Iran J Public Health, Vol. 45, No.7, July 2016, pp.917-925

Original Article

Air Quality Analysis by Using Fuzzy Inference System and Fuzzy C-mean Clustering in Tehran, Iran from 2009-2013

*Amir Abbas HAMEDIAN [1], Allahbakhsh JAVID [2], Saeed MOTESADDI ZARANDI [1], Yousef RASHIDI [3], *Monireh MAJLESI [1]*

1. Dept. of Environmental Health Engineering, School of Public Health,International Branch Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. School of Public Health, Shahroud University of Medical Sciences, Shahroud, Iran
3. Environmental Sciences Research Institute, Shahid Beheshti University, Tehran, Iran

***Corresponding Author:** Email: monireh_majlessi@yahoo.com

(Received 21 Aug 2015; accepted 12 Dec 2015)

Abstract

Background: Since the industrial revolution, the rate of industrialization and urbanization has increased dramatically. Regarding this issue, specific regions mostly located in developing countries have been confronted with serious problems, particularly environmental problems among which air pollution is of high importance.

Methods: Eleven parameters, including CO, SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2, benzene, toluene, ethyl-benzene, xylene, and 1,3-butadiene, have been accounted over a period of two years (2011-2012) from five monitoring stations located at Tehran, Iran, were assessed by using fuzzy inference system and fuzzy c-mean clustering.

Results: These tools showed that the quality of criteria pollutants between the year 2011 and 2012 did not as much effect the public health as the other pollutants did.

Conclusion: Using the air EPA AQI, the quality of air, and also the managerial plans required to improve the quality can be misled.

Keywords: Fuzzy c-mean clustering, Air quality, Fuzzy inference system, Iran

Introduction

Since the industrial revolution, the rate of industrialization and urbanization has increased dramatically. Regarding this issue, specific regions mostly located in developing countries have been confronted with serious problems, particularly environmental problems among which air pollution is of high importance (1). In order to manage this problem, several tools have been developed. Air quality indices are the most widely used tools which mainly indicate the quality of air in a specific region. In fact, air quality indices are widely used in air quality management schemes. Main air quality index being applied over the world is Air Quality Index or AQI (USEPA, 2009) developed by United States Environmental Protection Agency (2). In this index, one can calculate the quality of air by entering required values of major pollutants (i. e. CO, SO2, PM10, $PM_{2.5}$, O3, and NO2). Since the air quality indices have been deeply discussed in previously conducted studies, further details can be found (1, 3-8).

As it mentioned earlier, an air quality index not only shows the quality of air, but also can be used to design specific plans in handling air pollution. However, previously developed air quality indices have been mentioned to have main limitations. In a study (1) it was mentioned that the applicability of USEPA AQI in specific areas which are not similar to United States is limited. This index had been developed based on the quality of air and also the major pollutants in United States. Other limitations not related to the present study are discussed in detail in previously conducted studies (3, 9). Considering the fact that pollutants are being emitted into the air from various sources, it is wise to think that some of the pollutants are indigenous to specific regions. In addition to this, these pollutants are not included in abovementioned index or other air quality indices developed around the world. In conducted study by Sowlat et al. (1) an air quality, namely, fuzzy air quality index was developed. By comparing the result of FAQI with USEPA AQI, they indicated that the major part of pollution is resulted from specific pollutants which are normally ignored in routine assessment of air quality; in other words, the pollution was mostly caused by benzene, toluene, xylem, and ethyl-benzen. In other developed indices for assessing water quality, it has also been shown that the quality is mostly rooted from parameters not normally assessed. Considering the fact that an air quality index is a tool by which we are able to make plans so as to manage the problem when it is necessary, the mentioned problem can gravely affect our decision. The major pollutants which are indigenous to specific regions should be studied and indicated. Therefore, to the best of our knowledge, very few studies have been conducted aiming to indicate the local pollutants of different regions around the world.

