International Journal of Pure and Applied Mathematics
Volume 81 No. 3 2012, 505-515
ISSN: 1311-8080 (printed version)
url: http://www.ijpam.eu
ijpam.eu

CLUSTERING OF AIR QUALITY AND METEOROLOGICAL VARIABLES ASSOCIATED WITH HIGH GROUND OZONE CONCENTRATION IN THE INDUSTRIAL AREAS, AT THE EAST OF THAILAND

Kidakan Saithan[1], Jatupat Mekparyup[2 §]
[1,2]Department of Mathematics
Burapha University
169, Tambon Saensook, Amphur Muang, Chonburi, 20131, THAILAND

Abstract: Both Chonburi and Rayong provinces located on the eastern shore of the Gulf of Thailand. Their terrain is composed of coastline and numerous industrial factories therefore it plays an important role in air pollution events including ozone problem. Cluster analysis was applied to determine the spatial patterns of daily ground ozone maximum in two situations. Cluster of variables was firstly used to group various air quality variables and meteorological variables. If any of two or more variables were high association, they would be classified in the same group. Cluster of objects was secondly utilized to categorize the days of having high ground ozone concentration. The results demonstrated that there were four clusters of variables in the air, Cluster #1: air quality variables, Cluster #2: pressure, Cluster #3: temperature, sun radiation and wind speed and Cluster #4: rain, relative humidity and wind direction. For grouping of high ground ozone concentration days with similar physical characteristic, it exposed that there were three distinct clusters relating to the time

Received: September 11, 2012

[§]Correspondence author

of day: Cluster #1: days with dusk time, Cluster #2: days with during noon to afternoon and Cluster #3: days with evening interval.

AMS Subject Classification: 62H30
Key Words: cluster analysis, cluster of variables, cluster of objects, ground ozone

1. Introduction

Chonburi and Rayong are both of the most developed provinces and have high population density in Thailand. With the continuous economic development and the population expansion of both provinces due to industry located nearby complex coastline, many severe environmental pollution problems have attracted much attention in nowadays, e.g., air pollution, noise pollution, shortage of land resource, waste and sewage disposal, etc. Among of these problems, air pollution has a negative impact on the environment and human health through exposure to pollutants at high concentration levels existing outdoors, including ozone pollution.

Ground ozone presented in the troposphere in sufficient quantities is harmful to human. However, ozone is unique among air quality variables because it is not emitted directly into the air. It is a secondary air pollutant formed by chemical reaction and the formation process has been clarified in much of detail, for example: the reaction of sulfur dioxide (SO_2) or the apportionment of volatile organic compounds (VOCs), carbon monoxide (CO) and the oxides of nitrogen (NO_X) [1]. The variation in ozone concentration is also closely related to some meteorological variables such as temperature, wind direction, wind speed and sun radiation [2]. It is necessary, therefore, to understand the complicated relationships between ozone concentration and the several variables that cause or hinder ozone production.

The aim of this work was to classify of various variables in the air, both air quality and meteorological variables, and determine of the ozone patterns suggested "prototype day" for the industrial areas at the east of Thailand. The ozone patterns of interest are the spatial distributions of the daily maximum concentration because it is the daily maximum that determines whether or not standards are exceeded and what to degree. According to Air Quality Index (AQI) stipulated by the Thai Environmental Protection Department, the standard ozone level in ambient is 100 part per billion (ppb.) or 0.02 mg/m^3 [3].

The approach of classification uses cluster analysis regarding many variables. Cluster analysis searches for groups of points that are closed together in variable space. The variable space contains as many dimensions as there are variables under consideration. The distances between cases in variable space are defined in a fashion to the Pythagorean based methods of ordinary Euclidean space using the sums of squares of the differences in corresponding variable values. Many papers [4], [5], [6], [7], [8] used and applied cluster analysis in classification of meteorological variables and ozone as well as determining ozone and weather patterns. Few papers displayed classification of ozone with cluster analysis in Thailand such as [9] employed data mining technique to compare the data clustering models with daily ozone data as quantity variables.

The structure in this paper is as follows. In section 2, observational data used for analysis is given. In section 3, the statistical analysis is discussed and section 4 presents the results of classification of the observational data. Conclusion and discussion are made in section 5.

