# Air Pollution Analysis Using Enhanced K-Means Clustering Algorithm for Real Time Sensor Data

Kingsy Grace. R[1], Manimegalai. R[2], Geetha Devasena. M.S[1], Rajathi. S[1], Usha. K[1], Raabiathul Baseria. N[1]
Department of Computer Science and Engineering
[1] Sri Ramakrishna Engineering College
[2] Park College of Technology
Coimbatore, India
kingsygrace.r@srec.ac.in, mmegalai@yahoo.com, msgeetha@srec.ac.in

***Abstract*— Air pollution affects body organs and human systems in addition to the environment. Smart air pollution monitoring consists of wireless sensor nodes, server and a database to store the monitored data. Huge amounts of data are generated by gas sensors in air pollution monitoring system. Traditional methods are too complex to process and analyze the voluminous data. The heterogeneous data are converted into meaningful information by using data mining approaches for decision making. The K-Means algorithm is one of the frequently used clustering method in data mining for clustering massive data sets. In this paper, enhanced K-Means clustering algorithm is proposed to analyze the air pollution data. The correlation coefficient is calculated from the real time monitored pollutant datasets. The Air Quality Index (AQI) value is calculated from the correlation co-efficient to determine the air pollution level in a particular place. The proposed enhanced K-Means clustering algorithm is compared with Possibilistic Fuzzy C-Means (PFCM) clustering algorithm in terms of accuracy and execution time. Experimental results show that the proposed enhanced K-Means clustering algorithm gives AQI value in higher accuracy with less execution time for when compared to existing techniques.**

**Keywords— PFCM; Enhanced K-Means Clustering; Analysis; Air Pollution; AQI**

## I. INTRODUCTION

One of the major public health and environmental concern is air pollution. According to the report of the World Health Organization (WHO), air pollution is a significant risk factor for multiple health conditions including skin and eye infections, irritation of the nose, throat and eyes [1]. It also causes serious conditions like heart disease, lung cancer, pneumonia, bronchitis, difficulty in breathing and coughing due to aggravated asthma. The World Health Organization concludes that 2.4 million people die each year from causes due to air pollution. Air pollution not only has worse effect on people's health but also on the environment and can lead to acid rain, smog, deterioration of the ozone layer and global warming. So, it becomes very essential to monitor and control the air pollution. The best way to control air pollution is to monitor exceeding levels of air pollutants and by taking appropriate actions to control it. Wireless Sensor Nodes (WSN) are used for monitoring of pollutant concentration in air around the city by installing WSNs in the moving public transport vehicles and cars [2]. The particles in air pollution data such as gases, smoke and other pollutants is collected using sensors and the data is stored in database for further analysis.

Data mining refers to the mining or discovery of new information based on patterns and rules from vast amounts of data [3][4][5]. Classification, clustering, and association are some of the data mining techniques used to analyze and extract meaningful information from complex data. Clustering is the process of organizing data objects into a set of disjoint classes called clusters. Cluster analysis is one of the primary data analysis tool in data mining. The *K*-means clustering algorithm is a partitioning clustering method that separates data into *K* groups. The *K*-means clustering algorithm is more prominent with its intelligence to cluster massive data rapidly and efficiently. The quality of the final clustering results of the *K*-means algorithm depends on the random selection of the initial centroids [6]. In the original *K*-means algorithm, different clusters are obtained for different runs for the same input data. In the enhanced *K*-Means clustering algorithm, the original data points are sorted based on distance between the centroid and data points. The sorted data points are partitioned into *K* equal sets. In each set the middle points are taken as the initial centroids [6] for that particular set.

