Journal Pre-proof

Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana.

C. Gameli Hodoli, F. Coulon, M.I. Mead

PII: S0269-7491(22)01662-1

DOI: https://doi.org/10.1016/j.envpol.2022.120448

Reference: ENPO 120448

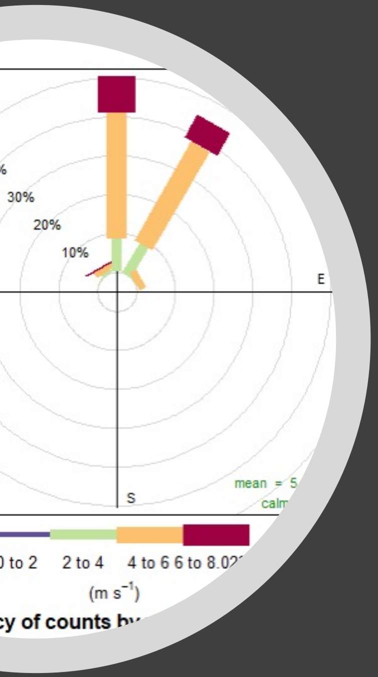
To appear in: *Environmental Pollution*

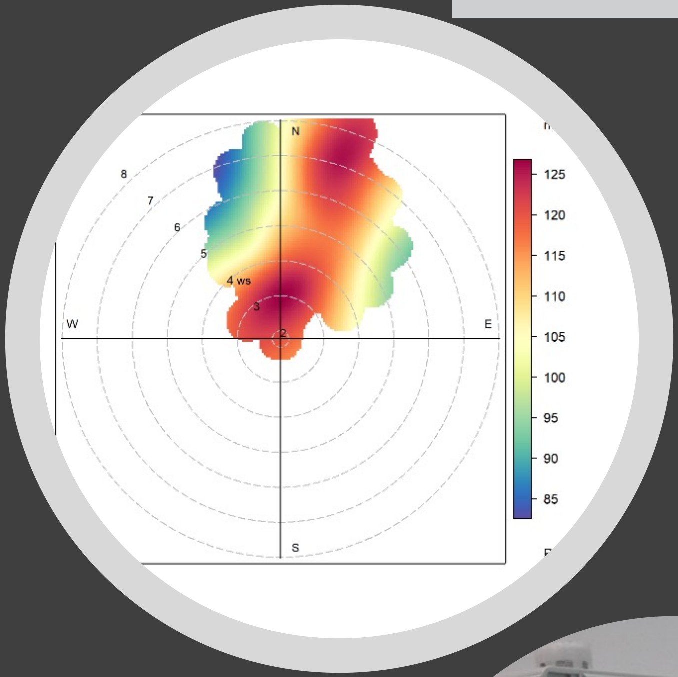
Received Date: 21 March 2022

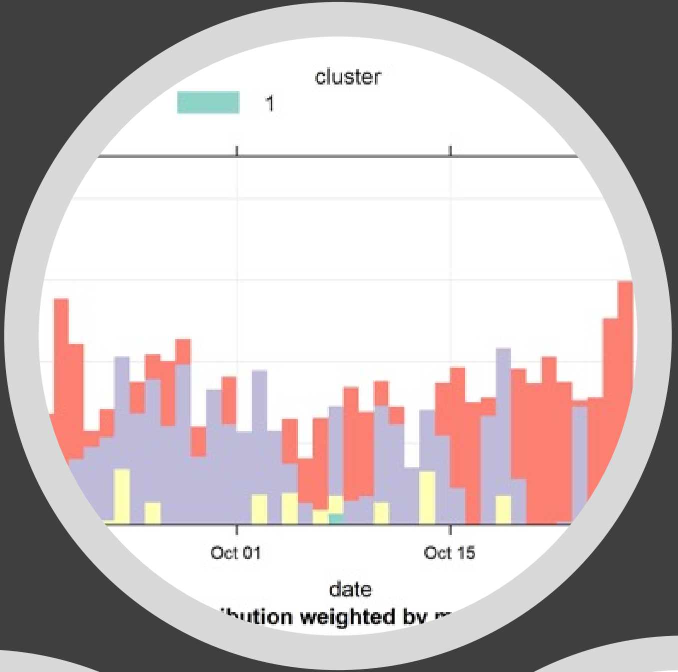
Revised Date: 7 October 2022

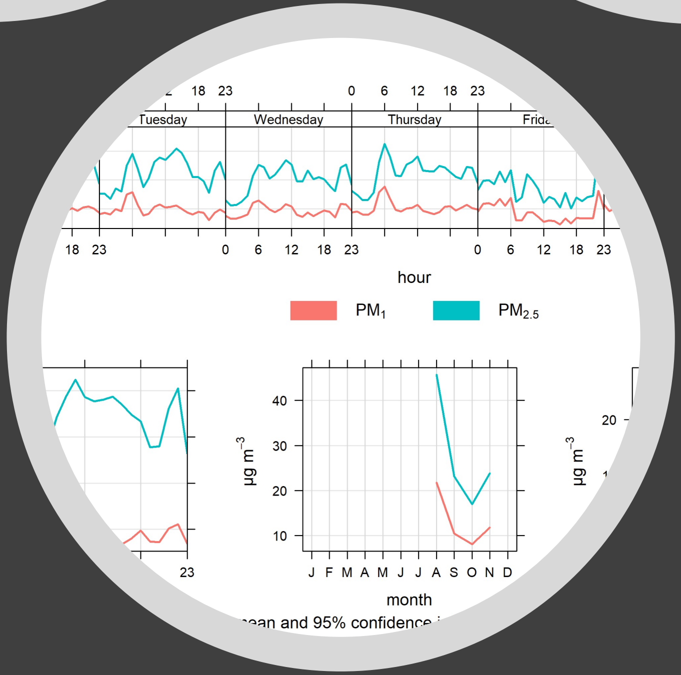
Accepted Date: 14 October 2022

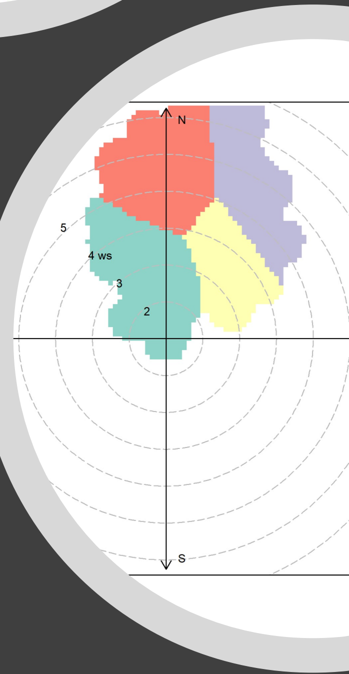
Please cite this article as: Hodoli, C.G., Coulon, F., Mead, M.I., Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana., *Environmental Pollution* (2022), doi: https://doi.org/10.1016/j.envpol.2022.120448.











Graphical abstract

Environmental Pollution, Volume 317, January 2023, 120448
DOI: 10.1016/j.envpol.2022.120448

Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana.

C. Gameli Hodoli[1,2], F. Coulon[3], M. I. Mead[1,3*].

1. Imperial College London, Environmental Research Group, London SW7 2AZ, UK.

2. University of Environment and Sustainable Development, School of Built Environment, PMB, Somanya, Eastern Region, Ghana.

3. Cranfield University, School of Water, Energy and Environment, Cranfield, MK43 0AL, UK.

[*]Corresponding author: m.mead@imperial.ac.uk

Abstract: The emergence of low-cost sensors for atmospheric observations presents a new opportunity for identifying atmospheric emission sources based on high-resolution data reporting. Low-cost sensors have been widely assessed for use in source monitoring and identification of hotspots of key atmospheric species in advanced countries (e.g., for CO, NOx, CO_2, SO_2, O_3, VOCs and PM (PM_{10}, $PM_{2.5}$ including emerging PM_1). In contrast, research in recent years has focused on their utility for real-time monitoring, understanding precision and associated calibration requirements in technologically lagging environments. This leads to limited evidence on the utility of high-resolution data from low-cost sensor networks for air pollution source identification in Ghana and more widely across the African continent. In this paper, we demonstrate the potential of low-cost sensors for emission source apportionment in urban areas of Ghana when used with analytical tools such as sectoral and cluster analysis. With a 14-week dataset from a low-cost sensor deployment study at Cape Coast in the Central Region of Ghana, we aimed to identify sources of particulate matter ($PM_{2.5}$ and PM_{10}). PM pollution was local (associated with increased PM at wind speeds of ≤ 2 ms^{-1}). High levels of PM during this study were associated with transport from the NNE. For coarse PM, hourly levels as high as 125 $\mu g/m^3$ were observed at higher wind speeds (7-8 ms^{-1}) indicating the importance of meteorology in the transport of PM. This study suggests that low-cost sensors could be deployed to (1) augment any existing sparsely distributed air quality monitoring and (2) undertake air quality monitoring for source apportionment studies in areas with no existing air quality observational capability (with appropriate calibration and operation in both cases). The urban landscape monitored in this study is typical of both Ghana and wider areas across Sub-Saharan Africa demonstrating the reproducibility of this study.

Keywords: Ghana; Sub-Saharan Africa; low-cost sensors; cluster analysis; source apportionment.

