Modelling and Evaluating Air Quality with Fuzzy Logic Algorithm-Ankara-Cebeci Sample

İsmail Atacak1, Nursal Arıcı2 and Dilem Güner3

Accepted: 24/11/2017 Published: 28/12/2017

Abstract: Air is one of the most important life sources for all living things. Gases that are present and absent in the composition of clean air also considered as pollutants in the atmosphere. If the pollutants rise above a certain concentration level, air pollution occurs. Air pollution damages all living things, especially human health. Accurate estimation of pollutant concentrations through air pollution modeling has an important effect in reducing the adverse effects of air pollution and taking necessary precautions. Conventional statistical models are widely used in air pollution forecasting and modeling. As a different approach, in this study, fuzzy logic algorithm (FLA), which has been increasingly successful in many field applications, is used to model air quality and air pollution analyses are made based on this model. Ankara province data is used in the sample of the research.

Keywords: Air pollution, fuzzy logic algorithm, air quality index, pollutant concentrations

1. Introduction

Air is a gas mixture in its composition that contains unchanging gases such as nitrogen and oxygen, changing gases such as carbon dioxide, ozone and water spring, and pollutants such as sulfur dioxide, carbon monoxide, ozone oxide and particle matter. Air is one of the indispensable sources of life for all living beings. Air pollution occurs when pollutants rise above a certain concentration value in air. Increases in the amount of pollutants in air are caused by natural sources such as forest fire, volcano eruption, and artificial sources such as heating devices, industrial establishments, vehicles [1 - 3]. Urbanization and industrialization have increased with the increasing population density and developing technology in the world. As a result, the air pollution has a significant impact on the global scale as well as on the local and regional scale. This has become a threat to all living things, especially to human health. Therefore, it is now an urgent necessity to determine the strategies and methods to improve air quality and prevent air pollution. Public authorities need a well-informed community support to succeed in their strategies [4]. Information on air quality status related to regions in all the countries of the world and its health effects are presented by public authorities via Air Quality Index (AQI). The index provides a viable flow of information to transform air pollution data into an understandable format and taking appropriate measures to protect public health [5]. AQI performs this function by dividing the air quality according to measured pollutants into classes such as good, moderate, unhealthy and hazardous, and assigning colours to each class including different definition intervals. AQI is defined according to each country's own methods and criteria. Accurately estimating the concentrations of pollutants by

1 Computer Eng., Technology Fac., Gazi Univ. Ankara – 06500, TURKEY
2 Computer Eng., Technology Fac., Gazi Univ. Ankara – 06500, TURKEY
3 Computer Eng., Sciences Inst., Gazi Univ. Ankara – 06500, TURKEY
* Corresponding Author: İsmail Atacak Email: iatacak@gazi.edu.tr

modelling the air quality has an important effect in reducing the adverse effects of pollution and taking the necessary precautions. In air quality estimation and modelling, traditional statistical models such as regression analysis, clustering analysis, variance analysis are widely used. However, these models do not provide the desired success in determining air quality due to the presence of complex and non-linear relationships between pollutants, and the inability to perceive large data sets [6]. In this study, an AQI model is designed to measure air quality status by the FLA, which is known to produce successful results in solving problems with nonlinear structure. The rest of the study is organized in the following way: Section 2 presents a literature review on the FLA-based air quality. The modelling of the air quality with the FLA is described in detail according to the pollutant data of Ankara Province, Cebeci region in Section 3. Section 4 presents the results and evaluations of the proposed model. In the last section, the conclusions we obtained are given.

2. Related Works on FLA-Based Air Quality

Studies on air quality over the last 10 years have shown that statistical and analytical methods cannot achieve the desired performance on modeling and estimating air quality. The nonlinear relationship between pollutants and air quality, and the inadequacies of working with large data sets of related methods are among the main reasons for this failure. Therefore, researchers needed new methods that make a sensitive assessment by establishing a relationship between contradictory observations and that correlate sensitively between nonlinear relations in environmental issues such as air quality assessment. The FLA proposed by Zadeh in 1965 is one of well-suited candidates to be used for solution of environmental problems. The first important application of FLA in environmental decision-making and imaging problems was performed by Silvert in 2000. In this study, Silvert applied FLA to environmental conditions and showed that this algorithm provides a powerful and useful formalism for classifying environmental conditions and defining anthropogenic changes [7].
The studies conducted in this area in the following years are presented in chronological order below.

