Modeling Emissions Sources and Characterization in a Dense Low-Income Township Near Industrial Zones and Mine Dumps (Case of Tsakane, South Africa).

Shonisani Singo
University of the Witwatersrand

Jean Mulopo (jean.mulopo@wits.ac.za)
Wits University: University of the Witwatersrand https://orcid.org/0000-0001-9786-6799

Research Article

Keywords:

Posted Date: December 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-999213/v1

MODELING EMISSIONS SOURCES AND CHARACTERIZATION IN A DENSE LOW-INCOME TOWNSHIP NEAR INDUSTRIAL ZONES AND MINE DUMPS (CASE OF TSAKANE, SOUTH AFRICA).

Shonisani Singo[1], Jean Mulopo [1,*]

Sustainable Energy and Environment Research Group, School of Chemical and Metallurgical Engineering, University of Witwatersrand, P/Bag 3, Wits, Johannesburg, 2050, South Africa[1]

Abstract

The sources of pollution in Tsakane township, which is situated within the City of Ekurhuleni in the province of Gauteng, South Africa, are investigated in this paper. The City of Ekurhuleni has the most industrial activities reported on South Africa's National Atmospheric Emission Inventory System (NAEIS), accounting for 40% of all listed activities in the country. The problem of suburban air pollution in South Africa is mainly associated with dense low-income areas like townships. The aim of this paper was to investigate atmospheric concentration correlation parameters, emissions roses, and probability modelling functions in order to analyse and classify significant emission sources affecting the township. Sulfur dioxide, nitrogen dioxide, ozone, and PM_{10} were the focus of the investigation. The probability functions for identifying and characterizing unknown or hidden sources of pollution were developed using hourly data. Furthermore, K-clustering algorithm analysis technique was used to provide graphical context for sources. PM_{10}, ozone, sulfur dioxide, and nitrogen dioxide have all been identified as having directional pollution sources that are problematic and the results provide baseline data for a detailed understanding of current emission levels and possible sources.

Keywords: air pollution, Tsakane township, South Africa, emissions sources, probability functions, k-means clustering techniques, Open-air, probability modelling functions

---

*jean.mulopo@wits.ac.za

1. Introduction

Tsakane township is located in the Gauteng Province of South Africa, in the City of Ekurhuleni. Gauteng Province is the country's economic powerhouse, accounting for 36% of the country's GDP and 10% of the African continent's GDP. Despite covering just about 1.5% of South Africa's geographical area, this province accounts for approximately 40.6% of the country's manufacturing industry. Gauteng Province still has the most people in South Africa [7, 13, 18]. The province's intensive manufacturing operations (depending on their systems, inputs, and outputs) emit a variety of contaminants into the atmosphere [15, 16]. The Gauteng Province currently has 380 defined mine residue areas, the majority of which are gold-mining residues [9]. These areas are sources of dust contamination, especially during secondary tailings reclamation, which requires the removal of protective layers of vegetation for the duration of the reclamation project. Due to a variety of factors, including rapid population growth, industrialization, and a relatively high living standard, the Gauteng Province also has a high-energy demand and motorisation.

Gauteng Province has 34 ambient air quality monitoring stations, 29 of which are operated by municipalities and by the national Department of Environment, Forestry, and Fisheries, respectively (DEFF). The majority of these monitoring stations are concentrated in three metropolitan areas: Ekurhuleni, Johannesburg, and Tshwane, with a smaller number in the Vaal Triangle, Airshed Priority Area, and West Rand District Municipalities. According to the Gauteng Province state of air report for the period of 1 November 2016 to 30 June 2017, the concentration of Particulate Matters of diameter less than 10 g/m^3 (PM_{10}) exceeded the daily average of 75 g/m^3 across the province, while sulfur dioxide exceeded the daily average of 125 (g/m^3) in both the City of Ekurhuleni and Sedibeng Municipality District, which have been identified.

On a global scale, 92% of the world's population lived in areas where World Health Organization (WHO) air quality recommendations were not followed in 2014 [1, 4, 16]. In 2012, air pollution in urban and rural areas was reported to have caused 3 million premature deaths worldwide [8]. Countries can reduce the burden of disease caused by stroke, heart disease, lung cancer, and chronic or acute respiratory disorders such as asthma by reducing levels of air pollution. It is worth noting that the WHO's ambient air quality requirements (SO_2= 20 g/m^3 on average for a 24-hour period and PM_{10}=20 g/m^3 on average per year) are stricter than South African standards (SO_2=125 g/m^3 on average for a 24-hour period and PM_{10}=40 g/m^3 on average per year).

South Africa has steadily worked to boost air quality while encouraging high industrialization over the years: Environmental

regulations, pollution inventories, dispersion modelling and concentration inventories, revision of ambient air quality limits, effective planning by local governments, and sector-specific controls have all been established as air quality management tools. The national policy of the South African government on integrated pollution and waste management was developed in the year 2000, with the vision and objectives outlined in the white paper on integrated pollution and waste management in South Africa. The change in policy and legislative direction after the passage of the National Environment Management Air Quality Act 39 of 2004 has received additional attention. Further consideration was given to national and international legislation and how it influenced decisions of mandatory air quality management functions within the spheres of government, especially in the last decade [14]. South Africa's national system for air quality management was developed in 2012.

However, the Gauteng Province has conducted very few air quality modelling studies. Air dispersion modelling is mostly done by manufacturing firms, government agencies, and, to a lesser degree, research councils in South Africa. For regulatory purposes, industrial companies are typically expected to perform dispersion modelling to ensure that their emissions are below standard concentrations. Current air quality monitoring policies are primarily focused on meeting set limit values for various key pollutants as specified by national directives [14] and calculated at fixed sites. While this technique may be useful for contaminants with homogeneous concentrations on a broad geographic scale, it does not allow for evaluating population exposure to pollutants

with high spatial variability, such as ultrafine particles. Furthermore, existing regulations do not provide monitoring of specifically exploitable criteria in terms of the health effects of air emissions or emission sources, despite the fact that the above information is critical for the implementation of appropriate public policies for improved air quality management. This is the motivation behind this work.

2. Study Area

Tsakane township is situated in the City of Ekurhuleni in the province of Gauteng, South Africa. This study first identified activities at the local area of the study as shown in Figure 1 i.e. activities that have a higher likelihood of influencing Tsakane township and the buffer is located at 20 kilometres radius away from Tsakane Ambient Monitoring Station to the circumference of the buffer zone.

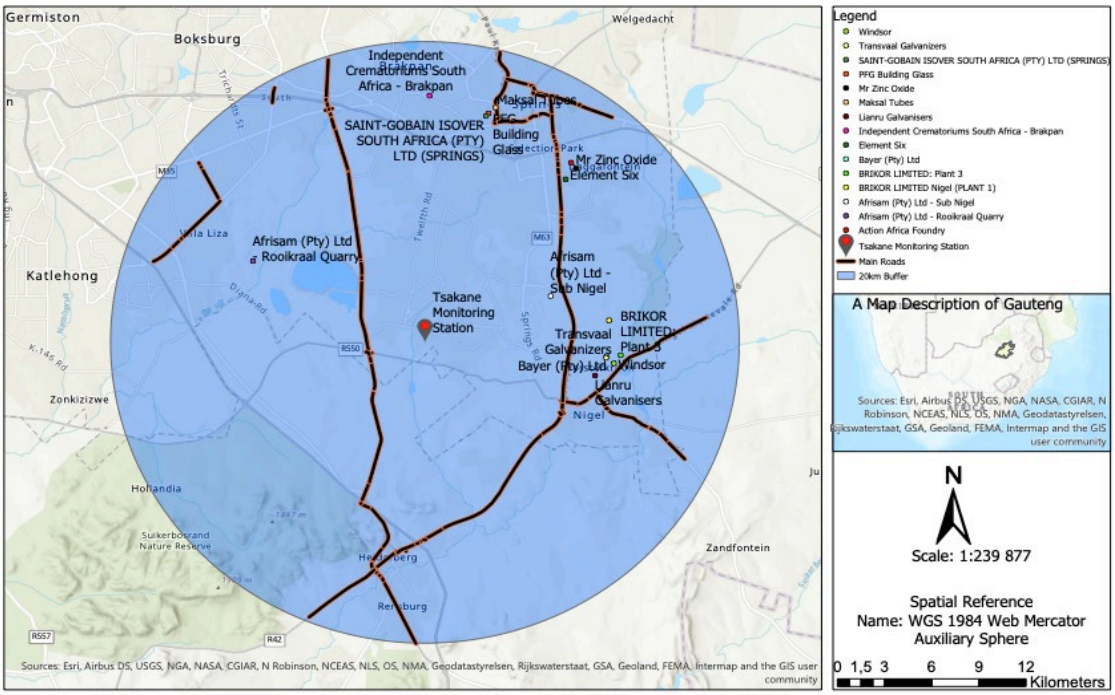


Figure 1: Tsakane township with NAEIS listed activities

3. Methodology

The City of Ekurhuleni established an ambient monitoring station in Tsakane township as a key place for monitoring the effects of air pollution on communities. Within a kilometre of the township, there are mine dumps, quarries, metallurgical factories, mineral processing industries, organic and inorganic chemical businesses, crematoriums, small boilers, traffic, and domestic fuel burning. The environmental monitoring station is set up to collect data on all of the parameters depicted in Figure 2. The following are the reference measurement

methods or principles to be used for conducting measurements of the designated pollutants listed in Table 1.