2. Observational Data

The analysis presented here is based on daily ground ozone maximum concentration and other various variables in the air, both air quality variables and meteorological variables, measured for the period 2006-2010 at two eastern monitoring stations. One site is at General Education Centre, Mueang District, Chonburi and the other is at Map Ta Phut Health Office, Mueang District, Rayong as shown in Fig. 1. Ten air quality variables were used as follows: concentration of carbon monoxide (CO), of nitrogen monoxide (NO), of nitrogen dioxide (NO_2), of oxide of nitrogen (NO_X), of sulfur dioxide (SO_2), of hydrocarbon (HC), of methane (CH_4), of non-methane hydrocarbons (NMHC), of particulate matter (PM_{10}) and of ozone (O_3). Seven meteorological variables were also used following: pressure, rain, relative humidity (RH), temperature (Temp), sun radiation (SR), wind direction (WD) and wind speed (WS).

3. Statistical Analysis

[10] were firstly proposed the term cluster analysis. The basic objective in cluster analysis is to discover data structures without a prior information on the data properties [11]. Many different algorithms and methods were used for combining variables (or objects) with similar properties into respective groups

Figure 1: Location of two eastern monitoring stations, Chonburi and Rayong (source: http://www.lonelyplanet.com/thailand)

in a way that the degree of association between two variables (or objects) is maximal if they belong to the same group and minimal otherwise.

There are basically two different algorithms of cluster analysis, hierarchical and nonhierarchical clustering [12]. Hierarchical technique moved by either a series of successive mergers or a series of successive divisions. Agglomerative hierarchical methods begin with the individual variables (or objects). Hence, there are initially as many as clusters as variables (or objects). The most similar variables (or objects) are firstly grouped and these initial groups are merged corresponding to their similarities. Finally, as the similarity decreases, all subgroups are blended into a single cluster. Divisive hierarchical methods perform in the opposite way. An initial single group of variables (or objects) is separated into two subgroups such that the variables (or objects) in one subgroup are far from the variables (or objects) in the other. These subgroups are then further separated into dissimilar subgroups. This process proceeds until there are as many subgroups as variables (or objects). A dendogram is the form of a two-dimensional diagram demonstrated the results of both agglomerative and divisive methods. The dendogram shows the mergers or divisions that have been made at successive levels. Nonhierarchical clustering technique, e.g. the K-means algorithm, need that the number of clusters is already known. This algorithm is then widely applied in cases where a priori information of the measurement is available.

In this paper, agglomerative hierarchical method is applicable since there is no a prior information on the number of the particular patterns in our data. Single linkage and squared Euclidean distance are then chosen for linkage methods and the distances between pairs of variables (or objects), respectively. The inputs to a single linkage algorithm are the distances between pairs of variables (or objects). Groups are formed from the individual entities by merging the smallest distance. Let the smallest distance in $\mathbf{D} = \{d_{ik}\}$ and merge the corresponding variables (or objects), say, U and V, to get cluster (UV). Then merge clusters U and V. The distance between (UV) and any other cluster W are calculated as Equation 1.

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\}, \quad (1)$$

where d_{UW} and d_{VW} are the distances between the smallest distance of clusters U and W and clusters V and W, respectively.

For the squared Euclidean distance is also computed as Equation 2.

$$d_{ik} = \sum_{j=1}^{p} (X_{ij} - X_{kj})^2, \quad (2)$$

where d_{ik} is the distance between variable (or object) X_{ij} and X_{kj}
$j$is number of variables (or objects); $j = 1, 2, \ldots, p$

4. Results

4.1. Classes of Various Variables in the Air

The agglomerative hierarchical methods showed amalgamation steps of variables with similarity and distance level as seen in Table 1. The dendogram shown in Fig. 2 also identified there are four typical classes of variables in the air. Cluster #1 referred to all ten air quality variables: CO, NO, NO_2, NO_X, SO_2, HC, CH_4, NMHC, PM_{10} O_3. In addition, all these variables could be particularly divided into 4 subgroups: (1) CO, O_3 and PM_{10}, (2) NO, NO_2 and NO_X, (3) HC, CH_4, and NMHC, and (4) SO_2. However, all seven meteorological variables were categorized in three following separated groups. Cluster #2 was only pressure. Cluster #3 mentioned three relevant variables: Temp, SR and WS. Cluster #4 was also relevant to three variables: Rain, RH and WD.

