EPH - International Journal of Agriculture and Environmental Research
ISSN (Online): 2208-2158
Volume 07 Issue 01 June 2021

DOI: https://doi.org/10.53555/eijaer.v5i2.61

MULTIVARIATE STATISTICAL ANALYSES FOR AIR QUALITY CONDITION IN A STEEL PLANT IN KAOHSIUNG CITY, TAIWAN

Edward Ming-Yang Wu[1]*, Shu-Lung Kuo[2]
*[1]*Doctoral Program of College of Maritime, National Kaohsiung University of Science and Technology, Taiwan*
[2]*Department of Technology Management, the Open University of Kaohsiung, Taiwan*

Corresponding Author:-
Email: singsuey@ms28.hinet.net

Abstract:-
Air pollution in Taiwan is mainly created domestically (by factories and power plants) or sent by other countries. Air quality is notably worse in the south compared to other areas due to the establishment of industrial areas and the northeast monsoon bringing pollutants that are not only slow to sediment, but can also result in the deterioration via air pollution. Also, it is not common to evaluate the characteristics and classification via air pollution in a region by the level of pollutants in Taiwan. Therefore, this study uses seven representative air quality variables from five air quality monitoring stations at a steel plant in Kaohsiung, Taiwan, and applies multivariate statistical analysis to discuss the actual situation of, while reflecting the differences in, air quality among the five stations. It then applies the results to the classification of air quality in Southern Taiwan. Through factor analysis under multivariate statistics, the eight air quality variables can be simplified and categorized into three main factors: photochemical polluting factors, organic polluting factors, and fuel factors. These three are the main factors that affect air quality in regions near the steel plant. Moreover, through cluster analysis, air quality in this particular area can be categorized into four clusters, with each cluster representing different characteristics and levels of pollution in the area. The results of this research can provide a reference for government agencies to propose and verify new air quality assessment models, formulate testing models of allowed increment limits of air pollutants, and determine the effectiveness of air quality improvement schemes.

Keywords:- *Multivariate statistical analysis, air quality, factor analysis, variable, steel plant*

1. INTRODUCTION

Air pollution is an important environmental issue in Taiwan and can be either locally produced or transported long distances from East Asia (Cheng et al., 2012; Chuang et al., 2017; Hsu and Cheng, 2019). The major domestic anthropogenic emissions are from urban areas, coal-fired power plants, crude oil refinery plants, industrial parks, and major highways and are emitted mostly in western Taiwan (Hsu and Cheng, 2016). In addition to emissions, meteorological conditions have been shown to play an important role in affecting air pollution dispersion in Taiwan (Lai, 2014; Kuo and Wu, 2018; Kuo and Ho, 2018).

In addition, air pollution is a well-known environmental problem associated with urban areas around the world. Various monitoring programs have been used to determine air quality by generating vast amounts of data on the concentration of each of the previously mentioned air pollutant in different parts of the world. The large data sets often do not convey air quality status to the scientific community, government officials, policy makers, and in particular to the general public in a simple and straightforward manner. This problem is addressed by determining the Air Quality Index (AQI) of a given area. AQI, which is also known as the Air Pollution Index (API) (Murena,2004) or Pollutant Standards Index (PSI) (EPA, 1994), has been developed and disseminated by many agencies in the U.S. Canada, Europe, Australia, China, Indonesia, Taiwan, etc (Cairncross et al., 2007).

This study uses data from five air quality monitoring stations at a steel plant in Kaohsiung, Taiwan, and applies multivariate statistical analysis to classify air quality according to levels of pollution, conditions of pollution, and characteristics of pollution. It also refers to the pollutant standards index (PSI), while discussing the relations among air quality parameters and the distribution characteristics of air pollution in the monitoring work stations. Hopefully, we can truly reflect the differences in air quality among the different stations and establish an evaluation model applicable to the characteristics and classification of all air quality monitoring work stations in Taiwan, thereby providing a reference for the management of air quality monitoring work stations.