In the study conducted by Pokrovsky, Kwok and Ng in 2002, the fuzzy sets and artificial neural network approaches were applied to air pollution parameters such as wind speed, wind direction, solar radiation, air temperature, and \( \text{O}_3, \text{NO}_2, \text{NO}, \text{SO}_2 \) to define the urban air pollution [8]. In the study conducted by Sowlat, Gharibi, Yunesian, Mahmodi and Lotfi in 2011, a FLA-based AQI was developed. In order to determine the air quality index, they used air pollutants such as \( \text{CO}, \text{SO}_2, \text{PM}_{10}, \text{O}_3, \text{NO}_2 \) and other harmful substances such as benzene, toluene, ethyl benzene, xylene known to have adverse effects on human health. Using the air quality data from five different monitoring stations in Tehran from January 2008 to December 2009, they proved that the FLA can successfully determine the AQI [9]. Another study, using FLA to determine air quality, was conducted by Mandal, Gorai and Pathak in 2012. In this study, the AQI was modelled with the FLA as in the previous study, and the results are compared with the traditional analytical hierarchy approach [10]. In another study conducted by Assimakopulos, Dounis, Spanou, and Santamouris in the same year, the AQI for indoor environments were evaluated with FLA according to \( \text{CO}_2, \text{PM}_{10}, \text{PM}_{2.5}, \text{PM}_1 \) pollutants and passenger counts measured in the Athens metro [11]. In 2014, Milla dos Santos and their friends conducted a study on health risk assessment by taking maximum and average concentrations of São Paulo city pollutant parameters \( \text{SO}_2, \text{NO}_2, \text{O}_3, \text{CO} \) and \( \text{PM}_{10} \) into account. In this study, they used both FLA and geographic information systems algorithm. They showed that this study can produce useful results in this area if two algorithms are used together [12]. In the study conducted by Dursun, Kunt and Taylan in 2015, the FLA has been utilized to show the influence of some environmental conditions on air quality and \( \text{SO}_2 \) pollution level. \( \text{SO}_2 \) concentrations and meteorological data were collected from 15 sites selected from Konya City for the estimation of the air quality. The results of the model were evaluated with fuzzy quality schedules and compared with the results obtained from Turkish Environmental Protection Association air quality standards [13].

3. Modeling Air Quality with FLA

The modeling of the air quality is carried out through air quality sampling data obtained from the web site of the Ministry of Environment and Urban Planning (EUP Ministry) in Ankara. This data consists of a data set for January-December of 2010-2016 which was watched in the Cebeci area [14]. The data set contains sulfur dioxide \( (\text{SO}_2) \), nitrogen dioxide \( (\text{NO}_2) \), carbon monoxide \( (\text{CO}) \), ozone \( (\text{O}_3) \) and particulate matter \( (\text{PM}_{10}) \) as pollutants. The model to be used in determining the air quality is created considering the data containing these matters. For the monitoring data above the boundary values in the data set, the maximum values determined by the EUP Ministry shown in Table 1 are taken into consideration. In addition, the data from the average values of the pollutants are used instead of missing data in the dataset so that the model can produce more healthy results.

In the first step of creating the FLA-based model, the input and output variables of the system are determined. In general, the air quality is determined by the AQI and the value of this parameter is decided according to pollutants in the air. Therefore, \( \text{SO}_2, \text{NO}_2, \text{CO}, \text{O}_3, \) and \( \text{PM}_{10} \) pollutants are selected as the input variables and the AQI parameter is selected as the output variable.

