EnvironmentAsia
The international journal by the Thai Society of Higher Education Institutes on Environment

EnvironmentAsia 11(2) (2018) 9-22
DOI 10.14456/ea.2018.19
ISSN 1906-1714; ONLINE ISSN: 2586-8861

Spatial Distribution and Source Apportionment of Air Pollution in Bahrain using Multivariate Analysis Methods

Majeed S. Jassim[a,*], Gulnur Coskuner[b], Hussain Marzooq[a], Ahmed AlAsfoor[a], Ahmed Abu Taki[a]

[a] *Department of Chemical Engineering, College of Engineering, University of Bahrain, Isa Town, P.O. Box 32038, Kingdom of Bahrain.*

[b] *Civil and Environmental Engineering Program in Bahrain, West Virginia University, hosted by Royal University for Women, Riffa, P.O. Box 37400, Kingdom of Bahrain*

* Corresponding author: drmajeed@gmail.com
Received: November 29, 2017;Accepted: February 26, 2018

Abstract

The objective of this study is to identify the most important air pollutants based on their individual contribution to Air Quality Index (AQI) and to determine the major air pollution sources in Bahrain. Data sets from seventeen air quality monitoring sites were evaluated using XLSTAT 2014 and Statistical Package for the Social Sciences (SPSS 22) over six-and-half-year between July 2006 and December 2012. Hierarchical Agglomerative Cluster Analysis (HACA) categorized the monitoring sites into three distinctive clusters based on similarities of air pollutants characteristics and meteorological parameters. Principal Component Analysis (PCA) identified major sources of air pollution in each cluster. Results demonstrated that dust storms, industrial activities, vehicular emissions, airport activities, power plants and filling stations were major air polluters. PCA analysis showed that temperature and wind speed have positive loading while relative humidity has negative loading. Multiple Linear Regression (MLR) analysis was applied to develop models for prediction of AQI for every cluster based on concentrations of key air pollutants. Results showed PM_{10} and $PM_{2.5}$ highly contributed to AQI values. MLR models exhibited good fit with adjusted R^2 value of 0.865, 0.794 and 0.842 for Clusters 1, 2 and 3 respectively. Standardized coefficient values for PM_{10} succeeded by $PM_{2.5}$ were the highest in each cluster.

Keywords: Air Quality Index; HACA; PCA; Multiple Linear Regression; Particulate matter.

1. Introduction

Air pollution provides insidious challenge to environment and public health. The long-term exposure to ambient air pollutants is associated with increased morbidity and mortality (Cohen *et al.,* 2017).

Monitoring and modeling of ambient air quality are necessary steps to evaluate source apportionment subsequently implementing air quality management programs (WHO, 1999). The status of ambient air quality is reported to the public using indices that explain the associated health risks in a simple manner. The USEPA defined AQI in terms of six key air pollutants: CO, NO_2, O_3, PM_{10}, $PM_{2.5}$, and SO_2 (EPA, 2009).

Bahrain is an archipelago of thirty-three natural islands and shoals located in West Asia. The kingdom is the most densely populated country in the region with a population density of 1,614 people per square kilometer (Bahrain Census, 2010). Its geographical location adjacent to arid and semi-arid areas of the Arabian Peninsula increased its exposure to dust storms (Tsiouri *et al.,* 2015). However, there is limited published air quality studies in Bahrain and that could be attributed to weaknesses with respect to air pollution subjects covered as part of environmental engineering curriculum in higher education institutions (Jassim and Coskuner, 2007). A recent analysis of ambient air quality showed an increase of particulate matter concentrations (Jassim *et al.,* 2018).

World Health Organization global ambient air pollution database ranked Bahrain in the top ten most urban ambient air polluted countries in the world based on high particulate matter concentrations (WHO, 2016). Similarly, an assessment of ambient air quality identified PM_{10} and $PM_{2.5}$ to be the most critical air pollutants with potential health implications based on calculated AQI values in Bahrain (Jassim and Coskuner, 2017).

Statistical analyses are utilized to complement large dataset generated by network of monitoring stations. HACA, PCA, and MLR are widely applied multivariate analysis methods in environmental issues including air quality studies (Dominick et al. 2012; Özbay *et al.,* 2011; Pavón-Domínguez *et al.,* 2014).

The purpose of this study is to identify spatial patterns of air pollution within study area by grouping air quality monitoring sites into similarly characterized clusters. Dataset on air pollutants concentrations and values of meteorological parameters between 2006 and 2012 were utilized. PCA was applied to identify the main sources of air pollution. MLR models were further developed to estimate contribution of key air pollutants to AQI values.

2. Materials and Methods

2.1 Air quality monitoring stations

The Supreme Council for Environment monitored seventeen sites utilizing five mobile ambient air quality stations between July 2006 and December 2012. The spatial arrangements of these sites are presented in Table 1.

2.2. Meteorological parameters

The meteorological measurement instruments are located in Bahrain International Airport in Muharraq Island and the meteorological data was received from Meteorological Directorate.

2.3. Pre-processing data

Multivariate analysis methods were applied on 5070 data from July 2006 to December 2012. These included monthly mean concentrations of ten pollutants (NH_3, NO_2, H_2S, SO_2, CO, O_3, PM_{10}, $PM_{2.5}$, C_6H_6, TNMHC) at five air quality monitoring stations and three meteorological parameters (ambient temperature, relative humidity and wind speed).