2. Methodology

2.1 The application of pollutant standards index

At present, the status of Taiwan's air quality is communicated to the public with the Pollution Standards Index (PSI) which is based on a similar system created by the US Environmental Protection Agency (EPA). Taiwan first used the PSI in 1993 to measure air pollution levels by the ROC Environmental Protection Administration. PSI calculates the sub-index of pollutants based on the influence of five pollutants: particulate matter with a particle size below 10 microns (PM_{10}), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3), all of which are measured on a daily basis. The maximum values of the daily sub-index then are used as the PSI value measured by the monitoring station. The main purpose is to monitor the integral air quality of central Taiwan and suggest areas for improvement. Through the evaluation of PSI, local air quality statuses can be fully understood. The concentration levels of the five air pollutants are used to determine PSI which is then relayed as a number between 0 and 500 and classified into Good (0~50), Moderate (51~100), Unhealthy (101~199), Very Unhealthy (200~299), and Hazardous (≥300) levels. The ranges for PSI and pollutant concentration levels as well as PSI are shown in Table 1.

Table 1 Comparison table of pollutant concentration and pollution sub-index.

Pollutant	PM_{10}	SO_2	CO	O_3	NO_2
Statistics	24-hour average	24-hour average	Maximum 8-hour average within a 24-hour period	Maximum and minimum within a 24-hour period	Maximum and minimum within a 24-hour period
Unit	μg/m^3	ppb	ppm	ppb	ppb
PSI					
50	50	30	4.5	60	
100	150	140	9	120	
200	350	300	15	200	600
300	420	600	30	400	1200
400	500	800	40	500	1600
500	600	1000	50	600	2000

2.2 Places of sample and data collection

The data used in this paper were obtained from the five air quality monitoring stations (referred to as stations A, B, C, D, and E) at a steel plant in Kaohsiung, Taiwan, between June 2017 and June 2019. Data were collected every day, except for the period during which the change of monitoring stations obstructed data collection and days with severe weather conditions that hampered data collection. A total of 610 complete sets of data were collected. Figure 1 shows the geographic location of the steel plant, and Figure 2 is the position chart of the steel plant; said plant is adjacent to the Port of Kaohsiung and occupies an area of approximately 550 hectares.

To ensure the completeness and diversity of the collected air quality data, this study chose the following seven air quality parameters: SO_2, NO_2, CO, PM_{10}, O_3, THC, and NMHC, to perform a factor analysis to find common factors. The concentration of said air quality parameters follows the standard set by Taiwan's Pollution Standards Index (PSI) (as shown in Table 1). The concentration of SO_2 is the average of data collected within a 24-hour period; the concentration of NO_2 is the maximum of data collected within a 24-hour period; the concentration of CO is the maximum of data collected within an 8-hour period; the concentration of PM_{10} is the average of data collected within a 24-hour period; the concentration of O_3 is the minimum of data collected within a 24-hour period; and the concentration of THC and NMHC is the average of data collected within a day.

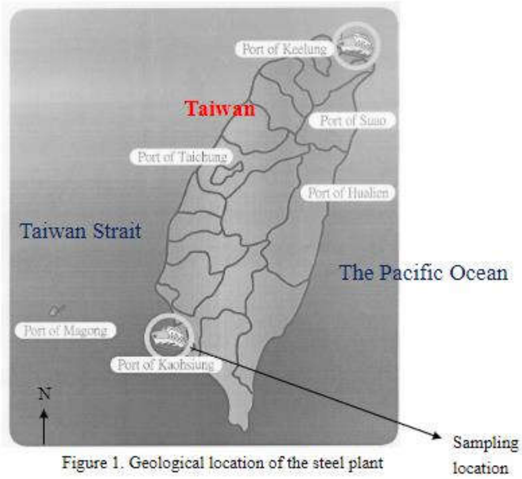

Figure 1. Geological location of the steel plant

Figure 2. The position chart of the steel plant

2.3 Multivariate statistical analyses—factor Analysis

For selecting the elements to be included in FA, a minimum of 70% of the samples needs to have measurable levels of the element. In principle, FA actually groups the elements whose concentrations fluctuate together from one sample to another and separates these elements into so called "factors" (Henry et al., 1984; Martinez et al., 2012). Factor analysis is used for source apportionment in environmental data, with the argument that those elements that fluctuate together have some common characteristics. Ideally, each extracted factor represents a source affecting the samples. Factor analysis has been performed using the Statgraphics plus program package (Statgraphics Manual 3.1, 1997). The initial components were rotated using the varimax method to obtain final eigenvectors with more representatives of individual sources of variation. Although there are no well-defined rules on the number of factors to be retained, usually either factors that are meaningful or factors with eigenvalues larger than 1 are retained. In theory, irrelevant factors have zero eigenvalues and eigenvalues less than 1 indicate that factor contributes less than a single variable. The physical meaning of the factors must be interpreted by observing which elements or variables display high ($\geq$0.25) loading within the factor. Loadings less than 0.25 in absolute value may be dominated by random errors.

