Source apportionment of fine atmospheric particles using Positive matrix factorization in Pretoria, South Africa.

Adewale Adeyemi[a,b*], Peter Molnar[c], Johan Boman[d], Janine Wichmann[a]

[a] School of Health Systems and Public Health, University of Pretoria, 31 Bophelo Road 00 01, Pretoria, South Africa.

[b] Department of Environmental Modeling and biometrics, Forestry Research Institute of Nigeria, Ibadan Nigeria.

[c] Occupational and Environmental Medicine, Sahlgrenska University Hospital & University of Gothenburg, Medicinaregatan 16A, SE-40530 Gothenburg, Sweden

[d] Department of Chemistry, University of Gothenburg Sweden.

abifaz@yahoo.com; peter.molnar@amm.gu.se; johan.boman@gu.se; janine.wichmann@up.ac.za

*Corresponding author: abifaz@yahoo.com

Abstract

In Pretoria South Africa, we looked into the origins of fine particulate matter ($PM_{2.5}$), based on one-year sampling campaign carried out between 18 April 2017 to 17 April 2018. The average $PM_{2.5}$ concentration was 21.1± 15.0 µg/m^3 (range 0.7 - 66.8 µg/m^3), with winter being the highest and summer being the lowest. The XEPOS 5 Energy dispersive X-ray fluorescence (EDXRF) spectroscopy was used for elemental analysis, and the US EPA PMF 5.0 program was used for source apportionment. The sources identified includes fossil fuel combustion, soil dust, secondary sulphur, vehicle exhaust, road traffic, base metal/pyrometallurgical, coal burning. Coal burning and secondary sulphur were significantly higher in winter and contributed more than 50% of $PM_{2.5}$ sources. The HYSPLIT model was used to calculate the air mass trajectories (version 4.9). During the one-year research cycle, five transportation clusters were established. North Limpopo (NLP), Eastern Inland (EI), Short-Indian Ocean (SIO), Long-Indian Ocean (LIO) and South Westerly-Atlantic Ocean (SWA). Local and transboundary origin accounted for 85%, while 15% were long-range transport. Due to various anthropogenic activities such as biomass burning and coal mining, NLP clusters were the key source of emissions adding to the city's PM rate. In Pretoria, the main possible source regions of $PM_{2.5}$ were discovered to be NLP and EI. Effective control strategies designed at reducing secondary sulphur, coal burning, and fossil fuel combustion emissions at Southern African level and local combustion sources would be an important measure to combat the reduction of ambient $PM_{2.5}$ pollution in Pretoria.

Keywords: Air pollution, Source apportionment, Particulate matter, Biomass burning, Transport cluster, South Africa

Introduction

Outdoor and indoor air pollution has been regarded as a serious issue in South Africa, with the emissions of various air pollutants and their resulting concentrations in the atmosphere being a major source of concern. Nearly 80% of the global population, for example, were subjected to air pollution levels that surpassed the World Health Organization's (WHO) air quality guidelines in 2011 (WHO, 2013). Pollution from a variety of sources has had a significant effect on air quality, posing a direct threat to the critical roles the environment plays in preserving and sustaining life by absorbing harmful ultraviolet radiation, warming the surface, and controlling the earth's

temperature (Hunter et al., 2007). Particulate matter (PM) suspended in the air for hours or days can travel a long distance, making it a long-range transported pollutant that is influenced by particle size, chemical composition, and other physical and biological characteristics (Zunckei, et al., 2008). $PM_{2.5}$ (particles smaller than 2.5 µm) has received a lot of attention recently because of the negative impact it has on human health, i.e. its potential to penetrate human lungs (Rupp, 2009). Furthermore, epidemiological studies have revealed a connection between PM and a variety of health problems (Brunekreef and Holgate, 2002; Schwartz, 2004; Kappos et al., 2004, Lokotola et al., 2020).

Carbonaceous aerosols have received a great deal of publicity because of their harmful effects on the environment, climate, agricultural production, public health, and visibility (Bisht et al. 2015; Li et al. 2016; Janssen et al. 2012). Black carbon (BC**)** is a type of carbonaceous material that is only generated in flames when carbon-based fuels are burned, with a unique set of physical properties that plays a crucial role in the Earth's climate system. BC stands out from other types of carbon and carbon compounds found in atmospheric aerosol because It absorbs visible light strongly, is refractory with a vaporization temperature approaching 4000 K, exists as a spherical aggregate, and is insoluble in water and ordinary organic solvents (Bond et al., 2013). There are two types of carbonaceous aerosols, namely elemental carbon (EC) and organic carbon (OC). EC are products of incomplete combustion of carbon-containing substances (such as fossil fuels and biomass). OCs, on the other hand, are emitted directly from soil re-suspension containing degraded biomass, biomass burning, vehicular and coal-based industrial exhaust etc. (Wang et al.,2018). In addition, OC can be classified as either primary organic carbon (POC) of secondary organic carbon (SOC) as a result of the heterogenous chemical reactions that occurs during the gas to particle conversion process (Mahilang et al. 2020).

A few papers and studies have attempted to investigate the air quality in South Africa's low-income regions. (Engelbrecht et al., 2000, 2002; Pauw et al., 2008), however these studies have spatial limitations and have not been able to characterize the major metropolitan centers where the population density is high. (Hersey et al., 2015). As a result, this research is required to define pollution sources and their contributions to atmospheric PM concentrations, as well as establish effective control strategies to combat pollution levels in the atmosphere in a long-term manner (Molnar et al., 2017). According to Scorgie et al. (2004), the major anthropogenic sources of PM pollution into the atmosphere of South Africa includes motor vehicles, factories, burning fossil fuels (without emission regulator equipment), and residential areas (where coal, firewood, and kerosene are used extensively for heating and cooking purposes). However, there is a scarcity of data on PM source apportionment in South Africa, making it difficult to establish successful strategies to address the problems that air pollution causes. Source apportionment is an important air quality management tool for providing information about source contributions required for pollution abatement strategies.

The recent review by Mathuthu et al. (2019), identified some studies on source apportionments in South Africa and the method used. Source apportionment of aerosol by Salma et al. (1994) in Kruger National Park, Formenti et al. (1998) in Soweto and Conradie et al. (2016) at different sites including Vaal Triangle, Amersfoort, Skukuza and Louis Trichardt used multivariate analysis, and principal component analysis respectively. Engelbrecht et al. (2002) used the Chemical Mass Balance method to compare PM sources of residential coal and low smoke fuels in Qalabotjha, and Van den Berg to identify PM sources in Kwadela township, while Tshehla and Djolov used Positive Matrix Factorisation (PMF) to identify pollution sources in an industrialised rural region of Limpopo. However, not many studies have applied air mass backward trajectory modelling with source apportionment

model analysis to investigate the sources of PM. It is important to consider the main origin of $PM_{2.5}$ and their contributions in order to design successful $PM_{2.5}$ reduction strategies (Gu et al., 2011). Therefore, this study is important in the source apportionment of aerosol in South Africa as it combines these two models, which will be useful to policy makers and stakeholders by providing important emission abatement strategies for the air quality management plan which is reviewed every five years and the recent one being reviewed currently.

The study's goals were to examine the chemical components of $PM_{2.5}$ and the influences of various sources in Pretoria, South Africa, using PMF, and to use back-trajectories cluster analysis to investigate source field contributions.

Methods

Aerosol sampling

The study site was located at the School of Health Systems and Public Health (SHSPH), University of Pretoria, and aerosol samples were collected on the roof of HW (6^{th} floor) Snyman South Building, Prinshof Campus (Fig 1). The sampling site (coordinates: -25.7S; 28.2E) was located close to the central business district (Pretoria CBD), which is approximately 1 km to 2 km away from a major road and about 5 -10 km away from the Pretoria CBD. This is an urban background site i.e. a "cleaner" option than industrial site. This study form part of the bigger project which was conducted concurrently during the same period of the year in other two cities, Thohoyandou (Novela et al., 2019) and Cape Town (Williams et al., 2020). This research used the same sampling methodology as these studies (Novela et al., 2019; Williams et al., 2020) and is outlined below.

Daily 24-hour samples were collected every third day from 9:00 am to 9:00 am the next day using GilAir-5 personal air samplers with 2.0e µm PFTE, PTFE supports (Zefluor) 37 mm filters (Zefon International, Inc Ocala, FL34474 USA) similar approach was used in the studies of Novela et al. (2019) and Williams et al. (2020). Between April 18th and April 17th, 2018, 147 $PM_{2.5}$ filter samples were obtained, with 25 duplicate samples, over a 122-day period. A duplicate was discarded due to faulty pumps leaving us with 24 duplicates. Meteorological data from weather station close to the sampling site was obtained from the South Africa Weather Services. All required quality assurance and quality control procedures were enforced throughout the sampling and analysis phases. To ensure correct data quality, sampling flow rate calibration, blank field tests, and instrument calibration were carried out according to the standard operating procedure.

Gravimetric and Chemical analysis of filter

Gravimetric analyses of $PM_{2.5}$ filters were carried out using 1 µg sensitivity microbalance (Mettler Toledo, XP6) under climate-controlled conditions (temperature and relative humidity were maintained at 21 ± 0.5°C and 50 ± 5%, respectively) at the Air Quality Laboratory, SHSPH before and after sampling. Black carbon (BC), UVPM (a proxy for organic carbonaceous particulate matter absorbing UV light at 370 nm) (Raja et al., 2016) and elemental concentrations were analysed for after gravimetric mass concentration determination. At the University of Gothenburg, BC and UVPM were measured using a Model OT21 Optical Transmissometer (Magee Scientific Corp., Berkeley, CA, USA). The presence of biomass burning is demonstrated by the additional absorption of UV light at 370 nm due to organics. (Sandradewi et al., 2008; Teich et al., 2017).

The elemental composition of aerosol particles on all filters were determined using an XEPOS 5 energy-dispersive X-ray fluorescence (EDXRF) spectrometer (Spectro analytical instruments GmbH, Germany) at the Department of Chemistry and molecular biology, Atmospheric Science Division, University of Gothenburg. The Spectro XRF Analyzer Pro program was used to process and quantify the EDXRF spectra. Every samples were analyzed for 3000 seconds overall, which was automatically divided among the four analytical setup conditions. The concentrations of the 18 elements Si, S, Cl, K, Ca, Ti, V, Fe, Ni, Cu, Zn, As, Se, Br, Sb, Ba, Pb, and U were determined. According to Molnar et al. (2014), the mean analytical precision was 5%, based on repeated analysis (N = 5) of two randomly chosen filters, one with a low mass loading and the other with a high mass loading.