Table 1: Measurement methods used for pollutants described in this work

Pollutants Analyser	Measurement Methods	Equipment	Description	Brand
Sulfur dioxide (SO_2)	Ultraviolet Fluorescence	Model 100E UV	Microprocessor controlled analyser that determines the concentration of sulfur dioxide (SO_2), in a sample gas drawn through the instrument's sample chamber where it is exposed to ultraviolet light, which causes any SO_2 present to fluoresce.	Teledyne
Nitrogen Oxide(NO_x)	Chemiluminescence	Model T200	Uses the proven chemiluminescence detection principle, coupled with state of the-art electronics to allow accurate and dependable low-level measurements for use as an ambient analyzer or dilution CEMS monitor.	Teledyne
Carbon Monoxide (CO)	Gas filter Correlation Infrared Absorption	Model T300	Measures low ranges of carbon monoxide by comparing infrared energy absorbed by a sample to that absorbed by a reference gas according to the Beer-Lambert law.	Teledyne
Ozone (O_3)	Ultraviolet Photometric	Model 400E	Microprocessor controlled analyzer that measures low ranges of ozone in ambient air using a method based on the Beer-Lambert law, an empirical relationship that relates the absorption of light to the properties of the material through which the light is traveling over a given distance.	Teledyne
Particulate Matter (PM_{10})	EN12341: 1999 Air Quality	Teom	Determination of the PM_{10} fraction of suspended particulate.	Teom
Particulate Matter ($PM_{2,5}$)	EN14907 Standard	Teom	Gravimetric measurement method for the determination of the $PM_{2.5}$ fraction of suspended particulate matter.	Teom

The South African Weather Service (SAWS), Gauteng Department of Agricultural and Rural Development, and the City of Ekurhuleni provided the data for this study. The data were processed and imported into the Open-air model using Microsoft Excel. The Open-air model is a statistical tool for analysing semi-empirical mathematical relationships between air pollution concentrations and other variables that may affect them [2]. Ambient concentration, wind direction, and wind speed can all be used to effectively distinguish between various pollution sources influencing the ambient station. The study was carried out using k-means clustering to distinguish between features that are identical or dissimilar, which can then be categorised and grouped

3.1 INVESTIGATION OF CORRELATION COEFFICIENT, R.

The (Pearson) correlation coefficient is a measure of the strength of the linear relationship between 2 variables. If there is perfect linear relationship with positive slope between the two variables, r=1. A correlation coefficient of zero means that there is no linear relationship between the variables [3].

$$r_{xy} = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; 1$$

3.2 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.

Pollution roses are polar diagrams that display air pollution depending 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. Pollution roses are used as a function of the wind direction for measuring the average concentration of a pollutant at a receptor from multiple emission sources over a span [3]. The construction of a pollution rose is easy when the wind direction and the concentrations of pollutants are averaged over one hour and therefore the data used in this analysis are at hourly averages. The pollution roses were determined according to the following:

$$C_{dd} = \frac{\sum_{j=1,n} p_j f_{dd,j} \alpha_j}{\sum_{j=1,n} f_{dd,j}\, \alpha_j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 2$$

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, p_j is the measured concentration on day j, $f_{dd,\,j}$ is the frequency of the wind from sector dd whereas α_j is some function based on the persistency of the wind vector during day j.

$$\sum_{dd=1,ndd} c_{dd} f_{dd,j} = p_j \sum_{dd=1,ndd} f_{dd,j} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ 3$$

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

3.3 INVESTIGATION OF POLLUTION SOURCES USING THE BIVARIATE POLAR PLOT AT RECEPTOR.

Wind speed and direction in a polar coordinate affect the concentration of a species to a particular degree at a specific location [12]. Wind direction and speed can be very useful in distinguishing between different pollution sources. The plots provide a useful graphical technique for providing directional information on sources by using polar coordinates. The mean concentration is determined for each bin after wind speed, wind direction, and concentration data are partitioned into wind speed-direction 'bins.'

Binning the data in this way isn't strictly necessary, but it's a reliable data reduction strategy that doesn't affect the plot's fidelity. Furthermore, due to the inherent wind direction variability in the atmosphere, data from several weeks, months, or years is normally diffuse and does not differ abruptly with wind direction or speed when used to create a bivariate polar map. The wind components, u and v are calculated:

$$u = \bar{u}.sin\left(\frac{2\pi}{\theta}\right), v = \bar{u}.sin\left(\frac{2\pi}{\theta}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 4$$

Where 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

provide u, v, and 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 [20]. A flexible framework for fitting a surface is to use a Generalized Additive Model (GAM). The GAM can be expressed as shown in Equation 5:

$$\sqrt{C_i} = \beta_o + s(u_i, v_i) + \epsilon_i \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; 5$$

Where C_i is the ith pollutant concentration, β_o is the overall mean of the response, $s(u_i, v_i)$ is the isotropic smooth function of ith value of covariate u and v, and ϵ_i is the ith residual. A penalized regression spline was used to model the surface as described elsewhere [6, 21]. Note that Ci is square root transformed as the transformation generally produces better model diagnostics e.g. normally distributed residuals. Moreover, the smooth function used is isotropic because u and v are on the same scales. The isotropic smooth avoids the potential difficulty of smoothing 2 variables on different scales e.g. wind speed and direction, which introduces further complexities.

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.

$$CBPF_{\Delta\theta,\Delta_u} = \frac{m_{\Delta\theta,\Delta_u I_{C\geq x}}}{n_{\Delta\theta}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; 6$$

Where $m_{\Delta\theta,\Delta_u}$ is the number of samples in the wind sector $\Delta\theta$ with wind speed interval Δ_u having concentration C greater than a threshold value x, $n_{\Delta\theta}$ is the total number of samples in that wind direction-speed interval.

3.4 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.

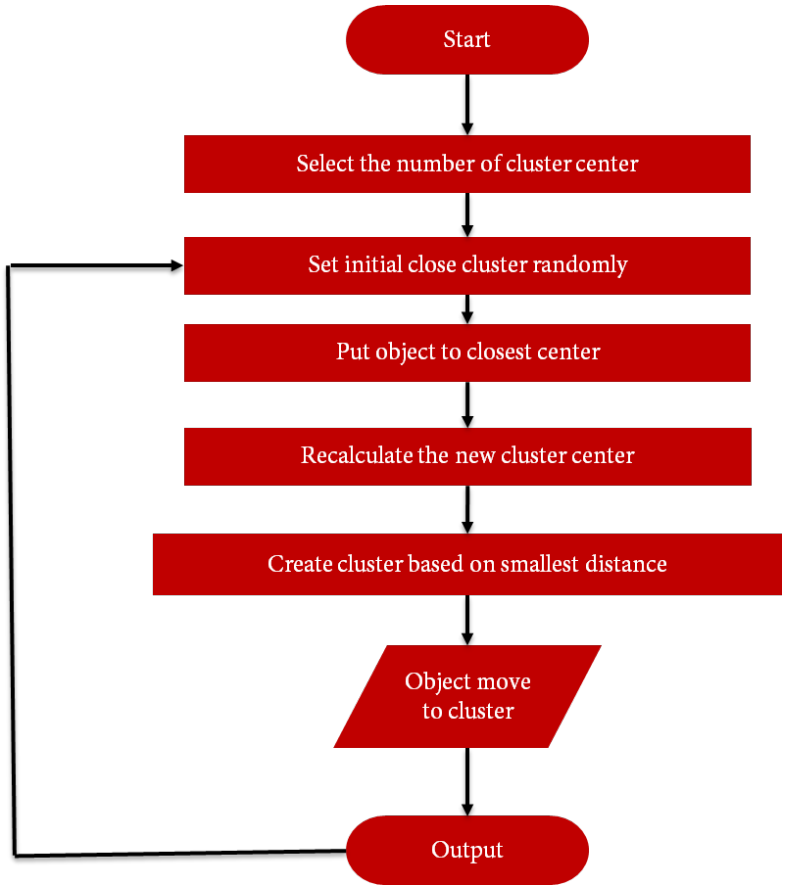


Figure 2: Flowchart of K-means clustering

The definition of distance, or any measure of similarity or dissimilarity between points, is central to the idea of clustering data. Clusters can be made up of points that are separated by small distances as compared to the distance between them. Three variables decide the similarity of concentrations shown in Figure 2: the u and v wind elements, as well as the concentration, C. All three variables are equally important in

describing concentration position data, but they are measured on different scales, such as a wind speed-direction measure and a concentration level. 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:

$$\sum_{k=1}^{k} \sum_{x1\epsilon_{ck}} \parallel X_i - \mu_k \parallel^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 7$$