2.4 Multivariate statistical analyses—cluster analysis

Cluster analysis is an exploratory data analysis tool for solving classification problems. Its objective is to sort cases into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may be abstracted through use from the particular to the general class or type. Hierarchical agglomerative clustering is the most common approach as it provides intuitive similarity relationships between any one sample and the entire dataset. It is typically illustrated by a dendrogram (tree diagram) (McKenna 2003). The dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. Additionally, cluster analysis helps in grouping objects (cases) into classes (clusters) on the basis of similarities within a class and dissimilarities between different classes. The class characteristics are not known in advance but maybe determined from the analysis. The results of CA help in interpreting the data and indicate patterns (Vega et al. 1998; Tobiszewski, et al., 2010).

3. Results and discussion

3.1 Selection of the results of factor analysis

This study adopts Varimax Rotation under factor analysis to carry out orthogonal rotation to explain the different characteristics of each factor. According to the analysis results, there are three factors with eigenvalues greater than one, as shown in Table 2. The total cumulative variance explained by these three factors is 73.568%, while their eigenvalues are 2.106, 1.685 and 1.231, respectively. As indicated separately in Table 3, the KMO value is 0.630, hence bigger than 0.5, which is suitable for factor analysis according to Kaiser's viewpoint. Furthermore, the χ^2 value obtained from Bartlett's Test of Sphericity is 2150.397 (the degree of freedom is 22); this value has reached the significance level, which means that common factors exist in the relevant matrix of the population, and it is thus suitable for conducting a factor analysis.

Table 2 Results of factor analysis and the variance explained

Components	Initial eigenvalues	% of total variance	Cumulative variance %
1	2.106	31.544	31.544
2	1.685	25.096	56.640
3	1.231	16.928	73.568
4	0.715	12.251	85.819
5	0.603	7.175	92.994
6	0..551	4.653	97.647
7	0.398	2.353	100.000

Table 3. KMO and Bartlett's test table of seven air quality parameters

Kaiser-Meyer-Olkin measure of sampling adequacy		.630
Bartlett test of sphericity	Chi-square distribution	2150.397
	Degree of freedom	22
	Significance	.000

3.2 Determining which factors to use

We can determine the number of main factors based on the number of factors with eigenvalues greater than one, as mentioned above. Through the component matrix that underwent orthogonal rotation, as shown in Table 4, we can choose variates among the factors. While the rotated matrix can explain the characteristics of each factor, the three factors can also be used to describe the relations and differences among air quality parameters.

Table 4. Loading matrix of factors

Parameters	Factors		
	1	2	3
O_3	0.865	-0.005	0.006
PM_{10}	0.803	0.026	-0.058
NO_2	0.710	-0.089	-0.084
SO_2	0.092	0.806	0.27
CO	-0.087	0.763	0.049
THC	0.036	0.248	0.924
NMHC	-0.015	0.098	0.862

3.3 Explanatory factors

Table 2 indicates that there are three main factors affecting air quality at the steel plant, while from the 3D scatter plot of factor distribution (Figure 3), we can see that there are three axes: O_3, PM_{10}, and NO_2 on the same axis and belonging to the first factor; SO_2 and CO are on the same axis and belong to the second factor; and NMHC and THC are on the same axis and belong to the third factor. The following paragraph offers a complete explanation of the characteristics of each factor.