PMF analysis and source identification

The Environmental Protection Agency program EPA-PMF5.0 was used to conduct the source apportionment study, which has been widely used by many studies to examine the origins of $PM_{2.5}$ (Martins et al., 2016; Molnar et al., 2017). PMF is a multivariate receptor model concept that uses a weighted least square approach to estimate source profiles and contributions (Paatero, 1994, 1997). In the Eq. (1), PMF model is used to obtain the unknown matrices, G and F, using an iterative least square method:

$$X = GF + E \quad (1)$$

$$Q(E) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{e_{ij}}{s_{ij}} \right)^2 \quad (2)$$

G is the source contribution to each sample *(size m x p)* for p factors, and F is the matrix of source profiles *(size p x n)*, and X is the data matrix *(size m x n)* consisting of n chemical components analyzed in m samples. The residual is represented by the matrix E. The main goal of iteration is to reduce the Q-value, which is defined as the number of squares of residuals (eij) weighted inversely with data point error estimates (sij) in Eq. (2). To run, the PMF model needs the concentration and its uncertainty. The following equation was used to measure the uncertainty (Reff et al., 2007; Norris et al., 2009)

$$\text{Unc} = \frac{5}{6} \text{ x MDL} \quad (3)$$

If the concentration exceeds the MDL, the uncertainty is determined using the following formula:

$$\text{Unc} = \sqrt{(Error\ Fraction\ x\ concentration)^2 + (MDL)^2} \quad (4)$$

In the case of missing data, the median was substituted, and the uncertainty was calculated using four times the median (Reff et al., 2007). $PM_{2.5}$ mass concentration was set as week by default while all other species were classified as ‘strong' (signal-to-noise ratio ≥ 2), ‘weak' (0.2 ≤ signal-to-noise ratio ≤ 2), or ‘bad' (signal-to-noise ratio < 0.2) species as done by Lee et al. (2011). (See Table S1). By tripling the uncertainty, the ‘weak' species were down weighted. No species with a signal-to-noise ratio of less than 0.2 was found in this analysis, so no species was classified as "bad." The number of runs in the model was set to 20 with 7 factors and a seed of 25. This is done to ensure that the findings are replicable. In addition, the best solution with the lowest Q value was

chosen, indicating that no outliers influenced the Q value, as in previous studies (Heo et al., 2008; Yu et al., 2013). Following that, the number of bootstrap runs was set to 100, with a minimum r^2 value of 0.6, and the average predicted (Y) and observed (X) values for $PM_{2.5}$ mass concentrations were calculated using the equation Y = 0.73X (R^2=0.63). The sources were identified by using major marker species that can be estimated based on a few references. The sources are also confirmed based on the source's percentage of the species (Yu et al., 2013). Furthermore, examining source inputs in relation to seasonal or meteorological factors aids in the comprehension of source characteristics.

Transport clusters

As in previous studies (Schwarz et al., 2016; Molnár et al., 2017; Tshehla and Djolov, 2018; Williams et al., 2020), the geographical origin of air masses passing through Pretoria, South Africa, was used as a proxy for long-range transport of air pollutants from distant sources and their composition. The Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) program was used to generate backward trajectories for the one-year sampling campaign (i.e. 18 April 2017 to 17 April 2018) (Draxler and Rolph 2003; Molnár et al., 2017, Williams et al., 2020). The model was motivated by the Global Reanalysis Meteorological Data from the National Centers for Environmental Prediction/National Centers for Atmospheric Research (NCEP/NCAR) web server at the National Oceanic and Atmospheric Administration Air Resources Laboratory (NOAA ARL). Every 6 hours (0:00, 6:00, 12:00, 18:00), an analysis field (resolution 2.5º x 2.5º and 17 vertical levels) was generated, and the wind field was interpolated linearly between each analysis; similar approaches were used in other studies. (Wichmann et al., 2014; Molnár et al., 2017; Williams et al., 2020)

Since a single backward trajectory has a significant uncertainty and is of little importance, as in previous studies (Wichmann et al., 2014; Molnár et al., 2017; Williams et al., 2020), this study used an ensemble of trajectories with a 500 m starting height and a fixed offset grid factor of 250 m (i.e. 250 m and 750 m). For cluster analysis, the average daily trajectories were estimated backwards for 72 hours. For the study period, 4380 backward trajectories were generated (24 h/6 h = 4, 4 x 3 heights x 365 days = 4380 trajectories) and used in the clustering analysis. The distance between a trajectory endpoint and the associated cluster mean endpoint was used to drive the clustering algorithm in HYSPLIT. The cluster analysis was done seasonally (autumn, winter, spring, and summer) owing to the drawback of using very large sample sizes in the HYSPLIT program's clustering function; similar approaches were used in other studies (Dorling et al., 1992; Wichmann et al., 2014; Molnár et al., 2017; Williams et al., 2020).

For the study, the seasons were categorised into the following dates: Autumn: 18 April – 31 May 2017 and 1 March – 17 April 2018; Winter (1 Jun – 31 Aug 2017); Spring (1 September – 30 November 2017) and Summer (1 December 2017 – 28 February 2018). Five transport clusters (North Limpopo (NLP), Eastern Inland (EI), Short-Indian Ocean (SIO), Long-Indian Ocean (LIO) and South Westerly-Atlantic Ocean (SWA)) were determined according to their mean pathways in the four seasons. Other researchers (Wichmann et al., 2014; Molnár et al., 2017; Williams et al., 2020) used a similar method to allocate each day of the study period to one of these geographical origins of air masses.

Results and Discussion

$PM_{2.5}$ concentration and seasonality

The average daily $PM_{2.5}$ concentration was 21.1 µg/m^3 (Table 1). The daily mean concentration in this study was higher than the South African National Ambient Air Quality Standards (20 µg/m^3) and the WHO yearly air quality guideline (10 µg/m^3). The WHO daily air quality guideline for $PM_{2.5}$ (25 µg/m^3) was found to be exceeded on 27 percent (33 days) of the sampling campaign days, with the most exceedance occurring during the winter season. The exceedances are indicative that the population of Pretoria and its surrounding are exposed to the adverse effect of exposure to ambient $PM_{2.5}$. The highest daily $PM_{2.5}$ level (66.8 µg/m^3) was recorded on 2 of June 2017 (winter season), this day correspond to cluster NLP which was observed to be associated with high level of $PM_{2.5}$ originating from north of Limpopo province. The anthropogenic activities that contributes to the sources of air pollution in this area includes coal power stations, mining, domestic fuel burning, agriculture and veld fires.

The concentration of the ambient $PM_{2.5}$ exhibited a distinct seasonal pattern, in winter, the mean level was significantly higher than autumn, spring, and summer: 35.5 vs 23.4, 14.3 and 10.7 µg/m^3, respectively ($p<0.001$), see Table 1. The mean $PM_{2.5}$ level for this study is significantly higher than the $PM_{2.5}$ levels from the other two cities of this bigger project, Cape Town (13.3 µg/m^3; Williams et al., 2020), and Thohoyandou (10.9 µg/m^3; Novela et al., 2019). In comparison with other African countries the mean $PM_{2.5}$ level were similar to the studies of Gaita et al. (2014) in Nairobi (21 µg/m^3), Arku et al. (2008) in Accra (22.7 µg/m^3) but lower to what were reported in Quagadougou Boman et al. (2009) 86 µg/m^3, Cairo Krzyzanowski et al. (2014) 80 µg/m^3 and Shaltout et al. (2020) (70 µg/m^3). There was no significant difference in results between weekdays and weekends, with mean levels of 20.7 and 22.1 µg/m^3, respectively ($p>0.05$), this may be attributed to other sources contributing to the pollutant levels apart from traffic, as it's expected that less traffic will be less over the weekend compared to weekdays. Another possible factor which may contribute to the increased levels of pollutants over the weekend is the South African tradition of barbequing, a similar observation was reported in the study of Williams et al. (2020). Fig S1 and Table S2 display a time series plot of temperature and relative humidity, as well as seasonal variations in meteorological data. The mean temperature, relative humidity and wind speed for full study are 17.6 °C, 54.3% and 1.6 m/s respectively. Four dominated wind direction were observed namely; westerly (27.4%), west- south westerly (16.4%), easterly (13.7%) and east north easterly (13.7%) respectively.

Correlation between air pollution and weather variables

Table 2 shows the spearman correlation between air pollution and weather factors. $PM_{2.5}$ revealed a weak positive significant correlation with wind speed ($r^2 = 0.21$) and a significant negative correlation with temperature ($r^2 = -0.32$) and relative humidity ($r^2 = -0.22$). Temperature revealed a slight negative significant association with relative humidity and wind speed. $PM_{2.5}$ had a positive correlation with BC and OC, while BC had a significant positive correlation with OC. The effect of meteorological conditions significantly influences air quality. The meteorological variables were compared to $PM_{2.5}$ concentrations to examine if there were any differences from what had been reported in the literature for seasonal variations. In winter, the mean temperature was lower with increased the amount of $PM_{2.5}$ concentration, whereas in summer, PM concentrations were low with increased temperature. This finding is supported by Cavanagh et al. (2009) and Ryu et al. (2013) studies, which found that air pollutants tend to increase with calm winds and stable atmospheric circumstances and vice versa. Sources of pollutants also play an important role in the concentration of ambient PM.

The enhanced conventional uplift and turbulence, which aid in the diluting and dispersion of pollutants, as well as reduced household burning operations, can help to minimize pollutant concentrations during warm season (Cavanagh et al. 2009; Yin et al. 2011).

Ambient PM has been found to be directly proportional to ambient temperatures, but most times negatively correlated with each other. Studies conducted in Japan (Wang and Ogawa, 2015) and China (Zhang and Jiang, 2018) respectively reported that temperature can either be positively or negatively correlated with ambient particulate matter depending on the geographical location, meteorological variable and other factors. The highest concentration of $PM_{2.5}$ was recorded when the relative humidity and wind speed were lower and calm, respectively and vice versa. Low temperature and wind speed can cause increased PM particles due to lack of dispersion, and high anthropogenic activities which includes biomass burning and space heating, these have been attributed to the high $PM_{2.5}$ concentration observed during the winter period (Chakraborty and Gupta, 2010; Ni et al. 2013). Precipitation, wet surface and higher relative humidity during summer could be responsible for the lower concentration of $PM_{2.5}$ during summer. This reduces the re-suspension of road and crustal dust which could have increased the concentration of PM in the atmosphere (Shah et al. 2012; Celo, and Dabek-Zlotorzynska, 2010).

Trace elemental composition

The annual mean concentration for trace elements is summarised in Tables S3. S had the highest average concentration of all the elements found, followed by Si, Fe, K, and Ca, in that sequence. The weekly variation showed that the highest concentrations of the elements were recorded on Thursdays, and lowest on Mondays for (K, Pb, and Cl), Tuesday for (S) and weekends for (Fe, Ca). The concentrations of Pb and Cl displays variation during weekdays, indicating variability in their sources, while the concentration of K, Ca, Fe and S do not display significant variation within the weekdays, thus indicating some consistency for their possible sources. Most of the species concentration followed the seasonal trend (winter>autumn>spring>summer) observed for $PM_{2.5}$. Major elements including S (1982.9 $ng/m^{3,}$ highest recorded), Si, Fe, Ca and K had the highest concentration in winter while V and Ba did not follow any pattern. Secondary aerosols, which are produced by the reaction and transformation of contaminants released (such as sulfate, nitrate and ammonium) to the atmosphere, have been confirmed to contain S in the form of sulphate (Viana et al., 2008; Yu et al., 2013). In coastal cities, Cl has been linked to sea salt. (Moon et al., 2008; Williams et al., 2020), other studies have linked it to coal combustion, which supports our findings since our research site is inland (Duan et al., 2006; Zheng et al., 2005). Results showed that the concentration was highest during autumn and winter which happens to be the cold period when coal is utilized for heating and power generation. Pb is not yet included in the ambient air quality criteria in South Africa. The annual $PM_{2.5}$ Pb level is far below the WHO annual guideline concentration of $500ng/m^3$ and well below the South African targeted ($250ng/m^3$) value that has yet to be promulgated. Since leaded gasoline was specifically phased out in South Africa after 2006, the low Pb concentrations observed are most likely due to this. The daily concentration of elements, e.g. S, K, Ca, Fe, Ni and Si were present in appreciable amounts, these were also used as marker for identification of relevant sources of $PM_{2.5}$

Metal- metal correlation during weekdays and weekend.