Air Quality Index (AQI) is calculated for measuring the concentration of pollutants in the air. The increase in AQI increases severe adverse health effects to the large percentage of population. AQI values are grouped into ranges. Each range is assigned the parameters such as (*i*) a descriptor, (*ii*) a color code, and (*iii*) a standardized public health advisory [7]. The standard AQI vales are shown in Table 1. The rest of the paper is organized as follows: The related work is presented in Section 2. Section 3 explains the proposed enhanced *K*-Means clustering algorithm for air pollution data analysis. The experimental results obtained using the proposed algorithm is presented in Section 4. Section 5 concludes the paper and gives directions for future work.

## II. RELATED WORK

This section provides some of the existing works in the literature related to air quality analysis. Akula et al. have proposed [8] a dispersion model for analysis of air quality data near roadways. This model analysis the experiments conducted by the U.S. EPA in July–August 2006. The

978-1-5090-2597-8/16/$31.00 ©2016 IEEE

proposed dispersion model is used to find Nitrogen Oxide (NO) and other pollutant concentration from vehicles emission. The various meteorological parameters considered are wind speed, wind direction, temperature, and humidity. This paper has also proposed a simple model with normal wind direction and infinite length of highway. The sensitivity tests are also conducted for the proposed system to identify observed meteorological variables are affected by vehicle emission within 20m. Nitrogen Oxide concentration is also affected by (*i*) the emission rate, (*ii*) the traffic flow rate, and (*iii*) the standard deviation of the vertical velocity fluctuations. Nurul et al. have calculated [9] Air Quality Index (AQI) for five stations in Selangor, Malaysia. The Malaysian Department of Environment (DOE) [7] have listed major pollutants such as (*i*) sulphur dioxide, (*ii*) Nitrogen Dioxide, (*iii*) Carbon Monoxide, (*iv*) Particulate Matter with 10-micron (PM10) size in diameter and (*v*) Ground-Level Ozone. The research in [9] focused on the urban air quality and human health effects.

**TABLE 1. AIR QUALITY INDEX (AQI) [7]**

| Air Quality Index (AQI) Values | Levels of Health Concern | Colors | Meaning |
|---|---|---|---|
| 0 to 50 | Good | Green | Air quality is considered to be satisfactory; air pollution poses little or no risk. |
| 51 to 100 | Moderate | Yellow | Air quality is acceptable; however, for some pollutants there may be a moderate health concern for a very small number of people who are unusually sensitive to air pollution. |
| 101 to 150 | Unhealthy for Sensitive Groups | Orange | Members of sensitive groups may experience health effects. The general public is not likely to be affected. |
| 151 to 200 | Unhealthy | Red | Everyone may begin to experience health effects; Members of sensitive groups may experience more serious health effects. |
| 201 to 300 | Very Unhealthy | Purple | Health alert: everyone may experience serious health effects. |
| 301 to 500 | Hazardous | Maroon | Health warning of emergency conditions. The entire population is more likely to be affected. |

Muhammad et al. have proposed smart environment monitoring system [2] by installing Wireless Sensor Network (WSN) on vehicles for pollution free smart cities. The WSN nodes are deployed around the city and other public transport vehicles for monitoring the air pollution. The LTE-M based mesh network is used to collect the data from WSN. Since the zigbee wireless sensors are connected in public vehicles, the pollutants data are gathered from there using USB or Ethernet. The gathered data are stored in database for further processing. Doreswamy et al. have analyzed [11] the air pollution data using *K*-means clustering algorithm in smart cities. The *K*-means clustering algorithm is applied on the pollution dataset generated from The CityPulse project [12]. The dataset contains pollutants such as (*i*) Ozone ($O_3$), (*ii*) Carbon Monoxide (CO), (iii) Particular Matter (PM), (*iv*) Sulphur Dioxide ($SO_2$) (*v*) Nitrogen Dioxide($NO_2$), and meteorological variables such as (*i*) timestamp, (*ii*) longitude, and (*iii*) latitude. The proposed system used to determine the healthy and unhealthy locations in The CityPulse project [23] for smart environment in smart city.