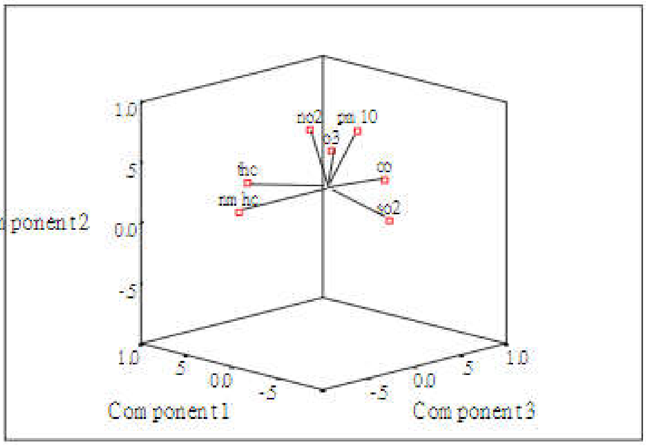

Figure 3. 3D distribution of main factors

3.3.1 The first factor

The first factor mainly consists of O_3, PM_{10}, and NO_2. The total cumulative, as shown in Table 2, reached 31.544%. From Table 4, we can see that O_3 has the highest factor loading of 0.865. While PM_{10} can be the factor resulting in poor air quality, O_3 remains the main contributing factor. In cities with heavy traffic, there is usually a high concentration of O_3, which can create a yellowish-brown photochemical smog because fuel consumption of cars and scooters produces NO_x, such as NO and NO_2; NO_2 can interact with O_2 under sunlight and produce O_3 and NO. O_3 will then interact with total hydrocarbons (THCs) emitted by cars or scooters, and produce photochemical smog; this process is the so-called photochemical reaction. O_3 is an essential air pollution indicator; it contributes largely to air pollution. PM_{10} and NO_2 churned out by the coking plant and sintering plant of the steel plant are the precursors of O_3; also, since there is a cluster of industrial areas adjacent to the steel plant, there is always a high concentration of O_3 in the atmosphere.

The factor loading of PM_{10} for the first factor in Table 4 is relatively high, reaching 0.803. PM_{10} is a crucial indicator that helps to determine the level of air pollution. PM_{10} in the atmosphere mainly comes from two sources: primary aerosol and secondary aerosol. Primary aerosol is usually produced by human activities (such as factories burning materials, emissions produced by vehicles, etc.), while other fugitive emission sources (such as street dust and soil dust) and sea salt droplets also contribute significantly to air pollution in different areas. These particles, especially those with a diameter smaller than 10μm, can enter the lungs and damage the respiratory system. If there is an excessive level of NO_2 and other irritant gases in the air, they will react with PM_{10} and form yellowish-brown photochemical smog, which serves as a warning sign of severe air pollution. Besides, among all the pollutants that result in poor air quality, PM_{10} is the primary indicative pollutant. PM_{10} in the air around the steel plant mostly comes from burning and coking coals of the coking plant and sintering plant.

Besides, the factor loading of NO_2 for the first factor in Table 4 is 0.710. NO_2 mostly comes from vehicle emissions, coal-fired power plants, and other forms of industrial burning, which create NO that later reacts with oxygen in the air and turns into NO_2. While NO_2 is a crucial indicator of air pollution, on a windless day, NO emitted by cars can accumulate in the air and trigger photochemical reactions, so it is also one of the pollutants causing photochemical smog. NO_2 in the air around the steel plant mostly comes from burning and coking coals of the coking plant and sintering plant.

To sum up the above analysis results, O_3, PM_{10}, NO_2 and other main pollutants and activities causing pollution are correlated to the photochemical reaction; this is why the first factor is referred to as the "photochemical pollutant factor."

3.3.2 The second factor

The second factor consists of SO_2 and CO; its total cumulative, as shown in Table 2, reaches 25.096%. From Table 4, we can see that SO_2 has the highest factor loading, 0.806. SO_2 is usually produced by burning minerals containing sulfur. Since diesel-fueled vehicles and industrial emissions are closer to receptors in cities, they are the key source of SO_2 in cities. In addition to harming plants and humans, SO_2 also causes local and regional acid rain (Chou, 2010). Furthermore, SO_2 is an important air pollution indicator. SO_2 in the air around the steel plant mostly comes from burning and coking coals of the coking plant, sintering plant, and converter plant.

The factor loading of CO for the second factor in Table 4 is 0.763. CO mostly comes from vehicle emissions and partially from the incomplete combustion of fuels in factories and power plants, so we can see that it is a kind of gas produced by incomplete combustion of fuels containing carbons. CO is an essential air pollution indicator as well, and has the highest concentration among other pollutants in the air (Kuo and Ho, 2018). The coking plant, sintering plant, blast furnace plant and steel plate plant in the steel plant also produce CO due to incomplete combustion during operations.