The Pearson correlation of metal-metal interaction during weekdays and weekends are summarized in Table 3 and 4. In this study, we used Mahilang et al. (2020) definition of strong positive correlation coefficients as values ranging from $r^2 \geq 0.5$ (highlighted in bold), indicating similar sources, and negative or weak correlation coefficients as values ranging from $r^2 \leq 0.5$ (highlighted in italic), indicating diverse metal aerosol sources. Elements associated with road and soil dust (Ca, Fe, K, Zn, Pb, Si, Ti and As) showed significant positive correlation with each other during weekdays and weekends suggesting the metals originates from the same sources. Br is among the elements that showed strong positive correlation with elements associated with road and soil dust (K, Ca, Fe, Zn, Pb, Si, Ti and As), also with Cl during weekdays and weekends During weekdays and weekends, respectively, the elements (V and U) and (V and Sb) showed no correlation with other metals. K and Ca were well correlated ($r^2 = 0.71$; $r^2 = 0.73$) during weekdays and weekends. K and Cl also showed strong significant correlation ($r^2 = 0.77$; $r^2 = 0.82$) during weekdays and weekends, respectively. This association was possibly due to biomass and coal burning since K has been attributed to biomass burning (Ryu et al. 2007) and Cl has been considered as an elemental tracer for coal combustion (Zheng et al. 2005). The presence of a weakly or negatively correlated relationship between other metals on weekdays or weekends suggests that different sources are contributing to the metal concentrations in the ambient air.

Sources of $PM_{2.5}$ identified by PMF

As previously mentioned, there are limited studies in South Africa that explored the sources of air pollution and used PMF in their investigation. The total daily $PM_{2.5}$ elemental composition dataset from the sampling campaign were used in the PMF model. Five to seven factors were investigated (Table S4) but the seven factors output is presented below. Fig 2 shows the specified source profiles, along with details on the mass contribution and percentage of the variables within each source, as done in Molnar et al. (2017). The mean $PM_{2.5}$ concentrations of the seven sources (i.e. are local and anthropogenic in origin) were named; Fossil fuel combustion, Soil dust, Secondary Sulphur, Vehicle exhaust, Road traffic, Base metal/pyrometallurgical, Coal burning can be found in Table 1 and Fig 2. Fossil fuel was found to be the major contributor with 22% of elemental mass in $PM_{2.5}$ concentrations on a yearly average. Coal combustion came in second, accounting for 18% of the total $PM_{2.5}$ mass. The combined contribution of soil dust and road traffic amounts to 4.0 $\mu g/m^3$ (22% of the total $PM_{2.5}$). Below is more detail on each of the source factors and their relative contributions.

Factor 1 – fossil fuel combustion

This factor is characterised by Cu, Se, Pb and U. Se, Pb and U are present in this factor as fossil fuel combustion and account for 4.1 $\mu g/m^3$ (22%) of the mass concentration. Increased concentration during the winter due to the use of coal for heating is consistent with Wåhlin et al. (2006) results, and U has been discovered to be a by-product of coal ash. When compared to an urban background, Jayasekher (2009) found that Se and Pb with other elements (such as Ni and As) increased in concentration near a coal-fired thermal plant. Possible source of lead in this factor may be from wear (tires, brake linings, clutch) rather than fuel since lead has been banned in fuel since 2006 in South Africa.

Factor 2 – Soil dust

The dust in the soil is typically wind-blown dust from the surrounding area. Si, K, Ca, Ti, and Fe are the main components of this factor. This component is known as soil dust because it accounted for 75, 45, 64, 57, and 27% of these elements, respectively. The annual mean contribution of this source to $PM_{2.5}$ was 1.4 μg/m^3, or 8%. In another study in South Africa, Tshehla and Djolov (2018) discovered that mineral dust, which includes soil and road re-suspended dust, is marked by the existence of Al, Si, Ti, and Ca. According to Gaita et al. (2014), the devoid of S, Zn, and Pb in this factor distinguishes mineral dust from re-suspended road dust in Kenya. It's also possible that regional transportation of atmospheric aerosols is a significant contributor to this factor. Ca was found to have a higher concentration than Ti and Fe in this study, implying that Ca-rich dust is a product of construction activities (Wang et al., 2015).

Factor 3 - Secondary Sulphur

This factor was found to contribute 2.7 μg/m^3 (15%) to the total $PM_{2.5}$ mass concentrations and is mainly characterised by the element S (80%) with traces of BC (17%). Secondary sulphur generally comes from various regional combustion processes. This study agrees with other studies (Cohen et al., 2010; Lee et al., 2011; Yu et al., 2013) that identified this factor in most urban PMF studies. The presence of S and BC in this factor also reveals that coal combustion could be an important contributor to the ambient PM levels.

Factor 4 – Vehicle exhaust

This factor is mainly characterised by Zn, Cl and Cu with a smaller amount of BC (22%). This is a mixed source of vehicle exhaust and industry emissions which contributes 2.5 μg/m^3, or 14% to the total $PM_{2.5}$ on an annual basis. Cu and Zn are major additives to lubricating oils. The presence of carbonaceous species is created by the vehicle exhaust tail pipe, while Cu is formed by both fuel/lubricant combustion and brake abrasions (Pant and Harrison, 2013). The presence of Cl in this factor may be due to coal combustion in industrial processes (Sun et al., 2013). The rapid development and car ownership in the city are continuous and therefore increase automobile exhaust emission which plays a crucial source of urban air pollution.

Factor 5 – Road traffic

This source group contains a broad variety of emissions arising from different vehicle types and processes. The Road traffic represented 14% of the total $PM_{2.5}$ mass. This source is characterised by the presence of V, Sb, Ba, Ca and Se. Ba, and also Cu and Zn, have been reported to be associated with tail pipe emissions, brake and tyre wear, and Sb is often used as a brake wear dust tracer. (Iijima et al., 2008, 2009).

Factor 6 – Base metal/ pyrometallurgical-related factors

This factor is described by Ni, Fe, V, and U (9%), which are due to a combination of pyrometallurgical and base metal processing factors. Except for Cr, the trace elements contained in this factor are identical to those found in a study by Venter et al. (2017), which defined this associated factor as pyrometallurgical-related factor. Fe and Cr have been linked to ferrochromium smelters in South Africa's Greater-Tubatse Municipality (GTM) and Bushveld Igneous Complex (Tshehla and Djolov 2018), while Ni has been linked to base metal smelters that refine base metals from PGM production processes (Venter et al., 2017). The presence of Ni (88%) and V (33%) in this factor is consistent with previous research, which found that Ni and V were used as residual oil combustion markers (Moreno et al., 2010; Lee et al., 2011). The mixture of these metals, on the other hand, can be traced back to metal

manufacturing. A similar factor in the source apportionment of $PM_{2.5}$ in Beijing was also identified by Wang et al. (2009).

Factor 7- Coal burning

Coal burning also contributed significant amount (3.3µg/m^3 or 18%) to the total $PM_{2.5}$ mass. In South Africa, coal has been found to be a major source both for electricity generation and household use. This factor is mainly characterised by BC, K, As, Br, Pb and Cl. The contribution of coal combustion from the coal fire plants in the province is close to the findings of a study by Tshehla and Djolov (2018), who reported 22% contribution to ambient PM. The existence of As and Pb in this source confirms Wåhlin et al. (2006) findings that As, Se, and Pb are common anthropogenic sources of coal combustion. Cl is used as a coal-burning marker (Owoade et al., 2015). Its existence in coal-burning has been documented to come from ammonium chloride, which is formed when ammonia and hydrogen chloride interact in the atmosphere. Elements found in this factor, such as As, Pb, K, and Br, have been linked to coal ash and have been linked to serious health issues in humans. (Lockwood and Evans, 2014).

Seasonal variations of $PM_{2.5}$ sources

Seasonal behaviour has been discovered in the known sources (see Table 1 and Fig 3). When compared to other seasons, the contribution of secondary sulphur, vehicle exhaust, base metal/pyrometallurgical, and coal burning to $PM_{2.5}$ concentration levels was found to be significantly higher during winter. In comparison to other seasons, levels of fossil fuel, soil dust, and road traffic were higher in autumn ($p<0.001$ for all tests). Since coal is the most common fuel type used by residents of informal settlements and a major source of power generation in South Africa, the increased contribution from coal burning of 18 percent during winter was anticipated. The increased secondary sulphur (15%) during winter was also expected since sulphate is among the product released from coal combustion. This observation is in contrast with other studies that reported high concentration of sulphate in summer (Yu et al., 2013; Huang et al., 2019). Lower relative humidity and wind speed in South Africa during the autumn and winter months, as expected, may also lead to soil dust contribution to PM sources, as this condition promotes dust formation and resuspension (Owoade et al., 2015).

The fossil fuel combustion contributed about 22% annually to the total $PM_{2.5}$ mass, this also formed significant fraction of the mass. The increased level observed during autumn and winter was expected as this is the onset of cold season respectively, in South Africa. The air masses from the neighbouring countries such as Mozambique, Zimbabwe, Zambia could also contribute to the elevated amount recorded during autumn over winter (Fig S2). This region has been characterised with biomass burning emissions which play an important role in regional transportation of PM to South Africa (Hersey et al., 2015). The use of coal and oil for heating and cooking, which leads to pollution levels, may be responsible for the higher level of $PM_{2.5}$ during this period. The most significant point sources of pollution in Gauteng province, where the study is being conducted, are three coal-fired power plants, two petrochemical refineries, five steel manufacturing facilities, and three cement factories (Hersey et al., 2015).

The increased levels of $PM_{2.5}$ from coal burning, secondary sulphur and base metal/pyrometallurgical during winter are most likely to be attributed to the activities carried out in the sources mentioned above. Coal, for example, is widely used in coal-fired power plants and is also used for cooking and space heating in some rural

areas of South Africa, with demand expected to increase during the winter. The findings corroborate South African research that found the highest levels of PM_{10} and $PM_{2.5}$ in townships and informal settlements during the winter, when domestic burning is most commonly used for heating and cooking. (von Schirnding et al., 2002; Wichmann and Voyi 2005; Williams et al., 2020). During the winter and autumn, increased contributions from vehicle exhaust and road traffic were also expected, as cold starts from vehicles would increase emissions.

