ISSN 2229-5518

Adequacy Assessment of Air Quality Monitoring Stations in Delhi, India using Fuzzy Similarity and Fuzzy c-Mean Clustering

Kamal Jyoti Maji, Anil Kumar Dikshit, Satish Salunkhe

Abstract— Air quality monitoring network (AQMN) has been designed by the air quality monitoring authorities to analyze the present air quality in metropolitan cities. Due to various factors including population growth, increasing energy demand, industrial growth, change in the land use pattern, and increase in number of vehicles; the existing AQMN may not always provide the best characterization of air quality in fast growing city. The objective of the study is to verify the fitness of existing air quality monitoring stations (AQMSs) within rapid environmental change for fast growing metropolitan city of Delhi, India. A useful formulation has been presented to identify monitoring stations in existing AQMN in Delhi, India which provide maximum air quality information including PM_{10}, NO_2, SO_2 and SPM; using cluster analysis which aims at reducing network density with a minimum loss of air quality data. In this technique, the existing air quality monitoring stations have been classified or grouped using fuzzy similarity measures. Non-linear mapping method is used to take decision for number of cluster and fuzzy c-mean (FCM) clustering method provides further classification of monitoring stations. Finally decision is taken based on the results of both fuzzy similarity measure and FCM.

Index Terms— Air quality monitoring network, Fuzzy similarity measures, Cosine amplitude method, Max-min composition, Fuzzy c-mean clustering.

1 INTRODUCTION

The monitoring of air pollutants in metropolitan city provides information about the present air quality scenario, to evaluate the impact of an air pollution control strategy, to identify and track episodic events (e.g. accident), to know long term trend of air quality in monitored area, to assess the risk due to air pollution and even to plan for future land use. So, the installation of AQMN in a metropolitan city becomes important task for central and state pollution control authorities. Design of AQMN in metropolitan city is based on high population density, highest pollution concentration in industrial/ residential area where violation of national ambient air quality standard is frequent. In most of the cases, the availability of suitable site is the only single consideration while deciding the location for air quality monitoring station. As such, there is no general acceptance on approaches for deciding the number of air quality monitoring stations and their locations. This is because air quality changes with place, time and year as the air pollutant concentration varies temporally and spatially. Other reasons are imprecision in measurement of air quality parametric concentration due to precision efficiency of monitoring instrument. There is epistemic uncertainty or fuzziness in describing the aggregate air quality status based on air quality indexing system with linguistic classification, as very poor, poor, good and alike by the domain experts.

- *Kamal Jyoti Maji is currently pursuing PhD degree in Center for Environmental Science and Engineering (CESE) in Indian Institute of Technology Bombay (IITB), Mumbai, India, PH-91-22-25767851/52. E-mail: kamaljm@iitb.ac.in*
- *Anil Kumar Dikshit is currently professor and head in Center for Environmental Science and Engineering (CESE) in Indian Institute of Technology Bombay (IITB), India, E-mail: dikshit@iitb.ac.in*
- *Satish Salunkhe is currently assistant professor in Computer Department, Mukesh Patel School of Technology Management & Engineering, Mumbai, India. E-mail: satishssalunkhe@gmail.com*

For designing AQMN, basically two methodologies has been described in literature (a) the statistical method and (b) the modelling method [1]. Like principle component analysis method to study sulphur dioxide (SO_2) monitoring network in St. Louis, United States [2] and a spatial sample stratification technique proposed based on the Monte-Carlo variance reduction method for air pollutant monitoring system [3]. A statistical technique was developed based on Fisher's Information Measure to design an optimal air quality monitoring network [4].

For a single pollutant, the concepts of Spatial Correlation Analysis (SCA) and principles of Minimum Spanning Tree (MST) was proposed [5]. Utility approach and sequential interactive approach was the extension of the MST algorithm for multi-objective optimization of AQMN [6]. Clustering method for analysis of ambient air quality data of Delhi city [7]. From economic point of view, both pollution and construction cost, a developed multi-attribute utility function method to allocate monitoring stations was developed [8].