The present study was aimed to indicate the major pollutants of Tehran, Iran, which is a highly populated area with a total population of approximately 10 million. Eleven parameters, including CO, SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2, benzene, toluene, ethyl-benzene, xylene, and 1,3-butadiene, have been accounted over a period of two years (2011-2012) from five monitoring stations located at Tehran. In fact, we selected these parameters to assess their concentrations and presence over a period of two years so as to find the major air pollutants of Tehran. At first, a set of input-output fuzzy parameters was defined for each mentioned parameter to have it is fuzzy based monthly quality; and also, fuzzy c-mean clustering was used to find the major part of each pollutant in the total quality of air in Tehran, Iran. Furthermore, the number of times, in monthly basis, which each pollutant's level had been higher than its set standard level are extracted in order to have a deep understanding of the quality of air in this area.

Materials and Methods

Fuzzy inference system

Fuzzy logic, suggested by Zadeh (10), as a tool for which is capable of thinking much more like the way human beings do. Over the past decade, fuzzy logic has been widely used to model and control complex systems (1, 11, 12). To develop a model by using this tool, we are able to use an individual's existing knowledge and experience as information (13). This property of fuzzy logic can be considered one of its main advantages not considered by other methods. There have been various models developed based on fuzzy logic (1, 11, 12, 14, 15).

We used fuzzy logic to have the fuzzy based quality of each parameter. This process requires knowing three main parts of the fuzzy inference system, including membership functions, fuzzy set operations and inference rules (13). Here, each parameter was assigned to a same output shown in Fig. 1.

A domain called the universe of discourse had been defined for each input and output (Set) divided into subsets and expressed by using linguistic terms. The linguistic terms shown in Fig. 1 were selected based on the terms used to define the air quality in USEPA AQI. We used trapezoidal membership functions based on the following equation and Table 1:

$$\text{Trapezoidal}: f(x;a,b,c,d) = \begin{Bmatrix} 0 & x < a \text{ or } d < x \\ \frac{(a-x)}{(a-b)} & a \le x \le b \\ 1 & b \le x \le c \\ \frac{(d-x)}{(d-c)} & c \le x \le d \end{Bmatrix}$$

In order to assign each subset to its corresponding subset in the output, fuzzy if-then rules were defined by extracting experts' knowledge. For each input-output relationship, we defined six rules. Totally 66 rules were defined in this study. The following rules are the examples of what we used in this study:

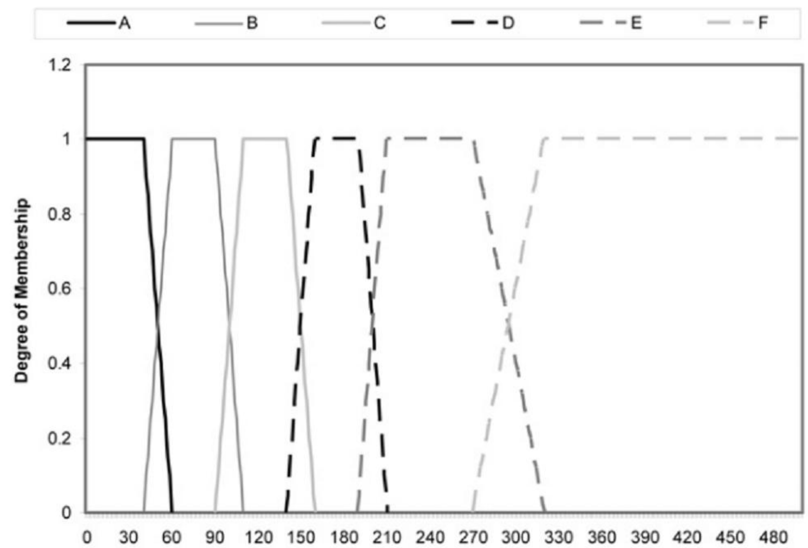

Fig. 1: Example of the trapezoidal membership functions for classifying the output.

