International Journal of Engineering & Technology, 7 (2.13) (2018) 287-291

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Highway tunnel air quality assessment using multivariate statistical classification on factor, cluster, and discriminant analysis

Shu-Lung Kuo [1], Edward Ming-Yang Wu [2] *

[1] Department of Technology Management, the Open University of Kaohsiung, No.436, Taiwan
[2] Department of Civil and Ecological Engineering, I-Shou University, Kaohsiung, Taiwan
**Corresponding author E-mail: edmywu@isu.edu.tw*

Abstract

The air quality monitoring points set up at the existing 10 tunnels (a total of 20 tunnel entrances) on Formosa Freeway in Northern Taiwan were used in this study to investigate the correlation among various types of air pollutants measured at these 20 tunnel entrances via a multivariate statistical analysis. This study aimed to determine the main factors that affect the extent of air pollution along the Formosa Freeway and its vicinity, and explored the interrelationships among various air pollutants to reflect the differences among the air pollutants found along the Formosa Freeway in Northern Taiwan, as well as to establish an evaluation mode for types and characteristics of air pollutants after their quantification. Two main factors were obtained from the factor analysis: the "photochemical reaction pollution factor" and the "vehicle fuel factor." Cluster analysis is used to classify air quality in Formosa Freeway in Northern Taiwan into five clusters to present different characteristics and pollution degrees of air quality in this study. The more that the air pollutants and samples are used when performing a factor analysis, the more effective the validity and reliability of the factor analysis! This research results can serve as a reference for those involved in the review of air quality management effectiveness and/or the enactment of management control strategies. In addition, it is also a good method to use in an air quality management program and it is expected that the results can serve as a reference for the management to prevent and control air pollution in the future.

Keywords: *Formosa Freeway; Multivariate Statistical Analysis; Reliability and Validity; Tunnel.*

1. Introduction

With the economic growth in Taiwan and the completion of various major construction projects, environmental protection and economic growth should be stressed for long-term development benefits. Among the infrastructures in Taiwan, the construction of freeways is the most common, especially transport plans concerning freeways connecting Northern and Southern Taiwan. As these freeways pass through several counties and cities, it is very important to explore the impact of the increase in emissions of various air pollutants on the air quality in all the counties and cities along the route.

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; Cheng et al., 1999].

Based on the reasons mentioned above, the air quality monitoring points set up at the existing 10 tunnels (20 tunnel entrances) on Formosa Freeway in Northern Taiwan were used in this study to investigate the correlation among various types of air pollutants measured at these 20 tunnels from January 2016 to December 2017, via multivariate statistical analysis. This study aimed to find the main factors that affect the extent of air pollution along the Formosa Freeway and its vicinity. In addition, this study performed a cluster analysis to find the differences among the air pollutants found along the Formosa Freeway in Northern Taiwan, as well as to establish an evaluation mode for types and characteristics of air pollutants after their quantification.

2. Methodology

A multivariate analysis is based on statistics, and can be used to analyze the observation data with two or more variables at the same time. This method makes complicated problems or phenomena into a reasonable and systematic collation, judging, evaluating and forecasting their features by using the statistical linear algebra of multivariate space. Since the 19th century, this method has been widely applied in various fields, such as socioeconomics, behavioral sciences and biostatistics, and recently has gradually been applied to environmental sciences. The main purpose of employing multivariate analysis is to investigate the status quo in order to control information. The analysis has broad applications, mainly in management science, such as marketing, business management,

Copyright © 2018 Shu-Lung Kuo, Edward Ming-Yang Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

international business, accounting, economics and industrial engineering, but most frequently in marketing research. The reason that best-selling products earn their fame lies in a cautious and in-depth multivariate statistical analysis.