Grey compromise programing method used for setting up new monitoring stations in a metropolitan city [9]. Study presented a linear programing model for optimal design of a multi-pollutant AQMN. A risk assessment technique to optimize the location and number of monitoring sites for ambient AQMN design has been proposed [10]. Fuzzy synthetic evaluation technique and fuzzy similarity measure technique also developed for AQMN [11, 12].

It is essential to estimate the fitness of existing AQMN and to decide whether new monitoring stations are required into the existing AQMN or relocation of existing monitoring stations is needed in the future to get better information about the air quality.

The aim of the present study is the identification of existing monitoring stations, providing maximum air quality info mation in context of air quality index, based on fuzzy similarty measure and fuzzy c-mean clustering technique within AQMN of the metropolitan city Delhi, India. Based on the identification by results, it is possible to reduce number of monitoring stations to bring down recurring expenditure or to take decision for relocating monitoring stations providing minimum information of air quality within study area.

2 METHODOLOGY

The analyst, in consultation with the domain experts, will first identify the most critical air pollution parameter(s) in order to arrive at the classification of the monitoring stations. It is generally agreed upon that air pollution in a city is due to auto exhaust, thermal power plant emission, solid waste burning, ongoing infrastructure development activity and the use of generators due to frequent power cuts. In most of the Indian cities, PM_{10} and NO_2 could be the governing parameters in the classification of AQMS.

The approach for achieving the study objectives primarily depends upon availability of parametric data with reasonable statistical accuracy. The procedure for the analysis of the data is as follows - air quality parametric data is divided into five parts, based on Indian air quality index (IND-AQI) compliance criteria for categories or groups of air quality as: good, moderate, poor, very poor and severe. Further, the number of data inputs in each group for each monitoring station are calculated and normalized. The generated $X \times Y$ matrix is on two universe. In this case, X refers to AQMS and Y refers to the compliance criteria or group. It is not possible to use, straightaway, fuzzy compositional rule of inference to estimate the similarity relation between the AQMS as the data is on two different universe. Therefore, relational formalism termed as cosine amplitude method has been used. The matrices generated using cosine amplitude methods are invariably fuzzy tolerance relations, meaning that the relation is reflexive and symmetric but not transitive. It is therefore necessary to transform fuzzy tolerance/ proximity/ compatibility relation to fuzzy equivalence relation using max-min composition method. After that fuzzy equivalence relation to crisp equivalence relation conversion can be achieved using the concept of alpha cut which signifies the possibility of the AQMSs in a particular group.

For each monitoring station, mean value of air pollution concentration data for all pollutants namely PM_{10}, NO_2, SO_2, SPM in worst case scenario for winter months (December and January) were taken. Fuzzy c-mean clustering and non-linear mapping were performed to calculate the best fit cluster and to identify the stations that are giving maximum information on air quality.

2.1 Technique used

Similarity Measures: Fuzzy relational calculus is one of the important facets in fuzzy set theory and has wide applications. Here various techniques used are explained in brief:-

Cosine Amplitude Method: - There are different ways to develop the numerical values that characterize a relation, but cosine amplitude method is one of the most dominant form of similarity measures to determining the relational values [13]. Cosine amplitude method is based on the relativity concept. Data samples of a set form a data array, say X and it is represented in the following equation for n data: $X = \{x_1, x_2, \cdots\cdots, x_n\}$. Each of the elements $x_1, x_2 \cdots, x_n$ in the data array X is itself a vector of length *m*, that is, $x_i = \{x_{i_1}, x_{i_2}, \cdots\cdots x_{i_m}\}$. Each element of a relation, r_{ij}, results from a pair wise comparison of two data samples, say x_i and x_j, where the strength of the relationship between data sample x_i and data sample x_j is given by $r_{ij} = \mu_R(x_i, y_j)$, such that $0 \le r_{ij} \le 1$. For the cosine amplitude method, r_{ij} is calculated by the following equation-