Rule 7: If $PM_{2.5}$ is "low" then Quality is "Moderate".
Rule 42: If SO_2 is "extremely high" then Quality is "Hazardous".
Rule 13: If PM_{10} is "moderate" then Quality is "Unhealthy for Sensitive Groups".
All computations were carried out using the "fuzzy logic toolbox" in MATLAB version 7.9.0

Fuzzy c-mean clustering

Fuzzy c-means (FCM), originally introduced by Bezdek (16), is a technique in each data point belongs to a cluster to some degree that is specified by a membership grade. By using this technique, we are able to group data points, in this case concentrations of abovementioned parameters, into a specific number of different clusters. In fact, the purpose of using this method is to have the best structure regarding the applied data; and also, this method provides us with the ability to partition the data set into groups with similar individuals. The raw concentration of each parameter (monthly average) was entered to the program. We used three cluster because of the trend in the concentration of parameters. In other words, most of the parameters' concentrations were following a same trend individually.

Case Study

The present work is a Cross-Sectional study. However, this study was not conducted on the residents of Tehran. In fact, we assessed the concentration and duration of the pollutants over a period of two years. Hence, the group studied in this work was included of eleven parameters (i. e. CO, SO_2, PM_{10}, $PM_{2.5}$, O_3, NO_2, benzene, toluene, ethyl-benzene, xylene, and 1, 3-butadiene). The concentration of these pollutants, which are routinely monitored by Air Quality Administration, was collected from five monitoring stations, namely, Aghdasieh, Poonak, Park Roz, Geophysics, and Shahre Rey. Tehran, the capital city of Iran, is one of the most populated cities around the globe with a total population of approximately 12 million. Based on the reports of previously conducted studies, the major sources of air pollution in this city are motor vehicles and industries. High concentrations of air pollutants have been detected in Tehran. Due mainly to prevailing west-to-east winds and also the location of main industries at the west part of this city, wide range of pollutants, especially those with industrial origin(i. e. BTEX and 1,3-butadiene) are present in the air of this region.

Results

Fig. 2 shows the monthly mean concentrations of criteria pollutants at five abovementioned sampling stations. As shown in this figure, monthly averages of these pollutants are mostly lower than set standard levels by EPA. For instance, CO concentrations (Fig. 2a) at different sampling stations, except Shahr-e-Rey, were mostly below the standard level of 9 ppm (USEPA, 1997). In addition, the monthly mean concentrations of other pollutants are shown in Fig. 3. It can be seen from this figure that the mean average concentrations of these pollutants exceed the standard levels in most months of study period. For instance, benzene concentrations (Fig. 3a) are dramatically higher than the standard level of 5 ppb at all sampling stations.