1 Introduction

A key issue in controlling air pollution is characterizing sources of emissions. Air pollution mitigation strategies can only be effective if appropriate measures are employed to understand and characterize emissions sources (Carslaw and Beevers, 2013). In general, atmospheric emission sources in urban settings are complex and variable. Those encountered in Ghana, which is typical of larger swaths of Sub-Saharan Africa (SSA) are both complex and variable as well as having limited regulation or controls. Characterising emission sources requires the collection of appropriate spatiotemporal air quality data (Hagan *et al*., 2019). These data are useful for air quality management, tracking of existing air quality standards and assessing human health exposure to air pollution (Hagan *et al*., 2019). Therefore, governmental agencies widely employ conventional air quality monitoring approaches consisting of relatively expensive and specialist instrumentation and support (i.e., requiring trained technical staff and significant infrastructure) to sample air quality data (Schwela, 2012a; Snyder *et al*., 2013; Rai *et al*., 2017). As a result, conventional air quality monitoring stations are mostly sparsely distributed in urban centres. In effect, spatiotemporal measurement of air quality species (e.g., $PM_{2.5}$, PM_{10}, CO, NOx and O_3) is limited (Moltchanov *et al*., 2015; Snyder *et al*., 2013; Rai *et al*., 2017; Hagan *et al*., 2019). The situation is more pronounced in low and middle-income countries such as Ghana and across wider swaths of SSA. This is due to the sporadic nature of air quality campaigns and lack of monitoring capabilities (Schwela 2012a; Petkova *et al*., 2013; Jovašević-Stojanović *et al*., 2015).

In addition to the above, source apportionment studies are challenging to undertake in part due to the limited availability of appropriate air quality data derived from the current regulatory air quality monitoring regime in Ghana (see Hodoli *et al*., 2020). In this study, the use of meteorological data with data from low-cost sensors is used to provide insight into understanding local PM pollution. These types of analysis can reveal source features if an appropriate amount and quality of data is gathered on the species of interest (Statheropoulos *et al*., 1998). Developing emission inventories e.g., based on the use of regression model slopes as applied to collected data is limited. This is in part due to the lack of appropriate local data. Where this work is undertaken, it is often the case that suitable data is unavailable and is substituted with data from previous studies, short-term air quality monitoring data or nearby location/ regional data (Grange *et al*., 2016). This is a feature of studies in Ghana and wider Africa (see Schwela, 2012a; Petkova *et al*., 2013) but Manoli *et al*., (2002) showed that to comprehensively understand the relationship between atmospheric variables, an appropriate and reliable data source is paramount.

The emergence and utility of low-cost sensors (LCS) have shifted the paradigm and offer a new approach to populating air quality data. Potentially, LCS provide an alternative approach and can be deployed in networks to provide additional air quality data. Instruments based on low-cost sensing are becoming increasingly commercially available (e.g., Mead *et al*., 2013, Snyder *et al*., 2013; Kumar *et al*., 2015; Castell *et al*., 2017) and have been shown to be robust with lower installation and operational overheads (Mead *et al*., 2013, Snyder *et al*., 2013; Kumar *et al*., 2015; Castell *et al*., 2017). A number of studies have demonstrated the potential of using LCS for air quality campaigns considering ease of deployment, the potential for relocation, routine data acquisition tied to internet-based platforms, and little need for human intervention (e.g., Kumar *et al*., 2015; Moltchanov *et al*., 2015; Snyder *et al*., 2013; Mead *et al*., 2013). These characteristics of LCS make them suitable for these types of studies as compared to traditional reference-grade instrumentation.

In recent years, the focus on the use of low-cost sensors has been on performance evaluation under real-world conditions, precision, developing appropriate data calibration and validation methodologies (see e.g., Mead *et al*., 2013; Baron and Saffell, 2017; Malings *et al*., 2019 in Europe and America; McFarlane *et al*., 2020; Subramanian *et al*., 2020; Raheja *et al*., 2022 emerging studies on LCS calibration studies in Africa and investigating how these low-cost devices can be used for real-time monitoring e.g., Wang *et al*., 2015; Jiang *et al*., 2016; Jiao *et al*., 2016; Rai *et al*., 2017). There has however been little information on applying LCS to expand our understanding of factors contributing to air pollution levels (for example identifying sources of atmospheric emissions) in environments with sporadic air quality monitoring approaches such as those encountered in low- and middle-income countries e.g., Ghana and wider parts of SSA (Lewis *et al*., 2018). In this study, we have demonstrated the application of LCS data for emission source identification by deploying low-cost multi-sensor nodes at Cape Coast in the Central Region of Ghana. This pilot study is to understand sources of PM using bivariate polar plots and cluster analysis. The application of these types of data from LCS for this type of work is to serve as an exemplar for emission source identification in these types of environments, specifically in Ghana and wider parts of SSA.