Transport clusters and potential source areas

Apart from local sources, air pollutants transport and regional sources have a significant impact on Pretoria's air quality. Five transport cluster pathways were identified in this study. Long-range transport (LRT) and local sources characterised these pathways. Cluster 1 North Limpopo (NLP) is a local source from the northern province of Limpopo (Fig 4), which is associated with a high concentration of $PM_{2.5}$ due to anthropogenic activities happening in the area, which include coal power stations, mining, domestic fuel burning, agriculture and veld fires. Although cluster 2 Eastern Inland (EI) has its origin from the Indian Ocean with minimal marine influence, the direction of air parcels passes over the Mozambique region via Mpumalanga which is associated with mining, biomass burning, waste burning, vehicles emission and industrial activities.

Clusters 3, 4 and 5 (Short Indian Ocean SIO, Long Indian Ocean LIO and South Westerly-Atlantic Ocean, respectively) depicts LRT sources. Clusters 3 and 4 are oceanic and associated with low concentration due to the oceanic effect while cluster 5 loses it marine influence as it travels over the Western Cape region. The transport clusters characterize unique transport routes and are connected to certain sources (Tang et al., 2014). For instance, the highest mean level for $PM_{2.5}$ (66.8 $\mu g/m^3$) that occurred on June 2, 2017 was found to be from a local source via NLP as mentioned earlier. Furthermore, about 78% of the days in which the WHO and SA standard were exceeded were of local source origin (cluster NLP and EI), while only 22% was attributed to LRT sources. The annual frequencies of the five long-range transport clusters (clusters 3, 4 and 5) were 11%, 4% and 6%, respectively.

In this study, air masses were distributed in three ways: in the winter, air masses were directionally distributed; in the spring, air masses came from the north and east, with contributions from the southwest and southeast; and in the autumn and summer, air masses came from the east and southeast (Fig S2). This supports the results of Hersey et al. (2015) in a South African study, which found that Gauteng air masses are directionally distributed in winter but have distinct directions in other seasons.

$PM_{2.5}$ source apportionment for each cluster

The combination of the PMF model's seven sources and the five transport clusters allows for a comparison of source intensity and separate transport clusters, Table 5 and Fig 5 demonstrate this. Regardless of cluster, vehicle exhaust/fossil fuel sources were the most significant contributors, followed by coal burning and secondary sulphur emissions. The Base metal/pyrometallurgical related factors had the highest concentration when their air masses originated nationally from Limpopo to the receptor site (NLP) followed by cluster EI and cluster SWA, respectively. Cluster SIO and cluster LIO were the lowest. The vehicle exhaust/fossil fuel peaked when the air masses originated from NLP and cluster SIO which are both national and LRT.

When air masses came from NLP, secondary sulphur had the highest concentration, followed by SIO and LIO, and SWA had the lowest. Soil dust was found in all transport clusters, with the highest concentration in NLP and

the lowest in cluster EI and SIO. Coal burning also had a significant contribution when air masses originated from cluster NLP, SWA, LIO, EI and SIO respectively. As air masses come from SWA, the highest concentrations of road traffic and vehicle exhaust are reported, while the lowest concentrations are recorded in cluster SIO and LIO, respectively.

$PM_{2.5}$ sources and transport clusters

According to Molnar et al. (2017), the mixture of source strength and frequency determines the average size of the impact from various source areas and source forms at a receptor point. The highest mean concentrations found in this study are attributed to air masses originating from SIO, despite the fact that concentrations were lower on days when air masses from LIO and EI were present (Table 3). LIO produced the least common air masses (4 out of 122, or 3.3% of the days), while EI and SIO recorded air masses on 40 and 41 days, respectively (33% and 34% of the days). When the mean contribution (Fig 5) was multiplied by the frequency of the air masses, a different result was obtained (Fig 6) (Molnar et al., 2017). On a daily average, NLP, EI, and SIO were the source regions that led to increased $PM_{2.5}$ levels in Pretoria, 3.3–6.6 µg/m^3, while LIO and SWA contributed just 0.3–0.8 µg/m^3. Regional and local sources continue to be the most important contributors.

Seasonality of the transport cluster

The highest mean $PM_{2.5}$ concentrations observed in this study during the winter indicate that the air masses originated from cluster 1 (Fig S1), i.e. highly contaminated areas where mining and smelting processes were prevalent, making them possible heavy metal transporters to the receptor site. This accounts for 51% of the transport clusters bringing pollutants to the sampling site. During autumn clusters 1 and 2 accounted for 87% of the local transport clusters responsible for the pollutants arriving at the sampling site. This is also responsible for the elevated mean concentration of $PM_{2.5}$ during the season. In spring and summer, cluster 2 with 18 and 20 days (60 and 67%), respectively were responsible for the local air masses reaching the site.

Emission control strategies

To effectively develop $PM_{2.5}$ reduction strategies, the sources of PM and their contributions, as well as the contribution from each source, must be understood using a combination of transport cluster analysis. The later provides additional information on the pathway through which the sources originated. We were able to categorised the contributions of the ambient PM concentration as of natural or anthropogenic origin, and either local, transboundary or LRT. The local sources in Pretoria includes fossil fuel combustion, secondary sulphur, vehicle exhaust, road traffic, base metal/pyrometallurgical, coal burning, while natural source to some extent is soil dust. Not only local sources increase the ambient PM levels, but also the regional/transboundary and LRT. Emission reduction strategies within the Southern African developing countries that focuses on mitigating industrial emission, road traffic emissions, coal burning, fossil fuel combustion will be of great benefits to Pretoria, Gauteng Province, and other surrounding provinces. On a local level, the city of Pretoria should concentrate on traffic, residential heating, power generation, and base metal/pyrometallurgical abatement strategies Immediate action to mitigate the effects of pollution in this region may include effective implementation of the Air Quality Act by authorities, electrification of areas where coal is used for heating and cooking, reduction of emissions from industrial stacks, and effective waste collection from households to avoid waste burning.

Conclusion

Due to chemical characterisation of $PM_{2.5}$, application of HYSPLIT, and PMF, the one-year sampling campaign offered a better understanding of the sources and strengths in the growth of $PM_{2.5}$ in urban background of Pretoria, South Africa. Seven sources which contributes to the $PM_{2.5}$ levels in the city, namely fossil fuel combustion, soil dust, secondary sulphur, vehicle exhaust, road traffic, base metal/pyrometallurgical, coal/biomass burning were identified. Secondary sulphur and coal burning showed the highest contribution during the cold seasons.

Local and regional sources of $PM_{2.5}$ accounted for more than 50% of the air masses reaching the site while 21% are LRT sources. The most frequent cluster was the short Indian ocean (34% of the year) followed by the Eastern Inland. The source areas for the highest $PM_{2.5}$ levels were known as Northern Limpopo and the short Indian Ocean, which occurred for 31 and 41 days, respectively. The NLP was found to be the highest daily mean contributor of $PM_{2.5,}$ which happens to be from a high polluted area.

The most successful interventions to combat the effects of $PM_{2.5}$ will be emission reduction policies aimed at reducing coal burning, secondary sulphur, and fossil fuel combustion from Southern Africa and local sources on a city-scale level. Lastly, the findings of the source apportionment study are important as this will support the reviewing and drafting of the air quality management plan of the country which currently under review.

Acknowledgements The authors gratefully acknowledge the NOAA Air Resources Laboratory for providing the HYSPLIT model and the NCEP Reanalysis datasets, as well as the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA**.** The authors would like to thank the South African Weather Service for the air pollution and meteorological data, University of Pretoria and the National Research Foundation for the doctoral scholarship of A.A.A.

Funding information. JW received funding from the South African National Research Foundation under the Collaborative Postgraduate Training Programme grant (CPT160424162937).

Conflict of interest. The authors declare that they have no competing interest.

Compliance with ethical standards. The Research Ethics Committee, Faculty of Health Sciences, University of Pretoria, approved ethics approval (reference 469/2017) in 2017.

Availability of data and materials. The datasets generated and/or analysed during the current study are not publicly available due to university's intellectual property right but are available from the corresponding author on reasonable request

Authors' contributions. JW conceptualize the research**.** The research design, methodology, analysis of findings, and writing of the manuscript were all the responsibility of all authors. The statistical analyses were carried out by AA and PM. JB was responsible for the chemical analysis.

References

Arku, R. E., Vallarino, J., Dionisio, K. L., Willis, R., Choi, H., Wilson, J. G., et al. (2008). Characterizing air pollution in two low-income neighborhoods in Accra, Ghana. *Science of the Total Environment*, *402*, 217-231. https://doi.org/10.1016/j.scitotenv.2008.04.042

Bisht, D.S., Dumka, U.C., Kaskaoutis, D.G., Pipal, A.S., Srivastava, A.K., Soni, V.K., Attri, S.D., Sateesh, M. and Tiwari, S., (2015). Carbonaceous aerosols and pollutants over Delhi urban environment: temporal evolution, source apportionment and radiative forcing. *Science of the Total Environment*, *521*, pp.431-445. http://dx.doi.org/10.1016/j.scitotenv.2015.03.083

Boman, J., Lindén, J., Thorsson, S., Holmer, B., & Eliasson, I. (2009). A tentative study of urban and suburban fine particles (PM2.5) collected in Ouagadougou, Burkina Faso. *X‐Ray Spectrometry: An International Journal*, *38*, 354-362. https://doi.org/10.1002/xrs.1173

Brunekreef, B., & Holgate, S. T. (2002). Air pollution and health. *The lancet*, *360*, 1233-1242.

Cavanagh, J.A.E., Zawar-Reza, P. and Wilson, J.G., (2009). Spatial attenuation of ambient particulate matter air pollution within an urbanised native forest patch. *Urban Forestry & Urban Greening*, *8*(1), pp.21-30. https://doi.org/10.1016/j.ufug.2008.10.002

Celo, V. and Dabek-Zlotorzynska, E., (2010). Concentration and source origin of trace metals in PM 2.5 collected at selected Canadian sites within the Canadian national air pollution surveillance program. In *Urban airborne particulate matter* (pp. 19-38). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12278-1_2

Chakraborty, A. and Gupta, T., (2010). Chemical characterization and source apportionment of submicron (PM1) aerosol in Kanpur region, India. *Aerosol and Air Quality Research*, *10*(5), pp.433-445. https://doi.org/10.4209/aaqr.2009.11.0071

Cohen, D. D., Crawford, J., Stelcer, E., & Bac, V. T. (2010). Characterisation and source apportionment of fine particulate sources at Hanoi from 2001 to 2008. *Atmospheric Environment*, *44*, 320-328. https://doi.org/10.1016/j.atmosenv.2009.10.037

Conradie, E. H., Van Zyl, P. G., Pienaar, J. J., Beukes, J. P., Galy-Lacaux, C., Venter, A. D., & Mkhatshwa, G. V. (2016). The chemical composition and fluxes of atmospheric wet deposition at four sites in South Africa. *Atmospheric environment*, *146*, 113-131. https://doi.org/10.1016/j.atmosenv.2016.07.033

Dorling, S. R., Davies, T. D., & Pierce, C. E. (1992). Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry—Results from Eskdalemuir, south Scotland. *Atmospheric Environment. Part A. General Topics*, *26*, 2583-2602. https://doi.org/10.1016/0960-1686(92)90111-W

Draxler, R.R., Rolph, G.D. (2003). HYSPLIT (Hybrid Single-Particle Lagrangian Integrated Trajectory) Model https://www.ready.noaa.gov/HYSPLIT.php.