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

$$d_{x,y} = \left(\sum_{j=1}^{n} \left(x_j - x_j\right)^2 \right)^{1/2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 8$$

Where n represents number of features as defined by Euclidean distance ($d_{x,y}$). Where **x** and **y** are 2 J-dimensional vectors, which have been standardised by subtracting the mean and dividing by the standard deviation. In the current case j is of length three i.e. the wind components u and v and the concentration C, each of which is standardised.

$$x_j = \left(\frac{x_j - \bar{x}}{\sigma_x} \right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 9$$

These sources could include major point sources that are far from receptor and minor sources that are difficult to detect. In order to propose relevant reductions strategy for poor air quality in Tsakane township, source identification. Distance of the source can be determined and wind speed can be obtained from the ambient station refer to Figure 13.

$$Time = \frac{Distance\ (between\ a\ source\ and\ station)}{Wind\ Speed} \ldots\ldots 10$$

4. RESULTS AND DISCUSSION

Data in Figure 3 display a very strong positive correlation for nitrogen dioxide and nitrogen monoxide. Nitrogen dioxide has strong positive correlation either with nitrogen monoxide. Ozone pollutant has weak positive correlation with wind speed. Sulfur dioxide concentrations displayed weak positive correlation with carbon monoxide. No correlations were observed between nitrogen dioxide and rain, wind speed and nitrogen monoxide, sulfur dioxide and wind speed, PM_{10} and Ozone, Toluene and Ozone, Benzene and Ozone, and solar radiation and rain. The highest negative correlation was observed between wind speed and carbon Monoxide.

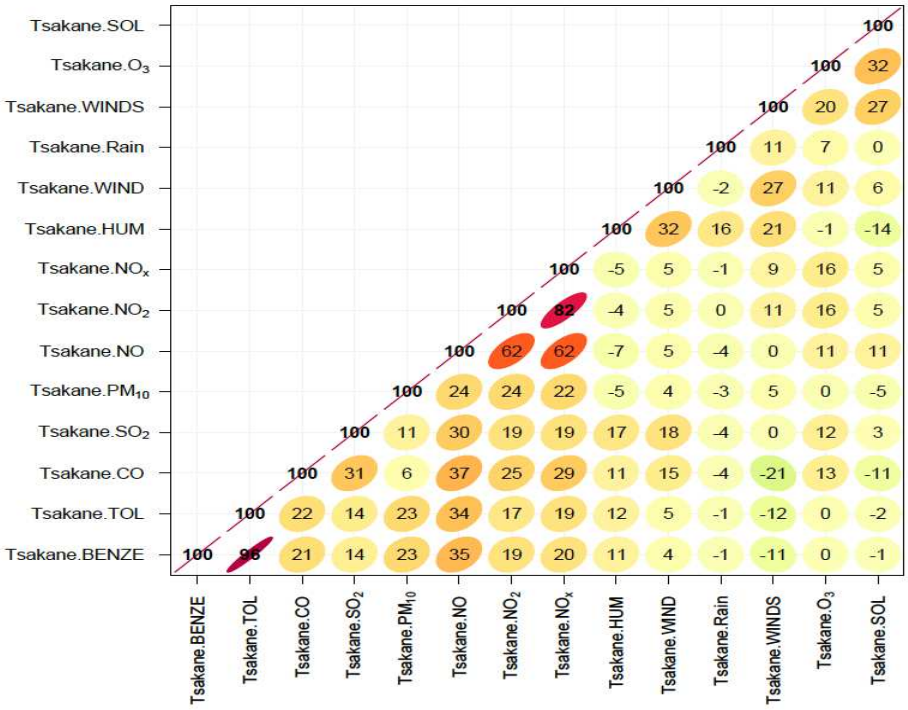


Figure 3: Tsakane Ambient Monitoring Station Parameters Correlations

Figure 4 shows strong emission sources in the North-East direction, which are higher between 15 and 20%, north at 15%, and North-West direction, which are higher between 10 and 15%. At Tsakane, the maximum atmospheric sulfur dioxide concentration was 966.37 g/m^3 of PM_{10} cubic metre. High pollution sources were located in the North-East, East, North, and North-West directions at ambient PM_{10} concentrations of 0 to 50 g/m^3. The results showed that weak pollution increased in the South-East, South, South-West lower, South-West higher, and South-East higher directions at the same above ambient PM_{10} concentrations.

Sulfur dioxide emissions from the East (between 15 and 20%), the South-East (between 10 and 15%), and the North (between 10 and 15%) are all shown in Figure 4. Sulfur dioxide concentrations in the atmosphere reached a high of 53.845 parts per billion. According to the results, emissions from the South-West were higher, while those from the South and South-West were lower, accounting for less than 5% of the total.

Strong sources of nitrogen dioxide emissions were found in the North-East, which accounted for 14%, the north between 12 and 14%, and the North-West, which accounted for 10 to 12%. The maximum nitrogen dioxide atmospheric concentration was 88.622 parts per billion, as shown in Figure 5. Poor emission sources are also found in the South-East, South, South-East, South-West, and South-West, according to the results. At atmospheric concentrations of 0 to 10 parts per billion, high pollution sources from the North-East, East, North-West, North, South-West lower, and South-West higher are all below 4%. Weak emission sources are also found in the South-East, South, South-West, and South-West, according to the results.

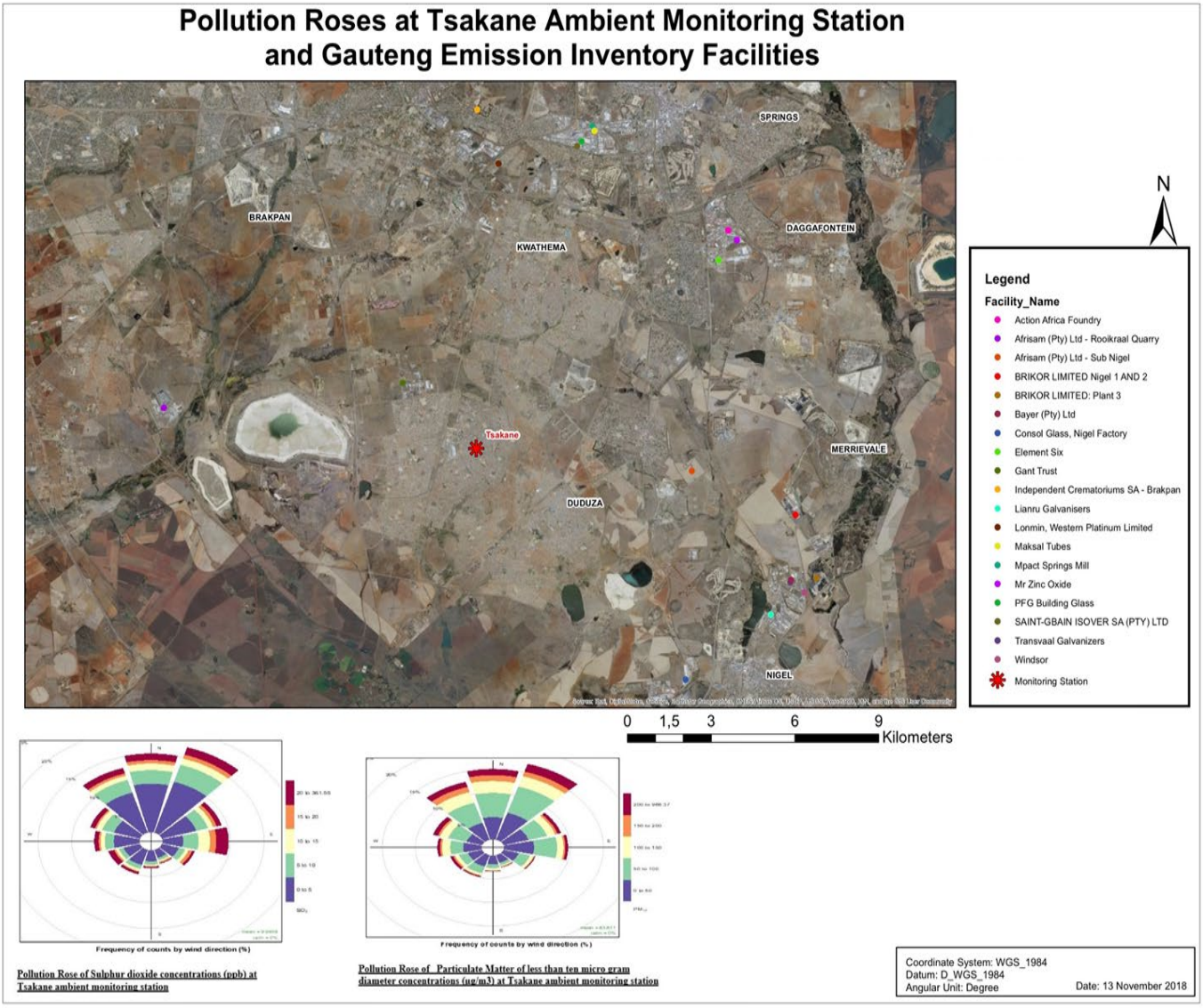
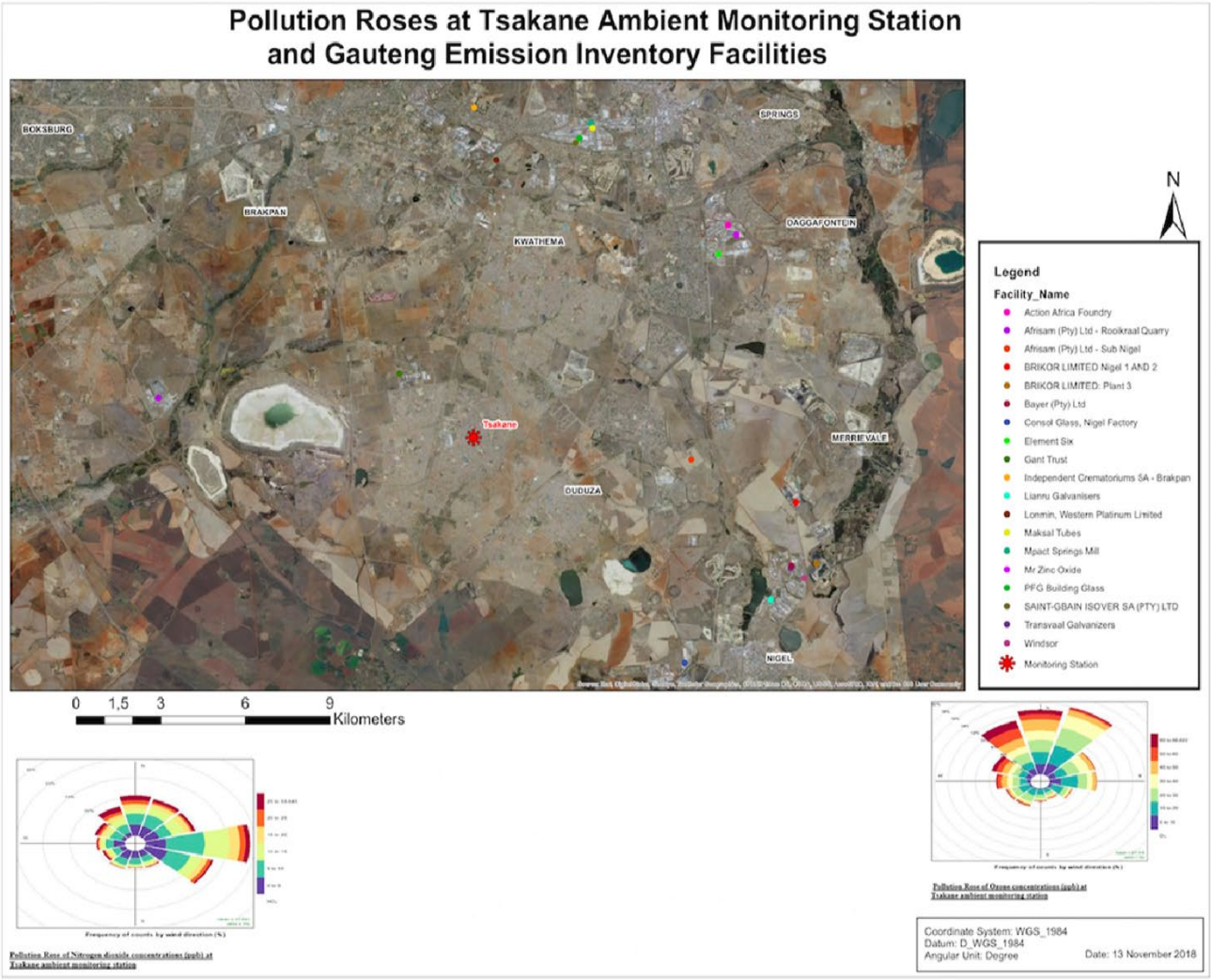