Step	Number of Clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	16	93.4324	0.131352	3	4	3	2
2	15	80.3972	0.392057	2	3	2	3
3	14	80.3961	0.392078	6	7	6	2
4	13	76.9046	0.461909	6	8	6	3
5	12	74.933	0.50134	14	15	14	2
6	11	72.7214	0.545571	1	10	1	2
7	10	72.234	0.55532	1	9	1	3
8	9	71.3742	0.572515	2	6	2	6
9	8	69.8913	0.602174	1	2	1	9
10	7	67.9451	0.641098	1	5	1	10
11	6	65.4471	0.691058	14	17	14	3
12	5	62.3031	0.753938	13	16	13	2
13	4	62.2334	0.755333	1	11	1	11
14	3	60.7338	0.785325	1	14	1	14
15	2	57.2944	0.854111	12	13	12	3
16	1	56.8791	0.862418	1	12	1	17

Table 1: Cluster of various variables in the air

4.2. Classes of High Ground Ozone Concentration Days

Important perspectives of the variations of ground ozone are discussed. The analysis was based on the 2,269 observations of ozone record for the period 2006-2010. It appeared obviously there was a seasonal variation of ozone concentration as shown in Fig. 3. That is, ozone concentration is highest in winter (mid October to mid February) then slightly decreases in summer (mid February to mid May) and finally drops to the lowest in rainy Season (mid May to mid October).

Only 33 days of exceeding ozone concentration (>100 ppb.) were analyzed in determination of the spatial patterns of daily ground ozone maximum. The time of all these days were during noon to dusk (12:00 pm. to 8:00 pm.). Also, most of these days (25 days) were in winter and the remaining were both in summer and rainy season. Cluster of objects were used to classify all these

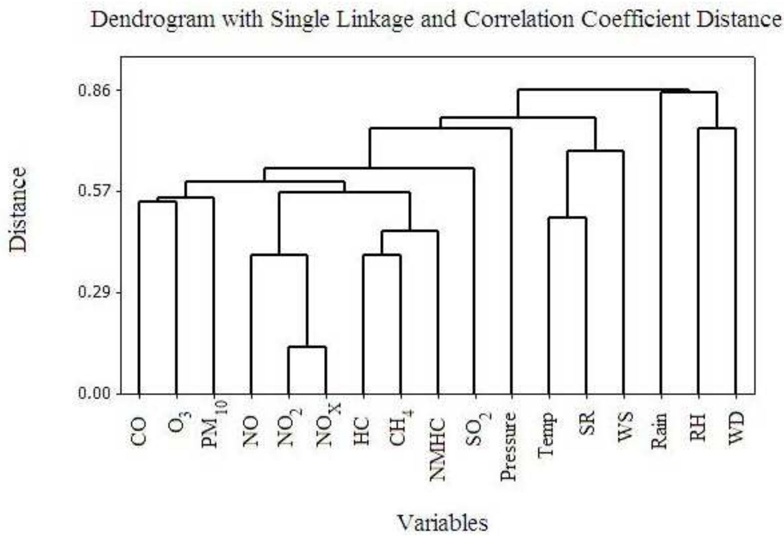

Figure 2: Dendogram of distances between various variables in the air

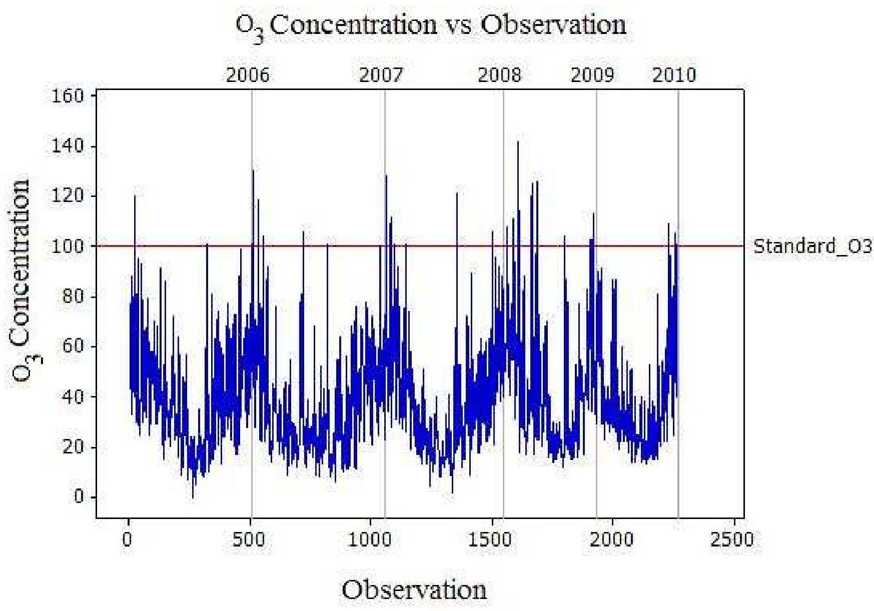

Figure 3: Ozone concentration for the period 2006-2010 at two eastern monitoring, Stations, Chonburi and Rayong