Approximately 8% of the data was not available from the complete data matrix. This percentage of missing data in a time-series multivariate application is acceptable because the imputation methodologies yield good results when the proportion of missing dataset is less than 10% (Junger and de Leon, 2015).

The multivariate analysis methods (HACA, PCA and MLR) require complete dataset (Kaiser, 2014). There are various imputation-based methodologies developed for treating missing data. Expectation-Maximization algorithm was utilized due to its simple implemantation using SPSS 22 (Donders *et al.*, 2006). This algoritm is advantageous because it produces valid and unbiased estimates.

2.4. Hierarchical Agglomerative Cluster Analysis (HACA)

This methodology hierarchically groups the air quality monitoring stations into clusters. The most similar monitoring stations are first grouped and these initial groups are merged according to their similarities. As the similarity decreases, all subgroups are combined into a single cluster (Johnson and Wichern, 2007). The classification of the clusters is illustrated in the form of a two-dimensional diagram known as a dendrogram (tree diagram), which shows similarity levels and illustrates the mergers that are made at successive levels. The vertical axis of the dendrogram represents the distance or dissimilarity between clusters and the horizontal axis represents the objects and clusters.

2.5. Principal Component Analysis (PCA)

PCA is an exploratory statistical method that is utilized for the identification of major sources of air pollution because it transforms a set of interrelated variables into a set of uncorrelated variables (Abdul-Wahab et al. 2005; Pires *et al.*, 2008). It highlights the meaningful variables that explains variance in data and excludes the less significant variables. The numbers of transformed variables or Principal Components (PCs) are equal to the number of independent variables. The first PC represents the largest proportion of variability of data and the second PC has the largest proportion of variability that has not been presented in first component (Jolliffe, 2002). The PCs generated by PCA are not readily available for interpretation therefore it is necessary to rotate them using orthogonal rotation method (varimax). The varimax rotation obtains new groups of variables called Varimax Factors (VFs) and this feature assists in identifying different possible sources of air pollution. The selection of PCs depends on Kaiser Criterion for eigenvalues and it defines the statistically significant PCs as those with eigenvalues greater than or equal to one (Kaiser, 1960).

Table 1. Air quality monitoring sites in Bahrain.

The color codes correspond to different clusters (Cluster 1 is pink, Cluster 2 is brown, Cluster 3 is green)

Monitoring Site	Station ID
Hamala	S1
Hamad Town	S2
Bahrain Fort	S3
Riffa	S4
Jaww	S5
Al-Areen	S6
Ras Hayan	S7
Manama	S8
Nabeeh Saleh	S9
Tubli	S10
Hidd	S11
Samaheej	S12
Arad	S13
Al-Busateen	S14
Maameer	S15
Sitra	S16
Salmabad	S17

Factor loadings indicate quantitatively the contribution of a variable to a particular PC and the similarity extent between variables. If factor loading is greater than 0.75 then it is considered "strong loading". If factor loading is between 0.75 to 0.50 then it is considered "moderate loading" and if it is between 0.49 to 0.30 then it is considered "weak loading" (Jolliffe, 2002).

2.6. *Multiple Linear Regression (MLR)*

MLR is a statistical method that allows prediction of variability between a dependent variable and independent variables (Jobson, 1991). This method is widely applied in atmospheric modelling for investigating statistical relationship between a dependent variable and several independent variables by fitting a linear equation to actual data and provides contribution percentage of each parameter to atmospheric pollution (Pai *et al.*, 2009). The performance indicators were utilized to evaluate goodness of fit for developed MLR. These indicators are coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), and Root Mean Square Error (RMSE).

3. Results and discussion

3.1 *HACA results*

This algorithm was utilized to study spatial variation of seventeen air quality-monitoring sites. It was performed on monthly mean values for ten air pollutants concentrations and three meteorological parameters as described earlier. Figure 1 shows a dendrogram grouping the air quality monitoring sites across Bahrain into three distinguishable clusters.

Cluster 1 includes Riffa (S4), Maameer (S15), and Salmabad (S17). These stations are exposed to air pollutants from various industrial activities. Maameer village is located in proximity of intensive industrial area including petroleum refinery, petrochemical plants, asphalts, ready-mix concrete, and other aluminum workshops. Salmabad is a small-scale industrial area for car garages, warehouses, and aluminum factories. Riffa is in proximity of power plant, petroleum refinery, and aluminum smelter.

Cluster 2 accommodates Hamad Town (S2), Bahrain Fort (S3), Jaw (S5), Al-Areen (S6), Ras Hayan (S7), Tubli (S10), Arad (S13) and Sitra (S16).