2 Methodology

2.1 Instrumentation and data acquisition

A relatively low-cost multi-sensor instrument was used for this study (model AS510, Atmospheric Sensors, UK). Technical characteristics of the static multi-sensor node are summarized in Table 1. The node contains a range of sensor technologies used for monitoring key pollutant species (i.e., CO, NO, NO_2, O_3, VOCs, CO_2, PM – PM_{10}, $PM_{2.5}$ and emerging PM_1), and environmental variables (temperature and relative humidity) but the focus of this study is on the PM data. The PM data was used because of 3 main reasons – 1. PM pollution sources are complex and varied in urban settings; 2. PM is used globally for setting air quality standards and 3. The Ghana EPA recognized PM as one key species for air quality management in Ghana and 4. On gaseous species, there were no resources available to undertake any form of data correction and studies have shown that out-of-the-box values for these types of species are not reliable compared to the PM species (e.g., Mead *et al*., 2013; Baron and Saffell, 2017; Crilley *et al*., 2018; Spinelle *et al*., 2017) Also, this paper focused on PM species but does point to the future use of other measured species if and when the infrastructure needed for reliable calibration becomes available. However, since this requirement has not yet been met, LCS data on gaseous species are not considered in the present analysis. Further to this, the technology underlying each of the monitored species is different (see Table 1) and calibration needs are similarly different. The node has integrated GPS for location logging and accurate timestamping as well as GPRS for data telemetry. The PM data for this study was obtained from the optical particle counter sensing technology (OPC-N2, manufactured by Alphasense, UK) which measures scattered light from the sample flow of aerosol particles to reconstruct particle mass levels (Hinds, 1999) (details in Table 2). Modelled wind data obtained from the Global Forecast System (GFS) repository was used in this study (NOAA, 2019; López and Schliep 2019). The GFS consists of datasets from the National Oceanic and Atmospheric Administration (NOAA) and the National Centres for Environmental Prediction (NCEP) (López and Schliep 2019). In this database, wind data since 2011 is saved at 3-hour intervals daily in velocity vector format with a resolution of 0.5 degrees and ~50 km at the Earth's surface.

Table 1: Characteristics of the static AS510 multi-sensor node

Technology (AS510)	**Purpose**
Optical particle counter (Alphasense OPC-N2)	PM_1, $PM_{2.5}$ and PM_{10}
Electrochemical cells (Alphasense A4)	CO, NO, NO_2 and O_3
Photo ionization detector (PID)	VOCs
Non-dispersive infra-red (NDIR)	CO_2
Global positioning system (GPS)	Timestamp and location
General packet radio service (GPRS)	Data telemetry
SD card	Local data storage
Capacitive	Temperature and relative humidity

Table 2: Characteristics of Alphasense OPC-N2 (AS510 Manual)

Alphasense OPC-N2		
Measurement	Particle range (µm)	0.38-17
	Number of software bins	16
	Size categorization	1.4-10
	Flow rate ($Lmin^{-1}$)	1.2
	Max particle count rate (s)	10,000
Key specifications	Digital interface	SPI (Mode 1)
	Laser classification	Class 1
	Temperature range (°C)	-20 to 50
	Humidity range (% rh)	0 to 95% (non-condensing)
	Weight (g)	< 105

2.2 Study area

Cape Coast is the capital of the Central Region and an emerging urban setting situated in the South of Ghana on the Gulf of Guinea. The preliminary report of the Ghana Statistical Service indicated that the Central Region has a population of 2,859,821 (GSS, 2021) The region is relatively humid with mean monthly relative humidity (RH) ranging between 85% and 99% as compared with a range of 77% to 85% in Accra (World Weather Online). The predominant wind direction at Central Region is from the south which has the potential to transport pollutants from across the region to Cape Coast as well as for onshore relatively clean air masses to be transported.

Details on background activities for the deployment area and period are described in Hodoli *et al*., 2020. Data from one node is used in this source apportionment study as it has been shown that the two nodes provided consistent measurements for PM species (see Hodoli *et al*., 2020). Typical sources in the area include unpaved roads (re-suspended dust), road-side food preparation (biomass and gas combustion), taxi ranks (vehicular) and roads used by private vehicles as well as heavy trucks and commercial vehicles (Schwela., 2012a, Lelieveld *et al*., 2015; HEI, 2019).

Figure 1: Overview of the deployment area at the University of Cape Coast (UCC) with background activities. The green circle shows the location of the deployed node (05°06′N 01°15′W) OpenStreetMap (and) Contributors, CC-BY-SA (accessed on 09.01.2020 using QGIS).

2.3 Analysis methodology

This section presents details on the analysis methodology used in this study specifically polar plots and cluster analysis including calibration methodology. As a proof-of-concept study, the application of these analyses was undertaken in the default mode and has been duly stipulated in the sub-sections below.