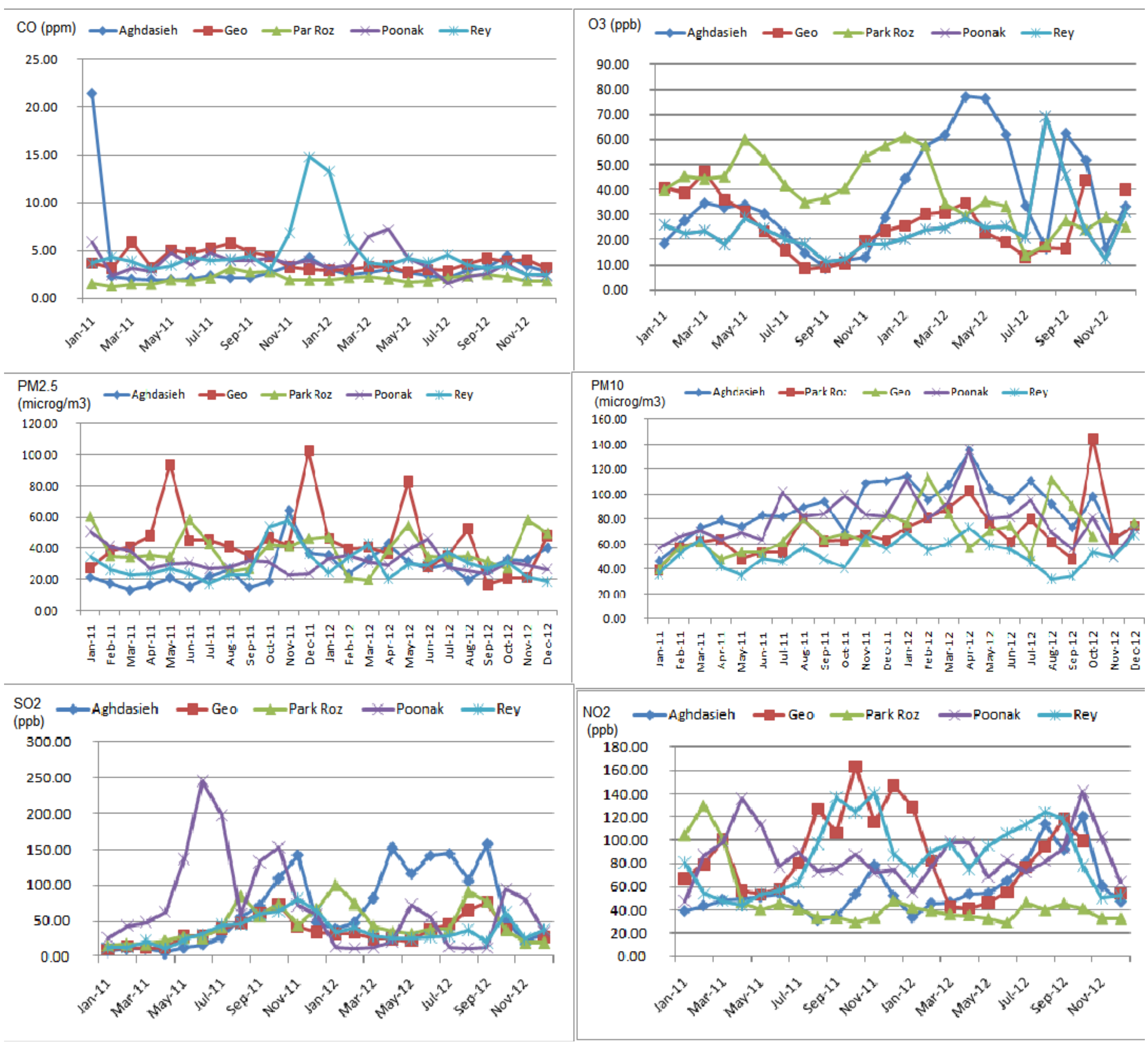

Fig. 2: Monthly mean concentrations of criteria air pollutants for the study period at different sampling stations

