World Academy of Science, Engineering and Technology
International Journal of Environmental and Ecological Engineering
Vol:11, No:06, 2017

Discovering the Effects of Meteorological Variables on the Air Quality of Bogota, Colombia, by Data Mining Techniques

Authors : Fabiana Franceschi, Martha Cobo, Manuel Figueredo

Abstract : Bogotá, the capital of Colombia, is its largest city and one of the most polluted in Latin America due to the fast economic growth over the last ten years. Bogotá has been affected by high pollution events which led to the high concentration of PM10 and NO2, exceeding the local 24-hour legal limits (100 and 150 □g/m3 each). The most important pollutants in the city are PM10 and PM2.5 (which are associated with respiratory and cardiovascular problems) and it is known that their concentrations in the atmosphere depend on the local meteorological factors. Therefore, it is necessary to establish a relationship between the meteorological variables and the concentrations of the atmospheric pollutants such as PM10, PM2.5, CO, SO2, NO2 and O3. This study aims to determine the interrelations between meteorological variables and air pollutants in Bogotá, using data mining techniques. Data from 13 monitoring stations were collected from the Bogotá Air Quality Monitoring Network within the period 2010-2015. The Principal Component Analysis (PCA) algorithm was applied to obtain primary relations between all the parameters, and afterwards, the K-means clustering technique was implemented to corroborate those relations found previously and to find patterns in the data. PCA was also used on a per shift basis (morning, afternoon, night and early morning) to validate possible variation of the previous trends and a per year basis to verify that the identified trends have remained throughout the study time. Results demonstrated that wind speed, wind direction, temperature, and NO2 are the most influencing factors on PM10 concentrations. Furthermore, it was confirmed that high humidity episodes increased PM2,5 levels. It was also found that there are direct proportional relationships between O3 levels and wind speed and radiation, while there is an inverse relationship between O3 levels and humidity. Concentrations of SO2 increases with the presence of PM10 and decreases with the wind speed and wind direction. They proved as well that there is a decreasing trend of pollutant concentrations over the last five years. Also, in rainy periods (March-June and September-December) some trends regarding precipitations were stronger. Results obtained with K-means demonstrated that it was possible to find patterns on the data, and they also showed similar conditions and data distribution among Carvajal, Tunal and Puente Aranda stations, and also between Parque Simon Bolivar and las Ferias. It was verified that the aforementioned trends prevailed during the study period by applying the same technique per year. It was concluded that PCA algorithm is useful to establish preliminary relationships among variables, and K-means clustering to find patterns in the data and understanding its distribution. The discovery of patterns in the data allows using these clusters as an input to an Artificial Neural Network prediction model.

Keywords : air pollution, air quality modelling, data mining, particulate matter
Conference Title : ICAPMST 2017 : International Conference on Air Pollution Management System and Technology
Conference Location : Toronto, Canada
Conference Dates : June 15-16, 2017