2.1. Factor analysis

In factor analyses, each variable, i.e. x_1~x_p, in a set of p variables is decomposed into q common factors, i. e. f_1~f_p ($q \leqq p$), which are linearly combined with special factors ϵ_i.
The model of factor analyses can be represented by:

$$X1=\mu_1+\ell_{11}f_1+\ell_{12}f_2+\ldots+\ell_{1q}f_1+\epsilon_1$$

$$X2=\mu_2+\ell_{21}f_1+\ell_{22}f_2+\ldots+\ell_{2q}f_2+\epsilon_2$$

$$Xi=\mu_i+\ell_{i1}f_1+\ell_{i2}f_2+\ldots+\ell_{iq}f_2+\epsilon_i$$

$$Xp=\mu_p+\ell_{p1}f_1+\ell_{p2}f_2+\ldots+\ell_{pq}f_q+\epsilon_p \quad (1)$$

Where: f_1, f_q are common factors contained in each variable x_i; ϵ_i is special factor contained only is the i^{th} variable (xi); ℓ_{ij} is the loading of i^{th} factor to the j^{th} common factor (f_j).

2.2. 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) [Einax et al., 1998; 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. Additinally, 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].

2.3. Discriminant analysis

Discriminant analysis is a method used to distinguish the type (group) of a sample (an interviewee) in a multivariate statistical analysis. It exhibits similarity to a cluster analysis, where similar samples (interviewees) are classified into one category (group). However, it differs from a cluster analysis in that information on classification in a cluster analysis cannot be obtained in advance. Discriminant analysis is based on the sample data used to infer one or a set of (discriminant) functions under the condition that the classification of research subjects is known. Meanwhile, a discriminant rule is specified to determine the category of the sample to be discriminated, in order to reach a minimum misjudgment rate. The common criteria used in a multi-parameter analysis are: 1) Fisher Discriminant Rule: it is based on Fisher's linear discriminant function, and is usually used for two sets of discriminant problems. The criteria require significant differences in the mean values of each group's variables. 2) Bayes Discriminant Rule: it is based on the prior probabilities of each population to minimize the average loss of misjudgment to discriminate, and is usually used for multiple sets of discriminant problems. This method requires that all groups of variables and three assumptions be satisfied: the multivariate normal distribution, the equal variance matrix of each group, and significant differences in the mean values of each group's variables.

2.4. Research methodology

2.4.1. The selection of air pollutants

Air pollutants chosen by this study were based on a total of 520 entries in the complete air quality monitoring data, which were collected from 10 tunnels (20 tunnel entrances) along the Formosa Freeway in Northern Taiwan between January 2016 and December 2017, using an air quality monitoring vehicle by an environmental monitoring corporation in Taiwan. The relevant locations of the 10 tunnels (20 tunnel entrances) chosen by this study are shown in Figure 1; they are in New Taipei City, Taipei City, and Taoyuan City in Taiwan.

2.4.2. The selection of air pollutant factor model analysis

To highlight the integrity and diversity of air pollutants, six major air pollutants detected by the air quality monitoring vehicle in the atmosphere were used for the factor analysis; they included suspended solids (particulate matter less than 10 micrometers in diameter, PM_{10}), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), total hydrocarbons (THC), carbon monoxide (CO), and ozone (O_3). The objective of the analysis was to identify the main factors that may affect the air quality of the Formosa Freeway in Northern Taiwan, and to find the interrelationships and differences among air pollutants.

3. Results and discussion

3.1. Results of the factor analysis

3.1.1. The selection of factors

In calculating the results of the factor analysis, this study mainly used varimax rotation to conduct orthogonal rotation to explain the number characteristics of the factors. According to the analysis results, there were two factors with eigenvalues greater than 1, as shown in Table 1 and Figure 1. The cumulative variance of these two common factors was 58.157%, and the eigenvalues of these two common factors were 2.456 and 1.523, respectively. Therefore, this study chose these two factors to explain the major factors that affect the air quality in this region.