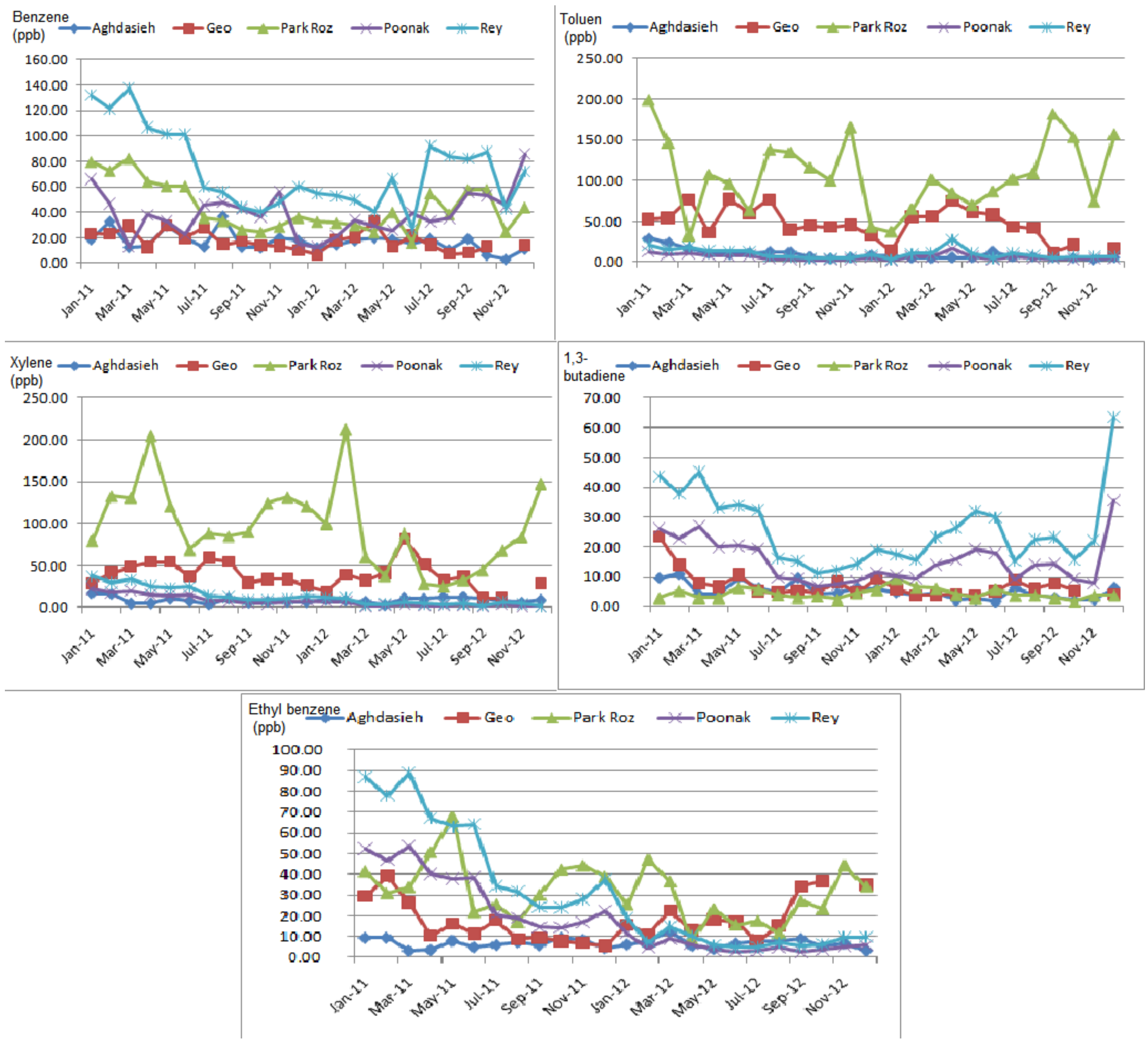

Fig. 3: Monthly mean concentrations of BTEX and 1, 3-butadiene air pollutants for the study period at different sampling stations

Table 1 indicates the individual quality of each pollutant in all the sampling stations derived by using fuzzy inference system. The fuzzy based quality of criteria pollutants cannot be considered as a threat to human health. However, the quality of other pollutants is considerably low. For instance, the quality of benzene and toluene is mostly E and/or F, indicating a hazardous condition.

Based on the results, almost all the clusters deduced for the criteria pollutants centered lower than the set standard level. We analyzed other pollutants as well, shown in Fig. 4, 5, 6, and 7. The concentrations of BTEX and 1,3-butadien are centered in almost all the sampling stations around the clusters representing the hazardous condition of air quality. In fact, these results show that relying on the index developed by USEPA cannot be a suitable option, since the main threat roots from the pollutants not included in that index.

Discussion

Based on the results of the present study, it can be implied that the application of EPA AQI is not suitable in regions with different air pollutants. In fact, the application of this index in regions with various sources of air pollutants and different range of pollutants could results in underestimating the quality of air. In other words, EPA AQI is included only with six parameters; whereas, there are many other pollutants emitted from varios sources (i. e. industry, and motor vehicles) which are abundantly present in mega cities like Tehran.