Figure 4: Pollution Roses for SO_2 and PM_{10} at the Tsakane Ambient Monitoring Station and registered NAEIS activities.



340

Figure 5: Pollution Roses for NO_2 and O_3 at Tsakane Ambient Monitoring Station and registered NAEIS activities.

In Figure 5, the wind speed is less than 2 metres per second, and the ambient air quality is higher in the northern areas of the station, at 18 parts per billion. The atmospheric concentrations of 11 parts per billion are found in the North-West, North, North-East, East, and South-East when the wind speed is between two and four metres per second. There are visible nitrogen dioxide sources on the East, South-East, South-West, and North-West of the Tsakane ambient station when the wind speed is between 4 and 10 metres per second, with a likelihood of an ambient concentration of 10 parts per billion.

The bivariate polar plot of PM_{10} in Figure 6 depicts local and distant sources around the station. When the wind speed is less than 4 metres per second, local sources on the North-West side of the station with an atmospheric PM_{10} concentration of 85 g/m^3 are very likely.

The bivariate polar plot of O_3 in Figure 7 shows local and distant sources around the station. When the wind speed is less than 2 metres per second, local sources on the North-West side of the station with an atmospheric concentration of 35 parts per billion of O_3 are a strong possibility.

The bivariate polar plot of sulfur dioxide concentrations shows local and distant origins of the gas. Figure 8 shows two major hot spots on the North-West and Eastern regions of the atmospheric station with ambient concentrations of 8 parts per billion when the wind speed is less than two metres per second. Sulfur dioxide sources with concentrations of 8 parts per billion are more likely on the South-West, East, and South-East of the ambient station when the wind speed is between two and six metres per second.

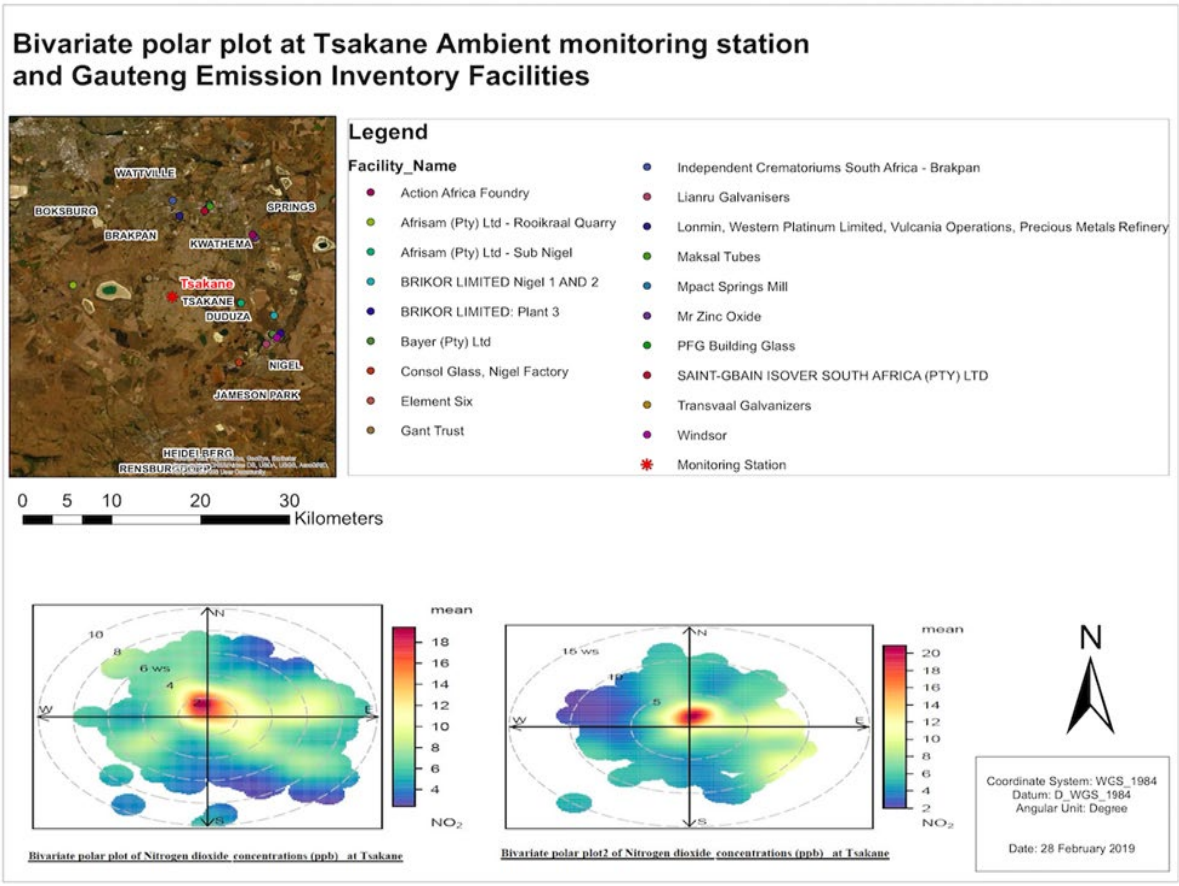


Figure 6: Bivariate polar plot for NO_2 at the Tsakane Ambient Monitoring Stations and registered NAEIS activities.