$$r_{ij} = \left|\sum_{k=1}^{m} x_{ik} x_{jk}\right| \Bigg/ \sqrt{\left(\sum_{k=1}^{m} x_{ik}^2\right)\left(\sum_{k=1}^{m} x_{jk}^2\right)} \tag{1}$$

where $i, j = 1, 2, \cdots, n.$

Computing all elements will form $n \times n$ fuzzy tolerance relation matrix. A fuzzy matrix relation must satisfy all three matrix conditions viz., reflexivity, symmetry and transitivity. Let R be a similarity relation and x, y be elements of a set X and $\mu_R(x, y)$ denote the grade of membership of the ordered pair (x, y) in R. Then R is a similarity relation in X if and only if, for all x, y and z in X, $\mu_R(x, x) = 1$ for all x in X (reflexivity), $\mu_R(x, y) = \mu_R(y, x)$ for all x and y in X (symmetry), and $\mu_R(x, z) \ge \max_{y} \in X\{\min\{\mu_R(x, y), \mu_R(y, z)\}\}$ for all x, y and z in X (transitivity) [13, 14]. If fuzzy relation matrix has only the properties of reflexivity and symmetry, then it is called fuzzy tolerance relation matrix. Before for defuzzification, fuzzy tolerance relation has to be converted to fuzzy equivalence relation by max-min composition method [16].

Max-min composition: - Considering $R_1(x, y), (x, y) \in X \times Y$ and $R_2(y, z), (y, z) \in Y, Z$ be two fuzzy tolerance relations. The max-min composition between R_1 and R_2 is then the fuzzy set is defined by following manner where $\mu_{R_1 \circ R_2}$ is the membership function of a fuzzy relation on fuzzy sets.

ISSN 2229-5518

$$\mu_{R_1 \circ R_2} = \left[(x,z), \max_y \left\{ \min\left\{ \mu_{R_1}(x,y), \mu_{R_2}(y,z) \right\} \right\} \right] \quad (2)$$

$x \in X, y \in Y, z \in Z$

Alpha cut method: - Begin by considering a fuzzy set R, then define a alpha-cut set R_α, where $0 \le \alpha \le 1$. The set R_α, is a crisp set called the alpha (α)-cut (or lambda (λ)- cut) set of the fuzzy set R, where $R_\alpha = \{x \mid \mu_{R(x)} \ge \alpha\}$. The α-cut set R_α is a crisp set derived from its parent fuzzy set, any particular fuzzy set R can be transformed into an infinite number of α-cut sets, because there are an infinite number of values α in the interval [0, 1]. Any element x in R_α belongs to fuzzy set R with a grade of membership that is greater than or equal to the value α.

Fuzzy c-mean clustering: - In fuzzy c-mean (FCM) clustering, a frequently used pattern recognition method, each sample point in a set of data belongs to more than one cluster with membership value between zero (completely unlike) and one (completely like) and the sum of all membership values of each sample point must be one [16].

The algorithm used in FCM cluster analysis for non-empty data set, fuzzy c-partition matrix U, for grouping a assembling of n data sets into c classes, is grounded on minimization of objective function J_m.

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^{m'} (d_{ik})^2 \quad (3)$$

$$d_{ik} = d(X_k - V_i) = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2 \right]^{1/2} \quad (4)$$

weighting parameter, m' is any real number ($1 \le m' \le \infty$) which controls the amount of fuzziness in the classification process and μ_{ik} is membership of k^{th} data point in i^{th} class. The Euclidean distance d_{ik} is between i^{th} cluster center and k^{th} data set (data point in m-space). i^{th} cluster center V_i has m features, which could be put in as vector form $V_i = \{v_{i1}, v_{i2}, \cdots v_{im}\}$.