In the second step of modelling, the boundary values of the universal sets for the selected input and output variables are determined and the fuzzy sets to be defined in these universes are detected for the fuzzification process. The boundary values of universal sets are decided by using the boundary values given in Table 1. Each variable is represented by three different fuzzy sets, “LOW”, “MODERATE”, “HIGH” in these universes. For these fuzzy sets, both the trapezoidal and triangular membership functions are used. The boundary values of “LOW”, “MODERATE” and “HIGH” fuzzy sets are determined in the form corresponding to the boundary values of the sets defined as “GOOD”, “MODERATE-SENSITIVE” and “UNHEALTHY-VERY UNHEALTHY-HAZARDOUS” by EUP Ministry, respectively. Table 2 shows the boundary values of universal sets and fuzzy sets for input variables.

| Index | AQI | \( \text{SO}_2 \) (\( \mu \text{g/m}^3 \)) | \( \text{NO}_2 \) (\( \mu \text{g/m}^3 \)) | \( \text{CO} \) (\( \mu \text{g/m}^3 \)) | \( \text{O}_3 \) (\( \mu \text{g/m}^3 \)) | \( \text{PM}_{10} \) (\( \mu \text{g/m}^3 \)) |
|-------|-----|-------------------|---------------------|------------------|-------------------|-------------------|
| GOOD  | 0 to 50 | 0 to 100 | 0 to 100 | 0 to 100 | 0 to 100 | 0 to 100 | 50 to 50 | 100 to 100 |
| MODERATE | 51 to 100 | 101 to 250 | 101 to 250 | 550 to 1000 | 121 to 1000 | 51 to 100 |
| SENSITIVE | 101 to 150 | 251 to 500 | 251 to 500 | 1000 to 2000 | 161 to 2000 | 101 to 200 |
| UNHEALTHY | 150 to 200 | 501 to 800 | 501 to 800 | 1600 to 2400 | 181 to 2400 | 180 to 200 |
| VERY UNHEALTHY | 200 to 300 | 851 to 1300 | 1001 to 2000 | 2400 to 3200 | 241 to 3200 | 261 to 320 |
| HAZARDOUS | >1100 | >2000 | >2000 | >3200 | >2400 | >2400 | >701 | >521 |

The membership functions for fuzzy sets belonging to input variables are defined according to the boundary values given in Table 2. Fig. 1 and Fig.2 show the triangular and trapezoidal type membership functions defined for \( \text{SO}_2, \text{NO}_2, \text{CO}, \text{O}_3 \) and \( \text{PM}_{10} \) pollutants, respectively.
Fig. 1. Triangular type membership functions defined for SO$_2$, NO$_2$, CO, O$_3$ and PM$_{10}$ pollutants: (a) for SO$_2$, (b) for NO$_2$, (c) for CO, (d) for O$_3$, (e) for PM$_{10}$.

Fig. 2. Trapezoidal type membership functions defined for SO$_2$, NO$_2$, CO, O$_3$ and PM$_{10}$ pollutants: (a) for SO$_2$, (b) for NO$_2$, (c) for CO, (d) for O$_3$, (e) for PM$_{10}$. 
The AQI, selected as an output variable, is represented by six fuzzy sets: “GOOD”, “MODERATE”, “SENSITIVE”, “UNHEALTHY”, “VERY UNHEALTHY” and “DANGEROUS”. Just as in determining the boundary values of the fuzzy sets defined for the input variables, the boundary values of these sets are also decided by considering the value ranges used by the EUP Ministry. Table 3 shows the boundary values of universal sets and fuzzy sets for the output variable.

### Table 3. The boundary values of universal sets and fuzzy sets for output variable

| Boundary values for universal set | Boundary values for triangular fuzzy sets | Boundary values for trapezoidal type fuzzy sets |
|----------------------------------|------------------------------------------|-----------------------------------------------|
| Output variable                  | Fuzzy sets                                |                                               |
| AQI 0-500                        | GOOD 0-40-60                              | 0-40-60                                       |
|                                  | MODERATE 40-75-110                        | 40-60-90-110                                  |
|                                  | SENSITIVE 90-120-170                      | 90-110-130-170                                |
|                                  | UNHEALTHY 130-175-220                    | 130-170-180-220                               |
|                                  | VERY UNHEALTHY 180-250-320               | 180-220-280-320                               |
|                                  | HAZARDOUS 280-320-500                    | 280-320-500                                   |

The triangular and trapezoidal membership functions defined for the AQI according to the boundary values given in Table 3 are shown in Fig.3 and Fig.4, respectively.