Table 1 Results Factor Analyses and the Variance Explained for the Formosa Freeway in Northern Taiwa

Components	Initial eigenvalues	% of total variance	Cumulative variance %
1	2.456	35.192	35.192
2	1.523	22.965	58.157
3	0.912	14.868	73.025
4	0.667	11.009	84.034
5	0.492	8.047	92.081
6	0.336	7.919	100.000

3.1.2. The determination of factors

Through the above-mentioned number of eigenvalues that were greater than 1, the number of the main factors could be determined. The parameters among various factors could be chosen from the number of the main factors of the component matrix after orthogonal rotation. Table 2 is the table of the component matrix after the orthogonal rotation. The matrix after rotation could be used to explain the characteristics of the two factors. Moreover, these two factors could be used to describe the main factors that affect the air quality in this region, as well as their differences.

Table 2 Matrix of Water Quality Factor Loadings for the Formosa Freeway in Northern Taiwan

Parameters	Factors	
	1	2
THC	0.903	0.291
O_3	0.817	-0.416
NO_2	0.747	0.214
PM_{10}	0.659	-0.095
CO	-0.321	0.812
SO_2	0.142	0.737

3.1.3. An explanation of the factors

1) The first factor

The first factor was mainly composed of the four air pollutants: THC, O_3, NO_2 and PM_{10}; the total variance could reach 35.192%, as shown in Table 1.

Table 2 shows that THC had the highest loading level of 0.903 within the first factor. After the irradiation with ultraviolet rays, the nitrogen oxides and hydrocarbons in the atmosphere underwent a photochemical reaction and formed photochemical smog. The hydrocarbons also underwent a photochemical reaction with nitrogen oxides, resulting in strong oxidizing substances, such as ozone (O_3) and peroxyacetyl nitrate (PAN), which are secondary pollutants. The ozone accounted for the highest amount of the secondary pollutants; it was the major component of photochemical smog. In addition, hydrocarbons are generally considered to be produced by the thermal decomposition of organic matter containing hydrocarbons during the combustion process; incomplete combustion and carbonization processes might also form hydrocarbons. For this reason, home cooking, fuel use, garbage incineration, and transportation exhaust emissions also generated hydrocarbons. Motor vehicles were the main source of PAH.

Table 2 shows that O_3 also had a high loading level of 0.817 within the first factor. In addition to the suspended particulates that caused poor air quality in Taiwan, the main substance of pollutants was ozone. In many busy cities, the concentration of ozone is very high, and it sometimes forms the yellow-brown "photochemical smog" mentioned above. The reason is that when cars and scooters burn gasoline, they produce nitrogen oxides, such as nitrogen monoxide and nitrogen dioxide. The nitrogen dioxide reacts with the oxygen under the sun, and releases ozone and nitrogen monoxide. The ozone in turn interacts with the hydrocarbons released by cars and produces photochemical smog, in the photochemical reaction. The ozone is an important indicator of air pollution as it contributes significantly to air pollution.

Table 2 shows that NO_2 had a loading level of 0.747 within the first factor. The main sources of NO_2 were vehicle exhaust and combustion in heavy industry. The pollutant produced in this aspect was mainly NO, which was then quickly oxidized by the oxygen in the air to NO_2. NO_2 is also the important indicator of air pollution sources. However, in the absence of wind, NO tends to accumulate in the atmosphere and produces photochemical oxidation reactions; it was therefore one of the pollutants that formed the photochemical smog. In this study, NO_2 emissions from vehicles were a precursor of ozone formation, and the concentration of NO_2 in the atmosphere was generally low.