Each of cluster coordinate for each class is calculated by following equation-

$$v_{ij} = \sum_{k=1}^{n} \mu_{ik}^{m'} . x_{kj} \Big/ \sum_{k=1}^{n} \mu_{ik}^{m'} \quad (5)$$

$1 \le i \le c;\ j = 1,2,\cdots\cdots,m.$

FCM is accomplished through an iterative optimization of the objective function, with the update of membership μ_{ik} and the cluster center v_{ij}. The steps in FCM algorithm are as follows:

1. Set up c value ($2 \le c < n$) and choose a value of parameter m'. Initial partition matrix, $U^{(0)}$ and each step in this algorithm will be marked as r, where $r = 0,1,2,\cdots$
2. Compute the c centers $\{V_i^{(r)}\}$ for each step.
3. Update partition membership matrix for r^{th} step, $U^{(r)}$ as equation

$$\mu_{ik}^{(r+1)} = \left[\sum_{j=1}^{c} \left(\frac{d_{ik}^{(r)}}{d_{jk}^{(r)}} \right)^{2/(m'-1)} \right]^{-1} \quad (6)$$

$1 \le k \le n; 1 \le i \le c$

4. Compare $U^{(r+1)}$ to $U^{(r)}$ in any convenient matrix norm. If $\left\| U^{(r+1)} - U^{(r)} \right\| \le \varepsilon_L$ (prescribed level of accuracy), stop; otherwise set $r = r + 1$ and return to step 2. This operation meets to a local minimum or a saddle point of J_m.

No general method is available to calculate the number of clusters in FCM. Basically, it is decided by domain expert or models using a suitable validity function. In this study a non-linear mapping (NLM) method has been used, based on the function partition coefficient F and the classification entropy H [16, 17]. The best number of cluster is chosen that gives the lowest H and highest F value. Formula of validity functions are shown in Table 1, where N is number of samples, C is number of clusters, u_{ki} is membership value for the k^{th} sample in the ith cluster. Disadvantages of the partition coefficient F is monotonic decreasing tendency when c is large [18].

TABLE 1
VALIDITY FUNCTION F, H AND THEIR LIMITING

Validity functional	Formula	Limits
Classification entropy	$F = \sum\sum\left[(u_{ki})^2/N\right]$	$1/C \le F \le 1$
Partition coefficient	$H = -\sum\sum\left[u_{ki} \times \log(u_{ki})/N\right]$	$0 \le H \le \log(C)$

3 CASE STUDY

The case study relates to Delhi, India where the Central Pollution Control Board (CPCB) has already installed AQMS at defined locations. Delhi, largest metropolis by area and the second-largest metropolis by population in India, is located at 28°22′48″N latitude and 77°7′12″E longitude and at an elevation of 216 m above mean sea level. The city is spread over 1,483 km2 of that 75.10% is urban and 22.9% is rural with a population of 16.75 million in 2011 [19]. With rapid population growth rate coupled with intensive infrastructure progress, heightened demand of energy, construction activities, industrial and transport sectors have resulted in significant air pollutants increase in Delhi.

In order to meet the energy demand, the city of Delhi has installed three big coal based thermal power plants – the Rajghat, the Indraprastha (IP) and the Badarpur, and three natural gas based power plants– the Indraprastha Power Generation Co Ltd. (IPGCL), the Pragati Power Station and Pragati-III Combined Cycle Power Plant.

From 1981 to 2011, the road length in Delhi increased from 14,316 to 31,183 km (2.18 times), whereas the number of registered vehicles increased from 0.52 to 6.93 million (13.3 times). Of 6.93 million vehicles, 94% are personal vehicles (31% cars, jeeps and 63% two wheelers) and 6% are commercial vehicles. Overall vehicular pollution or mobile transportation contributes 67% of the total air pollution loads in Delhi. There were about 148,680 industrial units during 2010-2011 [20]. These include engineering goods, textile, chemical, electronics, electrical goods, dyes and paints, steel, plastic, rubber and automobiles.

There exists a total of twelve manually operated AQMS in Delhi (Table 2). In this study daily average concentration from January 2008 to January 2010 for PM_{10}, NO_2, SO_2 and SPM in $\mu g/m^3$ were used.

Unfortunately, air quality parametric data was not available for three AQMS; therefore, such stations are excluded from the analysis. The excluded stations are: Ashok Vihar, Delhi College of Engineering and Bahadur Shah Zafar Marg (S10 to S12). In order to classify the air quality data, five linguistic descriptors: good, moderate, poor, very poor and severe were used to classify Indian air quality index [21].

