Open Access

Pollution control 2020: Investigation of emissions sources and characterization at Mamelodi Township, Gauteng, South Africa using conditional probability function modelling

Shonisani Norman Singo,

University of Witwatersrand, South Africa Email: shonisingo19@gmail.com

Abstract

This paper investigates pollution sources affecting Mamelodi Township within the City of Tshwane in Gauteng province, South Africa. Gauteng province has also the largest population in South Africa. The ambient pollution concentration at the vicinity of Mamelodi depends upon their output of gases from various activities emanates from biogenic and anthropogenic. Anthropogenic activities are the main made sources of emissions such as domestic fuel burning, industrial activities and transport emissions.It is therefore an objective of this study to assess and determine significant sources of emission which are affecting the Mamelodi Ambient Monitoring Station by investigating ambient concentration correlation parameters, pollution roses and probability functions modelling. Investigations will be focus on the following pollutants Sulphur dioxide, Nitrogen dioxide, Ozone and Particulate Matter of less than ten micro diameter. In this study k-means clustering techniques has been applied to bivariate polar plot to identify and group similar features.

Methodology and results: The following methodology will be used in the analyzing of correlation of pollutants at different receptors which are affecting Mamelodi ambient monitoring station. The study uses pollution rose polar coordinates plots to provide a useful graphical technique which provides directional information on sources. Ambient concentration, wind direction together with wind speed can be highly effective at discriminating different emission sources affecting the ambient station. Data collection dated first January 2012 to 30 April 2018. The data processed and imported in the open-air model are presented in comma separated value (csv) format which is extension files of Microsoft Excel. The study has divulge major and minor sources at the received environment. The results has display strong positive correlation of oxides of nitrogen and Nitrogen dioxide. Nitrogen dioxide and nitric oxide sources has display positive moderate correlation of ambient concentrations. The problematic areas emission sources were discovered for ozone and particulate matter less than ten micro diameter per cubic meter on the North west of Ambient monitoring station. Nitrogen dioxide and Sulphur dioxide has reveled local and far sources affecting the ambient monitoring station. The study has thoroughly investigate bivariate polar plot using K-means clustering to discriminate of similarity or dissimilarity between the features which can be identified and grouped.

This is partially presented at Pollution control 2020

This Work is presenting at 8th Global Summit and Expo on Pollution Control On August 24-25, 2020 Webinar

Introduction:

Identifying activities and emissions contributing to the ambient air pollution is vital in order to implement proper reduction measures to the source. Ambient air quality is defined as the physical and chemical measure

quality is reported to have deteriorated seriously, especially in urban areas, exposing populations to pollutant levels above the recommended limits (van der Leun et al.,1998). of pollutant concentrations in the atmosphere to which the general population is exposed. Concentrations are vary by wind speed and wind direction plot as continuous surface and surfaces are calculated through modelling using smoothing techniques. In most developing countries, ambient air

1.2 Study Area

Mamelodi township is located in the City of Tshwane in Gauteng Province, South Africa. Gauteng Province is not only the wealthiest province in South Africa (CDP, 2014) it is also its economic hub, accounting for 36 per cent of the country's Gross Domestic Product (GDP) and 10% of the African continent's GDP. It is important to understand potential activities at the vicinity of the study area before investigating ambient sources in Mamelodi Township. Identification of current sources is important factors in identifying unknown source that may affect the ambient monitoring station. The activities that has higher probability to affect Mamelodi ambient monitoring station which are located within 5,89 kilometers of the Mamelodi ambient station and there are various sources of emissions. There are three different facilities located on the west of south west of the ambient monitoring stations which are storing and handling of petroleum product such as diesel, petroleum and alumina paraffin. The facilities are located at less than 2,6 kilometers which are Engen Petroleum (Pty) Ltd, Chevron (Pty)Ltd, BP Southern Africa (Pty) Ltd and Total South Africa (Pty) Ltd. Transnet pipelines is for storage and handling of petroleum products located on the south of south west of the ambient monitoring station at 1,92 kilometers away. Above mentioned fuel storage facilities has permanent immobile liquid storage facilities at a same site with a combined storage tanks of greater than 1000 cubic meters. The atmospheric emissions expected from these sites are the vapours from the stored petroleum products and will largely be made up of various volatile organic compounds. There is Afrisam quarry on the north west of the ambient monitoring station located at 5,82 kilometres away and emission associated with quarry are particulate matters and total suspended particulates. On the south west of the ambient monitoring station there is facility called Silverton Engineering located at 3,79 kilometers away from Mamelodi Ambient Monitoring Station. Metallic objects are coated at Silverton Engineering in a layer of zinc by dipping them in molten zinc either in a continuous process or in batches. Pre-treatment of the articles ensure that the surface is in an optimum state to accept the coating involves a series of chemical baths. The first step is to clean the oil on the surface of the article from production which is knows as caustic dip, followed by rinsing to remove the caustic solution. Pickling in acid follows to remove any oxide scale on the surface, followed by fluxing to increase the bonding ability of the zinc and the metal. The articles are then dipped into the molten zinc bath and the final step in the process is quenching. These processes will emit metal fumes and hydrogen chloride and any acid related pollutant depending on the type of acid which requires extraction system to make the workplace safe. The Gauteng province also experiences a high-energy demand and motorization due a combination of factors

