

Identification of Regional Air Pollution Characteristic and the Correlation with Public Health in Taiwan

Cheng F. Lee*, Jen H. Hsiao, Shin J. Cheng, and Huey H. Hsieh

Department of Environment and Resources Engineering, Diwan University, 87-1, Nanshih Li, Madou, Tainan 721, Taiwan, ROC

*Correspondence to Dr. Cheng F. Lee; Email: kvnlee@dwu.edu.tw

Received: 10 January 2007 / Accepted: 30 April 2007 / Published: 30 June 2007

Abstract: This study aims to classify regions with different air pollution characteristics into groups in Taiwan, and further to evaluate and compare the air quality of various groups. A selected multivariate analysis technique, cluster analysis, is applied to the pollution monitoring dataset which including PM₁₀, SO₂, NO₂, CO and O₃. The obtained results have proved that the regions with similar air pollution characteristic can be appropriately grouped by applying cluster analysis. All 22 regions are classified into six groups, and the pollution pattern for each group is characterized as: Group 1 (high SO₂/NO₂; low PM₁₀), Group 2 (high PM₁₀), Group 3 (high SO₂/PM₁₀), Group 4 (low SO₂/NO₂/CO; high O₃), Group 5 (low CO/NO₂; high O₃) and Group 6 (low PM₁₀/SO₂/NO₂/O₃/CO). Results from air quality evaluation indicate that the regions in group 6 (Ilan, Hualien and Taitung) have the best air quality while the regions in group 3 (Kaohsiung and Kaohsiung City) have the worst air quality in Taiwan. The results from correlation analysis reveal that incidence of the respiratory system disease is significantly positively correlated with pollution of NO₂ and CO at 99% confidence level.

Keywords: Air quality, air pollution characteristic, public health, cluster analysis; correlation analysis

Introduction

Due to the complexity and large variance of environmental sets, common statistic methods are not sufficient for assessment of pollution state [1]. Multivariate statistical methods for classification, modeling and interpretation of large datasets from environmental monitoring programs allow the reduction of the dimensionality of the data and the extraction of information [2]. The application of multivariate analysis methods such as cluster analysis, principal component analysis and factor analysis is therefore recommendable, and becoming popular in environmental studies dealing with measurements and monitoring [3]. With regard to application of cluster analysis in environmental studies, Chen et al. employed hierarchical cluster analysis (HCA) to classify the Cu, Ni, Pb, and Zn concentrations in soil samples which were collected from 30 urban parks in Beijing City [4]. Their results indicated that the location of the parks appears to affect the heavy metal concentrations in the soil samples greatly. Zhang used multivariate analyses and GIS to classify the chemical elements in urban

soils and to identify elements influenced by human activities in Ireland. Cluster analysis (CA) and principal component analysis (PCA) were applied to classify the elements into two groups: the first group predominantly derived from natural sources, the second being influenced by human activities [5]. Owega et al. identified long-range aerosol transport patterns to Toronto via classification of back trajectories by cluster analysis and neural network techniques [6]. They found that both techniques illustrate the cleaner nature of northerly and northwesterly transport patterns in comparison to southerly and southwesterly ones, as well as the effect of near stagnant air masses. Li and Shue applied data mining to uncover the hidden knowledge of air pollution distribution in the voluminous data retrieved from monitoring stations in TAQMN [7]. The cluster analysis in their study was used for data pattern identification. Simeonov et al. presented the application of different multivariate statistical approaches for the interpretation of a large and complex data matrix obtained during a monitoring program of surface waters in Northern Greece. In the study, CA was used for site similarity analysis [8]. Facchinelli et al. used PCA and CA to predict

potential non-point heavy metals sources in soil on the regional scale [9].

This study aims to identify the air pollution characteristic of all 22 cities (counties) in Taiwan and to evaluate the correlation between air pollution and public health. A selected multivariate analysis technique, cluster analysis, is applied to the pollution monitoring dataset which including PM₁₀, SO₂, NO₂, CO and O₃. The obtained results allowed to determine groups of cities (counties) with similar pollution characteristic, and to compare air quality among these groups. Moreover, correlation analysis is also used to analyze the correlation between air pollution and disease of the respiratory system.

