (<0.001) | (<0.001) | (0.795) | (0.002) | (0.119) | (<0.001) | (0.594) | (0.181) |
| MAL | -16.25 | -15.56 | -2.18 | -2.73 | -0.80 | 0.48 | -3.14 | -1.22 |
| | (<0.001) | (<0.001) | (0.028) | (0.017) | (0.418) | (0.629) | (0.001) | (0.222) |
| JAL | 4.62 | 6.02 | 4.98 | 5.25 | 3.68 | 3.98 | -1.34 | -1.74 |
| | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (0.179) | (0.081) |
| COO | 26.79 | 25.53 | 1.59 | 2.09 | 0.50 | -0.62 | -0.33 | -1.77 |
| | (<0.001) | (<0.001) | (0.109) | (0.036) | (0.617) | (0.529) | (0.739) | (0.076) |
| DAR | | 10.34 | 1.65 | 2.45 | -2.06 | -1.63 | -46.65 | -46.70 |
| | (<0.001) | (<0.001) | (0.104) | (0.014) | (0.038) | (0.101) | (<0.001) | (<0.001) |
| ND | 20.17 | 26.97 | 4.03 | 3.67 | 6.26 | 7.26 | 3.79 | 6.25 |
| | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) | (<0.001) |

in spatio-temporal framework for modelling and forecasting spatio-temporal time series data. The proposed method also endeavors to fill the gap of the existing methods. It is observed that proposed method has given lower prediction error compared to the existing univariate Chen fuzzy time series method. Hence, it can be concluded that proposed method reveals its significant performance over the traditional Chen method in terms of forecasting accuracy. The key feature of the proposed method is that, it can be applied on both linear as well as non-linear data, it is useful when there is imprecise data like weather data, agricultural production etc.,it can provide better result in the presence of small number of observations. The proposed method can be extended to the other areas where imprecise data, linear or non-linear or combination of linear and non-linear data, and/or where the number of observations in spatio-temporal time series are less.

## REFERENCES

Chen, S.M. (1996). Forecasting enrollments based on fuzzy time series. *Fuzzy Sets and Sys.* **81**, 311-319.

Johnson, R.A. and Wichern, D.W. (1996). *App. Multivariate Stat. Ana.*, Prentice-Hall of India Private Limited.

Kandel, A. (1996). *Fuzzy Mathe. Tech. with Appl.*, Reading, MA: Addison-Wesley.

Kaufmann, A. and Gupta, M.M. (1988). *Fuzzy Math. Models in Engi. and Mana. Sci*., Amsterdam, The Netherlands: North-Holland.

Klir, G.J. and Folger, T.A. (1988). *Fuzzy Sets, Uncertainty, and Info.*, Englewood Cliffs, NJ: Prentice-Hall.

Klir, G.J. and Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Appl.*, Englewood Cliffs, NJ: Prentice-Hall.

Kuo, S.C., Chen, C.C. and Li, S.T. (2015). Evolutionary fuzzy relational modeling for fuzzy time series forecasting. *Int. J. Fuzzy Sys.*, **17**, 1295-1306.

Lee, H.S. and Chou, M.T. (2004). Fuzzy forecasting based on fuzzy time series.*Int. J. Comput. Math.*, **81**, 781-789.

Lee, W.J., Jung, H.Y., Yoon, J.H. and Choi, S.H. (2017). A Novel Forecasting Method Based on F-Transform and Fuzzy Time Series. *Int. J. Fuzzy Sys.*, **19(6)**, 1793-1802.

Liu, H.T. (2007). An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers. *Fuzzy Optim. Decis. Making*, **6**, 63-80.

Pfeifer, P. and Deutsch, S. (1980). A Three-Stage Iterative Procedure for Space-Time Modeling. *Technometrices*, **22(1)**, 35-47.

Pfeifer, P. and Deutsch, S. (1981). Variance of the Sample Space-Time Autocorrelation Function. *J. Royal Sta. Soc. Ser. B (Methodological)*, **43(1)**, 28-33.

Qu, H. and Chen, G. (2012). An improved method of fuzzy time series model. *Third Int. Conf. on Inte. Cont. and Info. Processing*, Dalian, 346-351.

Singh, S.R. (2007). A robust method of forecasting based on fuzzy time series. *Appl. Math. and Computation*, **188(1)**, 472-484.

Song, Q. and Chissom, B.S. (1993a). Forecasting enrollments with fuzzy time series-part I. *Fuzzy Sets and Sys*.**, 54**, 1-10

Song, Q. and Chissom, B.S. (1993b). Fuzzy time series and its models. *Fuzzy Sets and Sys.*, **54**, 269-277.

Song, Q. and Chissom, B.S. (1994). Forecasting enrollments with fuzzy time series - part II. *Fuzzy Sets and Sys.*, **62**, 1-8.

Zadeh, L.A. (1965). *Fuzzy sets,Inform. and Control*, **8**, 338-353.