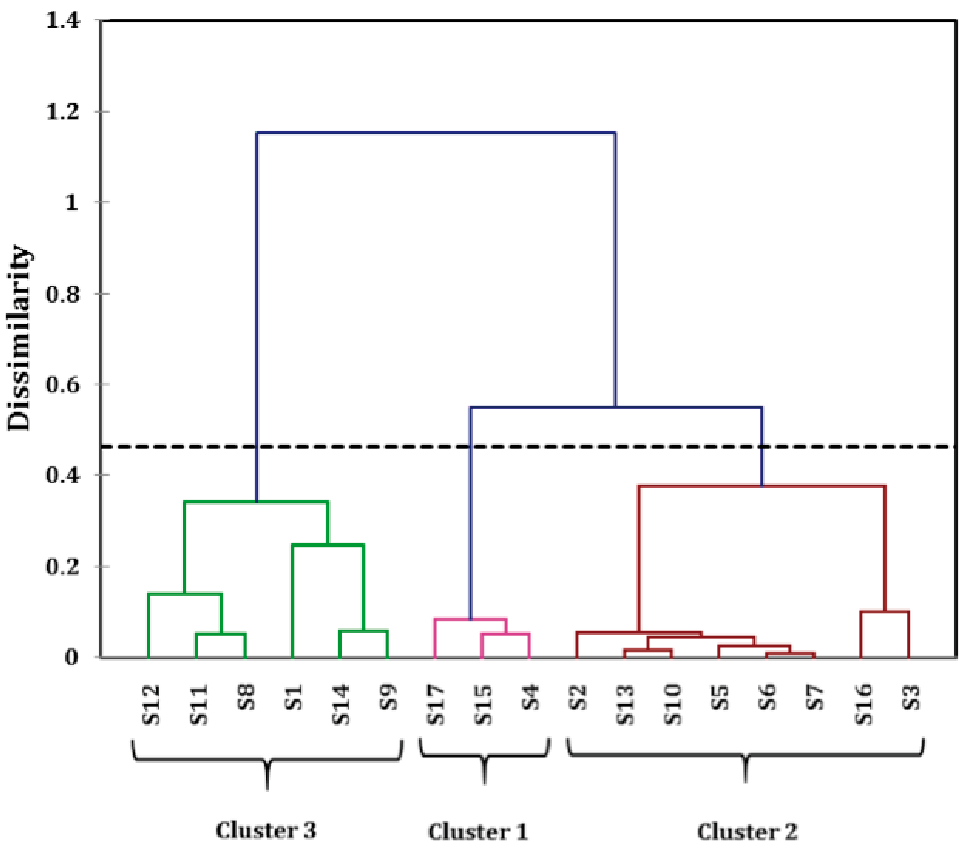


Figure 1. A dendrogram showing three clusters of air quality monitoring sites

These stations are mainly exposed to pollutants from congested highways. Hamad Town is situated adjacent to Shaikh Khalifa bin Salman highway. Tubli is situated near to Shaikh Isa bin Salman highway and close proximity to a major sewage treatment plant. Sitra is situated near to Shaikh Jaber Al Subah highway. Arad is near to Khalifa Al Kabir Highway and near to airport. Ras Hayan is near to Hawar highway and Jaw is near to King Hamad highway.

Hamala (S1), Manama (S8), Nabih Saleh (S9), Hidd (S11), Samaheej (S12) and Busateen (S14) are included in Cluster 3. These stations are very close to seashores with Hidd, Samaheej and Busateen located in Muharraq Island.

3.2 PCA Results

PCA was performed on concentrations of ten air pollutants and the three meteorological parameters. Analysis was implemented to reduce the number of parameters and to identify the major sources of variations in each cluster. Results of PCA loadings after varimax rotation are presented in Table 2. Only strong factor loadings (higher than 0.75) are considered as a source of variation. Table 2 also presents eigen-values, variability (%) and cumulative variability (%) for each cluster.

a) Cluster 1

PCA revealed six factors that explain 78.33% of total variance in Cluster 1. First and second factors (F1 and F2) account for 23.46% and 13.72% of total variance respectively.

F1 has strong positive loading on PM_{10} (0.87) and $PM_{2.5}$ (0.88) but strong negative loading on humidity (-0.88). F2 has strong positive loading on wind speed (0.89). These reflect the contribution of Aeolian processes in which sandstorms across the Arabian Peninsula transport coarse particulate matter to Bahrain.

F3 contributes 12.83% of total variance and has strong positive loading on NH_3 (0.82) and strong negative loading on SO_2 (-0.76). Ammonia emissions are from animal manure and production of nitrogen containing fertilizers due to proximity of Maameer (S15) to main slaughterhouse, feedlot of main livestock company and petrochemical industries (Sutton *et al.,* 2000).

Major source of SO_2 from anthropogenic sources are combustions of fuel that contain traces of sulfur in oil refinery at Maameer (S15) and the gas-fired electric power generation at Riffa (S4) (Mukhopadhyay and Forssell, 2005). These industries are responsible for the heterogeneous discharge of SO_2, NH_3 and $PM_{2.5}$ (Jaramillo and Muller, 2016). It is believed that $PM_{2.5}$ is originating from chimneys of industrial facilities in proximity to air monitoring sites within Cluster 1.

F4 explains 11.58% of total variance which has strong negative loading on O3 (-0.81) but strong positive loading on C_6H_6 (0.78). Ozone is a secondary pollutant that is produced from the reaction of NOx and TNMHC under the action of sunlight. Sources of these anthropogenic precursors are industrial and vehicular emissions (Abdul-Wahab et al. 2005). Main sources of C6H6 worldwide are gasoline and diesel vehicles (Karakitsios *et al.,* 2006). Traffic in Cluster 1 is congested because it is highly residential.