Ojeda-Magana et al. have proposed PFCM [13] [14] [15] clustering algorithm to analyze air pollution in real database of Salamanca (Mexico). The data are collected from three monitoring stations includes Cruz Roja, Nativitas, and DIF, are analyzed. The combined measure and also correlation between pollution and environmental variables are calculated. The proposed system uses two groups such as synthetic cloud and pollutant concentration. The PFCM clustering algorithm is applied after identifying the groups. The correlation coefficient is calculated to determine the strength and direction of the relationship between pollution and environmental variables. The PFCM clustering algorithm is not compared with any of the existing algorithm. Yajie et al. have monitored and mined [16] the air pollution data based on sensor grid in London. For monitoring and mining the air pollution, a distributive infrastructure consists of wireless sensor network and Grid Computing technology is used. The technology aims to provide low-cost and real-time data collection from road traffic in urban environment. TinyOS is used to simulate the operation of sensor network. The simulation is visualized using OMNet++ [17]. The proposed system provides high performance based on high quality mobile sensors.

Dogruparmak et al. have compared Principal Component Analysis (PCA) and Fuzzy C–Means Clustering (FCM) [18] for analyzing air quality. Both the proposed systems are implemented using the data obtained from a test bed of 39 air quality monitoring stations established in 11 provinces of Marmara Region. The monitoring stations are distributed in urban, traffic, industrial, and rural areas. The measured pollutants are $PM_{10}$, $SO_2$, NO, $NO_2$, NOX and $O_3$. Both $SO_2$ and $PM_{10}$ are considered for analysis because these two pollutants are measured in all the stations. When comparing the FCM with PCA, the pollutant concentration places are well identified using FCM than PCA [18]. Ghaemi et al. have proposed a Hadoop based air pollution prediction system using Support Vector Machine, namely, LaSVM [19]. The extracted support vectors are used in LaSVM technique. Instead of using existing training data, LaSVM trained the online data to increase the speed of training. The memory usage is also reduced in the proposed system. LaSVM is implemented for predicting the air pollution in Tehran for the next 24 hours. In the proposed algorithm, LaSVM, the efficiency is evaluated using the parameters such as accuracy, RMSE and RSquered estimators. Wang et al. have introduced plume based analysis for air pollutant emissions from high emitters [20]. Air pollutant measurements are taken in real-world conditions at Toronto, Canada, during 2013–2014. The obtained measurements were processed and analyzed using Igor Pro 6.34.

## III. ENHANCED *K*-MEANS CLUSTERING ALGORITHM FOR AIR POLLUTION ANALYSIS

The *K*-Means clustering algorithm is an unsupervised clustering algorithm in which the centroids are initialized randomly [21]. The *K*-Means algorithm is very impressionable to the initial starting points. It is necessary for *K*-Means to refine initial cluster centers. Several methods have been proposed in the literature for finding the better initial centroids. The proposed system uses enhanced *K*-Means clustering algorithm to analyze air pollution to improve both the accuracy and efficiency of the *K*-Means clustering algorithm. In enhanced *K*-Means clustering algorithm, the data points are sorted and divided into *K* equal partitions and the middle point of each partition is taken as the initial centroids [6].

The measurement of air pollutants such as Sulphur Dioxide ($SO_2$), Particular Matter ($PM_{10}$), Ozone ($O_3$), CO, NOx and $SO_2$ as well as environmental variables such as wind speed and wind direction. It cannot be easy to analyze the concentration level of gases separately for every minute, so it is analyzed in a combined measure using clustering analysis algorithm. Air Quality Index (AQI) is a number used by government agencies to communicate to the public how polluted the air currently is or how polluted it is going to be in near future [7]. Different countries have their own air quality indices, corresponding to different national air quality standards. This AQI is having six categories indicating the levels of health concern. An AQI value over 300 means hazardous air quality and below 50 means good air quality. [7].