i.e. rapid population growth, industrialization and relatively high living standard. Gauteng province has the largest amount of vehicles in South Africa from eNatis statistics in the year 2018. The petroleum fuel used by motor vehicles is made up of a mixture of aromatic hydrocarbons and paraffins which ignites at high temperatures in the air. The incomplete combustion of the petroleum in the engine of the vehicle results in exhaust emissions of carbon dioxide, carbon monoxide, volatile organic compounds ozone precursor gases, and particulate matter. Diesel fuel vehicle generally are known of less carbon dioxide which results on high particulate matter at the exhaust. The concentration of air pollution in a given environment depends on both the amount of pollution produced and the rate at which pollutants disperse. Wind direction together with wind speed can be highly effective at discriminating different emission sources.

Table 1: Facilities around Mamelodi Ambient Monitoring

		Wind Speed	**10 m/s**	**8 m/s**	**6 m/s**	**4 m/s**	**2 m/s**
	Kilometers	**Meters**	**Time (min)**	**Time (min)**	**Time (min)**	**Time (min)**	**Time (min)**
Radius	5,89	5890	10	12	16	25	48
Afrisam	5,82371	5823,71	10	12	16	24	49
Silverton Engineering	3,79102	3791,02	6	8	11	16	32
Transnet Pipelines (Pty)Ltd	1,92485	1924,85	3	4	5	8	16
Total South Africa (Pty) Ltd	2,47336	2473,36	4	5	7	10	21
BP Southern Africa (Pty) Ltd	2,55461	2554,61	4	5	7	11	21
Chevron (Pty) Ltd	2,34871	2348,71	4	5	7	10	20
Engen Petroleum (Pty) Ltd	2,20203	2202,03	4	5	6	9	18

2. Literature Review

The atmosphere is the earth's largest single shared resource, which protects and supports life through the absorption of dangerous ultraviolet solar radiation also its warms the surface and regulates temperature. However, these vital roles are under serious threat due to human-driven activities that result in the introduction of pollutants in the atmosphere (Hunter et al., 2002). These activities or drivers include industrialization, urban growth, population growth and changing consumption patterns. Significant sectors contributing to atmospheric degradation are transport, power generation, incineration, waste and biomass burning. This study will outline specific sources of emissions which causes high concentration of pollutant to the ambient air quality stations.

On a global stage, in 2014, 92 per cent of the world's population lived in locations where the WHO air quality guidelines were not adhered to. In 2013, there was an estimated 3 million premature deaths worldwide caused by air pollution in urban and rural areas (WHO, 2013). By reducing levels of air pollution, countries can reduce the burden of disease caused by stroke, heart disease, lung cancer and chronic or acute respiratory diseases such as asthma. It is worth noting that ambient air quality standards of the WHO (SO_2= 20µg/m^3 on average for a 24 hours period and PM_{10}=20µg/m^3 on average per year) is stricter compared to the South African standards (SO_2= 125µg/m^3 on average for a 24 hours period and PM_{10}=40µg/m^3 on average per year).legislative direction since the promulgation of the National Environment Management Air Quality Act 39 of 2004. Attention was drawn towards national and international legislation and how it affected the decisions of air quality management mandatory functions within the spheres of government, particularly in the last decade. The national framework for air quality management in South Africa was promulgated in 2012. Regulations regarding air dispersion modelling with series of mathematical simulations were promulgated in 2014.