F5 explains 8.65% of total variance that has strong positive loading on H_2S (0.75). Hydrogen sulfide is an industrial pollutant associated with natural gas processing and refining operations

(Lu and Schaefer, 2004). F6 contributes 8.08% of total variance and has strong positive loading on CO (0.90). Carbon monoxide is released from anthropogenic activities like incomplete combustions of fuel to produce electric power and vehicle emissions (Levy, 2015).

b) Cluster 2

Cluster 2 has five factors that explain 77.24% of total variance. F1, F2, F3, F4, and F5 explain 32.29%, 13.58%, 13.58%, 9.74%, and 8.05% of total variance respectively. F1 has strong positive loading on ambient temperature (0.84) but strong negative loading on humidity (-0.84). Growth of real estate sector and higher groundwater salinity resulted in the decline of agricultural activities and disappearance of green areas. These contributed to higher atmospheric air temperature with increase of air-conditioned high-rise buildings, sharp population growth and intensive energy consumption from fossil energy resources.

F2 has strong positive loading on NO_2 (0.77). A major source of NO_2 is fuel combustion from motorized traffic and industry (Ghazali *et al.,* 2009). Monitoring sites in this cluster are in proximity of congested residential areas and highways.

F4 has strong loading on PM_{10} (0.75) and F5 has strong positive loading on wind speed (0.86). These parameters were previously discussed in Cluster 1.

c) Cluster 3

Cluster 3 has six factors that explain 70.83% of total variance. F1, F2, F3, F4, F5 and F6 explain 20.45%, 14.03%, 11.23%, 9.0%, 8.30, and 7.82% of the total variance respectively. F1 has strong positive loading on PM_{10} (0.87) and F2 has strong positive loading on wind speed (0.88). Presence of PM_{10} and wind speed in this cluster confirm the importance of PM_{10} over the islands and the transportation of coarse particulate matter from nearby arid region.

F3 has strong positive loading on NO_2 (0.76) but strong negative loading on TNMHC (-0.75). Both are primary anthropogenic precursors in the presence of sun light to form ozone. Sources of these two pollutants are industrial and vehicular emissions (Abdul-Wahab *et al.,* 2005). F5 has a strong positive loading on NH_3 (0.88). Ammonia could be produced from human and animal waste. There is a large-scale poultry farm in the area of Hamala (S1). Animal traders are having barns near to Samaheej (S12).

F6 has a strong positive loading on H_2S (0.77) and on C_6H_6 (0.76). H_2S is associated with processes that utilize natural gas (Lu and Schaefer 2004). Major sources of C_6H_6 are filling stations, exhaust from motor vehicles and in dustrial emissions. Cluster 3 includes some of the most densely populated areas in Bahrain. S8 is part of Capital Governorate while S11, S12, S13 and S16 are part of Muharraq Governorate. The population density of Capital is 8671 people/km^2 and it is 3377 people/km^2 in Muharraq (Bahrain Census, 2010). Most of filling stations (Al Fateh, Al Mahooz, Saar) and automobile service stations (Salmabad, Toyota Plaza, Jawad Service Center) are located within Cluster 3 due to high population densities and heavy traffic.

Table 2. The factor loadings after PCA varimax rotation

Variables	Cluster 1						Cluster 2					Cluster 3					
	VF1	**VF2**	**VF3**	**VF4**	**VF5**	**VF6**	**VF1**	**VF2**	**VF3**	**VF4**	**VF5**	**VF1**	**VF2**	**VF3**	**VF4**	**VF5**	**VF6**
NH_3	-0.12	0.11	**0.82**	-0.36	-0.03	0.07	0.16	0.70	-0.25	0.09	0.21	-0.04	-0.13	-0.09	0.00	**0.88**	-0.01
NO_2	0.33	0.03	0.48	0.30	0.39	0.40	-0.09	**0.77**	-0.13	0.03	0.05	0.03	0.07	**0.76**	-0.15	0.38	-0.01
H_2S	0.01	-0.36	0.04	0.10	**0.75**	0.03	0.03	0.14	-0.54	-0.27	-0.33	0.16	-0.05	0.05	0.00	0.27	**0.77**
SO_2	0.20	0.19	**-0.76**	-0.19	0.05	0.35	0.03	-0.02	0.73	-0.14	0.00	0.02	-0.13	0.14	0.83	0.06	0.01
CO	-0.10	-0.04	0.10	0.08	-0.10	**0.90**	-0.31	0.36	0.41	0.41	-0.36	0.13	0.08	0.33	0.23	-0.13	-0.06
O_3	0.01	-0.02	0.28	**-0.81**	0.07	0.04	0.28	-0.17	0.59	-0.02	-0.12	0.02	-0.35	0.16	-0.73	0.08	-0.09
PM_{10}	**0.87**	0.22	-0.04	0.11	-0.01	-0.02	0.25	0.11	0.02	**0.75**	0.11	**0.87**	0.16	0.12	0.05	0.09	-0.03
$PM_{2.5}$	**0.88**	0.03	0.04	-0.01	0.03	-0.04	0.47	0.10	0.03	0.39	0.33	0.74	0.10	-0.03	-0.03	0.01	0.11
TNMHC	-0.01	0.38	0.08	0.09	-0.70	0.17	-0.28	0.46	-0.19	0.15	0.18	0.17	0.07	**-0.75**	-0.04	0.28	0.06
C_6H_6	0.08	-0.04	0.16	**0.78**	0.09	0.14	-0.20	0.30	0.25	-0.55	0.22	0.01	0.04	-0.17	0.07	-0.29	**0.76**
Temperature	0.61	-0.71	-0.09	-0.02	0.03	-0.09	**0.84**	-0.03	0.03	0.22	-0.33	0.64	-0.70	-0.05	-0.11	-0.02	0.03
Wind Speed	0.19	**0.89**	-0.06	0.02	-0.02	-0.06	-0.02	0.11	-0.01	0.00	**0.86**	0.22	**0.88**	0.05	0.05	-0.15	-0.01
Humidity	**-0.88**	0.16	0.22	0.03	-0.03	-0.02	**-0.84**	-0.13	-0.08	-0.17	-0.16	-0.86	0.14	0.10	-0.07	0.13	-0.10
Eigenvalue	3.05	1.78	1.67	1.51	1.13	1.05	1.77	1.77	1.77	1.27	1.05	2.66	1.82	1.46	1.17	1.08	1.02
Variability (%)	23.46	13.72	12.83	11.58	8.65	8.08	32.29	13.58	13.58	9.74	8.05	20.45	14.03	11.23	9.00	8.30	7.82
Cumulative (%)	23.46	37.18	50.02	61.60	70.25	**78.33**	32.29	45.87	59.45	69.19	**77.24**	20.45	34.48	45.71	54.71	63.01	**70.83**