![Fig. 3. Triangular type membership functions defined for AQI](image)

![Fig. 4. Trapezoidal membership type functions defined for AQI](image)

In the third step of modelling, the rule base representing the relationship between input variables and output variables is formed. Since there are 5 input variables in the AQI model and each input variable is represented by 3 fuzzy sets, a total of 243 (3^5) rules are used in the rule base. Rules are established by considering the highest values of pollutants. A cross-section from the rule base formed is given in Table 4.

### Table 4. Rule base

| Rule number | SO2 (μg/m3) | NO2 (μg/m3) | CO (μg/m3) | O3 (μg/m3) | PM10 (μg/m3) | AQI    |
|-------------|-------------|-------------|------------|------------|--------------|--------|
| 1           | LOW         | LOW         | LOW        | LOW        | LOW          | GOOD   |
| 2           | LOW         | LOW         | MODERATE   | MODERATE   | LOW          | MODERATE |
| 3           | LOW         | LOW         | LOW        | MODERATE   | HIGH         | UNHEALTHY |
| ..          | ..          | ..          | ..         | ..         | ..            | ..      |
| 242         | HIGH        | HIGH        | HIGH       | MODERATE   | VERY         | UNHEALTHY |
| 243         | HIGH        | HIGH        | HIGH       | HIGH       | HIGH         | HAZARDOUS |

In the last step of the modeling, the methods to be used in the inference and defuzzification process are decided. In general, there are 4 types of inference strategies used in the composition of rules in the inference procedure: max-min composition, max-product composition, max-drastic-product composition and max-bounded-product composition. Among them, the max-min and max-product inference strategies are widely used in FLA-based applications due to their computational simplicity and efficiency. The difference between the max-product and max-min inference strategies can be explained in the way that the max-product inference strategy produces a lower change in the output of the FLA than the max-min inference strategy because, in this strategy, the effect of the corresponding rule on the composition is decided by using the product operator. On the other hand, many methods such as the max-membership principle, centroid, weighted average, and center of sums are used in the defuzzification process. Among them, the centroid method produces a more accurate result than the other methods because it finds the value that corresponds to the center of gravity of the shape obtained as a result of inference as the defuzzified result. Thus, in this study, while the max-min inference strategy is used in the inference process, the centroid method is used in the defuzzification process.

### 4. Results and Discussions

In the FLA-based AQI model presented in Chapter 3, the input variables with the highest values are effective in determining the AQI value. Fig. 5 shows the surface graphics representing the changes in AQI with respect to double combinations of pollutants used as input variables (SO2 and NO2, SO2 and CO, SO2 and O3, SO2 and PM10, NO2 and CO, NO2 and O3, NO2 and PM10, CO and O3, CO and PM10, O3 and PM10). It is clearly seen from the graphics that the changes in the pollutant values given in double combinations are directly affect the AQI. The increases above the boundary values for the input variables led to a deterioration of the air quality by increasing the AQI value, while the changes below these values decrease ensures being good of the air quality by decreasing the AQI value.

In order to obtain the test results of the FLA-based AQI model, a data set containing 300000 data related to SO2, NO2, CO, O3 and PM10 pollutants is used. The evaluations are made on the AQI values calculated on hourly and daily basis for the FLA-based AQI models with triangular and trapezoidal membership functions. The graphics representing the distributions of index data by months, which is obtained from the FLA-based AQI model with both the trapezoidal and triangular membership functions, are shown in Fig. 6. In the graphics depicting the distributions of AQI between 2010 and 2016, the distributions represented by blue dots show the AQI values calculated by the AQI model with the trapezoidal membership functions, whereas the distributions represented by
red dots show AQI values calculated by the AQI model with the triangular membership functions. It can be concluded from the graphics that the AQI values of the Ankara-Cebeci region except for 2012 are reached to a high level during autumn and winter months in some years, and only winter months in some years. In addition to these months, the summer months are added in 2012, and even the AQI value reaches to the highest level of the last 6 years in the summer months.