Step	Number of Clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	32	99.9319	189.2	4	10	4	2
2	31	99.8656	373.3	15	23	15	2
3	30	99.8470	425.0	12	19	12	2
4	29	99.8071	535.5	21	30	21	2
5	28	99.7904	581.9	4	14	4	3
6	27	99.7594	668.0	26	32	26	2
7	26	99.7301	749.5	15	16	15	3
8	25	99.6965	842.7	12	18	12	3
9	24	99.6307	1025.4	5	13	5	2
10	23	99.5839	1155.4	4	24	4	4
11	22	99.5767	1175.4	5	20	5	3
12	21	99.5074	1367.9	6	15	6	4
13	20	99.4814	1440.1	12	17	12	4
14	19	99.4324	1576.2	26	29	26	3
15	18	99.407	1646.6	4	5	4	7
16	17	99.3876	1700.6	21	31	21	3
17	16	99.3852	1707.2	25	26	25	4
18	15	99.3128	1908.4	4	11	4	8
19	14	99.2654	2039.9	4	22	4	9
20	13	99.2234	2156.5	3	12	3	5
21	12	98.9667	2869.2	3	6	3	9
22	11	98.6712	3689.9	3	4	3	18
23	10	98.4711	4245.6	21	25	21	7
24	9	98.2548	4846.0	2	9	2	2
25	8	97.9382	5725.3	7	8	7	2
26	7	97.3844	7263.0	2	3	2	20
27	6	96.6350	9344.0	27	33	27	2
28	5	96.2289	10471.6	2	28	2	21
29	4	96.1493	10692.8	2	27	2	23
30	3	95.3256	12980.0	1	2	1	24
31	2	94.2270	16030.6	1	21	1	31
32	1	77.2740	63105.7	1	7	1	33

Table 2: Cluster of high ground ozone concentration days

days. The amalgamation steps of grouping the days with similar characteristic as well as similarity and distance levels are listed in Table 2.

Final Partition
Number of clusters: 1

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	33	968967	153.207	330.306

Three distinct classes of days with high ozone concentration were grouped by predominant the time of day as displaying of dendogram in Fig. 4. Cluster #1 contained 24 days preferring time in the dusk of day (6:00 pm. to 8:00 pm.). Relative humidity was over 50% while wind direction was greater than 200 degreeM., and wind speed estimated 2.4 m/s. Cluster #2 carried 7 days revealing the time of day during noon to afternoon (12:00 pm. to 3:00 pm.). It presented wind direction was between 150 and 164 degreeM. with wind speed ranged from 2.6 to 4.5 m/s. Cluster #3 hold only 2 days indicating the evening of day (4:00 pm. to 5:00 pm.) with wind speed was similar to cluster #1. but relative humidity and wind direction were the same as cluster#2.

5. Conclusion and Discussion

Cluster analysis is able to classify of various variables in the air and determine the ozone patterns suggested prototype day. Cluster of variables still grouped ozone in the same type of other air quality variables, especially highly associated to CO and PM_{10}. However, ozone was definitely separated from the other meteorological variables. When the days of exceeding ozone were considered, the prototype days disclosed typical characteristics as follows: having of PM_{10} ranged from 22 to 58 mg/m^3, of pressure valued 760 mmHg and of sun radiation extremely varied from 48 to 393 w/ m^3. Cluster of objects subsequently employed to categorize the days, three distinct clusters of days associating to the time of day (dusk, noon to afternoon and evening) relied on three variables, relative humidity, wind direction and wind speed. Because of the location of Chonburi and Rayong, land breeze and sea breeze are influential and related to all these variables. In other words, relative humidity, wind direction and wind speed were the important indicators to identify on days when ozone standard was exceeded. Finally, the identification of a specific pattern associated with high ozone levels in the industrial areas, at the east of Thailand gives us a guide-