3.3 MLR Results

MLR is applied to predict the AQI for each cluster. The AQI is regressed against concentrations of five key air pollutants: SO_2, CO, O_3, PM_{10} and $PM_{2.5.}$ However, NO_2 was not included in the analysis due to its low measured concentrations.

Regression coefficients explain the influence of each air pollutant parameter and ascertain its contribution. Developed AQI correlations along with regression coefficient of determination (R^2), adjusted regression coefficient of determination (R^2_{adj}) and Root Mean Square Error (RMSE) are shown in Equations 3, 4, and 5 respectively.

Cluster 1

$$\begin{aligned} &\text{AQI} \\ &= 22.06-0.08(SO_2) \\ &-8.70(CO)-0.18(O_3) \\ &+0.53(PM_{10})+0.55(PM_{2.5}) \end{aligned} \tag{3}$$

$R^2 = 0.873$

$R^2_{adj} = 0.865$

RMSE = 26.712

Cluster 2

$$\begin{aligned} &\text{AQI} \\ &= 38.00+1.31(SO_2) \\ &-1.28(CO)-0.46(O_3)+0.46(PM_{10}) \\ &+0.53(PM_{2.5}) \end{aligned} \tag{4}$$

$R^2 = 0.801$

$R^2_{adj} = 0.794$

RMSE = 62.216

Cluster 3

$$\begin{aligned} &\text{AQI} \\ &= -10.42+0.43(SO_2)+4.35(CO) \\ &+0.83(O_3)+0.35(PM_{10}) \\ &+0.91(PM_{2.5}) \end{aligned} \tag{5}$$

$R^2 = 0.847$

$R^2_{adj} = 0.842$

RMSE = 28.238

Performance indicators show that better model prediction is realized with close to unity for R^2 and R^2_{adj} and with lower RMSE value. For the three clusters, *p*-value of developed correlations was less than 0.05 indicating it is statistically significant.

Equation 3 shows that Cluster 1 has the highest R^2 with 0.873, R^2_{adj} with 0.865, and the lowest RMSE with 26.712. The five air pollutants in Cluster 1 contribute 86.5% to AQI. Average concentrations of SO_2, CO, and O_3 have negative influence on the AQI while average concentrations of PM_{10} and $PM_{2.5}$ have positive influence on AQI. Cluster 3 has the second highest R^2 with 0.847 and R^2_{adj} with 0.842.

Total contribution of the five air pollutants to AQI is 84.2%. The pollutants SO_2, CO, O_3, PM_{10} and $PM_{2.5}$ have positive influence on AQI. The lowest R^2 was observed in Cluster 2 with a value of 0.794. Average concentrations of CO and O_3 have negative influence on AQI while SO_2, PM_{10} and $PM_{2.5}$ have positive influence. Table 3 summarizes contribution of five air pollutants to AQI in each cluster.

Table 3 shows that AQI values is predominately influenced by PM_{10} and $PM_{2.5}$ within each cluster using performance indicators. High R^2 and low RMSE values statistically show that PM_{10} and $PM_{2.5}$ are the most influential on AQI. It shows that the effects of O_3, CO and SO_2 on AQI values are insignificant.

The relationships between air pollutants (PM_{10}, $PM_{2.5}$, CO, O_3, SO_2) and AQI values in Cluster 1 are illustrated as scatter plots in Figure 2. Results of R^2 with 95.4% for PM_{10} and 48.8% for $PM_{2.5}$ demonstrate that AQI values are predominately influenced by particulate matter.