Materials and Methods

Materials

Taiwan area is comprised by 7 cities and 15 counties as shown in Fig 1. The population and human activities are mainly centered in the west of Taiwan. The dataset which contains the yearly average concentration values of five selected air pollutants (PM₁₀, SO₂, NO₂, CO and O₃) in Taiwan’s 22 cities (counties) is quoted from “Air Quality Annual Report Taiwan Area in 2004” [10]. The report is based on the data of the Taiwan Air Quality Monitoring Network (TAQMN), which is operated by the Environmental Protection Administration, Taiwan.



Figure 1: Map of 22 cities (counties) in Taiwan.

Methods

Cluster Analysis

Cluster analysis is an exploratory data analysis technique for solving classification problems. This

technique comprises an unsupervised classification procedure that involves measuring either the distance or the similarity between objects to be clustered. The information obtained from the measured variables is used to reveal the natural clusters existing between the studied samples. Objects are grouped in clusters in terms of their similarity, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. The initial assumption is that the nearness of objects in the space defined by the variables reflects the similarity of their properties [2, 11, 12]. In our study, a hierarchical clustering by applying complete linkage method as the amalgamation rule and the squared Euclidean distance as metric were performed. Statistical calculations were performed by using the SPSS® software package, and the map of clustering results was produced using ArcView® software.

The similarities in this case were quantified through the squared Euclidean distance measurement. The distance between two objects (regional air pollution characteristics), *i* and *j*, is given as [3]:

$$d_{ij}^2 = \sum_{k=1}^m (Z_{ik} - Z_{jk})^2$$

Where,

d_{ij} : the Euclidean distance,

Z_{ik} : the standardized value of *X_{ik}*,

Z_{jk} : the standardized value of *X_{jk}*,

m : the number of pollutant kinds

Pollution Level

In this study prior to cluster analysis, the descriptor variables (concentrations of PM₁₀, SO₂, NO₂, CO and O₃) were standardized by means of z-scores to avoid any effects of units scale on the distance measurements by applying the equation [3]:

$$Z_{ik} = \frac{X_{ik} - \mu_k}{\sigma_k}$$

Where,

Z_{ik} : the standardized value of *X_{ik}*,

X_{ik} : the original value of measured parameter (the concentration of pollutant *k* in region *i*),

μ_k : the average value of pollutant *k*,

σ_k : the standard deviation of pollutant *k*.

The standardized value (*Z_{ik}*) of pollutant concentration in this case can be defined as “the pollution level of pollutant *k* in region *i*”. *Z_{ik}*>0 means that the pollution level is higher than the average value of all regions (*μ_k*), and the pollution state is relatively poor. On the contrary, *Z_{ik}*<0 means that the pollution level is comparatively low. *Z_{ik}*=0 means that the pollution state is at the average level.

Air Pollution Characteristic Analysis

In this study, the air pollution characteristic is constituted by the pollution levels of five air pollutants, i.e. PM₁₀, SO₂, NO₂, CO and O₃. The air pollution characteristic for region *i* (*C_i*) consequently can be represented as:

$$C_i = [Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5}]$$

Furthermore, the air pollution characteristic for group *t* is defined as:

$$C_t = \left[\frac{\sum_{i=1}^n Z_{i1}}{n}, \frac{\sum_{i=1}^n Z_{i2}}{n}, \frac{\sum_{i=1}^n Z_{i3}}{n}, \frac{\sum_{i=1}^n Z_{i4}}{n}, \frac{\sum_{i=1}^n Z_{i5}}{n} \right]$$

$$= [Z_{t1}, Z_{t2}, Z_{t3}, Z_{t4}, Z_{t5}]$$

Where,

- C_t*: the air pollution characteristic for group *t*,
- Z_{ik}*: the pollution level of pollutant *k* in region *i*,
- Z_{tk}*: the pollution level of pollutant *k* for group *t*,
- n*: the number of regions in group *t*.

Results and Discussion

Cluster Analysis Results

The dendrogram, Fig 2, reveals the results obtained from using hierarchical complete linkage clustering method and squared Euclidean distance as a criterion of similarity.