In a study conducted by Sowlat et al. (1) this is also insisted on. The index developed by them indicates a more stringent result, in comparison with that of EPA AQI. They reported that this difference is due to the difference in the number of pollutants included in each air quality index. The present study does not aim to consider the total air quality of Tehran, since previously conducted studies have shown that the current AQI index (EPA AQI) is not suitable for indicating the quality of air in this city.

We also tried to further analyze the quality of air between 2011 and 2012 by using fuzzy c-mean clustering. Based on the results, the monthly mean average concentrations of criteria pollutants in all the sampling stations are lower than the standard level. This is also in line with the results of the study conducted by Sowlat et al. (1). If one uses EPA AQI to assess the quality of air between the year 2011 and 2012, the quality of air resulted from using this index would not be what in the reality is. In other words, the results achieved for BTEX and 1, 3 butadiene which are shown in Fig. 4 and 5 for respectively Aghdasieh and Geophysics sampling stations confirm that the quality of air in these years is indeed affected by the parameters not included in the index developed by USEPA.

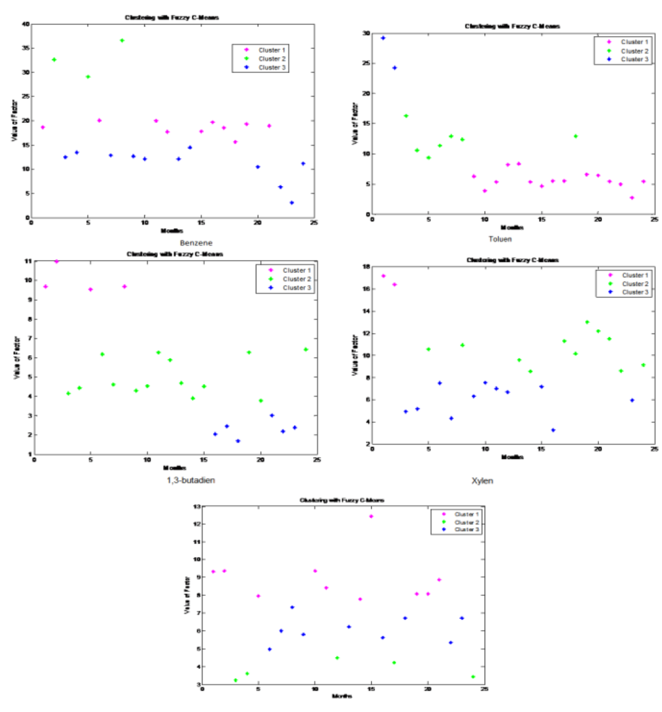

Fig. 4: FCM of the monthly mean concentrations of BTEX and 1, 3-butadien pollutants in Aghdasieh (sampling stations)