Duan, F. K., He, K. B., Ma, Y. L., Yang, F. M., Yu, X. C., Cadle, S. H., et al. (2006). Concentration and chemical characteristics of PM2. 5 in Beijing, China: 2001–2002. *Science of the Total Environment*, *355*, 264-275. https://doi.org/10.1016/j.scitotenv.2005.03.001

Engelbrecht, J. P., Swanepoel, L., Chow, J. C., Watson, J. G., & Egami, R. T. (2002). The comparison of source contributions from residential coal and low-smoke fuels, using CMB modeling, in South Africa. *Environmental Science and Policy*, *5*, 157-167. https://doi.org/10.1016/S1462-9011(02)00029-1

Engelbrecht, J. P., Swanepoel, L., Zunckel, M., Chow, J. C., Watson, J. G., & Egami, R. T. (2000). Modelling PM10 aerosol data from the Qalabotjha low-smoke fuels macro-scale experiment in South Africa. *Ecological Modelling*, *127*, 235-244. https://doi.org/10.1016/S0304-3800(99)00212-4

Formenti, P., Annegarn, H. J., & Piketh, S. J. (1998). Time resolved aerosol monitoring in the urban centre of Soweto. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *136*, 948-954. https://doi.org/10.1016/S0168-583X(97)00770-2

Gaita, S. M., Boman, J., Gatari, M. J., Pettersson, J. B., & Janhäll, S. (2014). Source apportionment and seasonal variation of PM2.5 in a Sub-Saharan African city: Nairobi, Kenya. *Atmospheric Chemistry and Physics*, *14*, 9977-9991. https://doi.org/10.5194/acp-14-9977-2014

Gu, J., Pitz, M., Schnelle-Kreis, J., Diemer, J., Reller, A., Zimmermann, R., et al. (2011). Source apportionment of ambient particles: comparison of positive matrix factorization analysis applied to particle size distribution and chemical composition data. *Atmospheric Environment*, *45*, 1849-1857. https://doi.org/10.1016/j.atmosenv.2011.01.009

Hersey, S. P., Garland, R. M., Crosbie, E., Shingler, T., Sorooshian, A., Piketh, S., & Burger, R. (2015). An overview of regional and local characteristics of aerosols in South Africa using satellite, ground, and modeling data. *Atmospheric Chemistry and Physics*, *15*, 4259-4278. https://doi.org/10.5194/acp-15-4259-2015

Heo, J. B., Hopke, P. K., & Yi, S. M. (2009). Source apportionment of PM2.5 in Seoul, Korea. *Atmospheric Chemistry and Physics*, *9*, 4957-4971. https://doi.org/10.5194/acp-9-4957-2009

Huang, F., Zhou, J., Chen, N., Li, Y., Li, K., & Wu, S. (2019). Chemical characteristics and source apportionment of PM2.5 in Wuhan, China. *Journal of Atmospheric Chemistry*, *76*, 245-262. https://doi.org/10.1007/s10874-019-09395-0

Hunter, D., Salzman, J., Zaelke, D. (2007). International environmental law and policy (Vol. 516). New York: Foundation Press.

Iijima, A., Sato, K., Fujitani, Y., Fujimori, E., Saito, Y., Tanabe, K., et al. (2009). Clarification of the predominant emission sources of antimony in airborne particulate matter and estimation of their effects on the atmosphere in Japan. *Environmental Chemistry*, *6*, 122-132. https://doi.org/10.1071/EN08107

Iijima, A., Sato, K., Yano, K., Kato, M., Kozawa, K., & Furuta, N. (2008). Emission factor for antimony in brake abrasion dusts as one of the major atmospheric antimony sources. *Environmental Science and Technology*, *42*, 2937-2942. https://doi.org/10.1021/es702137g

Janssen, N.A., Hoek, G., Simic-Lawson, M., Fischer, P., Van Bree, L., Ten Brink, H., Keuken, M., Atkinson, R.W., Anderson, H.R., Brunekreef, B. and Cassee, F.R., (2011). Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM10 and PM2.5. *Environmental health perspectives*, *119*(12), pp.1691-1699. https://doi.org/10.1289/ehp.1003369

Jayasekher, T. (2009). Aerosols near by a coal fired thermal power plant: Chemical composition and toxic evaluation. *Chemosphere*, *75*, 1525-1530. https://doi.org/10.1016/j.chemosphere.2009.02.001

Kappos, A. D., Bruckmann, P., Eikmann, T., Englert, N., Heinrich, U., Höppe, P., et al. (2004). Health effects of particles in ambient air. *International Journal of Hygiene and Environmental Health*, *207*, 399-407. https://doi.org/10.1078/1438-4639-00306

Krzyzanowski, M., Apte, J. S., Bonjour, S. P., Brauer, M., Cohen, A. J., & Prüss-Ustun, A. M. (2014). Air pollution in the mega-cities. *Current Environmental Health Reports*, *1*, 185-191. https://doi.org/10.1007/s40572-014-0019-7

Lee, H. J., Gent, J. F., Leaderer, B. P., & Koutrakis, P. (2011). Spatial and temporal variability of fine particle composition and source types in five cities of Connecticut and Massachusetts. *Science of the Total Environment*, *409*, 2133-2142. https://doi.org/10.1016/j.scitotenv.2011.02.025

Li, Y., Henze, D.K., Jack, D., Henderson, B.H., Kinney, P.L., (2016). Assessing public health burden associated with exposure to ambient black carbon in the United States. Sci. Total Environ. 539, 515–525. https://doi.org/10.1016/j.scitotenv.2015.08.129

Lockwood, A. H., & Evans, L. (2014). Ash in lungs: how breathing coal ash is hazardous to your health. A report from Physicians for Social Responsibility and Earthjustice. *https://earthjustice. org/sites/default/files/files/Ash_In_Lungs_1. pdf*.

Lokotola, C. L., Wright, C. Y., & Wichmann, J. (2020). Temperature as a modifier of the effects of air pollution on cardiovascular disease hospital admissions in Cape Town, South Africa. *Environmental Science and Pollution Research*, *27*, 16677-16685. https://doi.org/10.1007/s11356-020-07938-7

Mahilang, M., Deb, M.K., Nirmalkar, J. and Pervez, S., (2020). Influence of fireworks emission on aerosol aging process at lower troposphere and associated health risks in an urban region of eastern central India. *Atmospheric pollution research*, *11*(7), pp.1127-1141. https://doi.org/10.1016/j.apr.2020.04.009

Martins, V., Moreno, T., Minguillón, M. C., Van Drooge, B. L., Reche, C., Amato, F., et al. (2016). Origin of inorganic and organic components of PM2.5 in subway stations of Barcelona, Spain. *Environmental Pollution*, *208*, 125-136. https://doi.org/10.1016/j.envpol.2015.07.004

Mathuthu, M., Dudu, V. P., & Manjoro, M. (2018). Source Apportionment of Air Particulates in South Africa: A Review. *Atmospheric and Climate Sciences*, *9*, 100-113. https://doi.org/10.4236/acs.2019.91007

Molnár, P., Johannesson, S., & Quass, U. (2014). Source apportionment of PM2.5 using positive matrix factorization (PMF) and PMF with factor selection. *Aerosol and Air Quality Research*, *14*, 725-733. https://doi.org/10.4209/aaqr.2013.11.0335

Molnár, P., Tang, L., Sjöberg, K., & Wichmann, J. (2017). Long-range transport clusters and positive matrix factorization source apportionment for investigating transboundary PM2.5 in Gothenburg, Sweden. *Environmental Science: Processes and Impacts*, *19*, 1270-1277. https://doi.org/10.1039/C7EM00122C

Moon, K. J., Han, J. S., Ghim, Y. S., & Kim, Y. J. (2008). Source apportionment of fine carbonaceous particles by positive matrix factorization at Gosan background site in East Asia. *Environment International*, *34*, 654-664. https://doi.org/10.1016/j.envint.2007.12.021

Moreno, T., Querol, X., Alastuey, A., de la Rosa, J., de la Campa, A. M. S., Minguillón, M., et al. (2010). Variations in vanadium, nickel and lanthanoid element concentrations in urban air. *Science of the Total Environment*, *408*, 4569-4579. https://doi.org/10.1016/j.scitotenv.2010.06.016

Ni, T., Li, P., Han, B., Bai, Z., Ding, X., Wang, Q., Huo, J. and Lu, B., (2013). Spatial and temporal variation of chemical composition and mass closure of ambient PM10 in Tianjin, China. *Aerosol and Air Quality Research*, *13*(6), pp.1832-1846. https://doi.org/10.4209/aaqr.2012.10.0283

Norris, G., Vedantham, R., Wade, K., Brown, S., Prouty, J., Foley, C. (2008). EPA positive matrix factorization (PMF) 3.0 fundamentals & user guide. Prepared for the US Environmental Protection Agency, Washington, DC, by the National Exposure Research Laboratory, Research Triangle Park.

Novela, R., Gitari, M., Chikoore, H., & Wichmann, J. (2019). Quantification of fine particulate matter in ambient air, chemical characterization and the geographical origin of the air masses passing through Thohoyandou, South Africa. *Environmental Epidemiology*, *3*, 438. https://doi: 10.1097/01.EE9.0000610936.96119.4a

Owoade, K. O., Hopke, P. K., Olise, F. S., Ogundele, L. T., Fawole, O. G., Olaniyi, B. H., et al. (2015). Chemical compositions and source identification of particulate matter (PM2.5 and PM2.5–10) from a scrap iron and steel smelting industry along the Ife–Ibadan highway, Nigeria. *Atmospheric Pollution Research*, *6*, 107-119. https://doi.org/10.5094/APR.2015.013

Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non- negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, *5*, 111-126. https://doi.org/10.1002/env.3170050203

Paatero, P. (1997). Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems*, *37*, 23-35. https://doi.org/10.1016/S0169-7439(96)00044-5

Pant, P., & Harrison, R. M. (2013). Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: a review. *Atmospheric Environment*, *77*, 78-97. https://doi.org/10.1016/j.atmosenv.2013.04.028

Pauw, C. J., Friedl, A., Holm, D., John, J., Kornelius, G., Oosthuizen, R., & Van Niekerk, A. S. (2008). Air Pollution in Dense, Low-income Settlements in South Africa: Issue 1. *Report to the Royal Danish Embassy and the Dept. of Environmental Affairs and Tourism*, *6*.