### *A. Steps for enhanced K-Means Clustering Algorithm for finding AQI of pollutant data*

1. Pollutant concentration is measured by setting different sensors to collect air samples and measures the concentration on $SO_2$, NOx, $PM_{10}$, $O_3$ and CO. These sensors measure concentration, i.e. unitless proportions (e.g. parts per million) or mass per volume (e.g. micrograms per cubic meter). These measurements are stored in a dataset.
2. Then, the given dataset is checked for negative value attributes. If it contains negative value then transform it into positive value by subtracting each data point attribute with the minimum attribute value.
3. Calculate the distance from origin. Then, the data points are sorted based on with sorted distance. The sorted data points are partitioned into *K* equal sets. In each set, take the middle point as the initial centroid.
4. Then, calculate the distance between each data point to all initial centroids and assign data points to the cluster having closest centroids.
5. If this distance is less than or equal to the present nearest distance, the data point stays in the same cluster. Otherwise for each data point, the distance from all the centroids is calculated.
6. Step 4 and 5 are repeated until the convergence criterion is met.
7. The Correlation Coefficient *r* is calculated based on (1) [13]. The value of the correlation coefficient is ranged from +1 to -1.

$$r = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2}\sqrt{n(\sum y_i^2) - (\sum y_i)^2}} \quad (1)$$

8. The AQI scale is ranged from 0 to 500. The aim is to convert the pollution concentration into a number between 0 and 500. The AQIs of 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 are referred to as "breakpoints."
9. The pollution concentration between the breakpoints is linearly interpolated using (2) [22].

$$I_p = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}}(C_P - BP_{LO}) + I_{LO} \quad (2)$$

Where, $I_p$ is the index of the pollutant; $C_p$ is the rounded concentration of pollutant *p*; $BP_{HI}$ is the breakpoint greater or equal to *Cp*; $BP_{LO}$ is the breakpoint less than or equal to *Cp*; $I_{HI}$ is the AQI corresponding to $BP_{HI}$; $I_{LO}$ is the AQI corresponding to $BP_{LOW}$;

10. The AQI is calculated by the pollutant with the highest index. Each AQI breakpoint [23] corresponds to a defined pollution concentration and is given in Table 2[23] [24].

TABLE 2. BREAKPOINTS OF AQI [20]

| Breakpoints | | | | | | | AQI | Category |
|---|---|---|---|---|---|---|---|---|
| $O_3$ (ppm) 8-hour | $O_3$ (ppm) 1-hour | $PM_{25}$ ($\mu g/m^3$) | $PM_1$ ($\mu g/m^3$) | CO (ppm) | $SO_2$ (ppm) | $NO_2$ (ppm) | | |
| 0.000-0.064 | - | 0.0-15.4 | 0-54 | 0.0-4.4 | 0.000-0.034 | - | 0-50 | Good |
| 0.065-0.084 | - | 15.5-40.4 | 55-154 | 4.5-9.4 | 0.035-0.144 | - | 51-100 | Moderate |
| 0.085-0.104 | 0.125-0.164 | 40.5-65.4 | 155-254 | 9.5-12.4 | 0.145-0.224 | - | 101-150 | Unhealthy for Sensitive Groups |
| 0.105-0.124 | 0.165-0.204 | 65.5-150.4 | 255-354 | 12.5-15.4 | 0.225-0.304 | - | 151-200 | Unhealthy |
| 0.125-0.374 | 0.205-0.404 | 150.5-250.4 | 355-424 | 15.5-30.4 | 0.305-0.604 | 0.65-1.24 | 201-300 | Very Unhealthy |
| - | 0.405-0.504 | 250.5-350.4 | 425-504 | 30.5-40.4 | 0.605-0.804 | 1.25-1.64 | 301-400 | Hazardous |
| - | 0.505-0.604 | 350.5-500.4 | 505-604 | 40.5-50.4 | 0.805-1.004 | 1.65-2.04 | 401-500 | Hazardous |