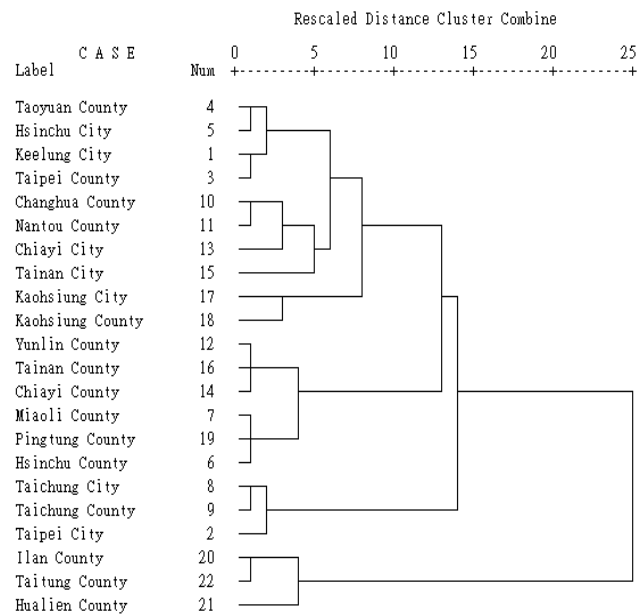


Figure 2: Dendrogram of the CA based on regional air pollution characteristic.

The all 22 regions in Taiwan can be classified into two major groups: from Taoyuan to Taipei City and from Ilan to Hualien as presented in Fig 2. Note that the second major group which is formed by Ilan, Taitung and Hualien is characterized by the biggest Euclidean distance to the other groups. This group corresponds to the cleanest area of Taiwan. The first major group is composed of the other 19 regions. The associations among regions in this group are quite complex, and these regions can be further classified into five subgroups. Thus, all 22 regions can totally be classified into six groups:

- Group 1: Taoyuan, Hsinchu City, Keelung and Taipei County.
- Group 2: Changhua, Nantou, Chiayi City and Tainan City.
- Group 3: Kaohsiung City and Kaohsiung County.
- Group 4: Yunlin, Tainan County, Chiayi County, Miaoli, Pingtung and Hsinchu County.
- Group 5: Taichung City, Taichung County and Taipei City.
- Group 6: Ilan, Taitung and Hualien.

The map of clustering results was produced using ArcView® software and is shown as Fig 3.

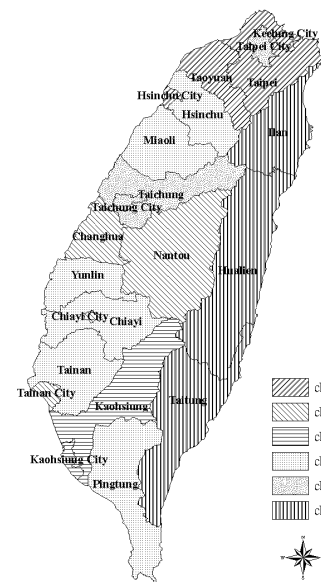


Figure 3: Map of the CA based on regional air pollution characteristic.

The four regions in Group 1 are all located at northern Taiwan, and are known as urbanized and populous regions. In Group 2, Changhua and Nantou are two important agricultural regions in middle Taiwan while Chiayi City and Tainan City are both urbanized regions where around two another primary agricultural counties (Tainan County and Chiayi County). Kaohsiung City and Kaohsiung County in Group 3 are famous for the heavy industries, and the former is also one of the primary cities in Taiwan. In Group 4, all six counties are agricultural regions in Taiwan. In Group 5, Taichung City and Taipei City are both primary

cities in Taiwan. Taichung County is adjacent to Taichung City. The three counties in the second major group (Group 6) are all located at eastern Taiwan where the economic activities are comparatively less.

Air Pollution Characteristics

The air pollution characteristics for all six groups are shown in Fig. 4, and are discussed as follows. The scale in the figure denotes the pollution level of pollutant *k* for group *t* (Z_{tk}).

Group 1

For group 1, the pollution levels of SO₂ and NO₂ are greater than 0 (i.e. the average level of all regions) while that of PM₁₀ is far lower than 0. Hence, this group may be characterized as high SO₂/NO₂ pollution and low PM₁₀ pollution.

Group 2

For group 2, the pollution levels of PM₁₀, NO₂, CO and O₃ are all located in the range of 0~1. The high PM₁₀ pollution state in this group is especially noticeable, and may be characterized as high PM₁₀ pollution.