Raja, S., Chandrasekaran, S. R., Lin, L., Xia, X., Hopke, P. K., & Valsaraj, K. T. (2016). Analysis of beta attenuation monitor filter rolls for particulate matter speciation. *Aerosol and Air Quality Research*, *17*, 14-23. https://doi.org/10.4209/aaqr.2016.03.0122

Reff, A., Eberly, S. I., & Bhave, P. V. (2007). Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. *Journal of the Air and Waste Management Association*, *57*, 146-154. https://doi.org/10.1080/10473289.2007.10465319

Rupp, G. (2009). Aerosol dynamics and health: Strategies to reduce exposure and harm. *Biomarkers*, *14*(sup1), 3-4. https://doi.org/10.1080/13547500903145458

Ryu, S.Y., Kwon, B.G., Kim, Y.J., Kim, H.H. and Chun, K.J., (2007). Characteristics of biomass burning aerosol and its impact on regional air quality in the summer of 2003 at Gwangju, Korea. *Atmospheric Research*, *84*(4), pp.362-373. https://doi.org/10.1016/j.atmosres.2006.09.007

Ryu, Y.H., Baik, J.J., Kwak, K.H., Kim, S. and Moon, N. (2013). Impacts of urban land-surface forcing on ozone air quality in the Seoul metropolitan area. *Atmospheric Chemistry and Physics*, *13*(4), pp.2177-2194. https://doi.org/10.5194/acp-13-2177-2013

Salma, I., Maenhaut, W., Cafmeyer, J., Annegarn, H. J., & Andreae, M. O. (1994). PIXE analysis of cascade impactor samples collected at the Kruger National Park, South Africa. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *85*, 849-855. https://doi.org/10.1016/0168-583X(94)95936-6

Sandradewi, J., Prévôt, A. S. H., Weingartner, E., Schmidhauser, R., Gysel, M., & Baltensperger, U. (2008). A study of wood burning and traffic aerosols in an Alpine valley using a multi-wavelength Aethalometer. *Atmospheric Environment*, *42*, 101-112. https://doi.org/10.1016/j.atmosenv.2007.09.034

Sandradewi, J., Prévôt, A. S., Szidat, S., Perron, N., Alfarra, M. R., Lanz, V. A., et al. (2008). Using aerosol light absorption measurements for the quantitative determination of wood burning and traffic emission contributions to particulate matter. *Environmental Science and Technology*, *42*, 3316-3323. https://doi.org/10.1021/es702253m

SAS (2003) Statistical software version 9.3, SAS Institute Inc., Cary, NC.

Schwartz, J. (2004). Air pollution and children's health. *Pediatrics*, *113*(Supplement 3), 1037-1043.

Schwarz, J., Cusack, M., Karban, J., Chalupníčková, E., Havránek, V., Smolík, J., & Ždímal, V. (2016). PM2.5 chemical composition at a rural background site in Central Europe, including correlation and air mass back trajectory analysis. *Atmospheric Research*, *176*, 108-120. https://doi.org/10.1016/j.atmosres.2016.02.017

Scorgie, Y., Annegarn, H., & Burger, L. (2004). Study to examine the potential socioeconomic impact of measures to reduce air pollution from combustion, Fund for Research into Industrial Development Growth and Equity (FRIDGE). *Trade and Industry Chamber, Johannesburg*.

Shah, M.H., Shaheen, N. and Nazir, R., (2012). Assessment of the trace elements level in urban atmospheric particulate matter and source apportionment in Islamabad, Pakistan. *Atmospheric pollution research*, *3*(1), pp.39-45. https://doi.org/10.5094/APR.2012.003

Shaltout, A. A., Boman, J., Hassan, S. K., Abozied, A. M., Al-Ashkar, E. A., Abd-Elkader, O. H., et al. (2020). Elemental Composition of PM2.5 Aerosol in a Residential–Industrial Area of a Mediterranean Megacity. *Archives of Environmental Contamination and Toxicology*, *78*, 68-78. https://doi.org/10.1007/s00244-019-00688-9

Sun, Y. L., Wang, Z. F., Fu, P. Q., Yang, T., Jiang, Q., Dong, H. B., et al. (2013). Aerosol composition, sources and processes during wintertime in Beijing, China. *Atmospheric Chemistry and Physics*, *13*, 4577-4592. https://doi.org/10.5194/acp-13-4577-2013

Tang, L., Haeger-Eugensson, M., Sjöberg, K., Wichmann, J., Molnár, P., & Sallsten, G. (2014). Estimation of the long-range transport contribution from secondary inorganic components to urban background PM10 concentrations in south-western Sweden during 1986–2010. *Atmospheric Environment*, *89*, 93-101. https://doi.org/10.1016/j.atmosenv.2014.02.018

Teich, M., Pinxteren, D. V., Wang, M., Kecorius, S., Wang, Z., Müller, T., et al. (2017). Contributions of nitrated aromatic compounds to the light absorption of water-soluble and particulate brown carbon in different atmospheric environments in Germany and China. *Atmospheric Chemistry and Physics*, *17*, 1653-1672. https://doi.org/10.5194/acp-17-1653-2017

Tshehla, C., & Djolov, G. (2018). Source profiling, source apportionment and cluster transport analysis to identify the sources of PM and the origin of air masses to an industrialised rural area in Limpopo. *Clean Air Journal*, *28*, 54-66. http://dx.doi.org/10.17159/2410-972x/2018/v28n2a18

Viana, M., Kuhlbusch, T. A., Querol, X., Alastuey, A., Harrison, R. M., Hopke, P. K., et al. (2008). Source apportionment of particulate matter in Europe: a review of methods and results. *Journal of Aerosol Science*, *39*, 827-849. https://doi.org/10.1016/j.jaerosci.2008.05.007

Venter, A. D., Zyl, P. G. V., Beukes, J. P., Josipovic, M., Hendriks, J., Vakkari, V., & Laakso, L. (2017). Atmospheric trace metals measured at a regional background site (Welgegund) in South Africa. *Atmospheric Chemistry and Physics*, *17*, 4251-4263. https://doi.org/10.5194/acp-17-4251-2017

Von Schirnding, Y., Bruce, N., Smith, K., Ballard-Tremeer, G., Ezzati, M., & Lvovsky, K. (2002). Addressing the Impact of Household Energy and Indoor Air Pollution on the Health of Poor: Implications for Policy Action and Intervention Measures (p. 52). Geneva: World Health Organization.

Wåhlin, P., Berkowicz, R., & Palmgren, F. (2006). Characterisation of traffic-generated particulate matter in Copenhagen. *Atmospheric Environment*, *40*,2151-2159. https://doi.org/10.1016/j.atmosenv.2005.11.049

Wang, H., Zhou, Y., Zhuang, Y., Wang, X., & Hao, Z. (2009). Characterization of PM 2.5/PM 2.5–10 and source tracking in the juncture belt between urban and rural areas of Beijing. *Chinese Science Bulletin*, *54*, 2506-2515. https://doi.org/10.1007/s11434-009-0021-x

Wang, J. and Ogawa, S., (2015). Effects of meteorological conditions on PM2. 5 concentrations in Nagasaki, Japan. *International journal of environmental research and public health*, *12*(8), pp.9089-9101. https://doi.org/10.3390/ijerph120809089

Wang, P., Cao, J. J., Shen, Z. X., Han, Y. M., Lee, S. C., Huang, Y., et al. (2015). Spatial and seasonal variations of PM2. 5 mass and species during 2010 in Xi'an, China. *Science of the Total Environment*, *508*, 477-487. https://doi.org/10.1016/j.scitotenv.2014.11.007

Wichmann, J., Sjöberg, K., Tang, L., Haeger-Eugensson, M., Rosengren, A., Andersson, E. M., et al. (2014). The effect of secondary inorganic aerosols, soot and the geographical origin of air mass on acute myocardial infarction hospitalisations in Gothenburg, Sweden during 1985–2010: a case-crossover study. *Environmental Health*, *13*, 1-16.

Wichmann, J., & Voyi, K. V. (2005). Air pollution epidemiologic studies in South Africa—need for freshening up. *Rev Environ Health*, *20*, 265-301.

Williams, J., Petrik, L., & Wichmann, J. (2020). PM2.5 chemical composition and geographical origin of air masses in Cape Town, South Africa. *Air Quality, Atmosphere and Health*, 1-12. https://doi.org/10.1007/s11869-020-00947-y

WHO. (2013). Air pollution www.who.int/topics/air_ pollution/en/

Yin, S., Shen, Z., Zhou, P., Zou, X., Che, S. and Wang, W., (2011). Quantifying air pollution attenuation within urban parks: An experimental approach in Shanghai, China. *Environmental pollution*, *159*(8-9), pp.2155-2163. https://doi.org/10.1016/j.envpol.2011.03.009

Yu, L., Wang, G., Zhang, R., Zhang, L., Song, Y., Wu, B., et al. (2013). Characterization and source apportionment of PM2.5 in an urban environment in Beijing. *Aerosol and air quality research*, *13*, 574-583. https://doi.org/10.4209/aaqr.2012.07.0192

Zhang, Y. and Jiang, W., (2018). Pollution characteristics and influencing factors of atmospheric particulate matter (PM2.5) in Chang-Zhu-Tan area. In *IOP Conference Series: Earth and Environmental Science* (Vol. 108, No. 4, p. 042047). IOP Publishing. https://doi.org/10.1088/1755-1315/108/4/042047

Zheng, M., Salmon, L. G., Schauer, J. J., Zeng, L., Kiang, C. S., Zhang, Y., & Cass, G. R. (2005). Seasonal trends in PM2.5 source contributions in Beijing, China. *Atmospheric Environment*, *39*, 3967-3976. https://doi.org/10.1016/j.atmosenv.2005.03.036

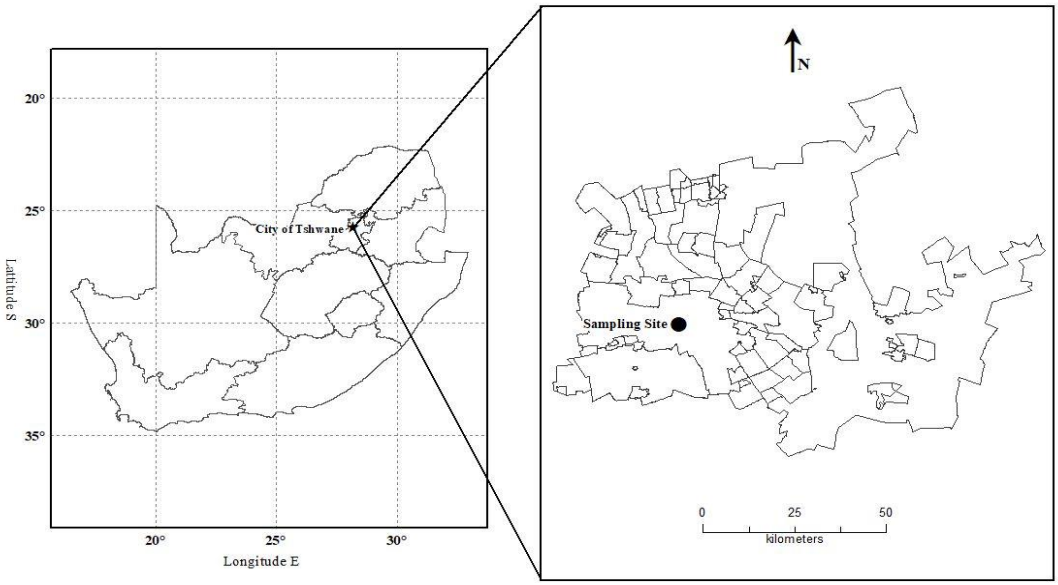

Fig. 1 Sketch map depicting the sampling location.