## IV. EXPERIMENTAL RESULTS

The experimental results are discussed in this section. The real time pollutant measurements used for investigation are

shown in Table 3. The proposed enhanced *K*-Means clustering algorithm is compared with PFCM algorithm [13] in terms of accuracy and execution time. After clustering operation, the correlation coefficient is calculated using PFCM and enhanced *K*-Means clustering algorithms. Table 4 presents the calculated correlation coefficient for enhanced *K*-Means clustering algorithm. The correlation coefficient is calculated between the air pollutants and Wind Speed (WS) and Wind Direction (WD). The corresponding AQI for the correlation coefficient is shown in Table 5.

The accuracy comparison between PFCM and enhanced *K*-Means algorithm is shown in Figure 1. The accuracy of the proposed algorithm enhanced *K*-Means clustering is calculated using (3).

$$Accuracy = \frac{Experimetalvalue - Theoreticalvalue}{Experimentalvalue} \quad (3)$$

TABLE 3. SOURCES OF DATASETS

| S. No | Dataset Name | Size of Dataset | Dataset Collection Date | Link and Place |
|---|---|---|---|---|
| 1. | OpenAir_example_data_long | 1109KB | 01-01-2008 | http://www.openair-project.org Los Angles |
| 2. | PollutionData 206025 | 1247KB | 01-08-2014 | http://iot.ee.surrey.ac.uk:8080/datasets/pollution/index.html Arhus City |
| 3. | PollutionData 209907 | 1303KB | 01-09-2012 | |
| 4. | PollutionData 204273 | 1.2MB | 01-10-2014 | |
| 5. | PollutionData 206422 | 1.5MB | 01-09-2013 | |

TABLE 4. CORRELATION COEFFICIENT FOR ENHANCED *K*-MEANS CLUSTERING ALGORITHM

| | Dataset 1 | | Dataset 2 | | Dataset 3 | | Dataset 4 | | Dataset 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | WS | WD | WS | WD | WS | WD | WS | WD | WS | WD |
| NOx | 0.14 | 0.68 | -0.13 | 0.94 | 0.09 | 0.12 | 0.13 | 0.29 | 0.09 | 0.12 |
| $NO_2$ | -0.09 | 0.02 | 0.24 | -0.21 | -0.03 | 0.58 | -0.19 | 0.9 | -0.03 | 0.58 |
| $O_3$ | 0.29 | 0.59 | 0.14 | -.042 | 0.09 | 0.59 | 0.23 | -0.22 | 0.09 | 0.59 |
| $PM_{10}$ | 0.48 | 0.15 | 0.47 | 0.07 | 0.29 | 0.13 | 0.12 | 0.2 | 0.29 | 0.13 |
| $SO_2$ | -0.09 | 0.33 | 0.92 | 0.9 | 0.52 | 0.15 | 0.8 | 0.12 | 0.52 | 0.15 |
| CO | 0.24 | 0.58 | 0.94 | 0.63 | 0.03 | 0.20 | -0.12 | 0.67 | 0.03 | 0.20 |

TABLE 5. AQI FOR DIFFERENT DATA SETS

| S. No | Dataset Name | AQI Value | Levels of Health Concern |
|---|---|---|---|
| 1. | OpenAir_example_data_long | 370.21 | Hazardous |
| 2. | pollutionData206025 | 347 | Hazardous |
| 3. | pollutionData209907 | 353 | Hazardous |
| 4. | pollutionData204273 | 74 | Moderate |
| 5. | pollutionData206422 | 164 | Unhealthy |

