O 29. EVALUATION OF THE AKSARAY PROVINCE AIR QUALITY: CONDITIONAL BIVARIATE PROBABILITY FUNCTION AND K-MEANS CLUSTERING

Ebru Koçak[1]*

[1]*Aksaray University, Faculty of Engineering, Department of Environmental Engineering, Aksaray Turkey*

E-mail: ebrukocak@aksaray.edu.tr

ABSTRACT: This study addresses the three major questions: (1) what are the emission sources of PM_{10} and SO_2 which are affecting the study area; (2) where do these emission sources come from; and (3) is there any temporal variation in the emission sources. In the current work K-means clustering techniques were applied directly to bivariate polar plots to identify and group similar features. The technique is analogous to clustering applied to back trajectories at the regional scale. When applied to data from a monitoring site with high source complexity it is shown that the technique is able to identify important clusters in ambient monitoring data. In Aksaray PM_{10} values follow a seasonal trend. The average PM_{10} concentration was recorded higher in the summer season and lower in the winter. It is observed that 50 μgm^{-3}, which is the 24-hour limit value of PM_{10}, was exceeded in both summer and winter months. The average SO_2 concentrations also was detected higher during the winter months due to domestic heating and there was a decrease in concentration in summer. The winter and summer SO_2 average concentrations were calculated as 7 and 2 μgm^{-3}, respectively. Looking at the SO_2 distribution over the months, it was seen that the normalized values are below 0.5 and the higher values were recorded in the period between November and February. Cluster analysis has been carried out for the PM_{10} and SO_2 surface for clusters between 2 and 10. The choice of the number of appropriate clusters is heuristic and is best determined by post-processing the data according to cluster. 5 and 4 clusters were considered for PM_{10} and SO_2, respectively. PM_{10} clusters were determined as 1 and 2- suburban emission, 3-traffic emission, 4-urban emission and 5-industrial emission. SO_2 clusters were identified as 1- suburban emission, 2- industrial emission, 3- urban emission and 4- mix of urban and suburban emission.

Keywords: PM_{10}, SO_2, temporal variation, CBPF, k-means clustering

1. INTRODUCTION

PM are solid and liquid particles suspended in the atmosphere. It is released into the atmosphere both by natural (volcanic eruptions, seismic activities and forest fires) and anthropogenic sources (all kinds of man-made combustion and some industrial processes). PM is one of the most important air pollutants that adversely affect human health. The chemical content and size of the particulate matter are important parameters that determine possible health effects. Particulate substances are divided into two according to their aerodynamic diameters; those whose aerodynamic diameters are less than 10 µm are classified as PM_{10} and those whose aerodynamic diameters are less than 2.5 µm are classified as $PM_{2.5}$. PM compositions vary as well as changes in size. This variation in the composition of particulate matter emerges as a result of the release of many different sources and the photochemical reactions observed in the atmosphere (Teixeira et al., 2012). According to the Air Quality Assessment and Management Regulation which came into force in 2008, PM_{10} limit values are determined as 50 $\mu g\ m^{-3}$ (not exceeding 35 in one year) and 40 $\mu g\ m^{-3}$ in 24 hours and annually, respectively.

Another important pollutant is SO_2 which is product of the combustion of sulfur compounds. Volcanoes and oceans are the main natural resources of SO_2 (Pereira et al., 2005; Carmichael et al., 2002; Garg et al., 2006; Reddy and Venkataraman, 2002). The anthropogenic sources of SO_2 are the burning of fossil fuels, especially coal and biomass, and the melting of sulfur-containing ores. SO_2 and its oxidation by-products are removed from the atmosphere by dry and wet precipitation. In addition to these transformation and removal processes, SO_2 can be transported over long distances, leading to global pollution. Many studies have been carried out to reduce SO_2 emissions. According to the Air Quality Assessment and Management Regulation which came into force in 2008, hourly, 24 hour and annual SO_2 emission limit values are 350 $\mu g\ m^{-3}$ (cannot be exceeded more than 24 in 1 year), 125 $\mu g\ m^{-3}$ (3 in

1 year) and 20 mg m^{-3}, respectively.
In this study, PM_{10} and SO_2 values of Aksaray were investigated. Aksaray is in a position to provide access to important cities and has borders to Konya, Ankara, Nevşehir and Niğde. In addition, geographical features also has a special place in Central Anatolia such as Turkey's second largest lake, Salt Lake and Hasan Mountain, Melendiz Mountains and Ekecik Mountain which are the old volcanic mountains.