Table [2] shows that PM_{10} had a loading level of 0.659 within the first factor. PM_{10} was an important indicator in distinguishing the extent of air pollution. The suspended particulates in the atmosphere, especially those whose diameter is less than 10 μm, can penetrate deeply into human lungs and damage the respiratory system. The suspended particulates can also affect visibility due to the scattering of sunlight. The phenomenon that is called "haze" is especially noticeable when the relative humidity is high. If the concentrations of NO_2 and other irritating gases are too high, they will form photochemical smog with the suspended particulates, which constitutes an important mechanism of photochemical reaction. In addition, Kumar et. al. (2009) argues that secondary aerosols derived from the photochemical reaction are the main reason for high concentrations of $PM_{2.5}$ and PM_{10} in Taiwan. Consequently, the relationship between the emissions of pollutants in the atmosphere and the concentration of PM_{10} in the environment is no longer causal; rather, the chemical reaction involved is more complicated.

Based on the results of analyzing the four air pollutants within the first factor, the loading levels in descending order were THC, O_3, NO_2 and PM_{10}, among which O_3 and PM_{10} were the major pollutants that cause serious air pollution in Taiwan. Although THC is not currently included in the air quality standards, there is a high possibility that it would entail subsequent photochemical reactions when its concentration is high, which then results in a high probability of producing the secondary pollutant, O_3. The region under investigation in this study is located at the Formosa Freeway in Northern Taiwan, where there is a large amount of vehicles coming and going on a daily basis. When the car exhaust emissions contain high concentrations of organic matter, they are more likely to produce photochemical reactions at noon when the sunlight is more intense. As for NO_2, although its concentration in the air is not high and its contribution to air pollution is not as bad as that of O_3 and PM_{10}, it is still one of the precursors of O_3. The four air pollutants mentioned above are related to photochemical reactions, and the findings of the analysis are the same as the results of Liao (2004) et al.'s air quality analysis of a steel mill in Kaohsiung, Taiwan; in the analysis, they argued that O_3, NO_2, and PM_{10} belonged to the same factor, and were closely related to the photochemical reactions in the atmosphere. Therefore, this factor could be called the "photochemical reaction pollution factor."

2) The second factor

The second factor was mainly composed of the two air pollutants: CO and SO_2; the total variance could reach 22.965%, as shown in Table 1.

As shown in Table 2, CO had a loading level of 0.812 within the second factor. The carbon monoxide came mainly from vehicle exhaust. It is estimated that 86% of the carbon monoxide in the atmosphere is emitted by cars and scooters, and a small part comes from the incomplete combustion of fuels in factories and power plants. Therefore, it is mainly caused by the incomplete combustion of fuels containing carbon. In addition, carbon monoxide is also an important indicator of the air pollution source, and is the most abundant pollutant in the air. In the study, the concentration of CO indirectly generated by fuel combustion increased due to the acceleration and deceleration of vehicles entering and leaving the tunnels, especially when they were close to the ground surface.

As shown in Table 2, SO_2 had a loading level of 0.737 within the second factor. Sulfur dioxide is mainly generated by the combustion of fossil fuels containing sulfur, and partly by the combustion of fuels from cars and scooters. In this study, the exhaust gas emitted by high-speed vehicles was the main source of SO_2 because the monitoring sites were located on both sides of the Formosa Freeway (Wang,1999). In addition to being harmful to plants and humans, SO_2 also causes the problem of regional acid rain.

From the above analysis results, it can be seen that the main pollution source of CO and SO_2 is the combustion of vehicle fuels. These two kinds of pollutants conform less to the photochemical reaction mechanism within the first factor, and their subsequent reactions are not as complicated as those within the first factor. Therefore, the second factor can be called the "vehicle fuel factor."

3.2. Cluster analysis: the analysis of air pollution characteristics

For the cluster analysis, this study adopts a two-staged clustering algorithm to acquire approximate cluster results via hierarchical methods and then different cluster numbers are tested with the K-mean method. Finally, this study chose four clusters to distinguish the differences in air quality of the region under investigation in this study since they were the most suitable ones. The relationship between clusters and factors is shown in Figure 2. Moreover, the study restored the air pollutants in the four clusters to their original monitoring values, as shown in Table 3, so that the impact of

air quality on each tunnel entrance would be more clearly understood. The study also used the pollution standard index (PSI) announced by Taiwan's Environmental Protection Agency (see Table 4) to study the extent of pollution of various sub-indicator values in various clusters, in order to precisely determine the actual full extent of air pollution.