TABLE 2
LOCATION OF AMBIENT AQMS

Station code	Location	Operated By	Land-use type	Station start year
S1	Mayapuri Industrial Area	NEERI[a]	Industrial	1995
S2	Shahdara	CPCB	Industrial	1988
S3	Shahzada Bagh	CPCB	Industrial	1988
S4	Janakpuri	CPCB	Residential	1988
S5	N.Y.School, Sarojini Nagar	NEERI[a]	Residential	1995
S6	Nizamuddin	CPCB	Residential	1988
S7	Pritampura	CPCB	Residential	2005
S8	Siri Fort	CPCB	Residential	1988
S9	Town Hall, Chandni Chowk	NEERI[a]	Residential	1995
S10	Ashok Vihar[b, c]	CPCB	Residential	1988
S11	Delhi College of Engineering[b]	CPCB	Residential	2005
S12	Bahadur Shah Zafar Marg[b]	CPCB	Residential	1988

[a] *National Environmental Engineering Research Institute;* [b] *Data not available;* [c] *Pritampura location was relocated from Ashok Vihar*

4 RESULTS AND DISCUSSION

Considering human health effects, the air pollutants PM_{10}, NO_2, SO_2 and SPM were selected for deciding AQMS providing maximum air quality information, but considering the magnitude of pollution level in Delhi, PM_{10} and NO_2 were used in methodology in section 2. Using the methodology outlined in section 2, first a contingency table of the number of occurrences of pollutant concentration in five linguistic class (IND-AQI descriptor) at each station were calculated and normalization of the data matrix for PM_{10} and NO_2 was done. After that cosine amplitude method was performed to get strength between stations (a, b in Fig. 1). But the corresponding matrices are fuzzy tolerance relational matrix, so in the process of defuzzyfication, fuzzy tolerance relational matrix is converted to fuzzy equivalence relational matrix by max-min composition for both pollutants (c, d in Fig. 1).

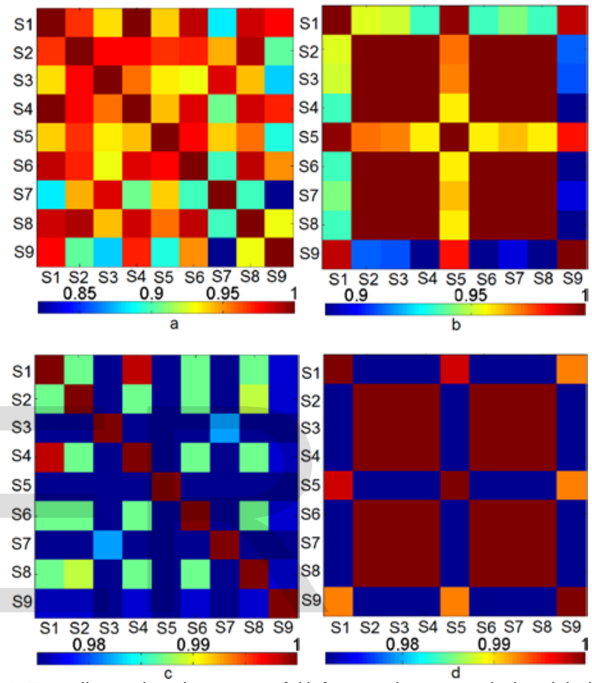

Fig. 1. Two dimensional contour of (I) fuzzy tolerance relation (similarity relation) matrix for (a) PM_{10} (b) NO_2 (II) fuzzy equivalence relation matrix for (c) PM_{10} (d) NO_2

After that in order to work out the similarity between nine stations, fuzzy equivalence relational matrix were converted to crisp equivalence relational matrix (Table 3 and Table 4) with alpha cut at 0.99.