Open air modelling has been developed by Insightful Corporation under the free Software Foundation's for General Public License. Open-air model as a package should be downloaded first in R software to ensure the availability of the package. The package can be downloaded from http://www.openair-project.org. Once the software is downloaded then the open-air model package is ready to be activated in R software by typing "library (open-air)". The tool is culpable to analyze big data in a form of computer programming which predicts temporary and spatial various pollutants accurately, not even Microsoft Excel. The R project is statistical computation language which improves air quality data analyses. Then air quality data to be analyzed can be inputted from computer files or imported from a monitoring station (Carslaw and Ropkins,2012). The open-air model is a tool for statistically analytically analyzing semi-empirical mathematical relationships between air pollution concentration and other factors that may affect it (Colls and Tiwary,2017). The application of open-air model for air quality monitoring analyses is still rare in Gauteng Province, however some research abroad has used open-air for analysis.

The data variable are measured on each Gauteng ambient monitoring station. It is interesting to establish relationship between two variable. There are three categories of correlation and are positive correlation means other variable has tendency to increase, negative correlation means other variable has tendency to decrease and no correlation means other variable does not tend to either increase or decrease. The correlation is an effect size and describe the strength of the correlation using the guide that Evans (1996) suggests for the absolute value of r.

a. 0 – 19% very weak
b. 20-39% weak
c. 40-59% moderate
d. 60-79% Strong
e. 80 -100% very strong

Pollutant roses are diagrams that show how air pollution depends on wind direction. In this study pollution rose is used in the investigations relationship directional sources for Sulphur dioxide, Nitrogen dioxide,

Ozone and Particulate Matter of less than ten micro diameter are measured from Mamelodi ambient monitoring. Pollution roses give valuable information on the location of the most important pollutant sources that affect the monitoring site. Pollution roses are used to calculate the average concentrations of a pollutant at a receptor from multiple emission sources over a period as a function of wind direction (McNaugh and Wilkinson, 2012). Location of emission sources and ambient monitoring stations as per Figure 1, will be helpful to determine area of concern. The pollution rose will be able to indicate concentration level and direction of origin.

Clustering is a division of data into groups of similar objects. Each group is called cluster which is consists of objects that are similar amongst themselves and dissimilar compare to object of other groups. Clusters are defined as areas of higher density within the data space compared to other regions. The drawback with these methods is that they expect some kind of density guide or parameters to detect cluster borders. Bivariate polar plot features can be identified and group using K-means clustering approach.

3. Methodology

The following methodology will be used in the analyzing of correlation of pollutants at different receptors which are affecting Mamelodi ambient monitoring station. The study uses pollution rose polar coordinates plots to provide a useful graphical technique which provides directional information on sources. Ambient concentration, wind direction together with wind speed can be highly effective at discriminating different emission sources affecting the ambient station. Data collection dated first January 2012 to 30 April 2018. The data processed and imported in the open-air model are presented in comma separated value (csv) format which is extension files of Microsoft Excel. The study has thoroughly investigate using K-means clustering to discriminate of similarity or dissimilarity between the features which can be identified and grouped.

Method 1: Investigation of correlation coefficient, r.

The (Pearson) correlation coefficient is a measure of the strength of the linear relationship between two variables. If there is perfect linear relationship with positive slope between the two variables, $r = 1$. If there is a perfect linear relationship with negative slope between the two variables $r = -1$. A correlation coefficient of zero (0) means that there is no linear relationship between the variables.

Method 2: Investigation of pollution roses

Pollution rose is defined as variation of wind rose that is useful for considering pollutant concentrations by wind direction and percentage time the concentration in a particular range.

The variation of wind speed and wind direction on different time is operated using the wind rose function. The output demonstrate concentration levels, percentage time the concentration in particular range, wind direction mean and per cent of calm. At the selected ambient monitoring station Particulate matter less than ten diameter micro gram per cubic meter ($\mu g/m^3$), Sulphur dioxide (ppb), Nitrogen dioxide (ppb) and Ozone (ppb).