Table 3 The Average and Extreme Values of Air Pollutants in Clusters

Items \ Clusters	First cluster	Second cluster	Third cluster	Fourth cluster
THC(ppb) (Daily average)	3.54 2.27~4.91	2.58 1.18~5.16	3.58 1.81~5.11	2.56 1.48~4.04
O_3(ppb) (Maximum hourly value)	18.14 6~47	22.31 7~66	23.22 6~75	28.69 18~62
NO_2(ppb) (Maximum hourly value)	35.2 19~93	39.55 15~91	27.02 16~73	40.27 3~109
PM_{10}(μg/m^3) (24-hour average value)	90.12 33~**289**	81.19 27~**214**	123.56 34~**703**	68.10 20~115
CO(ppm) (Maximum 8-hour average value)	2.39 0.5~6.3	1.41 0.6~3.7	2.02 0.3~5.1	0.98 0.4~2.6
SO_2(ppb) (24-hour average value)	16.57 8~38	10.86 3~27	12.54 5~21	6.99 2~28
Pollution standard index (PSI) and health impacts	Good~ unhealthful	Good~ unhealthful	Good~ hazardous	Good~ moderate
Compliance with air quality standards	no	no	no	yes

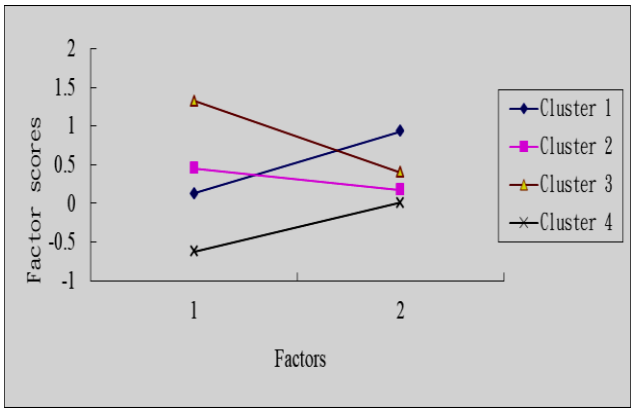


Figure 2 Relationship between clusters and factors

1) First cluster

Figure two shows that this cluster's factor score ranked the third highest within the first factor (the photochemical reaction pollution factor), and ranked the highest within the second factor (the vehicle fuel factor). The average concentrations of CO and SO_2 in this cluster were the highest among the four clusters, with CO concentrations ranging from 0.5 to 6.3 ppb and SO_2 from 8 to 38 ppb. However, in PSI in Table 4, the corresponding sub-PSI values were not high; they were therefore categorized as having a moderate impact on human health, without a negative impact on human health. In addition, this cluster ranked the third highest within the first factor. Except for three occasions where PM_{10} concentrations in PSI indicated a negative impact on human health (among which the highest concentration was 289 μg/m^3) and did not meet the air quality standards, none of the remaining air pollutants within the first factor showed any abnormally high concentrations. This cluster even showed a good impact on human health in terms of O_3 concentrations. Although the average concentration of THC in this cluster was the highest among others, it should not affect the immediate air quality since THC was not included in the assessment of the sub-indicators in Table 4. Based on the above analysis results, the cluster showed a relatively high concentration of pollutants within the fuel factor and a relatively low concentration of pollutants within the photochemical reaction pollution factor. Therefore, this cluster could be summarized as the "moderately polluted air cluster within the fuel factor."

