AIR QUALITY ASSESSMENT IN THE URBAN AREAS WITH MULTIVARIATE STATITICAL ANALYSIS AT THE EAST OF THAILAND

Kidakan Saithanu[1], Jatupat Mekparyup[2][§]

Department of Mathematics
Burapha University
169, Tambon Saensook, Amphur Muang
Chonburi, 20131, THAILAND

Abstract: Two statistical multivariate analysis techniques had been applied to assess and predict whether the AQI in the urban areas at the east of Thailand exceeded the standard level. Cluster analysis was utilized to group the five main pollutants also classify the monitoring stations. To evaluate the air quality group, discriminant analysis was then built to determine the prediction model. The results present that there are three clusters of the main pollutants. Cluster #1 mentions SO_2. Cluster #2 contains NO_2, O_3 and PM_{10} while only CO is in the cluster #3. For classifying of monitoring stations, it exhibits that there are three distinct clusters: Cluster #1 consists five stations: General Education Office, Laem Chabang Municipal Stadium, Field Crop Research Center, Health Promotion Hospital Maptaput and Agricultural Office. Ta Sit Subdistrict Administrative and Sriracha Municipal Youth Center are the only one station in the cluster #2 and the cluster #3, respectively. A correct classification rate (CCR) was then measured to evaluate performance of the predictive model. The discriminant model performs very well with the average CCR 80.10% for the training data set and 78.97% for the validation data set.

[§]Correspondence author

1. Introduction

The Air Quality Index (AQI) is an indicator of air quality relying on the five main pollutants (SO_2, NO_2, CO, O_3 and PM_{10}) that have harmful effects on human health and environment. All of these five pollutants are gauged, monitored and then pointed the AQI. Each of these pollutants has an air quality standard which is used to calculate the overall AQI for the city. If any dominant pollutant provides the largest AQI, the overall AQI should be reported with this specific responsible pollutant. According to AQI specified by the Thai Environment Protection Department, AQI (in number) is illustrated and also broken down into 6 classes, each color coded with the number scale as follows: 0-50 (blue) representing good air, 51-100 (green) standing for moderate air, 101-200 (yellow) counting unhealthy air for sensitive group, 201-300 (orange) signifying very unhealthy air, 301-400 and 401-500 painted with red representing hazardous air [1], [2]. The standard level of AQI represented satisfactory air is 100.

Rayong and Chonburi are the industrial centers and urban coastal provinces at the east of Thailand so abundant critical air pollution problems have been extensively increasing nowadays. The Pollution Control Department of Thailand also built more air quality monitoring stations in 21 provinces all over Thailand including Rayong because a large number of patients with respiratory disease had been rapidly increased [3]. Furthermore, the annual concentration report of Chonburi [4] and the air quality report of Rayong [5] notified O_3 and PM_{10} concentrations were frequently over the standard level. Therefore, an accurate AQI alert prediction is essential to issue warnings to the public before the AQI reaches a dangerous level.

Some papers mentioned cluster of air pollutant parameters and meteorological parameters such as [6] used cluster analysis to represent various wind patterns and assess which patterns were most likely to be coincident with the maximum ozone of the day in Houston, Texas. At the east of Thailand, [7] applied cluster analysis to determined the spatial patterns of daily ground ozone maximum and [8] build the principal component regression model for predicting the daily maximum concentration of ground level ozone from air pollutant and meteorological factors by factor analysis method. Few researchers in Thailand studied AQI, for examples, [9] found association between digital number format and concentration of five main pollutants corresponding to AQI with multiple regression analysis, [10] brought into linear regression analysis to create an air quality modeling map from a relationship between AQI and dispersion of pollution in Chonburi. However, discriminant analysis has been rarely applied to

predict whether the AQI exceeds the standard level.

2. Observational Data

Four monitoring stations in Rayong (Ta Sit Subdistrict Administrative: 28T, Health Promotion Hospital Maptaput: 29T, Agricultural Office: 30T and Field Crop Research Center: 31T) and three monitoring stations in Chonburi (Laem Chabang Municipal Stadium: 32T, Sriracha Municipal Youth Center: 33T and General Education Office: 34T) were the representatives of the eastern urban area monitoring stations in Thailand. The observational data analyzed here was collected and measured basing on monthly five main pollutants for the period 2008-2012 [11]. The concentration of SO_2 ($\mu g/m^3$), NO_2 (ppb.) and O_3 (ppb.) were gauged in term of the maximum of average 1-hour value while the concentration of CO (ppm.) and PM_{10} ($\mu g/m^3$) were judged in term of the maximum of average 8-hour value and average 24-hour value, respectively.