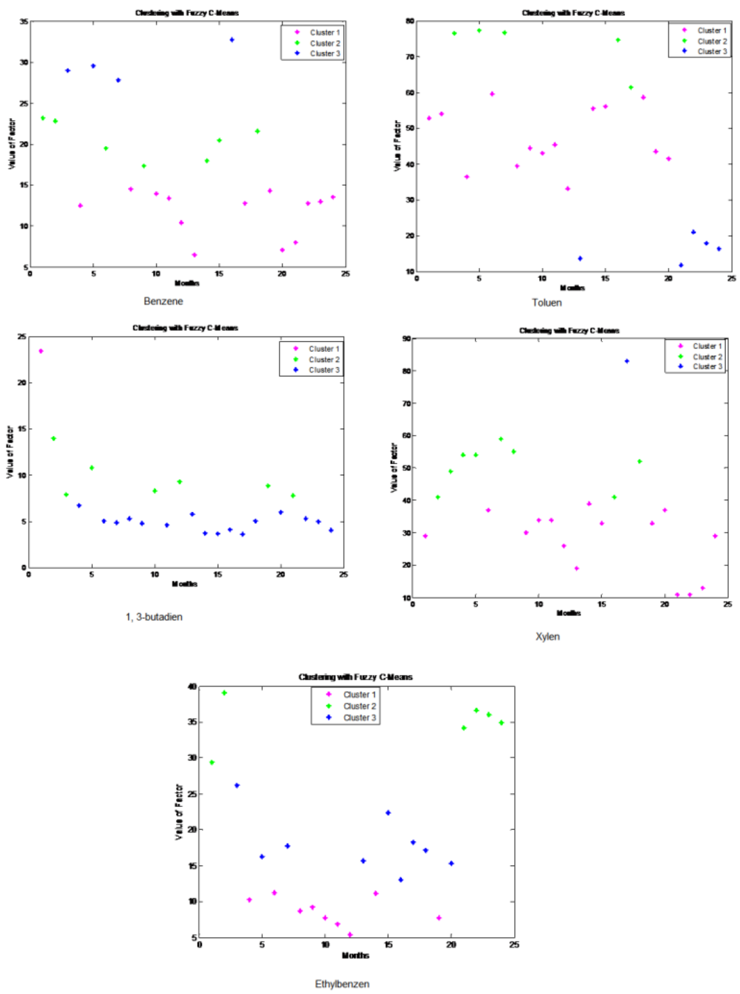

Fig. 5: FCM of the monthly mean concentrations of BTEX and 1, 3-butadien pollutants in Geophysics (sampling stations)

Considering the fact that an air quality index is developed to not only show the quality of air, but also be used as a tool to develop managerial plans, relying on the output of such index can results in completely different outcomes. The real problem can be missed, since the main pollutants are ignored in this index. Annually, a considerable amount of money is spent to improve the quality of air in Tehran. However, this all can be wasted if we do not know where the problem is rooted from.

Conclusions

The present study was conducted with the aim of assessing the main air pollutants of Tehran, Iran between 2011 and 2012. A broad range of air pollutants exist in the environments jeopardizing the health of human beings. It showed that the main pollutants of this city are not what general public believes. In fact, the main threats come from the pollutants which are not currently included in the air quality index developed by USEPA. In the present work, eleven parameters were assessed by using fuzzy inference system and fuzzy c-mean clustering. These tools showed that the quality of criteria pollutants between the year 2011 and 2012 did not as much effect the public health as the other pollutants did. Iit is clear that by using the air EPA AQI, the quality of air, and also the managerial plans required to improve the quality can be misled.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgment

This project is part of a PhD thesis, with the title of "modeling of the air pollutants distribution in Tehran by using hybrid fuzzy inference system and geographical information system (GIS) from 2009 to 2013", by Shahid beheshti University of Medical Sciences, Tehran, Iran.