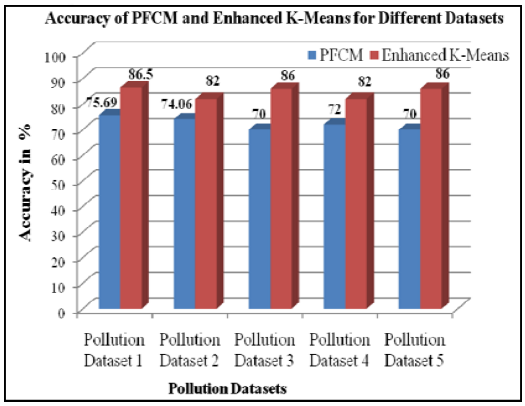


Fig 1. Accuracy Comparison of PFCM and Enhanced *K*-Means for Different Datasets

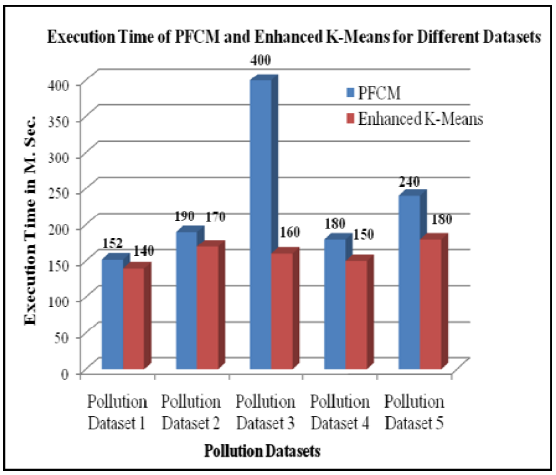


Fig 2. Execution Time of PFCM and Enhanced *K*-Means for Different Datasets

The execution time of PFCM and enhanced *K*-Means clustering algorithms for different datasets is calculated and plotted in Fig 2. From the figure it is clear that the execution time of enhanced *K*-Means clustering algorithm is less than when compared to PFCM for different datasets. It is evident that the proposed enhanced *K*-Means clustering algorithm provides AQI value in better accuracy but less execution time for different data sets.

## V. CONCLUSIONS

In this paper both PFCM and enhanced *K*-Means algorithms are implemented for different Datasets for finding Air Quality Index. Real time datasets are taken from different places. It is evident that the enhanced *K*-Means clustering algorithm gives AQI value in higher accuracy but less execution time when compared to PFCM Clustering Algorithm. The proposed enhanced *K*-means Clustering algorithm gives 40% more efficiency in terms of Accuracy and Execution time than PFCM Algorithm. A distributed

version of the *K*-means Clustering algorithm can be implemented where data or computational power is distributed. Efficiency can also be improved by using variable clusters instead of constant '*K'* number of clusters.

## *Acknowledgment*

The authors would like to thank the Management, Principal of Sri Ramakrishna Engineering College and Head of the Department of Computer Science and Engineering for their support in this work.

## *References*

[1] http://www.who.int/mediacentre/factsheets/fs313/en/

[2] Muhammad Saqib Jamil, Muhammad Atif Jamil, Anam Mazhar, Ahsan Ikram, Abdullah Ahmed, Usman Munawar, Smart Environment Monitoring System by employing Wireless Sensor Networks on Vehicles For Pollution Free Smart Cities, Humanitarian Technology: Science, Systems and Global Impact 2015, HumTech2015, Procedia Engineering, Vol.107, pp. 480 – 484, 2015.

[3] Margaret H. Dunham, Data Mining- Introductory and Advanced Concepts, Pearson Education, 2006.

[4] Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Pearson Addison Wesley, Boston, 2006.

[5] Berkhin, P.: A Survey of clustering data mining techniques. Technical Report, Accrue Software, San Jose, CA 2002.

[6] Ramzi A. Haraty, Mohamad Dimishkieh, MehediMasud, An Enhanced k-Means Clustering Algorithm for Pattern Discovery in Healthcare Data, International Journal of Distributed Sensor Networks Volume, Hindawi Publishing Corporation, pp.1-11, 2015.