3. Methodology

3.1. Calculation of AQI

The overall AQI for this study calculated by using the specific pollutant concentration could be indicated corresponding to Table 1 [2]. The AQI of each pollutant (I_p) calculated as of Equation 1 [1].

$$I_p = \left(\frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} \right) (C_p - BP_{Lo}) + I_{Lo}, \tag{1}$$

where I_p is the index for pollutant p, C_p is the rounded concentration of pollutant p, BP_{Hi} is the break down point that is greater than or equal to C_p, BP_{Lo} is the break down point that is less than or equal to C_p, I_{Hi} is the AQI value corresponding to BP_{Hi} and I_{Lo} is the AQI value corresponding to BP_{Lo}.

Since the AQI at the east of Thailand would never reached to 300 or above, there were only three groups of air quality to consider and classify in this study. If the AQI run from 0 to 50, the air quality was in the good air while the AQI ranged from 51 to 100 then it would be classified in the moderate air. Otherwise, it would be set in to the unhealthy air for sensitive people.

AQI	SO_2 ($\mu g/m^3$)	NO_2 (ppb.)	CO (ppm.)	O_3 (ppb.)	PM_{10} ($\mu g/m^3$)
0-50	0-65	0-85	0-4.48	0-51	0-40
51-100	66-300	86-170	4.49-9.00	52-100	41-120
101-200	301-800	171-600	9.01-14.84	101-203	121-350
201-300	801-1,600	601-1,202	14.85-29.69	204-405	351-420
301-400	1,601-2,100	1,203-1,594	29.70-40.17	406-509	421-500
401-500	2,101-2,620	1,595-1,993	40.18-50.21	510-611	501-600

Table 1: Concentration of five main pollutants equivalent to AQI

3.2. Cluster Analysis

To combine variables with similar attributes into respective groups in a way that the degree of association between two variables is maximal if they belong to the same group and minimal otherwise, cluster of variables in cluster analysis is employed. Since there was no a prior information on the number of the particular patterns in this study, the agglomerative hierarchical technique was applicable for both of clustering five main pollutants and seven monitoring stations. Groups of variables are figured from the individual entities by joining the smallest distance ($D = \{d_{ik}\}$) and merge the corresponding variables, say, U and V, to get cluster (UV). Then, cluster U and V are merged. The distance between (UV) and any other cluster W are computed as of Equation 2 [12].

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\}, \tag{2}$$

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

3.3. Discriminant Analysis

The objective of discriminant analysis is to obtain the combination of variables that best predicts the group to which a case belong. In the analysis, the group identification has to be known for each case. Two classification functions (linear or quadratic discriminant functions) are applied to categorize new cases whose group membership is not known. The difference of two discrimination methods is an assumption of equality of covariance matrices of independent variables for all groups. If the covariance matrices of independent variables for all groups are equal, an estimate of linear discriminant score is used. Otherwise, an estimate of quadratic discriminant score is substituted. The equations for calculation of linear and quadratic discriminant score are defined in [12].

Step	Number of Clusters	Similarity Level	Distance Level	Clusters Joined		New Cluster	Number of Obs. in New Cluster
1	4	78.1782	0.436436	2	4	2	2
2	3	68.8511	0.622977	2	5	2	3
3	2	67.9152	0.641695	1	2	1	4
4	1	66.0886	0.678227	1	3	1	5

Table 2: Steps for cluster of five main pollutants

The most technique often used in any classification particularly for measuring a performance of discriminant model is a correct classification rate (CCR) describing as of Equation 5 [13].

$$CCR = \frac{\sum_{k=0}^{C-1} CC_k}{n}, \tag{3}$$

where CC_k is the number of correctly classified observations and n is the number of observation in the group. The model with the large CCR is the one with satisfactory performance.