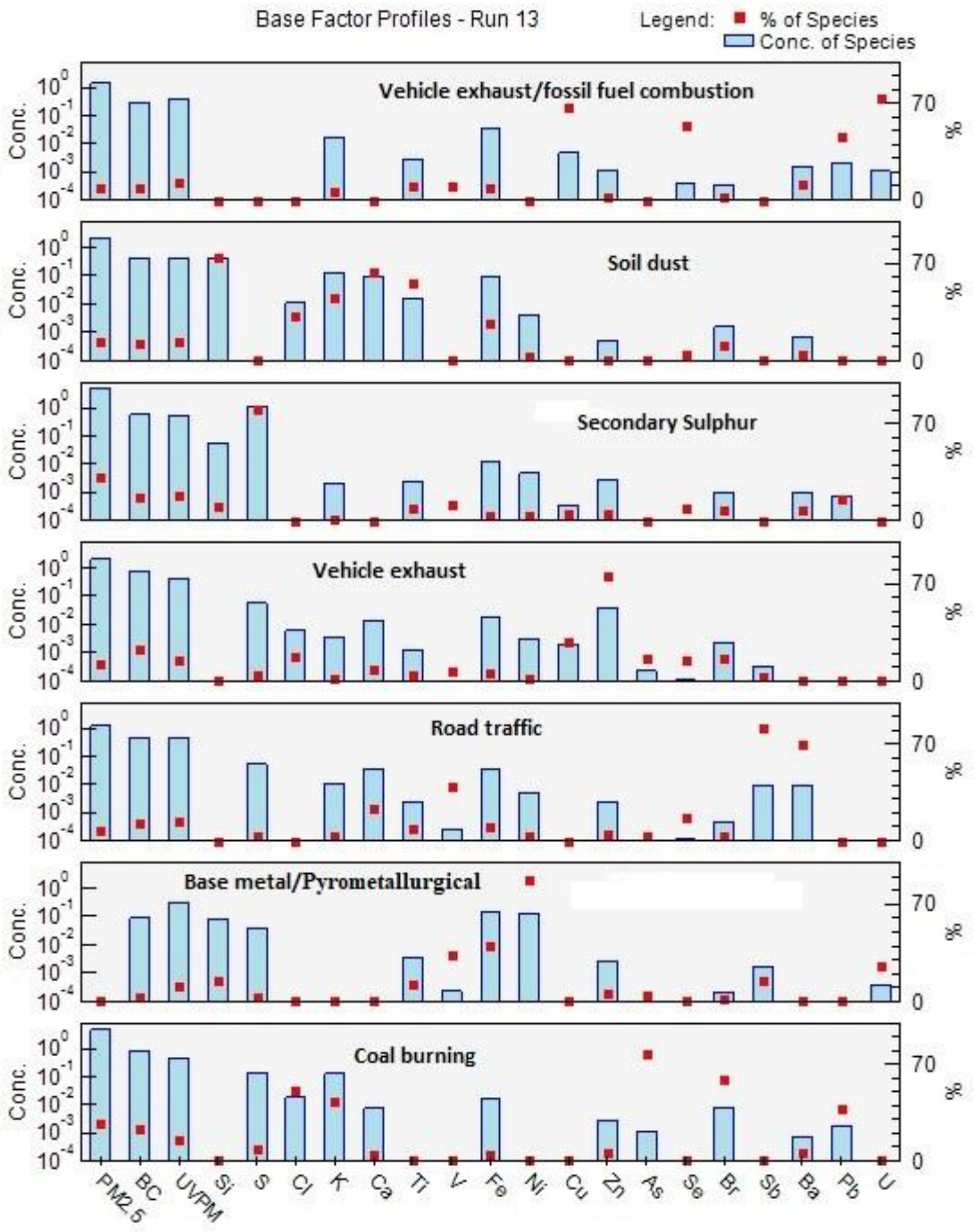

Fig. 2 PMF sources and mean concentration contribution (concentrations are in μg/m^3) between 18 April 2017 to 17 April 2018 in Pretoria, South Africa**.**

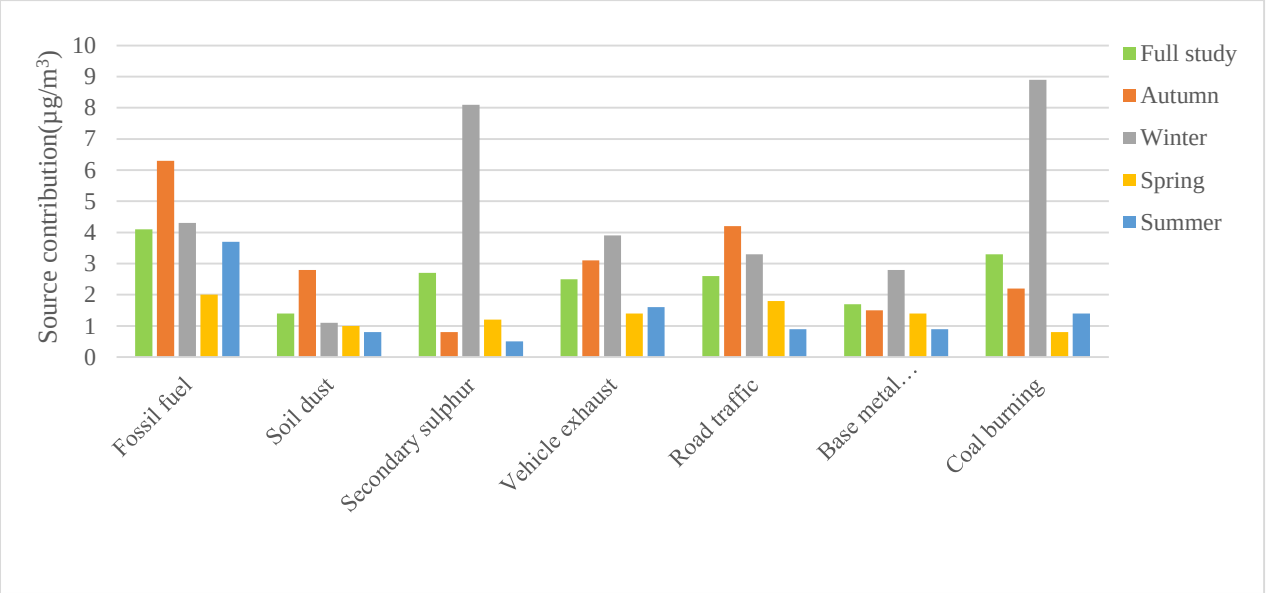

Fig. 3 Mean source contributions from the seven-factor positive matrix factorization model to the yearly mean $PM_{2.5}$ concentrations.

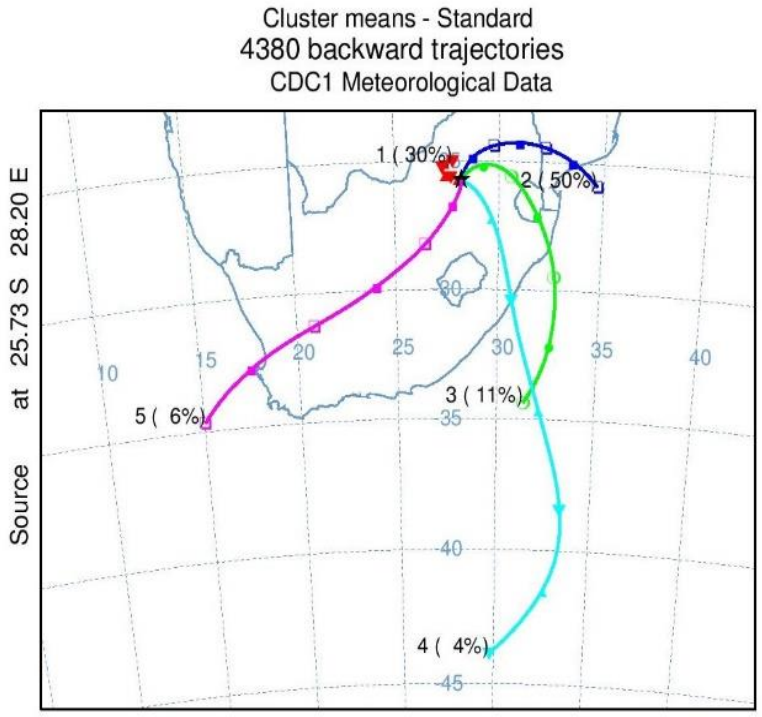

Fig. 4 Five transport pathways (cluster plots) arriving at the sampling site during 18 April 2017 and 18 April 2018

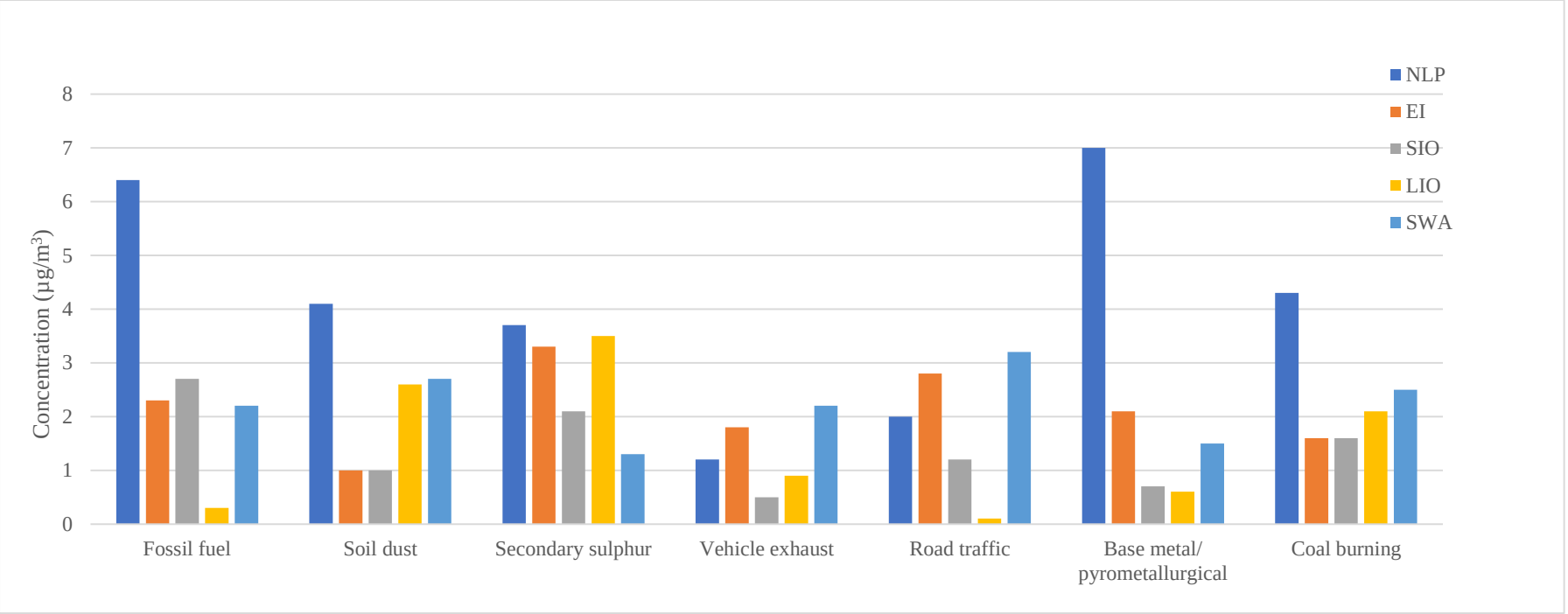

Fig. 5 Mean contribution to the $PM_{2.5}$ concentration per source and transport cluster from 18 April 2017 to 17 April 2018