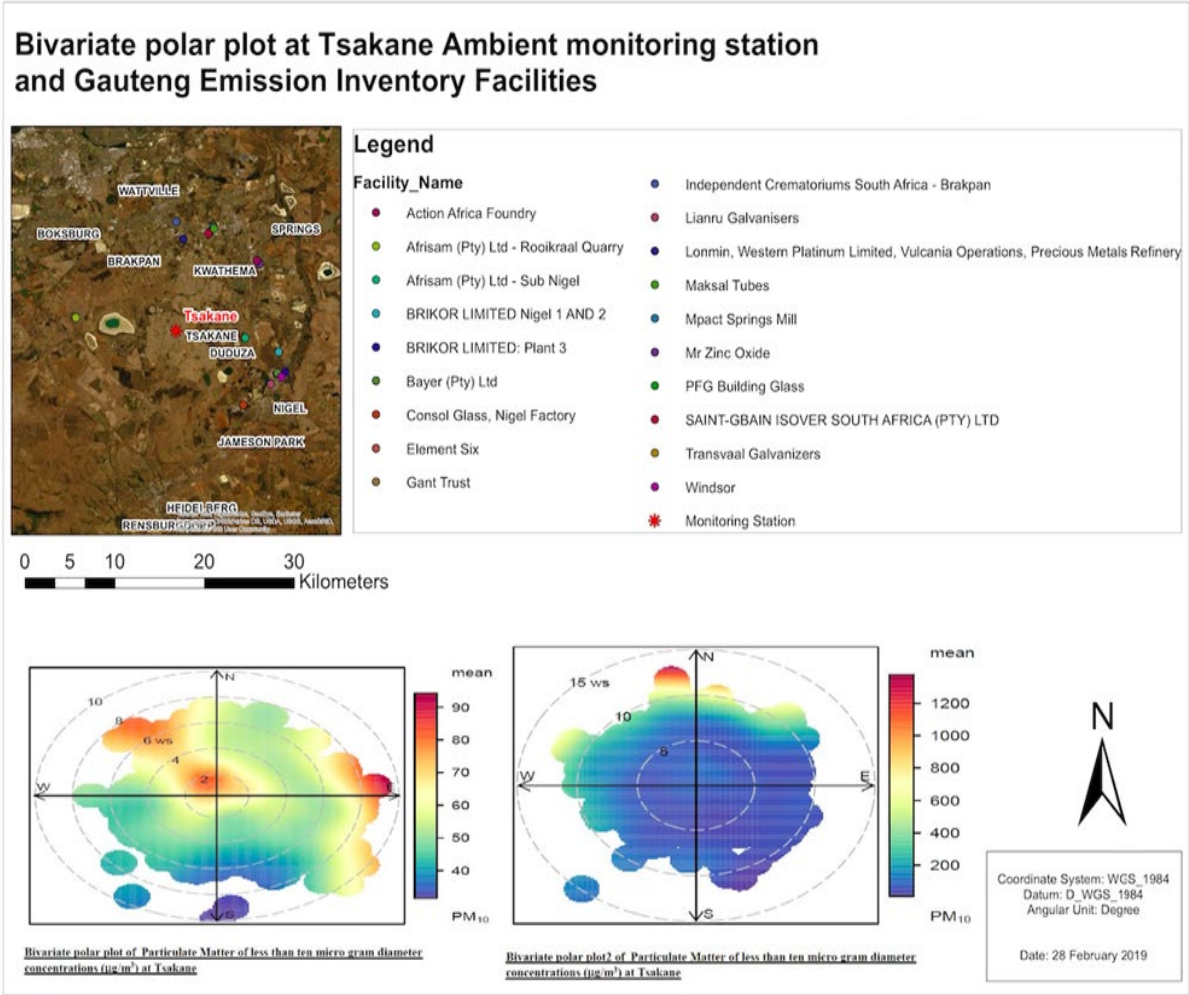


Figure 7: Bivariate polar plot for PM_{10} at the Tsakane Ambient Monitoring Stations and registered NAEIS activities.

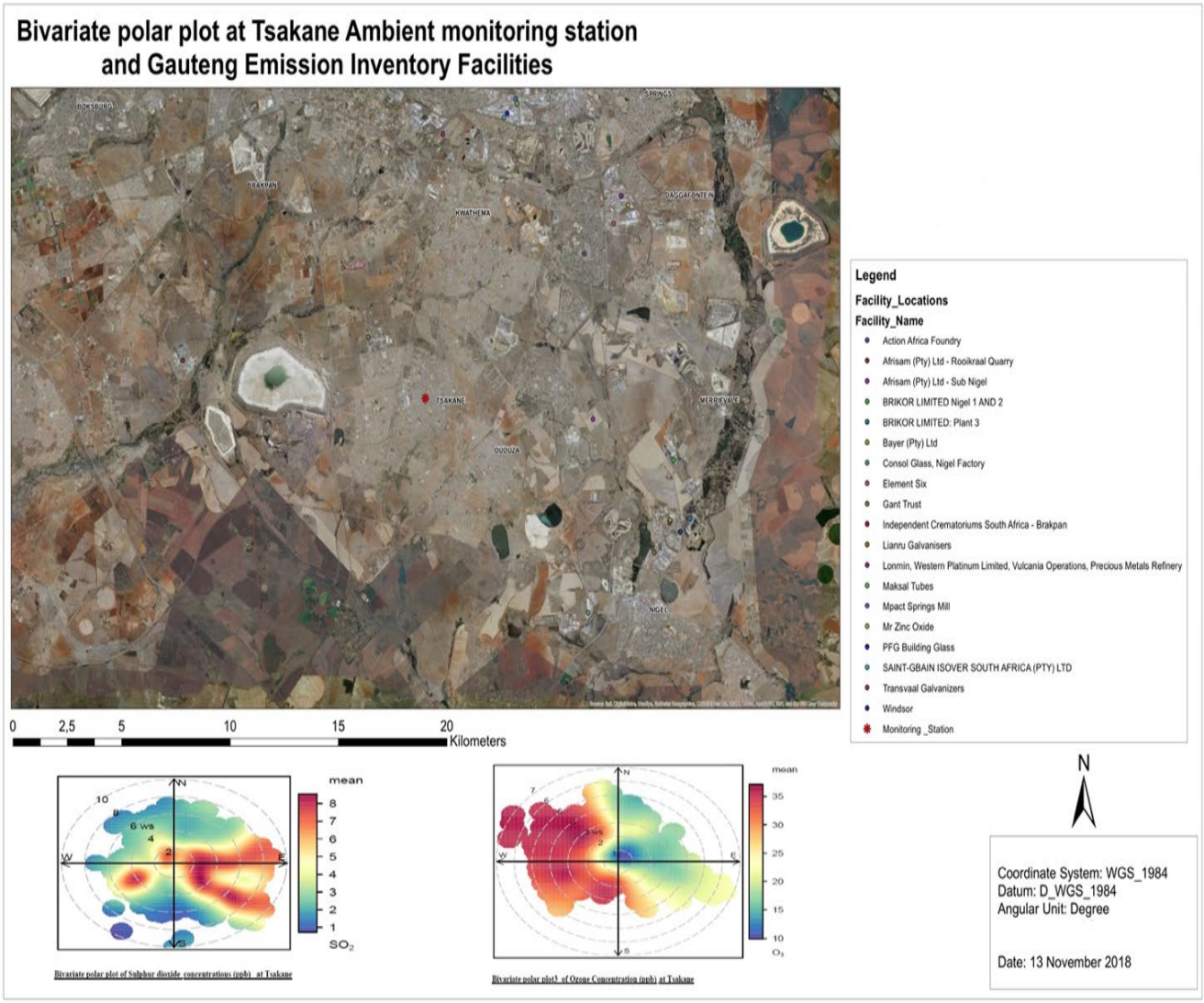


Figure 8: Bivariate polar plot SO_2 and O_3 at the Tsakane Ambient Monitoring Stations and registered NAEIS activities.

The findings reveal an interesting nitrogen dioxide trend. Cluster 1 has minimum emission levels of less than two parts per billion in the southern parts of the Tsakane atmospheric monitoring station. Previous work by Carslaw et al [2] showed that plumes from aircraft jets have high NO_x concentrations at high wind speeds when most other ground level sources show reduced concentrations.

Figure 9 in cluster 3 shows the lowest atmospheric nitrogen dioxide concentrations of less than 8.5 parts per billion. The findings show a thick layer of high nitrogen dioxide along the east and south coasts. When the wind speed is between 3 and 6 metres per second, Cluster 4 reveals other outlets in the south-west. In an hourly average, Tsakane ambient monitoring stations reported an appropriate value that was below the national ambient air quality limit for nitrogen dioxide.

At a wind speed of 4 metres per second, the k-means of ozone atmospheric concentrations at Tsakane show the highest cluster number 6 on the North and North-East areas. Cluster 6 had the highest atmospheric concentration of 60 parts per billion, which was lower than the national ambient air quality norm of 61 parts per billion over an eight-hour period with 11 frequency of exceedance (Figure 10). Cluster 2 had a concentration of less than 25 parts per billion in the ambient air. The atmospheric concentrations in Clusters 1 and 3 were less than 45 parts per billion. The maximum PM_{10} value at Tsakane was 90 parts per billion, indicating high exceedance. In the hourly average of 90 parts per billion, Figure 11 had the highest atmospheric PM_{10} concentrations.