2.3.1 Bivariate polar plots

The bivariate polar plots (Carslaw and Ropkins, 2012) can be used for source apportionment studies using data from LCS. This approach of having a mixing ratio of pollutants plotted in polar coordinates for source identification is not new (Carslaw and Ropkins, 2012) but is suited to small sensors as these devices are capable of acquiring air quality data at relatively high resolutions in both time and space if they are deployed as properly calibrated and operated networks (see e.g. Mead *et al*., 2013; Snyder et al., 2013; Rai *et al*., 2017). This approach has been demonstrated by source apportionment studies in an airport setting and in exploring characteristics of dispersion of pollutants in street canyons (Carslaw *et al*., 2006). Bivariate polar plots suggest potential sources of air pollutants based on wind speed and direction and pollutant level. The acquired measurement data (in this case concentrations of PM) are plotted as a function of wind speed and direction. These types of plots have been tested on a wide variety of data and it has been suggested that wind direction intervals of 10 to 30° capture enough detail of pollutant dispersion to allow for source identification (Carslaw and Beevers, 2013). The data were grouped into bins based on wind speed and wind direction using machine learning. Concentration means are then calculated for each of these bins (Carslaw and Ropkins, 2012; Carslaw, 2015). The combination of wind speed and direction is an efficient approach in differentiating varying sources of air pollution (Carslaw and Beevers, 2013). As a proof of concept, modelled wind directions data were approximated to 10° with typical surface measurements usually between 0-30 ms^{-1}. Wind speeds beyond 30 ms^{-1} were eliminated as these were proven to have been difficult to confirm as accurate (see Carslaw and Beevers, 2013). This aggregation of data provided a reduction technique without a bias analysis of the plots since wind component data is variable and tend to diffuse making raw data to yield limited results (Carslaw, 2015). In pollution forecasting the relationship between variables is non-linear but their interactions are important. To account for this, a surface fitting model, the Generalized Additive Model (GAM) (e.g., Hastie and Tibshirani, 1990; Wood, 2006) was applied to the polar plots (Carslaw and Ropkins, 2012) to provide a smoothing approach useful for pollution forecasting.

2.3.2 Cluster analysis

Cluster analysis is a useful tool for identifying features and source extraction from polar plots (details can be found here Carslaw, 2015). This is an advanced technique which selects groups with similar characteristics and maps them and provides means to understand source features as compared to the polar plots. In essence, bivariate polar plot interpretation is limited by the ability of the investigator which may lead to bias as investigators tend to focus on areas of interest in these types of plots. Further to this, some patterns may not be plotted in the selected intervals of the wind components (Carslaw, 2015). Cluster analysis using the *k*-means algorithm provides a better approach than bivariate polar plot analysis. This algorithm for clustering was introduced by Hartigan (Hartigan, 1975). This procedure allocates a number of observations into clusters using *k*-means clustering. The data allocation process groups the pollution data with similar diurnal patterns into one group using the *k*-means. With the *k*-means, (i.e., grouping the data by making in-group data points more similar to each other than to out-of-group data points), cluster analysis is performed in which features in bivariate polar plots are identified and categorized (Shi *et al*., 2014). This categorisation helps to identify records in the original time series data, enhancing post-processing for identification of potential source characteristics. Firstly, *k* points are randomly selected from the space represented by the objects that are being clustered into *k*-groups. These group points are then represented as centroids and every object is attached to a group with the closest centroid (Carslaw, 2015). The k centroids are then recalculated after assigning all objects; recalculation and group assignment is done until the centroids no longer move, in which case the objects are grouped to minimize the algorithm's metric. The three most important variables in this analysis are wind speed, direction measure and concentration. As per principle, more weight is given to the concentration rather than the wind components though this tends to identify clusters with similar concentrations with varying sources (Carslaw and Ropkins, 2012). Data binned by cluster can be presented as a stacked bar chart time series (Carslaw, 2015). This approach is often used to support outputs from the cluster analysis (Carslaw and Ropkins, 2012). A recent study using this type of approach has shown that *k*-means clustering is an effective way of extracting source features using data from LCS (Bousiotis *et al*., 2022). This approach however requires experts with sophisticated machine learning skills and application.

2.3.3 Calibration methodology

This study relied on factory calibration methodology for the use of the OPC-N2 data because past studies have found that relative humidity below 85% has a limited impact on the accuracy of the OPC-N2 PM data (see for example Crilley *et al*., 2018) and Spinelle *et al*., 2017 found no impact of temperature and relative humidity on the reported OPC-N2 PM data. This makes the factory-calibrated OPC-N2 PM data useful for understanding the complex sources of PM pollution in urban settings in SSA with appropriate data mining tools/ analysis. Per these standards, the reported PM data used in this study have been considered to meet the required quality control/ quality assurance protocols.