Fig. 5. Surface graphics representing the changes in AQI with respect to different pollutants used as input: (a) changes in AQI with respect to SO$_2$ and NO$_x$, (b) changes in AQI with respect to SO$_2$ and CO, (c) changes in AQI with respect to SO$_2$ and O$_3$, (d) changes in AQI with respect to SO$_2$ and PM$_{10}$, (e) changes in AQI with respect to NO$_2$ and CO, (f) changes in AQI with respect to NO$_2$ and O$_3$, (g) changes in AQI with respect to NO$_2$ and PM$_{10}$, (h) changes in AQI with respect to CO and O$_3$, (i) changes in AQI with respect to CO and PM$_{10}$, (j) changes in AQI with respect to O$_3$ and PM$_{10}$.

Fig. 6. Graphics representing the distributions of index data by months, which are obtained from FLA-based AQI model with both the trapezoidal and triangular membership functions according to measured pollutant data between 2010 and 2016: (a) for the year 2010, (b) for the year 2011, (c) for the year 2012, (d) for the year 2013, (e) for the year 2014, (f) for the year 2015, and (g) for the year 2016.
The statistical data, comparing the AQI results on hourly basis of FLA-based AQI model with the ones of the current method used by the EUP Ministry, is shown in Table 5.

Table 5. Comparison of the AQI results on hourly basis of FLA-based AQI model and the ones of the current method used by the EUP Ministry.

| Statistical data of EUP Ministry | Statistical data of FLA-based AQI model with | Statistical data of FLA-based AQI model with |
|---------------------------------|---------------------------------------------|---------------------------------------------|
|                                | trapezoidal membership functions            | triangular membership functions             |
| GOOD                            | %42.46                                      | %48.40                                      |
| MODERATE                        | %36.15                                      | %40.34                                      |
| SENSITIVE                       | %18.85                                      | %9.70                                       |
| UNHEALTHY                       | %2.08                                       | %3.28                                       |
| VERY                            | %0.31                                       | %0.08                                       |
| DANGEROUS                       | %0.16                                       | %0.00                                       |

When the results are analyzed in general, it is seen that, in some ranges, the AQI results produced by the FLA-based AQI model are closer to the ones produced by the current model whereas, in other ranges, these results are different from each other. The difference between the results of the current model and the FLA-based AQI model results from the dissimilarity in deciding the AQI. While the current model shows a sharper approach by taking into consideration the highest pollutant value to determine the AQI, the FLA-based AQI model provides a more flexible approach by determining this index with rules based on the human experience. When the results of both FLA based models are analyzed, it can be seen that the FLA-based AQI model with the trapezoidal membership function produces closer results to the current model than the FLA-based AQI model with triangular membership functions.

5. Conclusion

In this study, an AQI model based on the FLA that determines the air quality according to the pollutant data watched in Ankara province-Cebeci area between 2010 and 2016 was developed and a number of analyzes were conducted through the related model in order to test its performance. Analyzes on the FLA-based models constructed using triangular and trapezoidal type membership functions showed that they can successfully determine the air quality in the mentioned region. In addition, observations on the results of the proposed model and the current method used by the EUP Ministry showed that the model with trapezoidal type membership functions produces closer results to the current method. In conclusion, these results proved to us that the FLA has successfully determined the AQI and so it can provide important contributions to the solution of the problems encountered in air quality estimation.

Acknowledgment

This study was presented as a summary paper in the 1st International Symposium on Multidisciplinary Studies and Innovative Technologies, Tokat/TURKEY, November 2-4, 2017.