4. Results

4.1. Cluster Analysis

The agglomerative hierarchical technique presented amalgamation steps of main pollutants with similarity and distance level as seen in Table 2 corresponding to the dendogram in Figure 1. The dendogram of Figure 1 shows three typical classes of pollutants. Cluster #1 refers to SO_2 while three pollutants (NO_2, O_3 and PM_{10}) are in Cluster #2. Cluster #3 consists in only CO.

Table 3 displayed the amalgamation steps of grouping the monitoring stations with similar characteristic and distance levels. Three different classes of monitoring stations were grouped by the AQI value as shown in Figure 2. Cluster #1 contains 5 stations (29T, 30T, 31T, 32T and 34T) with various AQI values ranged from 32 to 162. Also, there are 86 observations of having AQI value beyond the standard level. Cluster #2 consists in 28T with moderate and high AQI value. There are 16 observations that the AQI values exceed 100. Cluster #3 is only 33T with low and moderate AQI value so most of observations have the AQI value run from 27 to 100.

Step	Number of Clusters	Similarity Level	Distance Level	Clusters Joined		New Cluster	Number of Obs. in New Cluster
1	5	93.8352	0.123295	5	6	5	2
2	5	91.3818	0.172364	5	7	5	3
3	4	88.9869	0.220261	3	5	3	4
4	3	88.6545	0.226910	1	3	1	5
5	2	85.1401	0.297198	1	4	1	6
6	1	77.3601	0.452799	1	2	1	7

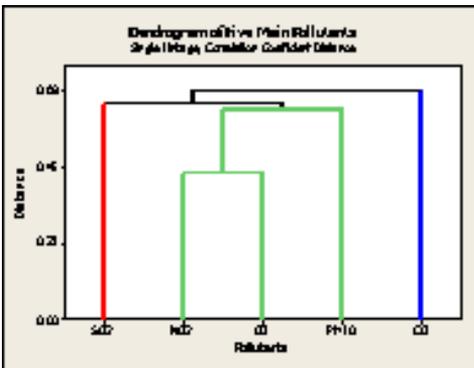Table 3: Steps for cluster of seven monitoring stations

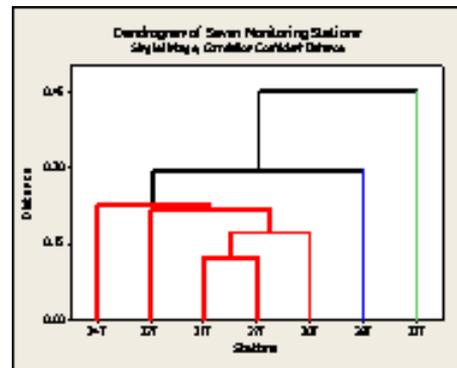Figure 1: Dendogram for cluster of five main pollutants

Figure 2: Dendogram for cluster of seven monitoring stations

4.2. Discriminant Predictive Model

Once the five main pollutant variables were assessed by the discriminant analysis, the linear discriminant classification function for each of the three air quality groups is determined as follows:

Good Air Quality:

$$\hat{Y}_0 = -30.749 + 0.110SO_2 - 0.002NO_2 + 2.791CO + 0.405O_3 + 0.084PM_{10}$$

Moderate Air Quality:

$$\hat{Y}_1 = -12.762 + 0.079SO_2 + 0.017NO_2 + 2.802CO + 0.230O_3 + 0.056PM_{10}$$

Unhealthy for Sensitive Air Quality:

$$\hat{Y}_2 = -5.906 + 0.081SO_2 + 0.019NO_2 + 2.943CO + 0.122O_3 + 0.022PM_{10}$$

Put into	Training Data Set			Validation Data Set		
	True Group			True Group		
Group	Good	Moderate	Unhealthy	Good	Moderate	Unhealthy
Good	112	22	0	112	26	0
Moderate	4	168	0	4	163	0
Unhealthy	0	55	47	0	56	47
Total of Obs.	116	245	47	116	245	47
CCR	0.966	0.686	1.000	0.966	0.665	1.000
Average of CCR	0.801			0.789		

Table 4: Classification result of air quality groups

The average CCR of the discriminant predictive model is demonstrated in Table 4. Both of training and validation data set clearly obtain the high average CCR approximated to 80%.