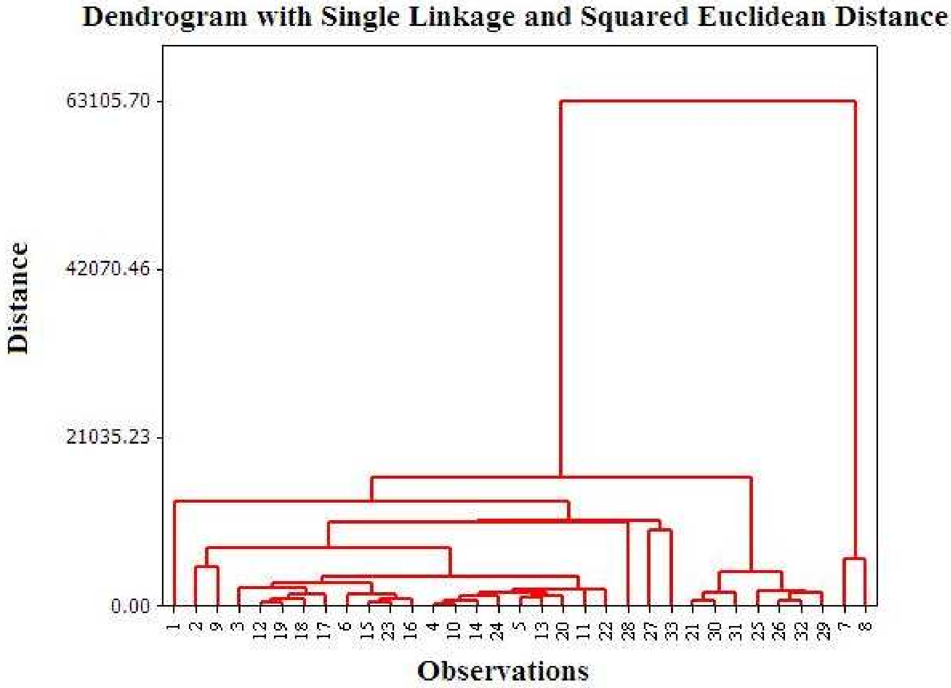

Figure 4: Dendogram of distances between high ground ozone concentration days

line to evaluate the effectiveness of control strategies somewhat independently from year to year incongruities of climate.

Acknowledgements

Authors wish to thank the Faculty of Science, Burapha University for their research grant. The authors are also grateful to Air Quality and Noise Management Bureau, Pollution Control Department for kindly providing all data.

References

[1] M. Trainer, D.D. Parrish, P.D. Goldan, J. Roberts, F.C. Fehsenfeld, Review of observation-based analysis of the regional factors influencing ozone concentrations, *Atmos. Environ.*, **34** (2000), 2045-2061.

[2] D. Klaus, A. Poth, M. Voss, Ozone distributions in Mexico City using principal component analysis and its relation to meteorological parameters, *Atmosfera*, **14** (2001), 171-188.

[3] Office of Natural Resources and Environmental Policy and Planning, Notice of the National Environment Committee No.28 (B.E. 2550) on Air Quality Standards. Retrieved October 25, 2011, from http://www.legalbase.ptit.org/Law.aspx?lid=3956

[4] F.L. Ludwig, J.Y. Jiang, J. Chen, Classification of ozone and weather patterns associated with high ozone concentrations in the San Francisco and Monterey Bay areas, *Atmospheric Environment*, **29(21)** (1995), 2915-2928.

[5] L.S. Darby, Cluster Analysis of Surface Winds in Houston, Texas, and the Impact of Wind Patterns on Ozone, *Journal of Applied Meteorology,* **44** (2005), 1788-1806.

[6] S. Beaver, A.N. Palazoglu, Cluster Analysis of Meteorological States to Understand the Weekday-Weekend Ozone Response in the San Francisco, CA, Bay Area, *Appl. Meteor. Climatol.*, **45** (2006), 1710–1726.

[7] J.C. Hains, *Ozone Sensitivity to NO_X Emissions*, a Dissertation, University of MD, USA (2007).

[8] O.A. Tarasova, C.A.M. Brenninkmeijer, P. Jöckel, A.M. Zvyagintsev, G.I. Kuznetsove, A climatology of surface ozone in the extra tropics: cluster analysis of observations and model results, *Atmos. Chem. Phys.,* **7** (2007), 6099-6117.

[9] V. Nuipian, P. Ammaruekarat, D. Thammasiri, N. Pho, P. Meesad, Ozone day Comparison Clustering with Data Mining Technical, *5-th edition, International Conference on Computing and Information* (2009).

[10] R.C. Tryon, *Cluster Analysis*, Edwards Brothers, MI, USA (1939).

[11] T. Hill, P. Lewicki, *STATISTICS: Methods and Applications*, StatSoft, OK, USA (2006).

[12] R.A. Johnson, D.W. Wichern, *Applied multivariate statistical analysis*, *5-th Edition*, Prentice-Hall Press, NJ, USA (1999).