Group 3

Fig. 4 indicates that the pollution levels of all five pollutants for group 3 are greater than 0, in which pollution states of SO₂ ($Z_{tk}=2.4$) and PM₁₀ ($Z_{tk}=1.2$) are the highest among all six groups. Note that Z_{tk} values of NO₂ and O₃ are also greater than most other groups. However, they are resulted from especially high NO₂ pollution state in Kaohsiung City ($Z_{tk}=1.4$) and high O₃ pollution state ($Z_{tk}=1.2$) in Kaohsiung County respectively. This group is characterized as high SO₂ and high PM₁₀ pollution.

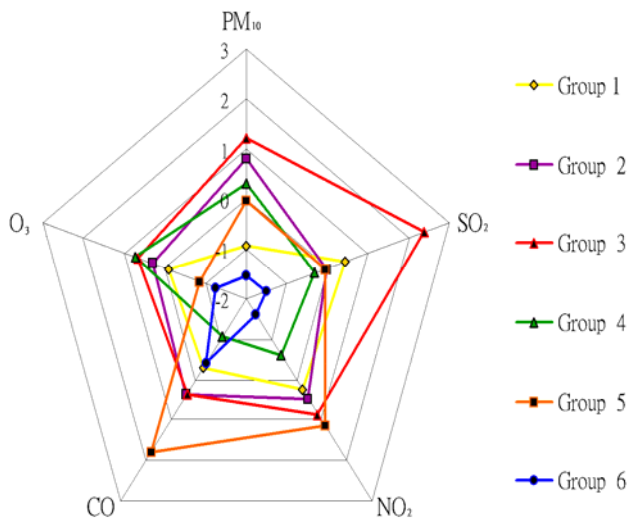


Figure 4: The air pollution characteristics of various groups.

Group 4

The pollution levels of SO₂, NO₂ and CO are under 0 while PM₁₀ and O₃ are both greater than 0. As presented in Fig. 4, the Z_{tk} value of CO is the smallest among six groups. This group may be characterized as low SO₂/NO₂/CO pollution and high O₃ pollution.

Group 5

For group 5, the pollution levels of CO ($Z_{tk}=1.8$) and NO₂ ($Z_{tk}=1.1$) are both the highest among six groups. The pollution level of O₃, however, is relatively low ($Z_{tk}=-0.8$). Hence, this group may be characterized as low CO/NO₂ pollution and high O₃ pollution.

Group 6

This group has the smallest Z_{tk} values of PM₁₀ (-1.5), SO₂ (-1.5), NO₂ (-1.6) and O₃ (-1.2). The Z_{tk} value of CO is also lower than most other groups except for group 4. The air pollution characteristic for this group may be located as low PM₁₀/SO₂/NO₂/O₃/CO.

Correlation Analysis Results

The Pearson correlation coefficient measures the strength of a linear relationship between two quantitative variables. Pearson’s correlation coefficients between air pollution and public health in Taiwan are depicted in Table 1. The results reveal that incidence of the respiratory system disease is significantly positively correlated with pollution of NO₂ and CO at 99% confidence level. While it shows only weak positive correlations with PM₁₀ and SO₂ pollution. Relevant studies by Guo [13, 14] also indicated that higher outdoor air pollution level, especially traffic-related pollutants, NO_x and CO, was associated with asthma prevalence in school children.

Table 1: Pearson’s correlation coefficients between air pollution and public health.

Items	PM ₁₀	SO ₂	NO ₂	CO	O ₃
Incidence of the respiratory system disease ^a	0.391	0.368	0.685*	0.566*	-0.100

* Correlation is significant at the 0.01 level.

^a Defined by the ratio of yearly outpatient number to population number in a region.

Conclusions

This study aims to classify regions with different air pollution characteristics into groups in Taiwan, and further to evaluate and compare the air quality of various groups. The obtained results have proved that the regions with similar air pollution characteristic can be appropriately grouped by applying cluster analysis. All 22 regions in

Taiwan are classified into two major groups. The first major group is formed by 19 regions, and these regions are all geographically located at western Taiwan. Due to the complex associations among regions in this group, the first major group may be further subdivided into five groups. The other three regions which are all located at eastern Taiwan form the second major group, and correspond to the cleanest area in Taiwan.