Table 3. The contribution of five air pollutants to AQI values

	Variables	**R^2**	**R^2_{adj}**	**Standard Error**	**RMSE**
Cluster 1	AQI – PM_{10}	0.859	0.858	0.0270	27.440
	AQI – $PM_{2.5}$	0.585	0.580	0.243	47.138
	AQI – O_3	0.019	0.008	0.636	72.454
	AQI – CO	0.013	0.002	17.034	72.672
	AQI – SO_2	0.020	0.008	1.416	72.440
Cluster 2	AQI – PM_{10}	0.764	0.762	0.0237	66.824
	AQI – $PM_{2.5}$	0.227	0.221	0.206	120.868
	AQI – O_3	0.000	-0.007	0.872	137.441
	AQI – CO	0.006	-0.002	21.590	137.063
	AQI – SO_2	0.000	-0.007	1.839	137.446
Cluster 3	AQI – PM_{10}	0.676	0.675	0.0270	40.441
	AQI – $PM_{2.5}$	0.599	0.596	0.0994	45.120
	AQI – O_3	0.012	0.006	0.370	70.792
	AQI – CO	0.001	-0.006	21.674	71.199
	AQI – SO_2	0.000	-0.006	1.229	71.224

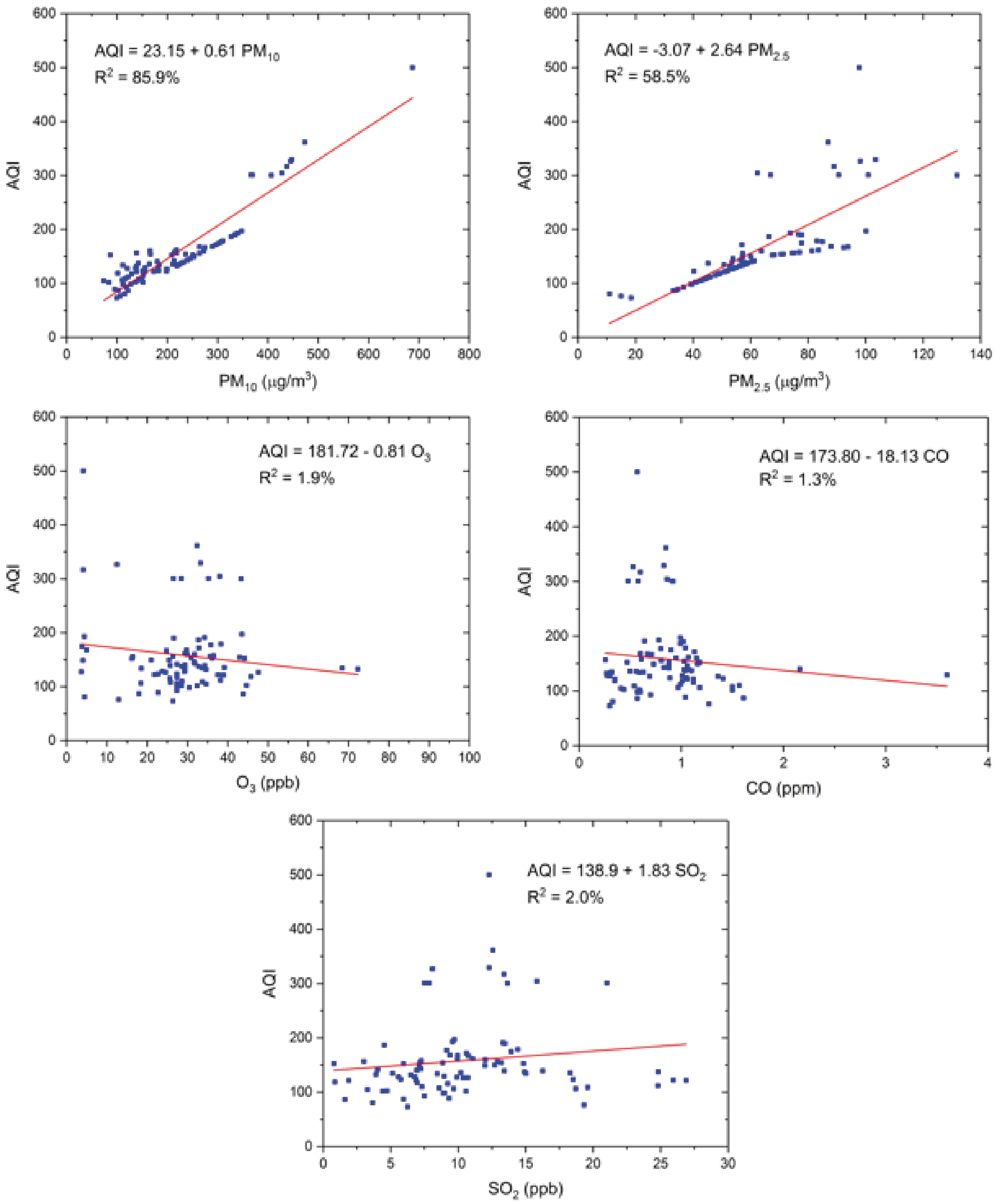
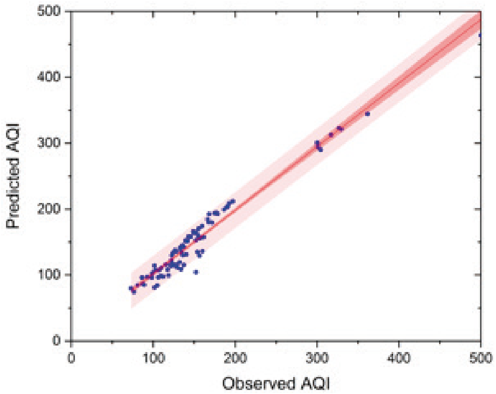