5. Conclusion and Discussion

Cluster analysis is capable to categorize of different five main pollutants and determine the AQI pattern of monitoring stations in the urban areas at the east of Thailand. Three main pollutants (NO_2, O_3 and PM_{10}) are grouped into the same cluster. It is in agreement with the results gained by Saithanu and Mekparyup [7], [14]. All of these three pollutants also often affect to the particular pollutant leading to indicate the AQI. Moreover, SO_2 and CO are completely different from each other also totally separate from the three pollutants previously mentioned so they are classified in to other distinct groups. For classifying monitoring stations, General Education Office, Laem Chabang Municipal Stadium, Field Crop Research Center, Health Promotion Hospital Maptaput and Agricultural Office are merged in cluster #1. It should be noted that all these stations are located in the industrial areas specially Laem Chabang Municipal Stadium and Health Promotion Hospital Maptaput which usually emit O_3 and PM_{10} into the air. Contradictory, Ta Sit Subdistrict Administrative and Sriracha Municipal Youth Center are separated into other clusters. Both of these stations are placed in the metropolitant area so the AQI should not be over the AQI standard. As of applying discriminant analysis, the predictive model gains a good performance especially it perfectly provides classification accuracy in classifying of unhealthy for sensitive air quality group for both of training and validation data set.

Acknowledgements

References

[1] United States Environment Protection Agency, *OAR Policy and Guidance Information*, Retrieved November, 26, 2011, from http://www.epa.gov /ttn/oarpg/t1/memoranda/rg701.pdf/

[2] Pollution Control Department, *Air Quality Index*, Retrieved November, 8, 2011, from http://www.pcd.go.th/info_serv/air_aqi.htm/

[3] Natural Resources and Environment, Thailand Development Research Institute, *Air pollution*, Retrieved November, 13, 2012, from http://www.thaienvimonitor.net/ Concept/priority5.htm/

[4] P. Khaenamkaew, P. Iamraksa, S. Raksawong, K. Wongsontam, C. Angwanisakul, S. Khuntong, Annual Concentration Report and Emission Sources Analysis of the Air Pollutants Measured by the AQM Station, Kasetsart University, Si Racha, Thailand, *Research Exhibition "Research in Kasetsart University 2011" in National Agricultural Fair* (2011).

[5] Thai Universities for Healthy Public Policies, 2011, Air quality around Map Ta Phut in Rayong, Thailand. Retrieved November 13, 2012, from http://www.ehwm.chula. ac.th/maptaphut/air-maptaphut.pdf/

[6] 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.

[7] K. Saithanu, J. Mekparyup, Clustering of Air Quality and Meteorological Variables Associated with High Ground Ozone Concentration in the Industrial Areas, at the East of Thailand, *International Journal of Pure and Applied Mathematics*, **3**, NO. 81 (2012), 505-515.

[8] S. Phansiri, K. Saithanu, Prediction of Daily Maximum Concentration of Ground Level Ozone with Principal Component Regression Model, *Burapha Sci. J.*, **18**, NO. 1 (2013), 144-152.

[9] A. Thanomsinrat, Remote sensing application for air pollution studies in Bangkok Metropolitan area, Mater's thesis, Technology of Information System Management, Mahidol University (2003).

[10] T. Intarat, Geoinformatics Application on Air Quality Assessment: A Case study in Chonburi Province, *Burapha Sci. J.*, **16**, NO. 1 (2011), 32-40.

[11] The Air Quality and Noise Management Bureau, Pollution Control Department,
Monthly report of Air Quality. Retrieved December, 11, 2012, from http://aqnis.pcd.go.th/data/monthly/

[12] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis, 6-th Edition*, Prentice-Hall Press, NJ, USA (2007).

[13] C. Oh, S. G. Ritchie, Recognizing vehicle classification information from blade sensor signature, *Pattern Recognition Letters*, **28**, No. 9 (2007), 1041-1049.

[14] K. Saithanu, J. Mekparyup, Assessment and Prediction of the Ground Level Ozone Concentration in the east of Thailand, *International Journal of Pure and Applied Mathematics*, **2**, NO. 84 (2013), 109-121.