TABLE 3
UNITS FOR MAGENTIC PROPERTIES

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	1	1	0	1	0	1	0	1	0
S2	1	1	0	1	0	1	0	1	0
S3	0	0	1	0	0	0	0	0	0
S4	1	1	0	1	0	1	0	1	0
S5	0	0	0	0	1	0	0	0	0
S6	1	1	0	1	0	1	0	1	0
S7	0	0	0	0	0	0	1	0	0
S8	1	1	0	1	0	1	0	1	0
S9	0	0	0	0	0	0	0	0	1

ISSN 2229-5518

TABLE 4
DEFUZZYFICATION RELATIONS FOR NO_2

	S1	S2	S3	S4	S5	S6	S7	S8	S9
S1	1	0	0	0	1	0	0	0	1
S2	0	1	1	1	0	1	1	1	0
S3	0	1	1	1	0	1	1	1	0
S4	0	1	1	1	0	1	1	1	0
S5	1	0	0	0	1	0	0	0	1
S6	0	1	1	1	0	1	1	1	0
S7	0	1	1	1	0	1	1	1	0
S8	0	1	1	1	0	1	1	1	0
S9	1	0	0	0	1	0	0	0	1

Inspection of the matrix (Table 3) that form after defuzzification of fuzzy equivalence relational matrix for PM_{10} shows that first, second, fourth, sixth and eighth columns are identical, i.e., S1, S2, S4, S6 and S8 are in one class; and columns the third, fifth, seventh and ninth are unique, indicate that S3, S5, S7 and S9 are stations each in their own class; these five different classes represent distinct stations. Similarly for NO_2 matrix, Table 4 shows that first, fifth and ninth columns are identical, i.e., S1, S5 and S9 are one class; and columns second, third, fourth, sixth, seventh and eighth are identical, i.e., S2, S3, S4, S6, S7 and S8 are in one class and these two different classes represent distinct stations.

For performing FCM, mean values of pollutant concentration were taken for PM_{10}, NO_2, SO_2 and SPM for nine stations in winter month (January and December 2008, January and December 2009, January 2010), considering the worst case situation. For the FCM study, m' as 2 and ε_L as 0.05 were taken. The number of clusters to be calculated must be specified in advance. Since the most appropriate number is not a priori known, solutions were calculated for cluster numbers between 3 and 7. The best solution was chosen by comparing the nonlinear maps as well as by calculating partition coefficient F and classification entropy H which indicate the robustness of the grouping from a mathematical viewpoint (Table 1) [16]. The best number of clusters in this sense is given by the lowest H and highest F value.

Cluster validity function are calculated (Table 5) and shown that six cluster solution is the best solution having lowest H and highest F value.

TABLE 5
VALIDITY FUNCTIONALS FOR THE CLUSTER SOLUTIONS FOR ALL DATA POINTS USING ALL VARIABLES. THE SCALED VALUES FOR F AND H ARE GIVEN. FROM THE MATHEMATICAL VIEWPOINT, THE SIX-CLUSTER SOLUTION IS THE BEST SOLUTION AS IT HAS THE LOWEST H AND THE HIGHEST F VALUE.

m'=2 and ε_L= 0.05			
No. of cluster	F	H	J_m
3	0.726	0.505	11614.8
4	0.754	0.512	6280.2
5	0.724	0.585	4994.1
6	0.827	0.393	2634.9
7	0.717	0.614	3778.8

Results of the FCM analysis (Table 6) for cluster six, showed that stations with similar characteristics came into same cluster with highest membership value.

TABLE 6
MEMBERSHIP VALUES OF NINE MONITORING STATIONS IN SIX CLUSTER

	S1	S2	S3	S4	S5	S6	S7	S8	S9
C1	0.698	0.016	0.016	0.009	0.007	0.000	0.000	0.021	0.276
C2	0.086	0.907	0.866	0.005	0.037	0.001	0.001	0.094	0.068
C3	0.060	0.025	0.024	0.030	0.014	0.997	0.000	0.039	0.122
C4	0.113	0.020	0.019	0.953	0.009	0.002	0.000	0.026	0.478
C5	0.018	0.011	0.026	0.001	0.115	0.000	0.995	0.178	0.024
C6	0.025	0.021	0.049	0.002	0.818	0.000	0.004	0.642	0.032

Highest membership value will determine the belongingness of the station to that specific cluster. Table 6 shows that S1 belongs to cluster number one (C1), consequently S2 and S3 are in second cluster (C2), S6 in third cluster (C3), S4 and S9 are in forth cluster (C4), S7 in fifth cluster (C5), S5 and S8 are in sixth cluster (S6).