3 Results and discussions

3.1 Trend analysis of PM

It is noteworthy that though the levels of PM_1 and $PM_{2.5}$ vary, the diurnal trends follow a common pattern which varies from PM_{10} (see Figure 2). Both PM_1 and $PM_{2.5}$ peak at the same period (at approximately 0600hrs) while PM_{10} peaks at approximately 1300hrs. This trend analysis has revealed that fine particles (PM_1 and $PM_{2.5}$) in this case are likely to be from a similar source (the site is heavily influenced by local traffic and re-suspended dust).

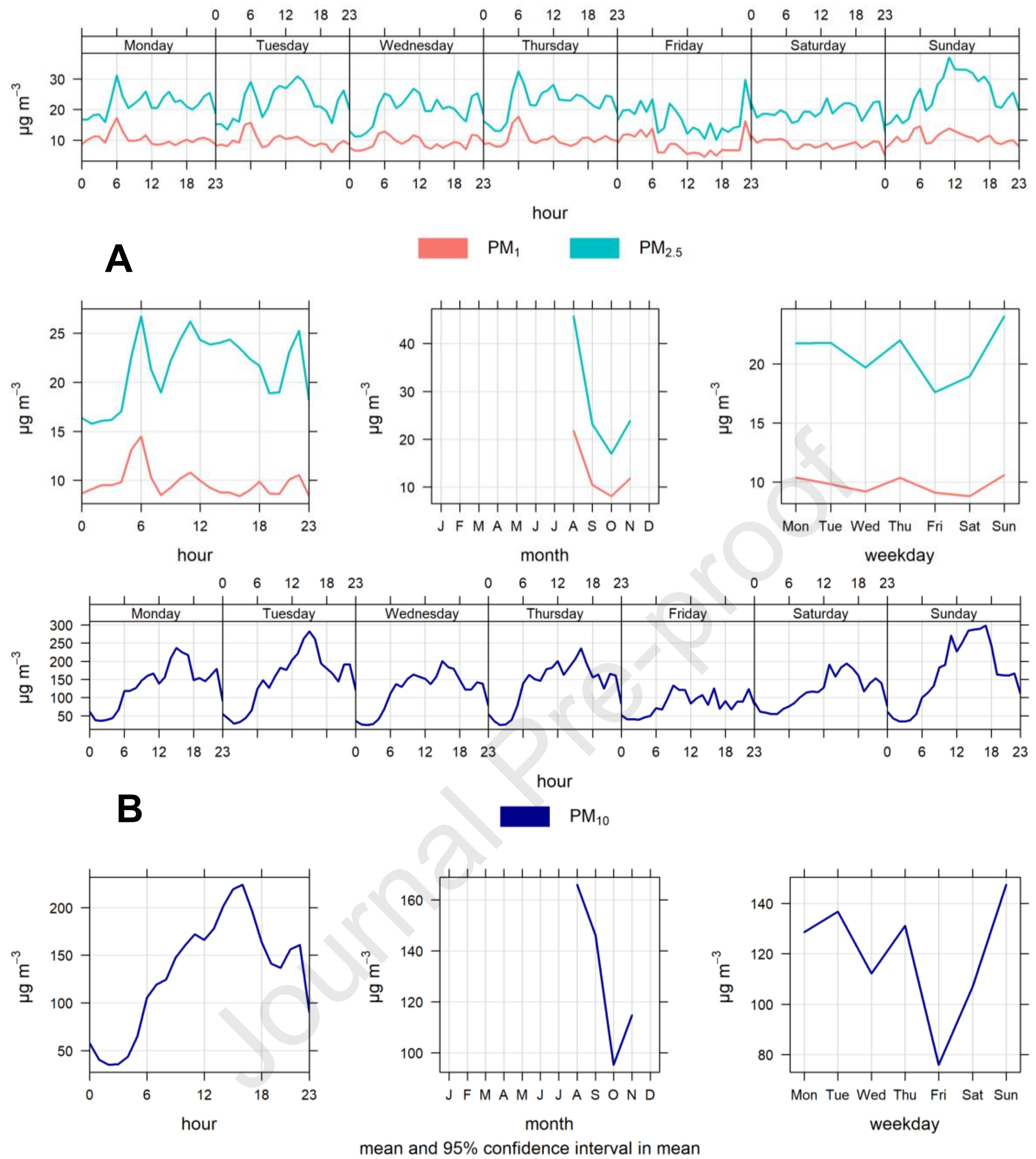


Figure 2: Trend analysis of PM (where A – shows PM_1 and $PM_{2.5}$ and B shows PM_{10}) at UCCSM. Top panels of A and B: daily variation; Bottom left panels of A and B: hourly variation; Bottom middle panels on A and B: monthly variation and Bottom right panels of A and B: day of the week variations of PM species