2. MATERIAL AND METHOD

2.1. Sampling location

Air quality monitoring station operated by the Ministry of Environment and Urban Development were used in the study. Aksaray has a station representing the city center. The urban and the industrial zone, which is considered as the center of Aksaray, are almost side by side. In the NE of the station, there is an industrial zone, in the east there is city center, in the south and SW regions there are partly residential areas, a bus terminal and a university campus.
PM_{10}, SO_2, wind speed and wind direction data have been taken from the National Air Quality Monitoring Stations website of the Ministry of Environment and Urbanization includes hourly data between March 1, 2017 and March 1, 2018. In order to make comparisons, data of border provinces (Ankara, Konya, Niğde and Nevşehir) were also used.

2.2. Conditional bivariate probability function

The ordinary conditional probability function (CPF) estimates the probability of the measured concentration exceeding a specified threshold criterion for a given wind sector (Ashbaugh et al., 1985).

$$KOF = {}^{m_{\Delta\theta}}/_{n_{\Delta\theta}} \quad (1)$$

Here, $\Delta\theta$ represents each wind sector, $n_{\Delta\theta}$ represents the entire number of hourly winds blowing from the wind sector and $m_{\Delta\theta}$ represents the number of hourly winds blowing from the wind sector by exceeding the specified threshold concentration.
The conditional bivariate probability function (CBPF) combines ordinary CPF and wind speed data into a third variable. The pollutant concentration is allocated to cells defined by wind direction and wind speed ranges rather than only wind direction sectors.

$$CBPF = {}^{m_{\Delta\theta,\Delta u}}/_{n_{\Delta\theta,\Delta u}} \quad (2)$$

Here, $m_{\Delta\theta}$ represents the number of hourly winds blowing in the wind speed range from the wind sector through the specified threshold concentration. $n_{\Delta Q,\Delta u}$ indicates the number of winds per hour in the wind direction-speed range. The two-variable function provides more information about the nature of sources because different types of sources may have different wind speed dependencies. The use of a third variable can therefore provide more information about the type of source. The third variable may be another variable which will provide a radial approach such as temperature independent of the wind speed. Bivariate polar plots show how a pollutant concentration changes in polar coordinates along with wind speed and wind direction. Polar drawings are an effective way to see how a pollutant is dispersed and to see the effects of specific pollutant sources. Wind direction is a method that can be used to decompose different emission sources with wind speed (Uria-Tellaetxe and Carslaw, 2014).

2.3. K-means clustering

K-means clustering is a method in which two-variable polar plot properties can be defined and grouped. The main purpose of this grouping of data is to aggregate the data in the original time series data to enable post-processing to better understand potential source characteristics. At the center of the clustering idea of data is the concept of distance, that is, similarities or differences between points. The similarities of the concentrations presented in the CBPF plots are defined by three variables: u and v components of wind and pollutant concentration. Basic k-means algorithm for k-clusters is obtained by minimizing;

$$\sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2 \quad (3)$$

$\|x_i - \mu_k\|^2$is selected distance, average of cluster- μ_k c_k
The distance measure is defined as the Euclidean distance.

$$d_{x,y} = \sqrt{\sum_{j=1}^{J}(x_j - y_j)^2} \quad (4)$$

x, and y are two J-dimensional vectors, standardized by subtracting the mean and dividing by the standard deviation. In the operations performed, J defines the length of the wind components u, v and concentration C, each is standardized. Standardization is necessary because u and v are located at different scales relative to C.