2) Second cluster

Figure [2] shows that this cluster's factor score ranked the second highest within the first factor (the photochemical reaction pollution factor), and ranked the third highest within the second factor (the vehicle fuel factor). This cluster showed similarity to the first cluster in terms of PM_{10} concentrations within the first factor, which were high in several data, while the average concentration of O_3 was generally higher than that of the first cluster, and the same as that of the third cluster. As for the second factor, the average concentration showed no abnormality among the four clusters, and none of them had poor effects on human health, as shown in Table 4. The difference between this cluster and the first cluster was that the concentrations of two air pollutants in this cluster were generally lower within the second factor than within the first factor; there was little difference between this cluster and the third and fourth clusters in terms of concentration range. Therefore, compared with the first cluster, this cluster was less affected by the fuel factor pollution. Also, in terms of the performance of the concentrations of the four air pollutants within the first factor, except for several data on PM_{10} that showed higher concentrations, the remaining pollutants did not show obvious abnormalities. This cluster could be summarized as the "lightly polluted air cluster."

Table 4 The Comparison Table of Pollutant Concentrations and Pollution Sub-indicator Values

Pollutants	PM_{10}	SO_2	CO	O_3	NO_2
Statistical method	24-hour average value	24-hour average value	Maximum 8-hour average value within 24 hours	Maximum hourly value within 24 hours	Maximum hourly value within 24 hours
unit	μg/m^3	ppb	ppm	ppb	ppb
PSI value					
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

3) Third cluster

Figure [2] shows that this cluster's factor score ranked the highest within the first factor (the photochemical reaction pollution factor), and ranked the second highest within the second factor (the vehicle fuel factor). The reason that this cluster's factor score was significantly higher than those of the other three clusters was because of the relatively high concentration of PM_{10}. During the monitoring process, especially in December 2017 (winter time), Taiwan was affected by smog from China. The concentration of PM_{10} at each tunnel entrance was relatively high (up to 703 μg/m^3), reaching the level of having a hazardous effect on human health, as shown in Table 4. As mentioned above, PM_{10} belongs to the first factor; the concentrations of THC, O_3, and NO_2 simultaneously monitored on the same day did not show relative increases in the cluster. The reason was that the PM_{10} concentration would be affected by the photochemical reaction time when it was generated, and would begin the photochemical reaction when the sunlight was more intense. Therefore, the ozone concentration monitored on that day had a delayed effect, and the O_3 concentration was not high. Secondly, dust storms mainly came from the sand and dust in the desert, which contributed little to other pollutants in the atmosphere. This fact could be verified with the monitoring data on the third cluster in Table 2. In addition, the average concentration within the second factor in this cluster was lower than that in the first cluster. If considered from the perspective of a single-pollutant sub-indicator, the concentrations of CO and SO_2 in this cluster all reached the level of having a good impact on human health. From the above analysis, it can be generally con-

cluded that as CO and SO_2 in the atmosphere are affected by their own concentration characteristics, their concentrations have to be high enough to reach the level of having an impact on human health(Wang, 2009; Tobiszewski et al., 2010). However, the region under investigation in this study is located along the Formosa Freeway in Northern Taiwan, where there is no pollutant from stationary sources, only from mobile sources. Therefore, the CO and SO_2 concentrations produced are generally not high. This cluster could be summarized as the "severely photochemical polluted air cluster."

4) Fourth cluster

Figure [2] shows that this cluster's factor score ranked the lowest within the first factor (the photochemical reaction pollution factor), and ranked the lowest within the second factor (the vehicle fuel factor). The concentrations of various air pollutants within either the first factor or the second factor in this cluster were almost always the lowest among all clusters. Taking PM_{10} concentration as an example, the maximum concentration in this cluster was only 115 $\mu g/m^3$. If considered from the perspective of a single-pollutant sub-indicator in Table 4, its impact on human health was moderate, and less serious as that in the other three clusters. Taking the SO_2 within the second factor as an example, its impact on human health was also moderate, and not significant on the immediate air quality. If considered from the perspective of the distribution of monitoring times in this cluster, most of them were non-holiday times. Due to the small number of vehicles on the Freeway during this period, the pollution effect between the two factors was reduced, and the concentrations of air pollutants were relatively low. This cluster could be summarized as the "good air quality cluster."