Finally, from every cluster the stations were chosen which has unique column after defuzzyfication method and form its own class for PM_{10} or NO_2 (Table 3 and Table 4) or the station have highest membership function in the cluster. For example in cluster two (C2) S2 and S3 both have highest membership value, but from that cluster station S3 was chosen because it forms unique class of its own after defuzzification. Consequently the resulting stations chosen from six cluster were S1, S3, S5, S6, S7 and S9. So in terms of Indian air quality index the pollution level in the area S1 (Mayapuri Industrial Area), S3 (Shahzada Bagh), S5 (N.Y.School), S6 (Nizamuddin), S7 (Pritampura), and S9 (Town Hall) are high and in area S2 (Shahdara), S4 (Janakpuri) and S8 (Siri Fort) are low. Sahahdara (S2) is an industrial area but there are all packaging industry, so the pollution level is low.

From the results, it can be summarized that based on PM_{10}, NO_2, SO_2 and SPM dataset, CPCB and NEERI need to continue monitor air quality in Delhi only at the following locations: S1 (Mayapuri Industrial Area), S3 (Shahzada Bagh), S5 (N.Y.School, Sarojini Nagar), S6 (Nizamuddin), S7 (Pritampura), and S9 (Town Hall, Chandni Chowk). Stations S2 (Shahdara), S4 (Janakpuri) and S8 (Siri Fort) which started air quality monitoring since 1988 should be relocated to other places based on the criteria like high population density, traffic polluted area, availability of place and power supply etc.

5 CONCLUSIONS

Fuzzy similarity measure and fuzzy c-means cluster analysis, on winter season air quality data set of worst case scenario, resulted in a grouping according to the linguistic value used in Indian AQI. In this study nine monitoring stations were classified into six clusters. As one expects same effects from similar class of objects, one can expect the same in the data and classify ambient air quality monitoring stations based on their similarity measures and FCM. Combination of both results outlines six monitoring stations out of nine providing maximum information of present ambient air quality. In similarity measure method the alpha cut value is high (0.99) providing finer classification of the data set and from FCM method, stations were taken with highest membership value. Using FCM method, it is possible to avoid the problem of choosing an element for a particular class having more than one elements in fuzzy similarity classification. Lack of direct connection to a geometrical property can be seen as a disadvantage of the partition coefficient F. From presented study, it has been demonstrated that not every monitoring station needs to monitor all air quality parameters like PM_{10}, NO_2, SO_2 and SPM. It can be concluded that combination of both similarity measure and FCM is good approach to enquire the fitness of monitoring stations in AQMN.

ACKNOWLEDGMENT

The authors gratefully acknowledge Central Pollution Control Board (CPCB), India for giving the permission to use air quality parametric data in this research. The first author thanks CSIR for providing the scholarship to carry out research at CESE, IIT Bombay.