3. RESULTS AND DISCUSSION

3.1. Temporal variations of PM_{10} and SO_2

Aksaray PM_{10} values follow a seasonal trend. Higher values were recorded in summer and lower values were recorded in winter. This is due to atmospheric scavenging in the winter and the removal of particulate matter and its reflection to the sampling. In the summer, because of stable weather conditions and suspended particulate matter, it is more collected in the sampling compared to winter months (Gramsch et al., 2006). Thus, median values of summer months were recorded as 57 μgm^{-3}, while winter median values were recorded as 37 μgm^{-3}. This effect can also be explained by the wind speed. Wind speed was high between 12:00 and 18:00, while low between 20:00 and 06:00. As shown in Figure 1, a decrease in PM_{10} concentrations is observed along with faster wind speeds, while there is an increasing trend in PM_{10} concentrations at low wind speeds. Looking at the distribution of weekdays and weekends, the expected trend is seen and Sat.-Sun. decline continues on an increasing line during the week.
High values for SO_2 concentrations were recorded in winter, while low values were recorded in summer. Winter and summer SO_2 median values of Aksaray province were calculated as 7 and 2 μgm^{-3}, respectively. When SO_2 emission sources are evaluated, this result is expected especially with the domestic heating factor. Looking at the distribution of SO_2 over the months, it is seen that the normalized values are below 0.5 between May-October and high values are recorded until February. Looking at the distribution of weekdays and weekends, there is a trend similar to PM_{10} and decreasing values of the Sat.-Sun. increased with the beginning of the week.
When hourly PM_{10} and SO_2 changes are examined, a decrease is observed after midnight. PM_{10} values increase during the hours of morning traffic and in the afternoon traffic decreases with decreasing values. PM_{10} and SO_2 concentrations are increased with the inversion of the atmosphere after 18:00 and the increase in emissions from domestic heating in winter. This increase in values decreases over time as the inversion loses its effect in the following hours and the emissions caused by domestic heating decrease.
The 24-hour PM_{10} limit value of 50 μgm^{-3}, which is determined in the regulation, is exceeded in both summer and winter average values (Figure 2). While 46% of data exceeded the limit value in winter months, this value increases to 64% in summer months.