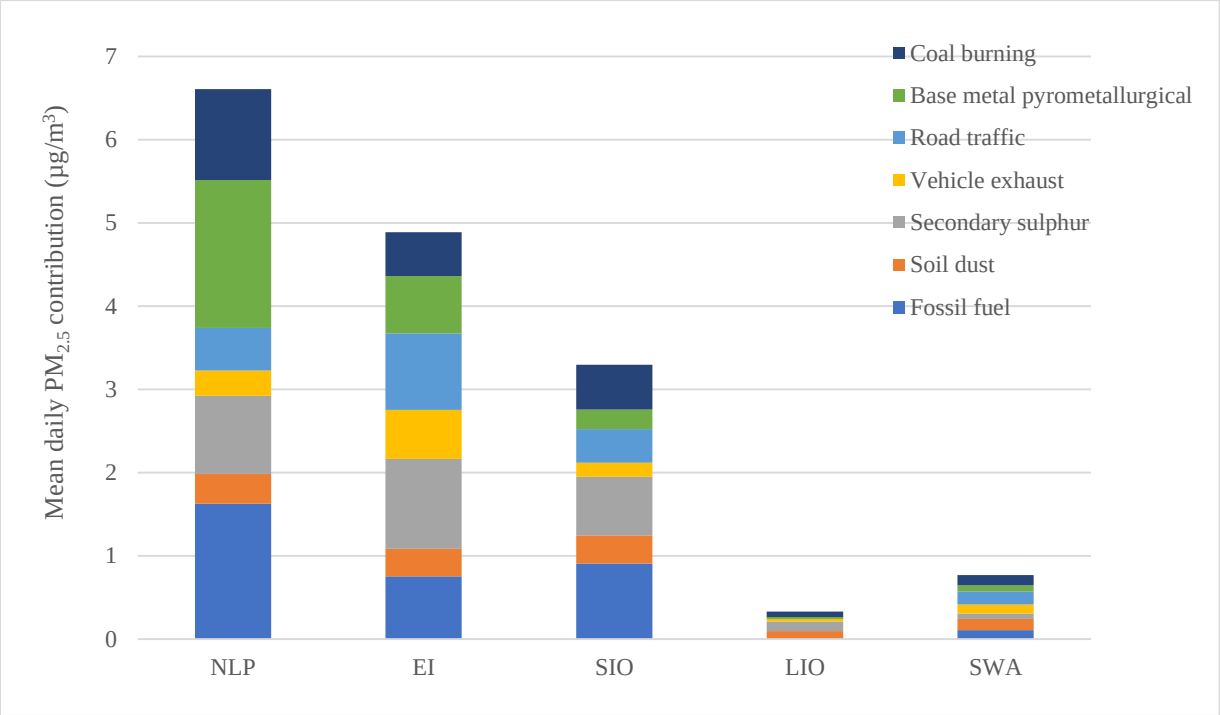

Fig. 6 Mean contribution from each transport cluster

Table 1 Descriptive statistics of $PM_{2.5}$ mass, explained mass from the positive matrix factorization model, and estimated source contributions for the seven sources. (All concentrations in µg/m^3)

Species	Full study		Autumn		Winter		Spring		Summer	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
$PM_{2.5}$	21.1	0.7–66.8	23.4	3.4–57.9	35.5	14.6–66.8	14.3	1.4–35.4	10.7	0.7–21.6
$PM_{2.5}$PMF	18.3		20.9		32.4		9.6		9.8	
Fossil fuel combustion	4.1	0–4.7	6.3	5.1–8.2	4.3	0–14.6	2.0	0–9.1	3.7	2.4–5.1
Soil dust	1.4	0.1–4.0	2.8	0.6–3.3	1.1	0–10.1	1.0	0–3.4	0.8	0–2.5
Secondary sulphur	2.7	0.3–6.5	0.8	0–4.9	8.1	0.6–13.5	1.2	0–4.9	0.5	0–2.2
Vehicle exhaust	2.5	0–5.9	3.1	0–4.4	3.9	0–12.0	1.4	0–3.7	1.6	0.1–3.8
Road traffic	2.6	0–4.7	4.2	0–5.8	3.3	0–9.1	1.8	0–4.2	0.9	0.2–1.4
Base metal/Pyrometallurgical	1.7	0–3.5	1.5	0–4.1	2.8	0–9.4	1.4	0–6.7	0.9	0–3.0
Coal burning	3.3	0.6–6.0	2.2	0.1–5.4	8.9	1.4–14.5	0.8	0–5.1	1.4	0.2–3.5

Table 2 Correlation between air pollution and weather variables

Variable	$PM_{2.5}$	BC	OC	Temp	RH	WS
$PM_{2.5}$	1.00					
BC	0.76*					
	0.00					
OC	0.76*	0.98*				
	0.00	0.00				
Temp	−0.32*	−0.49*	−0.46*			
	0.00	0.00	0.00			
RH	−0.22*	−0.06	−0.12	−0.39*		
	0.03	0.56	0.24	0.00		
WS	0.21*	0.18	0.21*	−0.39*	−0.04	1.00
	0.04	0.07	0.04	0.00	0.67	

$PM_{2.5}$ particulate matter with an aerodynamic diameter of less than 2.5 μm, *BC* black carbon, *Temp* temperature, *RH* relative humidity, *OC* organic carbon, *WS* wind speed

*Significant ($p < 0.05$)

Table 3 Correlation between metals during week days in Pretoria

	S	Cl	K	Ca	Fe	Zn	Pb	Si	Ti	V	Ni	Cu	As	Se	Br	Sb	Ba
Cl	*0.38**																
K	*0.47**	**0.77***															
Ca	*0.43**	**0.77***	**0.71***														
Fe	*0.32**	**0.66***	**0.75***	**0.63***													
Zn	*0.28**	**0.65***	*0.49**	**0.62***	*0.49**												
Pb	*0.25**	**0.50***	*0.39**	*0.43**	*0.32**	**0.57***											
Si	**0.58***	**0.78***	**0.78***	**0.78***	**0.69***	*0.40**	*0.25**										
Ti	**0.51***	**0.72***	**0.71***	**0.79***	**0.67***	*0.43**	*0.32**	**0.85***									
V	−0.03	0.04	−0.06	−0.07	0.00	0.05	−0.04	−0.12	−0.12								
Ni	0.07	*0.25**	*0.39**	0.11	**0.71***	0.21	0.03	*0.26**	0.16	−0.03							
Cu	*0.26**	*0.48**	*0.29**	*0.46**	*0.35**	**0.54***	**0.51***	*0.30**	*0.38**	0.12	0.07						
As	*0.31**	0.60*	0.42*	0.49*	0.48*	0.60*	0.74*	*0.39**	0.39*	0.04	0.20	**0.55***					
Br	*0.43**	**0.77***	**0.72***	**0.69***	**0.68***	**0.68***	**0.58***	**0.58***	**0.56***	0.01	*0.36**	*0.48**	**0.64***	*0.29**			
Sb	−0.03	0.00	−0.00	−0.08	0.00	−0.03	−0.08	−0.01	−0.04	*0.30**	0.06	-0.04	0.00	-0.20	-0.04		
Ba	*−0.22**	−0.02	0.12	0.15	0.07	−0.00	0.06	−0.08	−0.17	0.16	0.04	-0.10	-0.12	-0.10	0.07	0.05	
U	0.01	0.04	0.14	−0.01	0.09	−0.06	0.02	0.10	0.04	0.02	0.10	-0.11	-0.05	0.07	0.05	0.19	0.04

*Significant ($p < 0.05$)

Table 4 Correlation between metals during weekends in Pretoria

	S	Cl	K	Ca	Fe	Zn	Pb	Si	Ti	V	Ni	Cu	As	Se	Br	Sb	Ba
Cl	*0.34**																
K	*0.47**	**0.82***															
Ca	*0.40**	**0.63***	**0.73***														
Fe	0.11	**0.75***	**0.65***	**0.60***													
Zn	0.19	**0.64***	*0.46**	**0.52***	**0.63***												
Pb	*0.39**	**0.66***	**0.64***	**0.60***	**0.56***	**0.60***											
Si	**0.55***	**0.81***	**0.89***	**0.73***	**0.68***	*0.43**	**0.64***										
Ti	*0.46**	**0.71***	**0.77***	**0.79***	**0.70***	**0.52***	**0.54***	**0.86***									
V	−0.06	−0.15	0.06	−0.15	−0.19	−0.26	−0.09	−0.12	−0.26								
Ni	−0.15	**0.54***	*0.42**	0.23	**0.74***	0.26	0.21	*0.38**	0.34	0.10							
Cu	0.08	**0.59***	**0.56***	**0.53***	**0.62***	**0.56***	*0.40**	**0.54***	**0.57***	−0.19	**0.52***						
As	*0.42**	**0.63***	**0.57***	*0.50**	**0.51***	**0.64***	**0.85***	*0.50**	*0.46**	−0.09	0.11	0.33					
Se	*0.38**	*0.40**	*0.42**	*0.43**	*0.43**	*0.44**	**0.58***	**0.55***	*0.41**	−0.15	0.07	*0.46**	*0.44**				
Br	0.20	**0.88***	**0.79***	**0.59***	**0.79***	**0.75***	**0.66***	**0.67***	**0.68***	−0.06	**0.55***	**0.62***	**0.69***	*0.41**			
Sb	0.02	0.04	−0.05	0.12	0.16	−0.04	0.19	−0.08	0.04	0.24	0.18	0.01	0.11	−0.03	0.05		
Ba	0.09	−0.16	0.14	0.18	−0.24	−0.18	0.03	−0.04	−0.06	0.29	−0.31	*−0.38**	0.02	−0.06	−0.09	0.16	
U	−0.14	−0.12	−0.08	−0.24	0.02	−0.11	−0.26	−0.02	−0.01	−0.04	0.10	0.11	−0.35*	0.15	−0.06	−0.19	-0.09

*Significant ($p < 0.05$)

Table 5 Mean concentrations for $PM_{2.5}$ and for the seven estimated PMF sources by the five different transport clusters in ($\mu g/m^3$)

Transport cluster	$PM_{2.5}$	Fossil fuel combustion	Soil dust	Secondary sulphur	Vehicle exhaust	Road traffic	Base metal pyro-metallurgical	Coal burning
1. NLP (31 days)	25.0	6.4	4.1	3.7	1.2	2.0	7.0	4.3
2. EI (40 days)	14.9	2.3	1.0	3.3	1.8	2.8	2.1	1.6
3. SIO (41 days)	26.2	2.7	1.0	2.1	0.5	1.2	0.7	1.6
4. LIO (4 days)	9.8	0.3	2.6	3.5	0.9	0.1	0.6	2.1
5. SWA (6 days)	15.7	2.2	2.7	1.3	2.2	3.2	1.5	2.5