By calculating group mean pollution levels, air pollution states of all six groups are characterized individually. Group 1 is characterized as high SO₂/NO₂ pollution and low PM₁₀ pollution. Group 2 is high PM₁₀ pollution. Group 3 is high SO₂ and high PM₁₀ pollution. Group 4 is low SO₂/NO₂/CO pollution and high O₃ pollution. Group 5 is low CO/NO₂ pollution and high O₃ pollution. Group 6 is low PM₁₀/SO₂/NO₂/O₃/CO pollution. Besides, results of air quality evaluation find that group 6 (Ilan, Hualien and Taitung) has the best air quality while group 3 (Kaohsiung City and Kaohsiung County) has the worst air quality in Taiwan. The areas with better air quality are geographically distributed at the east of Taiwan, and those with worse air quality are located at the south part of Taiwan.

The results from correlation analysis reveal that incidence of the respiratory system disease is significantly positively correlated with pollution of NO₂ and CO at 99% confidence level.

For the sake of the similarities in this study were quantified through the squared Euclidean distance measurement, the results of air pollution characteristic analysis indicate that there still exist some differences of characteristic among regions in a same group. This may be improved through application of different clustering methods in the future work.

References

- Einax, J. W.; Zwanziger, H. W.; Geiss, S.: Chemometrics in Environmental Analysis. *Wiley, Weinheim*, **1997**.
- Massart, D. L.; Kaufman L.: The interpretation of analytical chemical data by the use of cluster analysis. *Wiley, New York*, **1983**.
- Kowalkowski, T.; Zbytniewski, R.; Szpejna, J.; Buszewski, B.: Application of chemometrics in river water classification. *Water Research*, **2006**, *40*(4), 744-752.
- Chen, T. B.; Zheng, Y. M.; Lei, M.; Huang, Z. C.; Wu, H. T.; Chen, H.; Fan, K. K.; Wu, K. Y.; Tian, Q. Z.: Assessment of heavy metal pollution in surface soils of urban parks in Beijing, China. *Chemosphere*, **2006**, *60*(4), 542-551.
- Zhang, C. S.: Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environmental Pollution*, **2006**, *142*, 501-511.
- Owega, S.; Khan, Badi-Uz-Zaman; Evans, G. J.; Jervis, R. E.; Fila, M.: Identification of long-range aerosol transport patterns to Toronto via classification of back trajectories by cluster analysis and neural network techniques. *Chemometrics and Intelligent Laboratory Systems*, **2006**, *83*(1), 26-33.
- Li, S. T.; Shue, L. Y.: Data mining to aid policy making in air pollution management. *Expert Systems with Applications*, **2004**, *27*(3), 331-340.
- Simeonov, V.; Stratis, J. A.; Samara, C.; Zachariadis, G.; Voutsas, D.; Anthemidis, A.; Sofoniou, M.; Kouimtzi, T.: Assessment of the surface water quality in northern Greece. *Water Research*, **2003**, *37*, 4119-4124.
- Facchinelli, A.; Sacchi, E.; Mallen, L.: Multivariate statistical and GIS-based approach to identify heavy metals sources in soils. *Environ. Pollut.*, **2001**, *114*, 313-324.
- EPA: Air Quality Annual Report Taiwan Area in 2004, *Environmental Protection Agency, Taipei, Taiwan*, **112**, **2005**.
- Einax, J. W.; Truckenbrodt, D.; Kampe, O.: River pollution data interpreted by means of chemometric methods. *Microchem. J.*, **1998**, *58*, 315-324.
- Frías, S.; Conde, J. E.; Rodríguez-Bencomo, J. J.; García-Montelongo, F.; Pérez-Trujillo, J. P.: Classification of commercial wines from the Canary Islands (Spain) by chemometric techniques using metallic contents. *Talanta*, **2003**, *59*, 335-344.
- Guo, Y. L.: Air Pollution and Childhood Asthma, *National Science Council, Taipei, Taiwan*, **2004**.
- Guo, Y. L.: The Relationship between Outdoor Air Pollution and Respiratory Diseases in Genetically Susceptible Population, *National Science Council, Taipei, Taiwan*, **2004**.