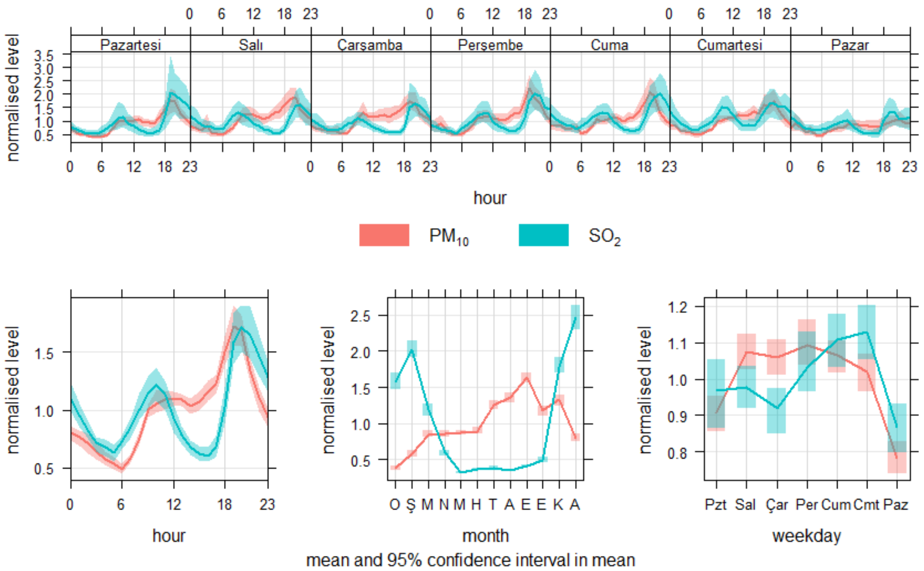

Figure 1. Temporal variations of PM_{10} and SO_2

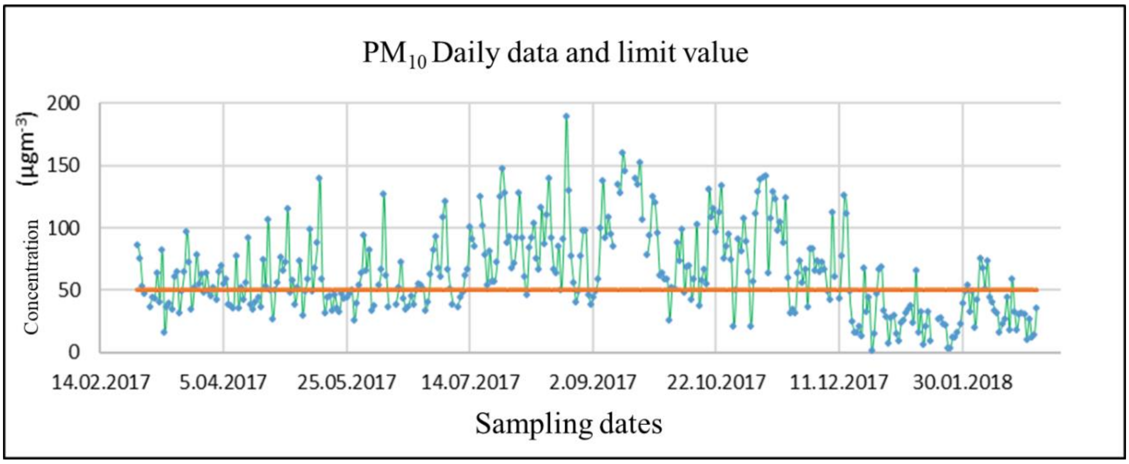

Figure 2. Aksaray-PM_{10} daily data and 24-hour PM_{10} limit value specified in the regulation

3.2. Conditional bivariate probability function results

2017-2018 PM_{10} and SO_2 values and pollution percentage plots and CBPF plots are shown in Figures 3 and 4. It is observed that PM_{10} concentrations of 150 µgm^{-3} and above are mainly caused by pollutant emissions from east side of the station and PM_{10} concentrations in the 100-150 µgm^{-3} range are caused by pollutant emissions coming from south side. Especially when looking at the eastern part of the station, it is seen that there is Ankara-Adana road and urbanization. Therefore, it can be concluded that there is a PM_{10} transport to the station in this direction with high wind speeds. When the percentage pollution plot is considered, it is seen that most of the PM_{10} is carried by eastern winds. When the SO_2 CBPF plot is examined, it is seen that SO_2 transport (>9 µgm^{-3}) is from the pollutant sources in the west direction with low wind speeds. The low wind speed and the SO_2 transport at these concentrations can be explained by the fact that the emission sources are very close and the emission is high. It could be concluded that SO_2 emissions from the organized industrial zone near the station contribute to this result. It is also seen that 5-7 µgm^{-3} SO_2 transport in the NNW and SSE direction is at high wind speeds.