To sum up the above analysis results, SO_2 and CO mostly come from burning fuels, which can be referred to as "fuel factors."

3.3.3 The third factor

The third factor consists of NMHC and THC; its total cumulative, as indicated in Table 2, is 16.928%. Also, from Table 4, we know that the factor loading of NMHC is 0.924, while that of THC is 0.862. The factor loadings of these two factors are utterly high, from which we can see that the two pollutants are highly correlated. These volatile organic compounds mainly stem from industrial activities or emanate into the air when filling gasoline in vehicles. NMHC can form photochemical smog, O_3, and other pollutants through the photochemical reaction under sunlight, or even produce secondary suspended particles, having a profound impact on the human respiratory system (Murena, 2004). The coking plant, blast furnace plant and hot rolling mill produce NMHC and THC of higher concentrations during operation.

Since NMHC and THC are both volatile organic compounds, the third factor can thus be referred to as "organic pollutant factor."

3.4 An analysis of air pollution characteristics: the results of cluster analysis

Before conducting cluster analysis, this study produced a standardized component score coefficient matrix in analyzing the above three factors, and then carried out a linear combination. After timing the component score with the value of the original variables, the factor score of each factor was obtained. The factor scores of the three main factors were used as the samples' measurement, i.e., "clustering variables", to classify air quality into clusters, while homogenizing them at the same time. The two-stage clustering method was applied during clustering. After acquiring a general clustering result using a hierarchical clustering method, the K-mean method was used to test different clusters. In the end, four clusters were chosen to differentiate air quality around the steel plant. The relationships among clusters and factors are demonstrated in Figure 4, while Figure 5 is a distribution graph of the four clusters from each air quality monitoring work station.

In Figure 4, we observe that the three factors in the 4th cluster all have a high factor score, and they are the worst among the four clusters in terms of air quality. The score of the first factor is exceptionally high. In the second cluster, the first and second factors have the lowest points. Transforming the four clusters into original air quality monitoring items will make it easier for us to understand the characteristics of air quality around the steel plant, as indicated in Table 5.

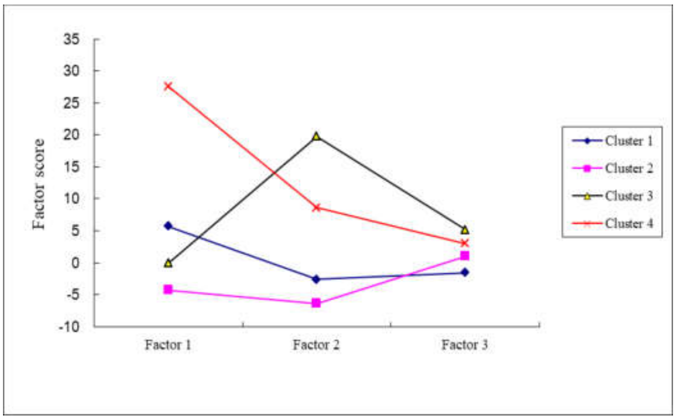

Figure 4. Relationships among clusters and factors

Table 5. Average and maximum values of clusters and air pollution parameters

variables \ clusters	1st Cluster	2nd Cluster	3rd Cluster	4th Cluster
O_3(ppb)	31.58 3.16~61.3	28.92 0.21~97.63	50.12 5.32~97.06	58.24 4.46~103.22
PM_{10}(μg/m^3)	88 38~163	72 21~145	122 50~233	182.15 93~501
NO_2(ppb)	**357.65** 198.32~523.6	43.21 0.36~206.18	51.38 1.46~111.45	86.91 2.45~271.55
SO_2(ppb)	18.88 9.16~29.46	23.22 0.43~92.17	110.54 40.69~341.02	35.16 3.47~129.36
CO(ppm)	1.05 0.49~2.36	1.26 0.06~9.57	1.54 0.80~15.63	1.50 0.28~27.69
NMHC(ppb)	0.38 0.21~0.88	0.65 0.021~6.47	0.81 0.30~2.11	0.74 0.080~2.38
THC(ppb)	1.86 1.05~2.58	1.96 0.46~8.02	2.16 1.31~4.23	2.69 0.50~5.01
Air Quality Classification	Good to average	Good to average	Average to poor	Average to harmful
Stations at which data were collected	Stations D and E	Stations A, B, C, D, and E	Stations B and C	Stations A, B, C, D, and E
Station(s) with the worst air quality	--	--	Station C	Stations A, B, and E
Total days	22	298	73	217