REFERENCES

[1] R. E. Munn, *The Design of Air Quality Monitoring Network*. MacMillan Publishers Ltd, London, 1981.
[2] J. T. Peterson., "Distribution of sulfur dioxide over metropolitan St. Louis, as described by empirical Eigenvectors, and its relation to meteorological parameters," *Atmospheric Environment*, vol. 4, no. 5, pp. 501–518, 1970. doi:10.1016/0004-6981(70)90020-X
[3] Y. Nakamori, S. Ikeda, and Y. Sawaragi, "Design of air pollutant monitoring system by spatial sample stratification," *Atmospheric Environment*, vol. 13, no. 1, pp.97–103, 1979.
[4] T. Husain, and S. M. Khan, "Air monitoring network design using Fisher's information measures—A case study," *Atmospheric Environment*, vo. 17, no. 12, pp. 2591–2598, 1983. doi:10.1016/0004-6981(83)90087-2
[5] P. M. Modak, and B. N. Lohani, "Optimization of ambient air quality monitoring networks : (Part I)," *Environmental Monitoring and Assessment*, vol. 5, no. 1, pp. 1–19, 1985. doi:10.1007/BF00396391
[6] P. M. Modak, and B. N. Lohani, "Optimization of ambient air quality monitoring networks : (Part II)," *Environmental Monitoring and Assessment*, vol. 5, no. 1, pp. 21–38, 1985. doi:10.1007/BF00396392
[7] S. Saksena, V. Joshi, and R. S. Patil, "Cluster analysis of Delhi's ambient air quality data," *Journal of Environmental Monitoring*, vol. 5, no. 3, pp. 491, 2003. doi:10.1039/b210172f
[8] Y. Kainuma, K. Shiozawa, and S. Okamoto, "Study of the optimal allocation of ambient air monitoring stations," *Atmospheric Environment*, vol. 24, no. 3, pp. 395–406, 1990. doi:10.1016/0957-1272(90)90047-X
[9] N. B. Chang, and C. C. Tseng, "Optimal evaluation of expansion alternatives for existing air quality monitoring network by grey compromise programing," *Journal of Environmental Management,* vol. 56, no. 1, pp. 61–77, 1999.
[10] R. W. Baldauf, D. D. Lane, and, G. A. Marote, "Ambient air quality monitoring network design for assessing human health impacts from exposures to airborne contaminants," *Environmental Monitoring and Assessment*, vol. 66, no. 1, pp. 63–76, 1999.
[11] F. I. Khan, and R. Sadiq, "Risk-Based Prioritization of Air Pollution Monitoring Using Fuzzy Synthetic Evaluation Technique," *Environmental Monitoring and Assessment*, vol. 105, no. 1-3, pp. 261–283, 2005. doi:10.1007/s10661-005-3852-1
[12] K. J. Maji, A. K. Dikdhit and A. Deshpande, "Can fuzzy set theory bring complex issues in sizing air quality monitoring network into focus?," Int J Syst Assur Eng Manag, Dec.14, 2014. DOI 10.1007/s13198-014-0327-1
[13] T. J. Ross, *Fuzzy Logic with Engineering Applications*, (*2nd Edition)*, John Wiley & Sons Ltd, 2004.
[14] H. J. Zimmermann, *Fuzzy Set Theory—and Its Applications (4th Edition),* Kluwer Academic Publishers, Boston, Dordrecht, London, 2001.
[15] L. A. Zadeh, "Similarity relations and fuzzy orderings," *Information Sciences*, vol. 3, no. 2, pp. 177–200, 1971. doi:10.1016/S0020-0255(71)80005-1
[16] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984. doi:10.1016/0098-3004(84)90020-7
[17] S. P. Vriend, P. F. M. Van Gaans, J. Middelburg, and A. De Nijs, "The application of fuzzy c-means cluster analysis and non-linear mapping to geochemical datasets: examples from Portugal," *Applied Geochemistry*, vo. 3, no. 2, pp. 213–224, 1988. doi:10.1016/0883-2927(88)90009-1
[18] X. L. Xie, and G. Beni, (1991). "A validity measure for fuzzy clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vo. 13, no. 8, pp. 841–847, 1991. doi:10.1109/34.85677
[19] Census of India, "Office of the Register General and Census Commissioner," 2011. URL. http://censusindia.gov.in. Accessed Jun 14. 2014
[20] SAD, "Statistical Abstract of Delhi (SAD) 2012," *Directorate of Economics and Statistics, Government of National Capital Territory of Delhi (GNCTD), New Delhi, 2012* URL. http://delhi.gov.in/DoIT/DES/Publication/abstract/SA2012.pdf.
[21] M. Sharma, M. Maheshwari, B. Sengupta, and B. P. Shukla, "Design of a website for dissemination of air quality index in India," *Environmental Modelling and Software*, vol. 8, no. 5, pp. 405–411, 003. doi:10.1016/S1364-8152(03)00003-3