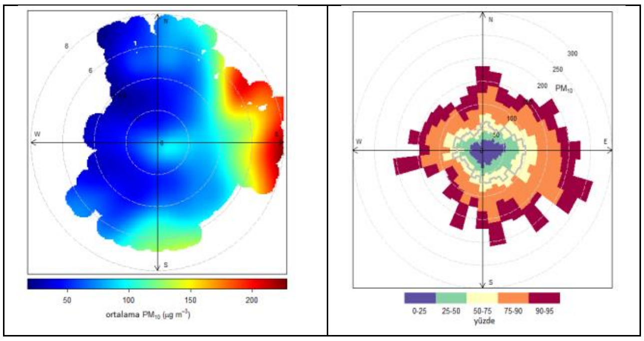

Figure 3. PM_{10} CBPF and percentage pollution plot

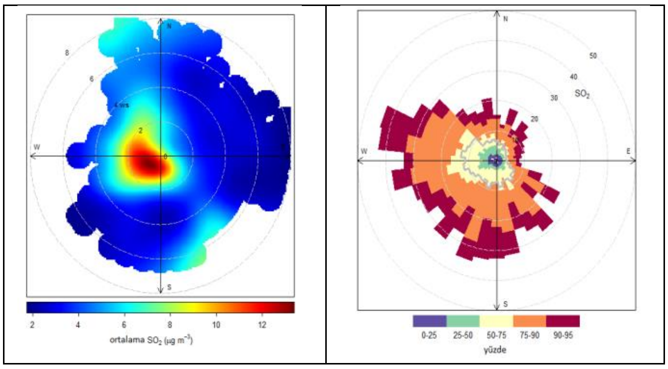

Figure 4. SO_2 CBPF and percentage pollution plot

3.3. K-means clustering results

In this study, the clustering study was conducted for PM_{10} and SO_2 pollutants with one year data. To determine the appropriate number of clusters, different clusters were used in the range of 2-10. Different methods to determine the optimum number of clusters have been presented in the literature (Everitt et al., 2011). However, these methods do not give effective results when applied to CBPF plots. At this point, it is best to select the appropriate number of clusters by selecting the appropriate number of clusters by post-processing the data with different clusters and evaluating the possible emission sources. There are two important clusters to be detected for PM_{10} in the east and south directions. Figure 5 shows the clusters between 2-10 and k-means clustering plots prepared for PM_{10}. Looking at the figure showing the result of 5 clusters, it is seen that the east and south clusters are splitting and this feature does not change as the number of clusters increases. When the results of the 5 clusters were examined, it was concluded that clusters 1 and 2 originated from the background emissions of the city, cluster 3- traffic emissions, cluster 4-city center emissions and cluster 5- industrial emission.

To determine the optimal number of clusters for SO_2, attention should be paid to the south, south-west and north-west directions, which carry highly concentrated pollutants, especially at low wind speeds. When the clusters numbers 2-10 and k-means clustering plots prepared for SO_2 are examined in Figure 6, it is seen that the number of clusters 2 and 3 could not separate the sources effectively. With the 4 clusters, it is seen that both the emission source belonging to the low wind speed and the emission source in the direction of SSE are separated. There is no change in the cluster structure formed in the low-speed region from the number of clusters 5 and more. In the number of clusters 9 and 10, it is seen that this region is also divided into two. Since the aim of the clustering study was to keep the number of clusters at the optimum number, the cluster number for SO_2 was chosen as 4 for this study. When we look at the distribution for the number of clusters selected, cluster 1 represents the city's background emission profile. It was concluded that cluster 2 represents the emission profile arising from the industrial zone close to the station, cluster 3 points to the city center emission profile, and lastly cluster number 4 represents the emission profile caused by atmospheric transport from the city and also from the surrounding provinces (especially Niğde).

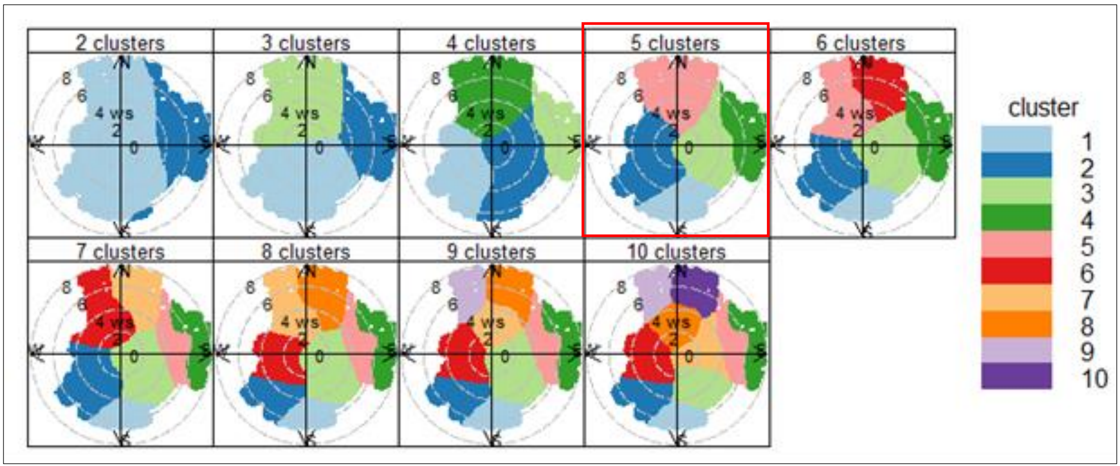

Figure 5. Clusters identified at Aksaray for PM_{10} concentrations for 2 to 10 clusters