3.3. Discriminant analysis

Through various tests of the discriminant parameter combinations, this study obtained a "discriminated cluster" based on the level classified by a previous cluster analysis as the "actual cluster," and identified which cluster a monitoring vehicle belonged to by calculating the probability of each monitoring vehicle that belonged to each group in accordance with the discriminant function coefficients. A cross-analytical comparison between the "discriminated cluster" and the monitoring vehicle values of the "actual cluster" obtained from a previous cluster analysis made the difference between them clear.

Table 5 shows the discriminant test results of four clusters via a discriminant analysis, as well as the coincidence rate (i.e. the percentage of correct discrimination) between the "discriminated cluster" obtained from discriminant function and the "actual cluster" obtained from a cluster analysis. Each of the clusters showed a high coincidence rate: the first cluster reached 96.78%, the second cluster 92.71%, the third cluster 93.55%, and the fourth cluster 93.75%. The average coincidence rate was up to 94.01%. This means that the four clusters could make accurate judgments, and the misjudgment rate was very low.

Table 5 The Verification Table of a Discriminant Analysis of Clusters

Discriminant cluster / Actual cluster	First cluster	Second cluster	Third cluster	Fourth cluster	Percentage of correct discrimination %
First cluster	60	1	1	0	(60/62)*100=96.78
Second cluster	1	89	2	4	(89/96)*100=92.71
Third cluster	1	3	58	0	(58/62)*100=93.55
Fourth cluster	0	3	1	60	(60/64)*100=93.75
Total	62	96	62	64	(267/284)*100=94.01

4. Conclusion

In this study, the air quality monitoring points set up at the existing 10 tunnels on the Formosa Freeway in Northern Taiwan were used to analyze the main factors that affect changes in concentrations of six air pollutants in the same region, as well as the air quality via a multivariate statistical method. Two main factors were obtained from the analysis: the "photochemical reaction pollution factor" and the "vehicle fuel factor." In the cluster analysis, the region was divided into the "moderately polluted air cluster within the fuel factor," the "lightly polluted air cluster," the "severely photochemical polluted air cluster," and the "good air quality cluster"; each cluster represented the extent and characteristics of air pollution in the region under investigation in this study. The reason that this study used the data on several air pollutants generated for a period of up to two years was mainly because when more samples were used in a factor analysis, the results were more stable, and the validity and reliability of the factor analysis could be effectively improved. Although the air pollutants emitted by the moving vehicle exhaust in the tunnel are not significant for the air quality in the neighboring areas, but it is expected that the results can serve as a reference for the management to prevent and control air pollution in the future.

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.

[2] Cheng, W., Chen, Y., Zhang, J., Lyons, T.J., Pai, J., and Chang, S. (1999). Comparison of revised air quality index with the PSI and AQI indices. Sci. Total Environ. 382, 191.

[3] Chiu, J. L., Kuo, S. L. and Ou, M. G. (2004). Multivariate Statistical Analysis Applied to Air Quality Characteristics of a Steel Plant in Kaohsiung, The Seventeenth Environmental Planning and Management Seminar, Taiwan.

[4] Einax, J.W., Truckenbrodt, D., and Kampe, O. (1998). River pollution data interpreted by means of chemometric methods. Microchem. J. 58, 315.

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

[6] Kumar, U., Prakash, A. and Jain, V.K. (2009). A Multivariate Time Series Approach to Study the Interdependence among O_3, NO_x and VOCs in Ambient Urban Atmosphere. Environmental Modeling and Assessment 14, 631.

[7] McKenna, Jr. J.E. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. Environmental Modeling & Softwar 18, 205.

[8] Murena, F. (2004). Measuring air quality over large urban areas: development and application of an air pollution index at the urban area of Naples. Atmos. Environ. 38, 6195.

[9] Tobiszewski, M., Tsakovski, S., Simeonov, V., and Namieśnik, J. (2010). Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. Chemosphere 80, 740.