3.4.1 The first cluster

This cluster, as indicated in Table 5, has the highest NO_2 concentration among the four clusters; however, it only ranks second in terms of the photochemical pollutant factor score in Figure 4. This is because no correspondent index can be used to measure the concentration of NO_2 in Taiwan, as shown in the pollutant standards index (PSI) mentioned in Table 1; under PSI, the concentration of any substance only ranges from 1 to 100. However, if the concentration of NO_2 reaches 600 ppb, the corresponding value under PSI should be 200. The highest concentration of NO_2 observed in this steel plant was 523.6 ppb, which is considered relatively high. However, since PSI could not adequately present it, the impact of photochemical pollutant on this cluster is weaker when compared with that on the 4th cluster. In addition, this cluster has the third-highest factor scores in terms of organic and fuel factors, with an average concentration of NMHC reaching 0.38 ppb and an average concentration of THC reaching 1.86 ppb, respectively. Under the fuel factor, the average concentration of SO_2 is 12.88 ppb, while that of CO is 1.05 ppm. The factor scores of this cluster mostly rank between the second and the third-highest; its air quality is classified as good to average. While the concentration of NO_2 in this cluster is generally higher than that in other clusters, poor or harmful air quality among the five stations was not observed. NO_2 concentration reaching 523.6 ppb was observed in Station D; however, it is not classified as poor to harmful air quality. From the above analysis results, the air quality under this cluster can be classified as “air with medium level photochemical pollution”, with Stations D and E being representative of this cluster.

3.4.2 The second cluster

This cluster, as shown in Figure 4, has the lowest factor scores in terms of the photochemical pollutant and fuel factors. It also has the second-lowest factor score in terms of organic pollutant factor. The results mean that air quality in this cluster is relatively good. In Table 5, we can see that no matter under what parameters, this cluster generally has a lower concentration than other clusters; concentration values do not increase sharply, either. Therefore, the air quality of this cluster falls between good and average, with average air quality being the majority; the overall air quality is better than that of the first cluster. This cluster is mostly affected by the concentration of PM_{10}. Air quality will be considered average if the concentration of PM_{10} falls between 50 and 150 $\mu g/m^3$; the concentration of PM_{10} in this cluster mostly remains above 50 $\mu g/m^3$. However, since the concentration of the other six air quality parameters are not considered to be high, the air quality of this cluster is mainly classified as average. Data in this cluster were collected between Stations A, B, C, D, and E. The days of data being collected were almost the same. For instance, the cluster includes data collected on 64 days at Station A, 57 days at Station B, 55 days at Station C, 62 days at Station D, and 60 days at Station E. In conclusion, the air quality of this cluster can be classified as “slightly polluted.”

3.4.3 The third cluster

This cluster, as indicated in Figure 4, has the highest factor score in terms of the fuel factor and the third highest factor score in terms of the photochemical pollutant factor. From Table 5, we also know that the average concentration of SO_2 in the fuel factor under the third cluster reaches as high as 110.54 ppb, which is the highest among all the clusters. On the other hand, the average concentration of CO under the same cluster is 1.54 ppm, which is the second-highest among them all. The concentrations of SO_2 are between 40.69~341.02 ppb, which indicates a considerable effect of fuel factor on this cluster. Moreover, from Figure 4, we understand that this cluster only has the third-highest factor score in terms of the photochemical pollutant factor. The average concentration of NO_2 under this cluster is 51.38 ppb, that of PM_{10} is 122 $\mu g/m^3$, and that of O_3 is 50.12 ppb. From the above results, we can see that the concentration of photochemical pollutants at monitoring stations with higher fuel factor loading did not increase significantly. The air quality of this cluster is classified as average to poor, with average air quality being observed the most. The worst air quality was observed at Station C. It was winter, and the concentration of SO_2 reached as high as 341.02 ppb. This cluster also includes days with PM_{10} concentration above 150 $\mu m/m^3$, which resulted in poor air quality. Data under this cluster were mainly collected at Stations B and C, with 64 days at Station C and only nine days at Station D. To sum up the above results, the air quality of this cluster can be classified as “seriously polluted by the fuel factor.”