3.2 Source apportionment of PM

Wind roses were generated for the period of deployment (Carslaw and Ropkins, 2012). These showed the frequency of wind contribution in percentage (represented as counts) by wind direction. Intervals of 10% were used to establish wind conditions and highlight the sources of atmospheric emissions (as shown in Figure 3). This showed that NNE winds were the main sources of atmospheric emissions at the Cape Coast (UCCSM) study area. These were shown in the polar plots where the resulting analysis suggests local sources – see Figures 5 a, b and c were used to identify the sources of monitored species based on the high-resolution data from the low-cost devices deployed for this study. The relationship between PM_1 and PM_{10} suggests that an important source of particulate matter is located towards the NNE of the deployment site. This source is either biased towards lighter particles or that larger particles are removed before arriving at the monitoring site. The data also potentially points towards a more local source of lighter particulates nearer to the monitoring site which has an important role in composition at lower wind speeds. Under still conditions, it seems there is no significant local source for coarse particles. Overall PM levels were relatively high (20 µg/m^3 for PM_1, 35 µg/m^3 for $PM_{2.5}$ and 220 µg/m^3 for PM_{10} as compared to the past recommended 25 µg/m^3 and 50 µg/m^3 and the new (15 µg/m^3 and 45 µg/m^3) limits of the WHO for $PM_{2.5}$ and PM_{10} respectively - WHO, 2021). Locally PM_1 and $PM_{2.5}$ concentrations were high while high PM_{10} concentrations were observed at higher wind speeds.

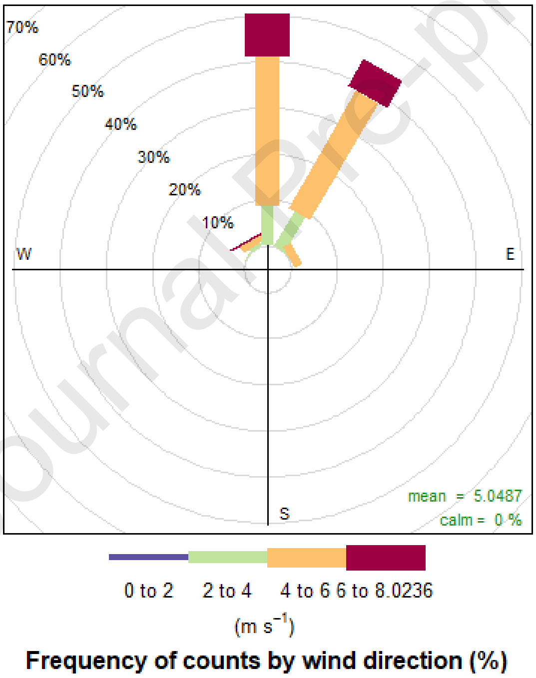


Figure 3: wind rose of meteorological data at UCCSM (Cape Coast) for the period of deployment

At lower wind speeds (i.e. ≤ 2 ms^{-1}), elevated levels of PM were observed, implying that there are potential local sources which heavily contributed to fine particles (PM_1 and $PM_{2.5}$) concentrations. This is expected due to the nature of the deployment site integrated with background activities specifically vehicular emissions and roadside food cooking. For PM_1 and $PM_{2.5}$ higher concentrations were experienced at westerly and northwesterly winds whereas the lowest concentrations were experienced at northeasterly winds (Figures 4a and b). For PM_{10}, higher concentrations were observed at northerly and northeasterly winds (Figure 4c). Using the cluster analysis to extract source features, cluster 4 (associated with northerly winds with speeds from 4-7 ms^{-1} in Figures 5a, 6a and 7a) contributed to PM levels as high as 11 μg/m^3 for PM_1; 24 μg/m^3 for $PM_{2.5}$ and 125 μg/m^3 for PM_{10}.