PM_{10} concentrations of up to 90 g/m^3 were found in Clusters 2 and 5. Cluster 1 had an atmospheric particulate matter concentration of less than 10 g/m^3. The k-means results from the Tsakane Ambient Monitoring Station revealed high ambient sulfur dioxide concentrations of 8 parts per billion in cluster 3 on the North-East, East, and South-East zones. On the North, North-West, West, and South-West sides of the ambient monitoring station, Cluster 4 had a higher probability of 8 parts per billion. Figure 12 depicted the maximum atmospheric sulfur dioxide concentrations in the hourly average of less than 8 parts per billion at cluster 1, which was significantly lower than the national ambient air quality levels for sulfur dioxide of 134 parts per billion for an 88 frequency of exceedance. Figure 13 shows the data availability for parameters measured at the Tsakane Ambient Monitoring Station.

Air pollution involves multiple contaminants emitted from a variety of sources to the Tsakane Ambient Monitoring Station. In order to understand the complexities involved, study investigated atmospheric concentration correlation parameters, emissions roses, and probability modelling functions to analyse and classify significant emission sources affecting the township. The probability functions identified and characterized unknown sources of pollution were investigated using hourly data. K-clustering algorithm analysis technique was used in providing graphical context for sources.

The results are applicable in the development of an Air Quality Management Plans for National, Provincial and or municipality as

contemplated in Section 15 of the National Environment Management Air Quality Act 39 of 2004. The Priority Area air quality management plans, as contemplated in Section 19 of National Environment Management Air Quality Act 39 of 2004. The atmospheric impact reports, as contemplated in Section 30 of the National Environment Management Air Quality Act 39 of 2004. In the application for an atmospheric emission licence (AEL) for such documentation and information as may be required by the licensing authority as contemplated in Section 37 (2)(b) of the National Environment Management Air Quality Act 39 of 2004. Specialist air quality impact assessment report as is prescribed in the National Framework for Air Quality Management in the Republic of South Africa.

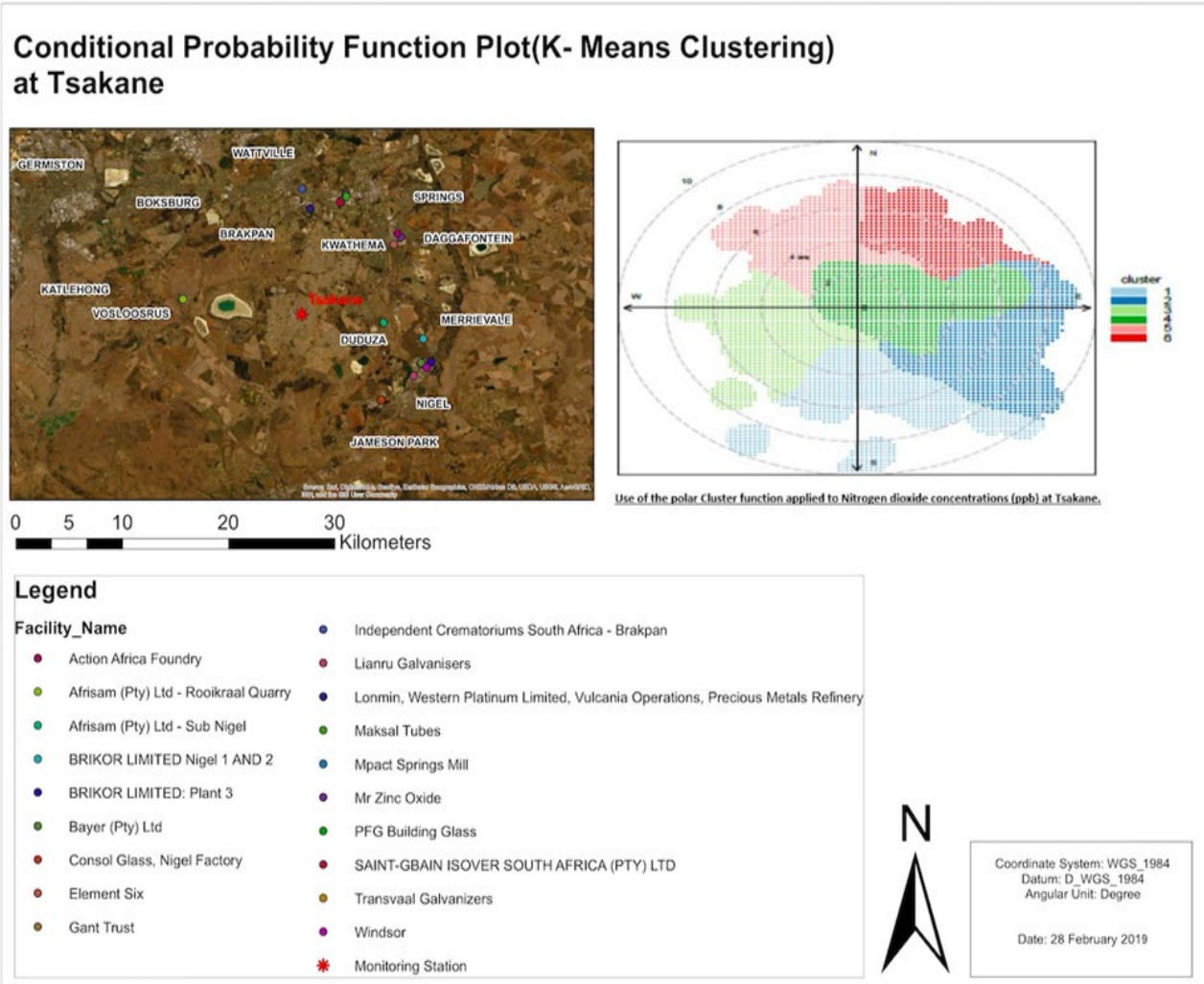


Figure 9: Conditional Probability Function Plot (NO_2) (K-Means Clustering)

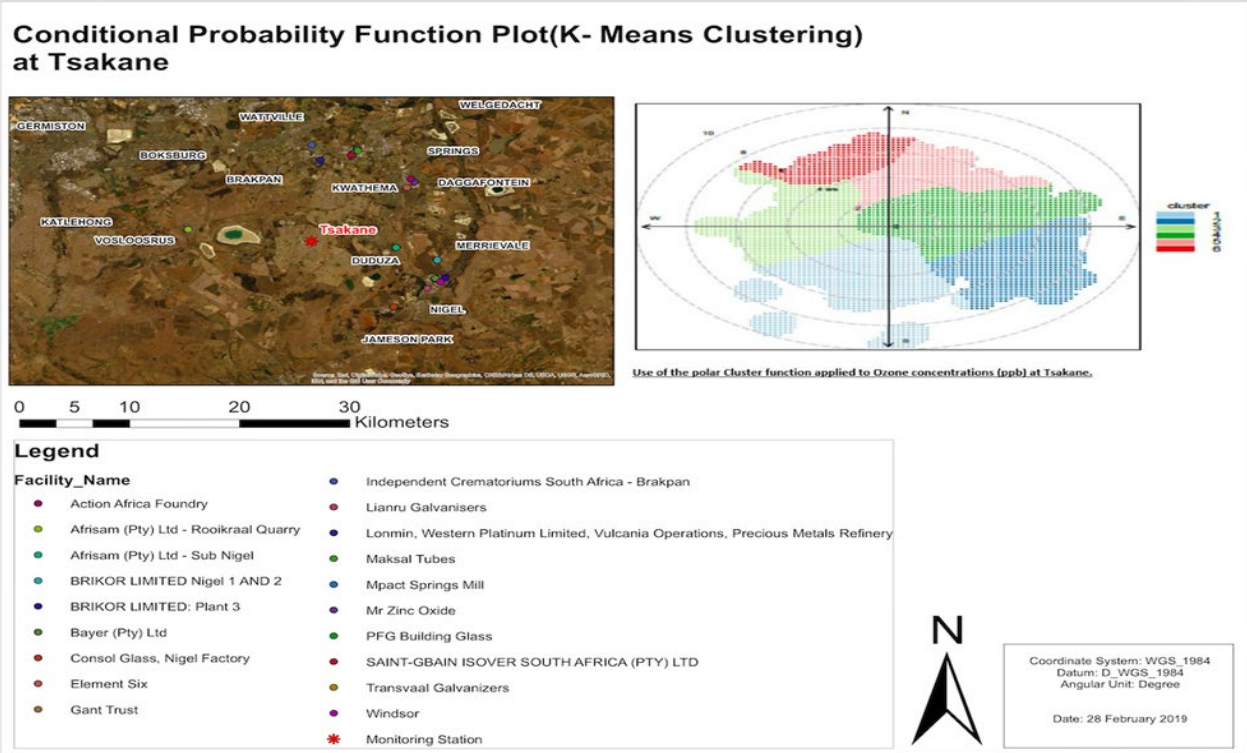


Figure 10: Conditional Probability Function Plot (O_3) (K-Means Clustering)