3.4.4 The fourth cluster

This cluster, as shown in Figure 4, has the highest factor score in terms of the photochemical pollutant factor, and the second-highest factor score in terms of the organic pollutant and fuel factors. The results indicate that the level of pollution under this cluster is the highest among all the clusters. In Table 5, we can see that the average concentration of O_3 under the photochemical pollutant factor is 58.24 ppb, which is the highest among all the clusters; the average concentration of PM_{10} is 182.15 $\mu g/m^3$, which is also the highest among all the clusters; and the average concentration of NO_2 is 86.91 ppb, which is the second-highest among all the clusters. The steel plant is a place easily affected by air pollution due to its location and the type of industry in the area. PM_{10} of high concentration plays a significant role in keeping the air pollution level high. To illustrate, in this cluster, the concentration of PM_{10} is always above 150 $\mu g/m^3$, and the number is especially high during the period of late autumn and early spring in the following year, with the maximum concentration reaching 501$\mu g/m^3$. The main factor causing this phenomenon is the external pollutants brought by the northeast monsoon, namely sandstorms, which keep the air pollution level at red alert, meaning that it is harmful to humans. Moreover, this cluster also has a high factor score in terms of the fuel factor, with an average concentration of SO_2 reaching 35.16 ppb, which is the second-highest among all the clusters. Its average concentration of CO is 1.50 ppm, which is also the highest among all the clusters. However, the maximum concentration of CO within eight hours of data collection at Station D reached as high as 27.69 ppm, which was harmful to humans. This cluster also obtains the second-highest factor score in terms of the organic pollutant factor among all the four clusters, with the highest average concentration of THC of 2.69 ppb and the second-highest average concentration of NMHC of 0.74 ppb. However, since organic pollutants are not clearly defined under PSI in Taiwan, while having smaller impacts on air pollution, their factor scores in Figure 4 are relatively low. The air quality of this cluster ranges between average and harmful, with the air quality at a harmful level being observed the most. Data of this cluster were collected from Stations A, B, C, D, and E. The cluster includes data collected at Station A on 37 days, Station B on 32 days, Station C on 49 days, Station D on 40 days, and Station E on 59 days. The harmful-level air quality could be observed mainly at Stations A, B, and E. The air quality of this cluster can be referred to as "air with severe photochemical pollution."

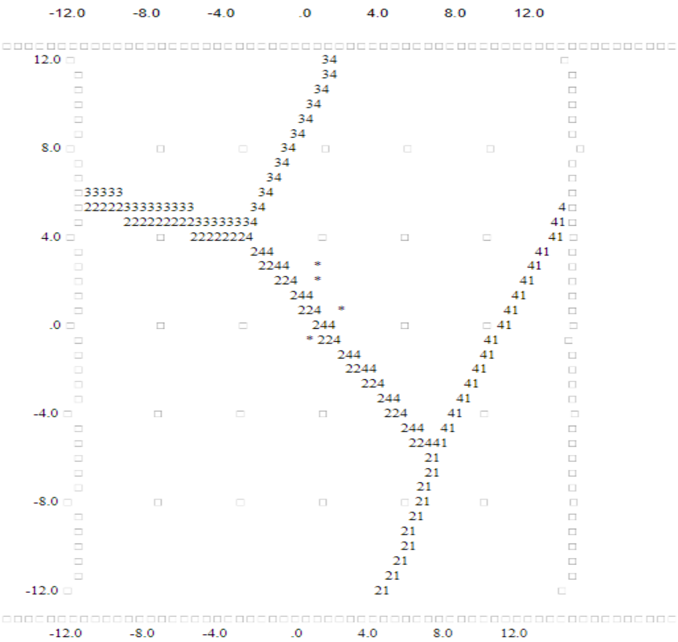

Figure 5. The distribution of the four clusters at the five air quality monitoring work stations