[10] Vega, M., Pardo, R., Barrado, E., and Deban, L. (1998). Assessment of seasonal and polluting effects on the qualityof river water byex ploratory data analysis. Water Res. 32, 3581.

[11] Wang, C. W. (2009). Effects of automobile tailpipe ein the Hsuehshan tunnel of on the air quality of neighboring areas using ADMS model. The Master's Thesis of Environmental Engineering Research Institute, National Sun Yat-Sen University, Kaohsiung, R. O. C. (in Chinese).

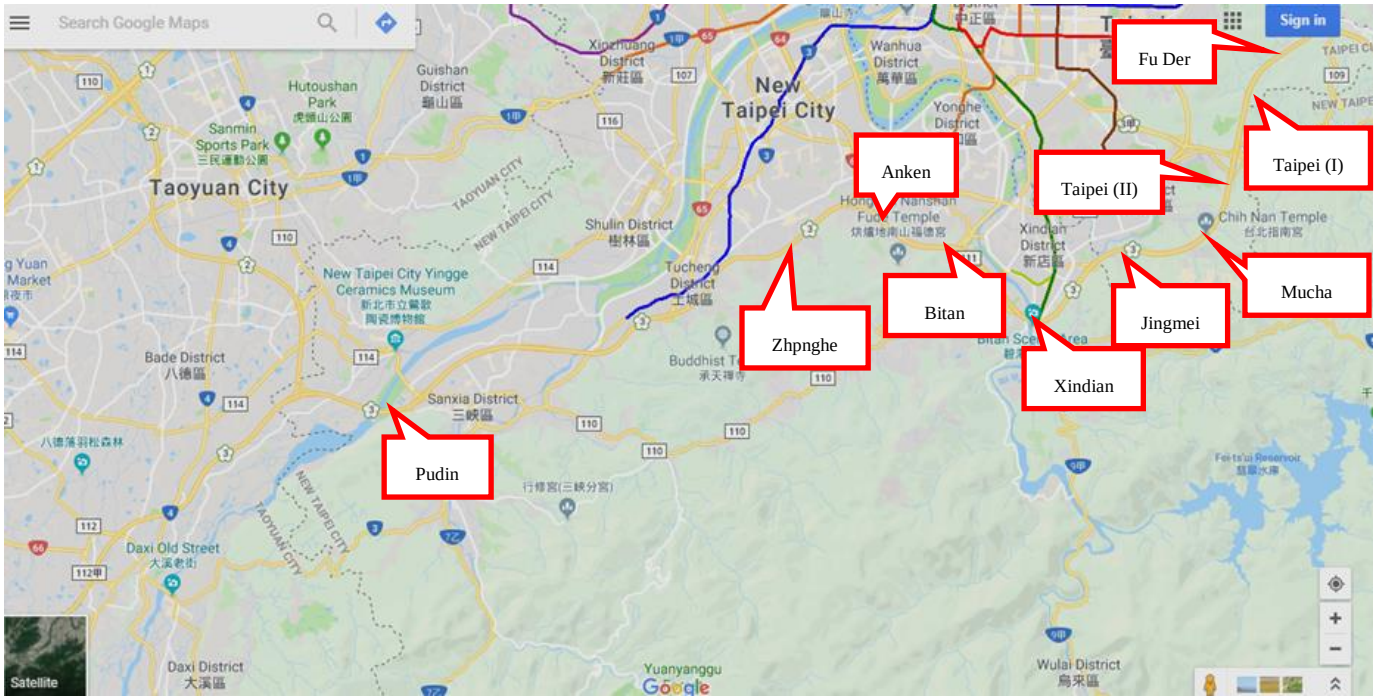


Figure 1 Air quality monitoring locations at tunnel entrances on Formosa Freeway in Northern Taiwan

Copyright © 2018 Shu-Lung Kuo, Edward Ming-Yang Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.