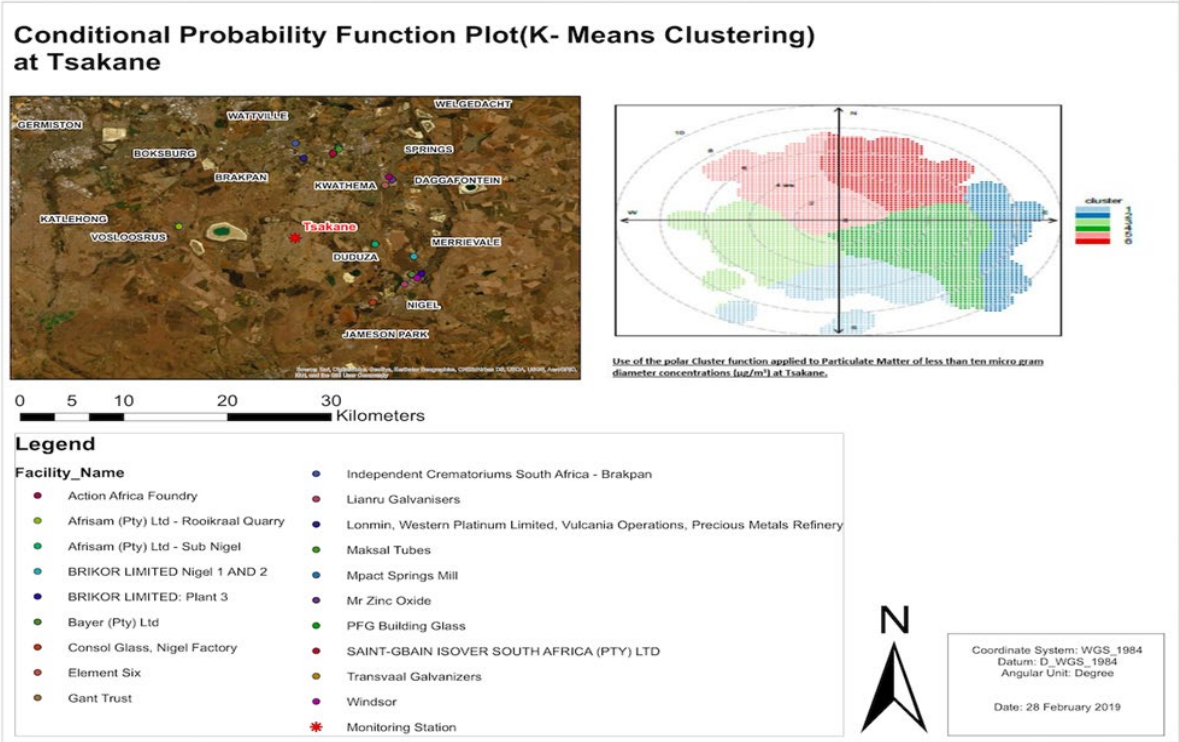


Figure 11: Conditional Probability Function Plot (PM_{10}) (K-Means Clustering).

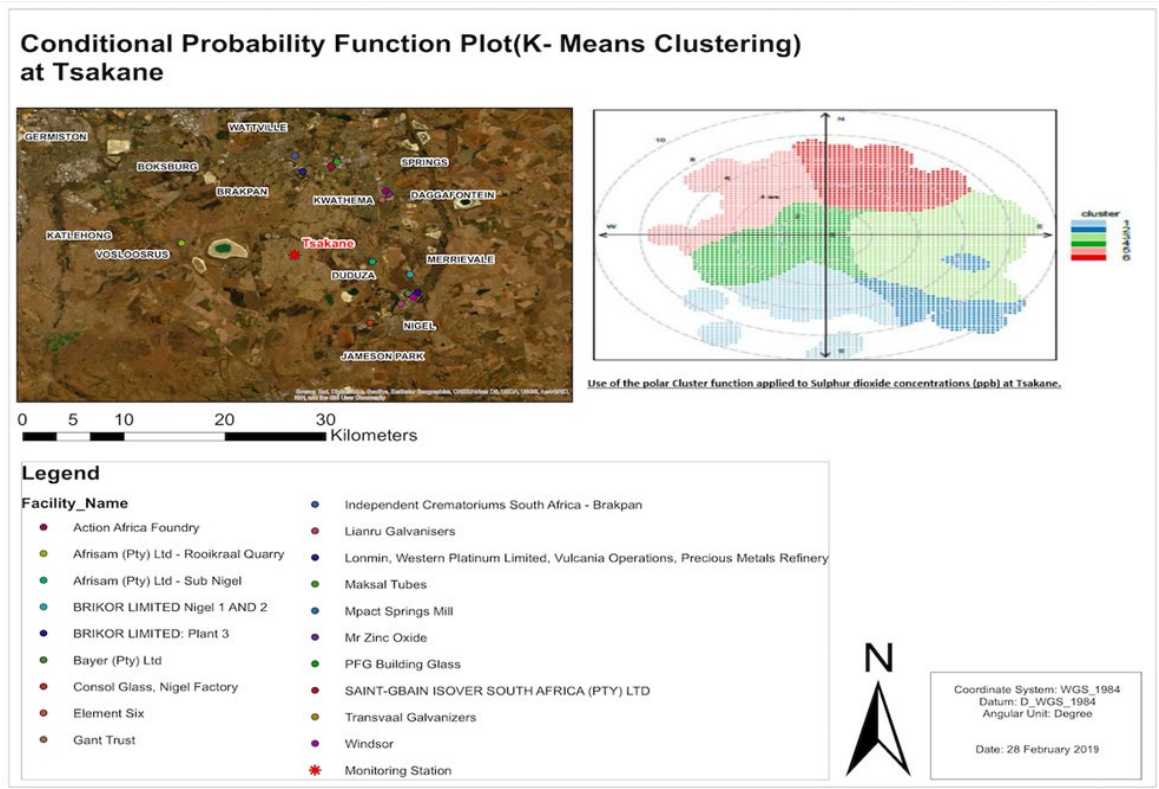


Figure 12: Conditional Probability Function Plot (SO_2) (K-Means Clustering).

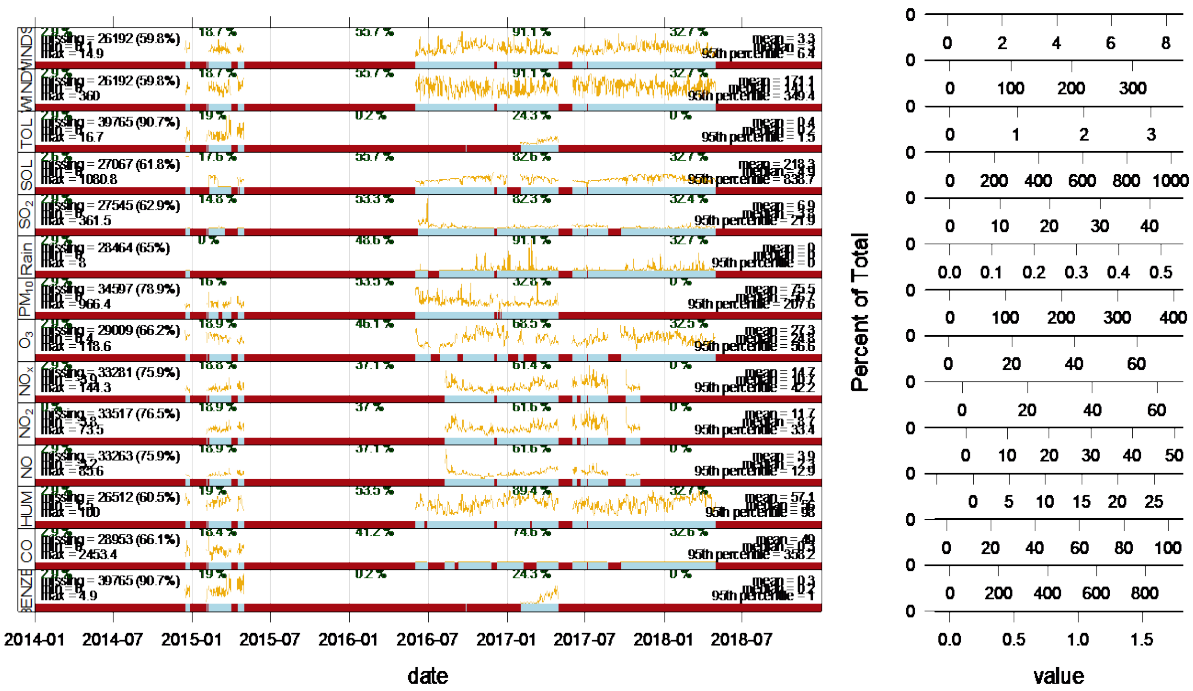


Figure 13: Tsakane ambient air quality data summary

5. Implications for South African Air Quality Policy

The 1965 Atmospheric Pollution Prevention Act (APPA) No.45 informs and regulates South African air quality management plans, whereas the 2004 National Environmental Management: Air Quality Act (NEMAQA) No.39 became law on February 19, 2005. Each provincial or national agency is responsible for drafting an environmental implementation plan in accordance with the spirit of the preceding law. Environmental management plans must have an air quality management strategy (AQMS) and an integrated development strategy (Chapter 5 of

the Municipal Systems Act). A close reading of the preceding legislation indicates that Atmospheric Dispersion Modelling is a critical component of the AQMP process depicted in part 3 of Figure 14 below, which integrates pollutant source knowledge with meteorological data to predict receptor locations. A wide range of dispersion models have been investigated in South Africa, with less attention placed on investigation models such as conditional probability functions. We believe that air pollution modelling allows for the assessment of exposure and risk, impact areas, and forecasting, and that conditional probability functions, in particular, could aid in identifying sources and pollutants. The significance of the source and its location can then be mapped in the development of an AQMP.