References

1. Sowlat MH, Gharibi H, Yunesian M, Tayefeh Mahmoudi M, Lotfi S (2011). A novel, fuzzy-based air quality index (FAQI) for air quality assessment. *Atmos Environ, 45*: 2050-2059.
2. USEPA (2009). Air Quality Index, A Guide to Air Quality and Your Health.
3. Cheng WL, Chen YS, Zhang J, Lyons TJ, Pai JL, Chang SH (2007). Comparison of the revised air quality index with the PSI and AQI indices. *Sci Total Environ*, 382: 191-198.
4. Mayer H, Holst J, Schindler D, Ahrens D (2008). Evolution of the air pollution in SW Germany evaluated by the long-term air quality index LAQx. *Atmos Environ,* 42: 5071-5078.
5. Kyrkilis G, Chaloulakou A, Kassomenos PA (2007). Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects. *Environ Int,* 33: 670-676.
6. van den Elshout S, Léger K, Nussio F (2008). Comparing urban air quality in Europe in real time a review of existing air quality indices and the proposal of a common alternative. *Environ Int*, 34(5): 720-726.
7. Cairncross EK, John J, Zunckel M (2007). A novel air pollution index based on the relative risk of daily mortality associated with short-term exposure to common air pollutants. *Atmos Environ*, 41: 8442-8454.
8. Murena F (2004). Measuring air quality over large urban areas: development and application of an air pollution index at the urban area of Naples. *Atmos Environ*, 38: 6195-6202.
9. Silvert W (2000). Fuzzy indices of environmental conditions. *Ecol Model,* 130: 111-119.
10. Zadeh LA (1965). Fuzzy sets. *Inform Control,* 8(3): 338-353.
11. Gharibi H, Sowlat MH, Mahvi AH, Mahmoudzadeh H, Arabalibeik H, Keshavarz H, N Karimzadeh, Hassani GH et al. (2012). Development of a dairy cattle drinking water quality index (DCWQI) based on fuzzy inference systems. *Ecol Indic*, 20: 228-237.
12. Gharibi H, Mahvi AH, Nabizadeh R, Arabalibeik H, Yunesian M, Sowlat MH (2012). A novel approach in water quality assessment based on fuzzy logic. *J Environ Manage,* 112: 87-95.
13. Ross TJ (2004). *Fuzzy Logic with Engineering Applications*. New York: John Wiley & Sons.
14. Lermontov A, Yokoyama L, Lermontov M, Augusta M (2009). River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. *Ecol Indic*, 9: 1188-1197.
15. Chau KW (2006). A review on integration of artificial intelligence into water quality modelling. *Mar Pollut Bull*, 52(7): 726-733.
16. Bezdek JC (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers, Norwell, MA, USA. ISBN: 0-306406713.

Table 1: Fuzzy sets and linguistic terms for the inputs and the output

Intervals	1, 3-butadiene (ppb) 0-1500				Benzene (ppb) 0-50				Toluene (ppb) 0-1				Ethylbenzene (ppb) 0-30			
Linguistic Variable	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***
Good	0	0	200	250	0	0	2	3.5	0	0	0.004	0.006	0	0	0.4	0.6
Moderate	200	250	460	500	2	3.5	5	6	0.004	0.006	0.01	0.011	0.4	0.6	1	1.5
Unhealthy for Sensitive Groups	460	500	600	650	5	6	8	10	0.01	0.011	0.015	0.017	1	1.5	3	4
Unhealthy	600	650	1000	1100	8	10	12	15	0.015	0.017	0.03	0.037	3	4	6	8
Very Unhealthy	1000	1100	1500	1550	12	15	25	30	0.03	0.037	0.1	0.12	6	8	15	20
Hazardous	1500	1550	2000	2000	25	30	50	50	0.1	0.12	1	1	15	20	30	30

Table 1: Continued

												Criteria Group											
Xylene (ppb) 0-150				$PM_{2.5}$ (µg/m^3) 0-180				PM_{10} (µg/m^3) 0-600				CO (ppm) 0-50				O_3 (ppb) 0-600				SO_2 (ppb) 0-1000			
a	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***a***	***b***	***c***	***d***
0	0	12	18	0	0	20	30	0	0	25	75	0	0	3	5	0	0	50	90	0	0	15	45
12	18	30	35	20	30	50	60	25	75	125	175	3	5	8	10	50	90	115	135	15	45	120	160
30	35	45	50	50	60	80	90	125	175	225	275	8	10	11	13	115	135	155	175	120	160	200	240
45	50	60	70	80	90	110	120	225	275	325	375	11	13	14	20	155	175	190	210	200	240	280	350
60	70	100	110	110	120	140	150	325	375	405	435	14	20	25	35	190	210	380	430	280	350	550	650
100	110	150	150	140	150	180	180	405	435	600	600	25	35	50	50	380	430	600	600	550	650	1000	1000

Table 1: Continued

NO_2 (ppb) 0-2000				Output Quality 0-500			
a	***b***	***c***	***d***	***a***	***b***	***c***	***d***
0	0	6	120	0	0	40	60
60	120	180	240	40	60	90	110
180	240	280	340	90	110	140	160
280	340	600	700	140	160	190	210
600	700	1100	1400	190	210	270	320
1100	1400	2000	2000	270	320	500	500