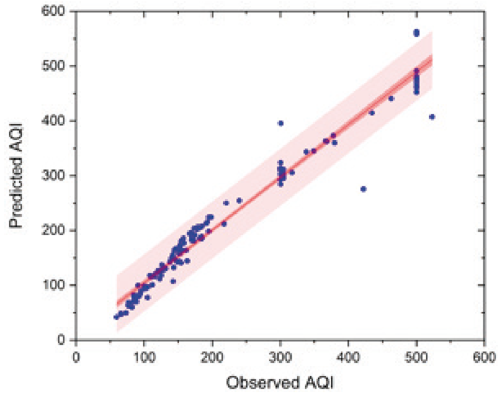
Figure 2. Relationships between air pollutants and AQI values in Cluster 1

Parity plots are useful tools to analyze overall accuracy for developed correlations. Figure 3 shows a comparison between observed (actual) and predicted (correlated) AQI values for Clusters 1-3. Coefficient of determination (R^2) for regression lines within Clusters 1, 2, and 3 are 96.4%, 96.2% and 92.1% respectively. The 95% confidence interval (CI) and 95% prediction interval (PI) bands show good agreement between observed and predicted AQI values.

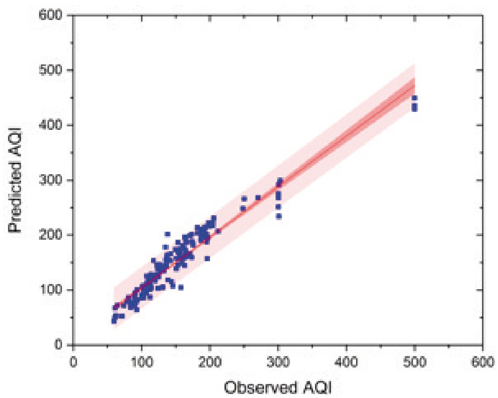
Standardized coefficient values were utilized to compare the relative influence of air pollutants on the AQI values and to explain strength of association. The most important variable has the highest absolute value of the standardized coefficient. Figure 4 shows that PM_{10} has the highest standardized coefficient value followed by $PM_{2.5}$ in all clusters.



Δ Cluster 1

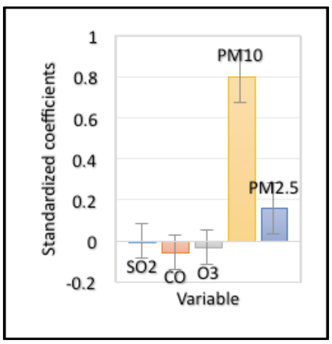


Δ Cluster 2

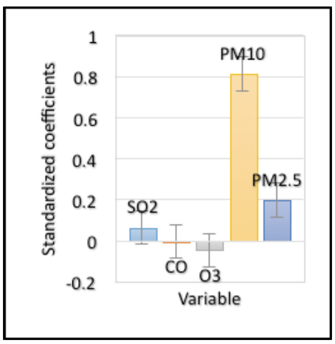


Δ Cluster 3

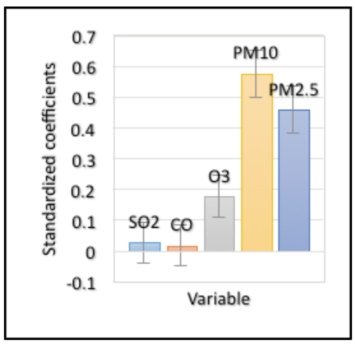
Figure 3. Parity plots of observed versus predicted AQI values for clusters



Δ Cluster 1



Δ Cluster 2



Δ Cluster 3

Figure 4. Bar charts of standardized coefficients for air pollutants

4. Conclusions

Multivariate analysis methods were utilized to evaluate spatial distribution and source apportionment of air pollution in Bahrain. HACA spatially categorized seventeen air quality-monitoring sites into three distinctive clusters.

Results of PCA analysis determined major sources of air pollutants for each cluster. High positive loadings for PM_{10} and $PM_{2.5}$ in Cluster 1 suggest that major sources of air pollution are regional sandstorms and heterogeneous emissions released from extensive industrial activities. Vehicular emissions and combustion of fuel were major sources of air pollution in Cluster 2.

Industrial activities, man-made islands, vehicular emissions from heavy traffic, filling stations and automobile service stations, international airport activities, power and desalination plants are considered major sources of air pollution in Cluster 3. It is concluded that PM_{10} is the most significant source of air pollution throughout Bahrain. MLR analysis demonstrated that PM_{10} is the highest contributor to AQI values followed by $PM_{2.5}$ for all clusters.

Acknowledgements

Any opinion, findings, conclusions, or recommendations expressed herein are those of the authors. They are grateful to the Supreme Council for Environment for providing ambient air quality monitoring data and to Meteorological Directorate for providing meteorological data.

References

Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environmental Modelling & Software 2005; 20(10):1263-71.