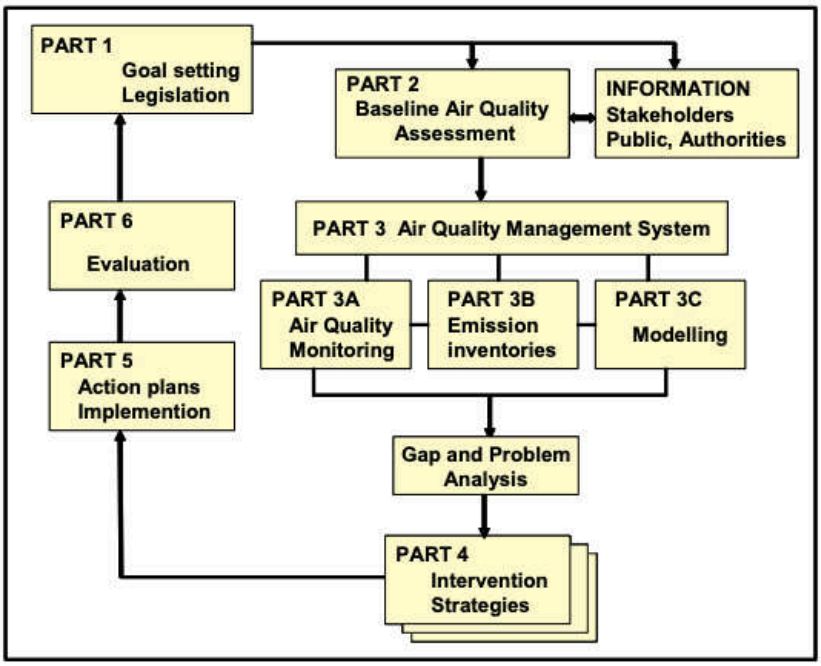


Figure 14: The AQMP method

Most environmental agencies across the world utilise models for regulatory purposes, especially when issuing emission permits or conducting environmental impact studies [17]. In these applications, models must offer the spatial distribution of high episodic concentrations as well as long-term averaged concentrations that may be compared to air quality standards or guidelines.

Environmental policy is currently confronted with world scale challenges (global warming, ozone depletion) [11]. Acidification, eutrophication, photo-oxidant production, urban air pollution, and the problem of air toxics are all key policy issues relating to the atmosphere. The goal of this study was to assess hidden significant sources of emission within and at the boundary of Gauteng Province, the country's most polluted region, using probability functions modelling in order to help identify sources that are causing exceedance to ambient air quality standards regardless of listed activity or not. Air pollution models will be essential for optimizing air pollution abatement techniques and assisting local environmental policymakers in forecasting the impact of abatement initiatives and will remain growing valuable tools for assessing emission reduction efforts, calculating ambient concentrations, and better understanding the economic aspects of air pollution [5].

6. CONCLUSION

This paper looked into the sources of air pollution in Tsakane township. The data identified potential sources near the ambient monitoring

stations. Clustering algorithms were used to identify features in a polar plot that had similar qualities. The data were provided by the Tsakane Air Quality Monitoring Station. The information was gathered between January 1, 2014, and April 30, 2019. The study looked at sulfur dioxide, nitrogen dioxide, ozone, and PM_{10}. The data were saved and then imported into the Open-air model. Hourly data were used to create classification probabilities and define hidden or undiscovered pollution sources within and around the Tsakane township boundary using probability modelling functions, assisting in the identification of sources that violate ambient air quality criteria. During the investigation, both major and minor sources were uncovered. PM_{10}, ozone, sulfur dioxide, and nitrogen dioxide all have directional emission sources that are problematic. The K-clustering algorithm analytical technique was used in class characterization and source recognition to give appropriate graphical context for sources. These findings can be used to construct Air Quality Management Plans for National, Provincial, Municipal, and Priority Area policies. The data can be further used for atmospheric impact reports, licensing authority information, and specialist air quality impact assessment reports.

Air pollution models will be critical for optimizing such abatement measures in order to help local environmental policymaking. In terms of policy support, models are needed to forecast the effect of abatement efforts, which may necessitate the model producing trustworthy results for future pollution scenarios. Air pollution models continue to be useful tools for evaluating emission reduction strategies, estimating ambient concentrations, and better understanding the economic aspects of air

pollution. Models of air pollution play an important role in decision-making.

References

1. Bondarouk, E., Liefferink, D., & Mastenbroek, E. (2020). Politics or management? Analysing differences in local implementation performance of the EU Ambient Air Quality directive. *Journal of Public Policy*, 40(3): 449-472.
2. Carslaw, D. C., Beevers, S. D., Ropkins, K., & Bell, M. C. (2006). Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. *Atmospheric Environment*, 40(28): 5424-5434.
3. Carslaw, D. C., & Ropkins, K. (2012). Openair—an R package for air quality data analysis. *Environmental Modelling & Software*, 27: 52-61.
4. Coker, E., & Kizito, S. (2018). A narrative review on the human health effects of ambient air pollution in Sub-Saharan Africa: an urgent need for health effects *studies*. *International journal of environmental research and public health*, 15(3): 427.
5. Conti, G. O., Heibati, B., Kloog, I., Fiore, M., & Ferrante, M. (2017). A review of AirQ Models and their applications for forecasting the air pollution health outcomes. Environmental Science and Pollution Research, 24(7): 6426-6445.
6. Evans, J. D. (1996). Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.
7. Greyling, T., & Tregenna, F. (2017). Construction and analysis of a composite quality of life index for a region of South Africa. *Social Indicators Research*, *131*(3): 887-930.
8. Harding, E. (2020). WHO global progress report on tuberculosis elimination. *The Lancet Respiratory Medicine*, 8(1):19.
9. Kamunda, C., Mathuthu, M., & Madhuku, M. (2016). Health risk assessment of heavy metals in soils from Witwatersrand Gold Mining Basin, South Africa. *International Journal of Environmental Research and Public Health*, 13(7): 663.

10. Mannucci, P. M., & Franchini, M. (2017). Health effects of ambient air pollution in developing countries. *International journal of environmental research and public health,* 14(9): 1048.
11. Martens, P. (2013). Health and climate change: modelling the impacts of global warming and ozone depletion. Routledge.
12. McNaught, A. D. (1997). Compendium of chemical terminology (Vol. 1669). Oxford: Blackwell Science.
13. Mushongera, D., Zikhali, P., & Ngwenya, P. (2017). A multidimensional poverty index for Gauteng province, South Africa: evidence from Quality of Life Survey data. *Social Indicators Research*, *130*(1): 277-303.
14. Naiker, Y., Diab, R. D., Zunckel, M., & Hayes, E. T. (2012). Introduction of local Air Quality Management in South Africa: overview and challenges. *Environmental science & policy*, 17: 62-71
15. Nkosi, V., Wichmann, J., & Voyi, K. (2017). Indoor and outdoor PM 10 levels at schools located near mine dumps in Gauteng and North West Provinces, South Africa. *BMC public health,* 17(1):1-7.
16. Panichev, N., Mokgalaka, N., & Panicheva, S. (2019). Assessment of air pollution by mercury in South African provinces using lichens Parmelia caperata as bioindicators. *Environmental geochemistry and health*, 41(5): 2239-2250.
17. Placet, M., Mann, C. O., Gilbert, R. O., & Niefer, M. J. (2000). Emissions of ozone precursors from stationary sources:: a critical review. *Atmospheric Environment*, 34(12-14): 2183-2204.
18. Retief, F. P., Mlangeni, G., & Sandham, L. A. (2011). The effectiveness of state of the environment reporting (SoER) at the local government sphere–a developing country's experience. *Local Environment*, 16(7): 619-636
19. Tiwary, A., & Williams, I. (2018). Air pollution: measurement, modelling and mitigation. CRC Press.
20. Uria-Tellaetxe, I., & Carslaw, D. C. (2014). Conditional bivariate probability function for source identification. *Environmental modelling & software*, 59: 1-9.
21. Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 62(2): 413-428.

Declarations

Funding

Not applicable

Conflicts of interest/Competing interests

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript

Availability of data and material

Data will be made available upon request

Authors' contributions

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version

Ethics approval

Not applicable