4. Conclusion

This study used data from five air quality monitoring work stations at a steel plant in Kaohsiung, Taiwan, and applied multivariate statistical analysis to classify air quality according to levels of pollution, conditions of pollution, and characteristics of pollution. This study also referred to the pollutant standards index (PSI), while discussing the relations among air quality variables and the distribution characteristics of air pollution at the monitoring work stations. The results of factor analysis under multivariate statistical analysis methods indicated three factors with an eigenvalue greater than 1; first factor: "organic pollutant factor", second factor: "photochemical pollutant factor" and third factor: "fuel factor." The total cumulative explained is 77.568%. Air pollution caused by the three factors in sequential accordance with severity is: photochemical pollutant factor> fuel factor> organic pollutant factor. Moreover, this study tested different clusters using K-mean methods and divided them into four different clusters. The first cluster is "air with medium level photochemical pollution", the second is "air that is slightly polluted," the third is "air seriously polluted by the fuel factor", and the fourth is "air with severe photochemical pollution." Furthermore, since the data on methane and part of the data we had were incomplete, we were unable to conduct a multivariate statistical analysis on the data. If we can fill this gap in the future, we could acquire more accurate analysis results. The results of this research can serve as a reference for government agencies to propose and verify new air quality assessment models, formulate testing models of allowed increment limit of air pollutants, and determine the effectiveness of air quality improvement schemes.

References

[1]. Cairncross, E.K., John, J., and 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.

[2]. Cheng, F.Y., Chin, S.C., Liu, T.H. (2012). The role of boundary layer schemes in meteorological and air quality simulations of the Taiwan area. Atmos. Environ. 54: 714–727.

[3]. Chou, T.G. (2010). The research of air quality management in fine particulate matters. Central Office of Administration, Academia Sinica Research weekly 1276, Central Office of Administration, Academia Sinica, Taiwan. (in Chinese)

[4]. Chuang, M.T., Chou, C.C.K., Lin, N.H., Takami, A., Hsiao, T.C., Lin, T.H., Fu, J.S., Pani, S.K., Lu, Y.R., Yang, T.Y. (2017). A simulation study on $PM_{2.5}$ sources and meteorological characteristics at the northern tip of Taiwan in the early stage of the Asian haze period. Aerosol Air Qual. Res. 17: 3166-3178.

[5]. EPA (Environmental Protection Agency) (1994), Measuring air quality: the pollutant standards index, EPA 451/K-94-001, U.S.A..

[6]. Henry, R.C., Lewis, C.W., Hopke, P.K., and Williamson, H.J. (1984). Review of receptor model fundamentals. Atmos. Environ., 18, 1507-1515.

[7]. Hsu, C.H. and Cheng, F.Y. (2016). Classification of weather patterns to study the influence of meteorological characteristics on $PM_{2.5}$ concentrations in Yunlin County, Taiwan. Atmos. Environ. 144: 397-408.
[8]. Kuo, S.L., Ho, C.L. (2018). The assessment of time Series for an entire air quality control district in southern Taiwan using GARCH model. International Journal of Engineering & Technology 7, 119-124.
[9]. Kuo, S.L., Wu, E.M.Y. (2018). Highway tunnel air quality assessment using multivariate statistical classification on factor, cluster, and discriminant analysis. International Journal of Engineering & Technology 7, 287-291.
[10]. Hsu, C.H., Cheng, F.Y. (2019). Synoptic weather patterns and associated air pollution in Taiwan. Aerosol and Air Quality Research 19, 1139-1151.
[11]. Lai, L.W. (2014). Relationship between fine particulate matter events with respect to synoptic weather patterns and the implications for circulatory and respiratory disease in Taipei, Taiwan. Int. J. Environ. Health Res. 24: 528-545.
[12]. Martinez, M.A., Caballero, P., Carrillo, O., Mendoza, A., Mejia, G.M. (2012). Chemical characterization and factor analysis of $PM_{2.5}$ in two sites of Monterrey, Mexico. Journal of the Air & Waste Management Association 62, 817-827.
[13]. McKenna, Jr. J.E. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modeling & Softwar 18, 205-220.
[14]. 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.
[15]. Tobiszewski, M.; Tsakovski, S.; Simeonov, V.; Namieśnik, J. (2010). Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. Chemosphere 80, 740-746.
[16]. Vega, M.; Pardo, R.; Barrado, E.; Deban, L. (1998). Assessment of seasonal and polluting effects on the qualityof river water by exploratory data analysis. Water Res. 32, 3581-592.