Pollution roses are polar diagrams that display air pollution depends on wind direction. If an ambient air quality monitoring station is markedly influenced by a source of the pollutant measured, the pollution rose shows a peak towards the local source. The construction of a pollution rose is simple when both wind direction and pollutant concentrations are averages over one hour and data used on this study data was at hourly averages for except particulate matter of less than ten micro diameter per cubic meter which was on daily average. These pollution roses were computed according to:

where *Cdd* is the average concentration for wind sector *dd*, *n* is the number of days in the period for which the rose is constructed, *pj* is the measured concentration on hourly or day *j*, *fdd*, j is the frequency that the wind came from sector *dd* and αj is some weight function based on the persistency of the wind vector during day j (Thiessen, L. and Y. Lenelle, 1991).

$$\sum_{=1,} \quad , = \sum_{=1,} \quad , \qquad 3$$

Where $\sum_{=1,} \quad ,$ the sum of frequencies of the wind direction over the *ndd* different bins during day j, i usually returns values for the concentrations that, for successive wind direction bins, oscillate between large positive values and large negative values.

Method 3: Investigation of pollution sources using the bivariate polar plot at receptor.

Concentration of a species varies to specific degree at the certain area, influenced by wind speed and wind direction in a polar coordinates. Wind direction together with wind speed can be highly effective at discriminating different emission sources. By using polar coordinates the plots provide a useful graphical technique which can provide directional information on sources. Bivariate polar plots are constructed in the following way. First, wind speed, wind direction and concentration data are partitioned into wind speed-direction ‘bins’ and the mean concentration calculated for each bin.

Binning the data in this way is not strictly necessary but acts as an effective data reduction technique without effecting fidelity of the plot itself. Furthermore, because of the inherent wind direction variability in the atmosphere, data from several weeks, months or years typically used to construct a bivariate polar plot tends to be diffuse and does not vary abruptly with either wind direction or speed. The wind components, u and v are calculated With u is the mean hourly wind speed and ū is the mean wind direction in degrees with 90 degrees as being from the east. The calculations above provides a u, v, concentration (C) surface. While it would be possible to work with this surface data directly a better approach is to model the surface to describe the concentration as a function of the wind components u and v to extract real source features rather than noise. A flexible

framework for fitting a surface is to use a Generalized Additive Model (GAM) e.g. (Wood, 2006). The GAM can be expressed as shown in Equation 7:

Finally conditional bivariate probability functions polar plots showing how species concentration vary jointly with wind speed and wind direction in polar coordinates will be formulated for all activities listed or not impacting ambient air quality in order to identify pollutants that cause exceedance.

Method 4: Identification of similar characteristics sources using k-means clustering method. K-means clustering is one method in which bivariate polar plot features can be identified and grouped. The main purpose of grouping data in this way is to identify records in the original time series data by cluster to enable post-processing to better understand potential source characteristics. Figure 2 display k-means flowchart of k-means clustering below.

*Flowchart of k-means clustering*Central to the idea of clustering data is the concept of distance i.e. some measure of similarity or dissimilarity between points. Clusters should be comprised of points separated by small distances relative to the distance between the clusters. The similarity of concentrations shown in Figure 2 is determined by three variables: the u and v wind components and the concentration, c. All three variables are equally important in characterising the concentration location information, but they exist on different scales i.e. a wind speed-direction measure and a concentration. Let X = fxig; i = 1; n be a set of n points to be clustered into K clusters, C =fck; k = 1;Kg. The basic k-means algorithm for K clusters is obtained by minimising:

Where $\| - \|^2$ is a chosen distance measure, is the mean of cluster ck. The distance measure is defined as the Euclidean distance:

CONCLUSION:

Investigations of sources and characterization using correlation ambient station parameters at the station helps in knowing pollution configuration behaviour at the monitoring station. Pollution rose was used to determine wind direction and concentration of the selected pollutant. The wind rose will identify wind direction with concentration thereof. Concentrations are vary by wind speed and wind direction plot as continuous surface and surfaces are calculated through modelling using smoothing techniques. This paper will employ Openair for air pollution analysis package based on the programming statistical package R. The package has many tools for manipulation of data Openair package has valuable and function tool that be employed successfully in supporting the air quality management systemThe maximum and mean ambient concentrations for Sulphur dioxide at Mamelodi ambient monitoring station was 435.58 and 5.2454 parts per billion respectively. The results has also display weak pollution roses in order from south, south east lower and south west lower and north. At the ambient concentrations of 0 to 2 less than ten micro gram diameter concentrations per cubic meter reveals strong pollution rose from south west higher and north east higher less than four per cent.