[7] https://en.wikipedia.org/wiki/Air_quality_index

[8] Akula Venkatram, Vlad Isakov, Eben Thoma, Richard Baldauf," Analysis of air quality data near roadways using a dispersion model", Atmospheric Environment, Vol. 41 pp. 9481–9497, 2007.

[9] Nurul Ashikin Mabahwi, Oliver Ling Hoon Leh, Dasimah Omar, "Urban Air Quality and Human Health Effects in Selangor, Malaysia, Procedia - Social and Behavioral Sciences , Vol. 170, pp. 282 – 291, 2015.

[10] Department of Environment (2012, November 13). Department Of Environment, Air pollutant index. Department of Environment Malaysia. Retrieved November 14, 2012, from http://www.doe.gov.my/apims/index.php

[11] Doreswamy, Osama A.Ghoneim, B R Manjaunath, "Air Pollution Clustering Using K Means Algorithm in Smart City", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Special Issue. 7, pp. 51-57, 2015.

[12] "CityPulse project" and the URL is http://iot.ee.surrey.ac.uk:8080/ datasets.html.

[13] B. Ojeda-Magaña1, M. G. Cortina-Januchs, J. M. Barron-Adame, J. Quintanilla-Domínguez, W. Hernandez , A. Vega-Corona, R. Ruelas1 and D. Andina, "Air pollution Analysis with a PFCM Clustering Algorithm Applied in a Real Database of Salamanca (Mexico)", Procedia Engineering, Vol. 15, pp.4147-4151, 2010.

[14] Aruna Bhat, "Possibility Fuzzy C-Means Clustering For Expression Invariant Face Recognization", International Journal on Cybernetics & Indormatics (IJCI), Vol. 3, No. 2, pp. 35-45, 2014.

[15] N.R. Pal, S.K. Pal, J.M. Keller and J.C. Bezdek. "A possibilitic fuzzy c-means clustering algorithm". IEEE Transactions on Fuzzy Systems, Vol. 13, No. 4, pp. 517–530, (2005).

[16] Yajie Ma, Mark Richards, Moustafa Ghanem, Yike Guo and John Hassard, "Air Pollution Monitoring and Mining Based on Sensor Grid in London", Vol. 8, pp. 3601-3623, 2008.

[17] OMNet++ Homepage. http://www.omnetpp.org/.

[18] S.C. Dogruparmak, G. A. Keskin, S. Yaman, A. Alkan, "Using principal component analysis and fuzzy c–means clustering for the assessment of air quality monitoring, Atmospheric Pollution Research, Vol. 5, pp. 656 – 663, 2014.

[19] Z. Ghaemi, M. Farnaghi, A. Alimohammadi, Hadoop-Based Distributed System for Online Prediction of Air Pollution Based on Support Vector Machine, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-1/W5, pp. 215-219, 2015.

[20] J. M. Wang, C.H. Jeong, N. Zimmerman, R. M. Healy, D. K. Wang, F. Ke, and G. J. Evans, Plume-based analysis of vehicle fleet air pollutant emissions and the contribution from high emitters, Atmospheric Measurement Techniques, Vol. 8, pp. 3263–3275, 2015.

[21] Abdolmajid Dejamkhooy, Ali Dastfan, and Alireza Ahmadifard, "K-Means Clustering and Correlation Coefficient Based Methods for Detection of Flicker Sources in NonRadial Power System", ISSN 1068-3712, Russian Electrical Engineering, Vol. 85 No. 4, pp. 251–259, 2014.

[22] https://stimulatedemissions.wordpress.com/2013/04/10/how-is-the-air-quality-index-aqi-calculated/

[23] David Mintz, Guidelines for the Reporting of Daily Air Quality – the Air Quality Index (AQI), U.S. Environmental Protection Agency Research Triangle Park, North Carolina. URL: https://www3.epa.gov/ttn/caaa/t1/ memoranda/rg701.pdf

[24] http://www.ciese.org/curriculum/bus/docs/Breakpoints.pdf