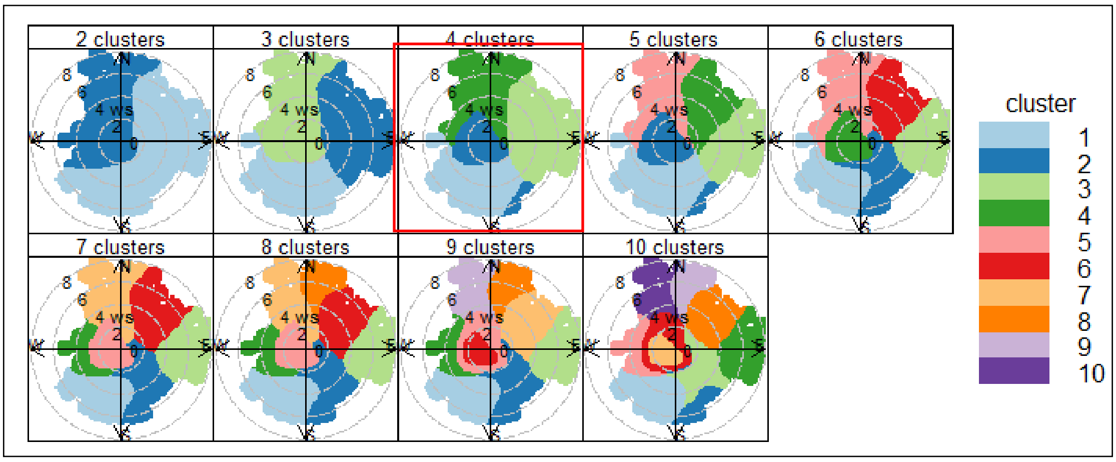

Figure 6. Clusters identified at Aksaray for SO_2 concentrations for 2 to 10 clusters

3.4. Correlation results of PM_{10} data with neighboring provinces

Due to its geographical location, Aksaray province is open to particulate pollutants caused by local-anthropogenic pollutants and regional transport. The 24-hour limit, 50 µgm^{-3}, was frequently exceeded. As the correlation factor R^2 value increases, it could be said that PM_{10} emission is regional. In the light of the findings, the strongest correlation was found with Nevşehir and the lowest correlation was with Ankara. In this case, the contribution of PM_{10} emission with atmospheric transport from Nevşehir can be mentioned. Considering that only the correlation factor remains at a maximum of 0.37, it is seen that the PM_{10} emission source of Aksaray Province is local pollutant.

REFERENCES

Ashbaugh, L. L., Malm, W. C., & Sadeh, W. Z. 1985. A residence time probability analysis of sulfur concentrations at Grand Canyon National Park. *Atmospheric Environment* (1967), 19(8), 1263-1270.

Carmichael, G.R., Streets, D.G., Calori, G., Amann, M., Jacobson, M.Z., Hansen, J., Ueda, H., 2002. Changing trends in sulfur emissions in Asia: implications for acid deposition, air pollution and climate. *Environmental Science & Technology* 36 (22), 4707–4713

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. 2011. Cluster analysis: Wiley series in probability and statistics.

Garg, A., Shukla, P.R., Kapshe, M., 2006. The sectoral trends of multigas emissions inventory of India. *Atmospheric Environment* 40 (24), 4608–4620

Gramsch, E., Cereceda-Balic, F., Oyola, P., & Von Baer, D. 2006. Examination of pollution trends in Santiago de Chile with cluster analysis of PM10 and ozone data. *Atmospheric Environment*, 40(28), 5464-5475

Reddy, M.S., Venkataraman, C., 2002. Inventory of aerosol and sulphur dioxide emissions from India: I-Fossil fuel combustion. *Atmospheric Environment* 36 (4), 677–697

Teixeira, Elba Calesso, Agudelo-Castaneda, Dayana M,, Fachel, Jandyra Maria, Guimaraes, Leal, Karen Alam, de Oliveira Garcia, Karine, Wiegand, Flavio, 2012. “Source identification and seasonal variation of polycyclic aromatic hydrocarbons associated with atmospheric fine and coarse particles in the Metropolitan Area of Porto Alegre, RS, Braz”, *Atmospheric Research*, 118, 390-403

Uria-Tellaetxe, I., & Carslaw, D. C. 2014. Conditional bivariate probability function for source identification. *Environmental Modelling & Software*, 59, 1-9.