References

[1] Z. Gemici, O. Kale and B.K. Çatal, “Hava kalitesi izleme sürecinde mevcut durum ve cafe direktifi gerekliklerine göre kenteşmevede hava kalitesi unsuru,” 6. Ulusal Hava Kirliliği ve Kontrolü Sempozyumu, İzmir, Türkiye, 2015, s. 910-923.
[2] B. Aydmar, H. Güven and S. Kırksekti, “Hava Kirliliği Nedir, Ölçüm ve Hava Kalite Modelleme Yöntemleri Nelerdir,” SÜ Fen Bilimleri Enstitüsü Çevre Mühendisliği Bölümü Hava Kirliliği ve Modellemesi Dergisi, 2009. [Online]. Available: www.Sahakk.sakarya.edu.tr/documents/hava%20kirliliği%20ve%20modellemesi%20I.pdf, Accessed on: Nov. 01, 2017.
[3] E. Kardeşoğlu, M. Yalçın and Z. İslak, “Hava Kirliliği ve Kardiyovasküler Sistem,” TAF Preventive Medicine Bulletin, cilt. 10, sayı. 1, s. 97-106, 2011.
[4] M. Sharma, R. Pandey, M. Maheshwari and B. Sengupta, “Interpretation of air quality data using an air quality index for the city of Kanpur,” India Journal of Environmental Engineering and Science, vol. 2, no. 6, pp. 453-462, 2003.
[5] G. Kyrkilis, C. Arhontoulia and A. K. Pavlos, “Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects,” Environment International, vol. 33, no. 5, pp. 670-676, 2007.
[6] İ. Büyük, “Hava kirliliğinin yapay zeka teknikleri ile belirlenmesi,” Y. L. Tezi, Fen Bilimleri Enstitüsü, Bahçeşehir Üniversitesi, İstanbul, Türkiye, 2010.
[7] W. Silvert, “Fuzzy indices of environmental conditions,” Ecological Modelling, vol. 130, no. 1, pp. 111-119, 2000.
[8] O. Pokrovsky, R. Kwok and C. Ng, “Fuzzy logic approach for description of meteorological impacts on urban air pollution species: a Hong Kong case study,” Computers & Geosciences, vol. 28, no. 1, pp. 119-127, 2002.
[9] M. Sovlat, H. Gharibi, M. Yusnesian, M. Tayefeh Mahmoudi and S. Lotﬁ, “A novel, fuzzy-based air quality index (FAQI) for air quality assessment,” Atmospheric Environment, vol. 45, no. 12, pp. 2050-2059, 2011.
[10] T. Mandal, A. K. Gorai and G. Pathak, “Development of fuzzy air quality index using soft computing approach,” Environmental Monitoring and Assessment, vol. 184, no. 10, pp. 187-196, 2012.
[11] M. Assimakopoulous, A. Dounis, A. Spanou and M. Santamouris, “Indoor air quality in a metropolitan area metro using fuzzy logic assessment system,” Science of The Total Environment, vol. 449, pp. 461-469, 2012.
[12] A. Milla dos Santos, A. Passuello, M. Schuhmacher, M. Nadal, J. Domingo, C. Martinez, S. Segura-Muñoz and A. Takayanagi, “A Support Tool for Air Pollution Health Risk Management in Emerging Countries: A Case in Brazil,” Human and Ecological Risk Assessment: An International Journal, vol. 20, no. 5, pp. 1406-1424, 2014.
[13] S. Dursun, F. Kurt and O. Taylan, “Modelling sulphur dioxide levels of Konya city using artificial intelligent related to ozone, nitrogen dioxide and meteorological factors,” International Journal of Environmental Science and Technology, vol. 12, no. 12, pp. 3915-3928, 2015.
[14] Ankara iline alt 1 Ocak 2010-31 Ekim 2016 tarihleri arasındaki hava kalitesi verileri, T.C. Çevre ve Şehircilik Bakanlığı WEB sitesi, [Online]. Available: http://www.havaizleme.gov.tr/Default.ltr.Aspx Basic Accessed on: Jan. 12, 2016.