Austin E, Coull BA, Zanobetti A, Koutrakis P. A framework to spatially cluster air pollution monitoring sites in US based on the PM 2.5 composition. Environment International. 2013; 59: 244-54.

Bahrain Census. Central Informatics Organization Kingdom of Bahrain. 2010.

Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R and Feigin V. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. The Lancet 2017; 389(10082): 1907-1918.

Dominick D, Juahir H, Latif MT, Zain SM, Aris AZ. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmospheric Environment. 2012; 60: 172-81.

Donders AR, van der Heijden GJ, Stijnen T, Moons KG. A gentle introduction to imputation of missing values. Journal of Clinical Epidemiology. 2006; 59(10): 1087-1091.

EPA (United States Environmental Protection Agency). Technical Assistance Document for Reporting of Daily Air Quality – the Air Quality Index (AQI). EPA-454/B-09-001. U.S. Environmental Protection Agency, 2009; North Carolina 27711.

Ghazali NA, Ramli NA, Yahaya AS. A study to investigate and model the transformation of nitrogen dioxide into ozone using time series plot. European Journal of Scientific Research. 2009; 37(2): 192-205.

Jaramillo P, Muller NZ. Air pollution emissions and damages from energy production in the US: 2002–2011. Energy Policy. 2016; 90: 202-11.

Jassim M, Coskuner G. Environmental engineering education (E3) in the Gulf Co-operation Countries. European Journal of Engineering Education. 2007; 32(1): 93-103.

Jassim MS, Coskuner G. Assessment of spatial variations of particulate matter (PM_{10} and $PM_{2.5}$) in Bahrain identified by air quality index (AQI). Arabian Journal of Geosciences. 2017; 10(1): 19.

Jassim MS, Coskuner G, Munir S. Temporal analysis of air pollution and its relationship with meteorological parameters in Bahrain, 2006–2012. Arabian Journal of Geosciences. 2018; 11(3): 62.

Jobson JD. Applied multivariate data analysis: volume I: Regression and Experimental Design. 4th edn. Springer Science & Business Media; 2012.

Johnson RA, Wichern DW. Applied multivariate statistical analysis. 6th ed. New Jersey: Prentice-Hall; 2007.

Jolliffe IT. Principal component analysis and factor analysis. Principal component analysis. John Wiley & Sons, Ltd. 2002: 150-66.

Junger WL, de Leon AP. Imputation of missing data in time series for air pollutants. Atmospheric Environment. 2015; 28; 102: 96-104.

Kaiser HF. The application of electronic computers to factor analysis. Educational and psychological measurement. 1960; 20(1): 141-51.

Kaiser J. Dealing with missing values in data. Journal of Systems Integration. 2014; 5(1): 42.

Karakitsios SP, Papaloukas CL, Kassomenos PA, Pilidis GA. Assessment and prediction of benzene concentrations in a street canyon using artificial neural networks and deterministic models: their response to "what if" scenarios. Ecological Modelling. 2006; 193(3): 253-70.

Levy RJ. Carbon monoxide pollution and neurodevelopment: a public health concern. Neurotoxicology and Teratology. 2015; 49: 31-40.

Lu Y, Schaefer L. A solid oxide fuel cell system fed with hydrogen sulfide and natural gas. Journal of Power Sources. 2004; 135(1): 184-91.

Middleton NJ. Dust storms in the Middle East. Journal of Arid Environments. 1986; 10(2): 83-96.

Mukhopadhyay K, Forssell O. An empirical investigation of air pollution from fossil fuel combustion and its impact on health in India during 1973–1974 to 1996–1997. Ecological Economics. 2005; 55(2): 235-50.

Özbay B, Keskin GA, Doğruparmak ŞÇ, Ayberk S. Multivariate methods for ground-level ozone modeling. Atmospheric Research. 2011; 102(1): 57-65.

Pai TY, Sung PJ, Lin CY, Leu HG, Shieh YR, Chang SC, Chen SW, Jou JJ. Predicting hourly ozone concentration in Dali area of Taichung County based on multiple linear regression method. International Journal of Applied Science and Engineering. 2009; 7(2): 127-32.

Pavón-Domínguez P, Jiménez-Hornero FJ, de Ravé EG. Proposal for estimating ground-level ozone concentrations at urban areas based on multivariate statistical methods. Atmospheric Environment. 2014; 90: 59-70.

Pires JC, Sousa SI, Pereira MC, Alvim-Ferraz MC, Martins FG. Management of air quality monitoring using principal component and cluster analysis—Part I: SO_2 and PM_{10}. Atmospheric Environment. 2008; 42(6): 1249-60.

Sutton MA, Dragosits U, Tang YS, Fowler D. Ammonia emissions from non-agricultural sources in the UK. Atmospheric Environment. 2000; 34(6): 855-69.

Tsiouri V, Kakosimos KE, Kumar P. Concentrations, sources and exposure risks associated with particulate matter in the Middle East area—a review. Air Quality, Atmospheric and Health. 2015; 8(1): 67–80.

WHO (World Health Organization). Monitoring ambient air quality for health impact assessment. Regional Office for Europe, Copenhagen. 1999.

WHO (World Health Organization). Burden of disease from Ambient Air Pollution for 2012. 2014.

WHO (World Health Organization). WHO Global Urban Ambient Air Pollution Database. 2016.