The analysis has shown that higher wind speeds (7-8 ms^{-1}) from the north contributed to elevated PM. Regardless of the wind speed, PM_1 is highest when winds are from the NNW direction (Figure 5a). For $PM_{2.5}$, higher levels were experienced at low wind speeds and from NNW (Figure 6a). However, in the case of PM_{10}, it is highest with both low wind speed from the north (associated with cluster 1; Figure 7a) and high wind speed from the northeast (associated with cluster 4) (Figure 7a). This indicates a likely source to the NNW for finer particles but more distant sources of coarse particles to the NNE. An examination of the background environment has shown that this type of result is expected as N is mainly unpaved flat land (lorry/ taxi park) adjacent to the main road which is usually dusty though paved; E is the main road which is often used by taxis, commercial vehicles and heavy-duty cars (diesel engines); NE is the mini-market with roadside food vendors including cooking; S is mainly office complexes with the coast about 3 km away; W is the main road but mostly unpaved about 100 m away from the deployment and NW is similar to NE except for high levels of wind-blown dust from the unpaved road. To demonstrate this, a time-domain analysis was carried out using the clusters identified. The categorical bar chart obtained with this analysis (Figures 5b, 6b and 7b) have shown that the peaks of PM are associated with cluster 4 (Figures 5a, 6a and 7a). At relatively higher wind speed (4-8 ms^{-1} associated with cluster 4), the effects of meteorology on atmospheric pollution changes are reduced (Carslaw and Beevers, 2013) which accounted for the agreement in the sources of PM as linked to the temporal variation plot. It is also important to note that this cluster analysis consistently grouped pollution levels from potentially the same source/ wind direction which was not shown in the bivariate polar plots. The downward trend shown in the PM species with high local levels (Figures 4a, b and c) and the contribution of the clusters from NNE winds as shown in the temporal variation plots (Figures 5b, 6b and 7b) is influenced by road traffic emissions as previous studies have reported similar scenarios (e.g., Kim *et al*., 2014) there are intermittent high sources.

3.3 Cluster analysis for source identification and extraction

In this study, this analysis is used to explore the effects of wind components on the measured concentrations of PM at Cape Coast over time (Figures 5a, 6a and 7a).

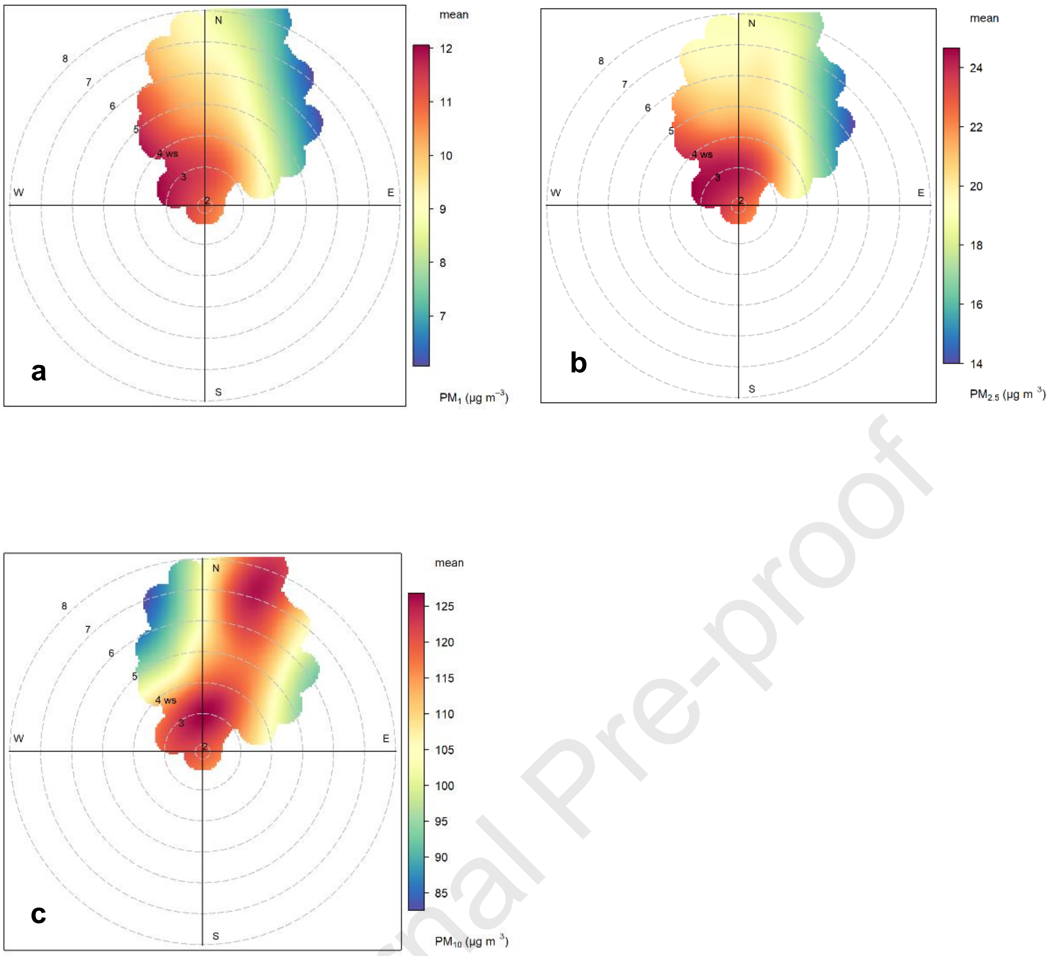


Figure 4: Hourly bivariate polar plot of PM (a - PM_1, b - $PM_{2.5}$ and c - PM_{10}) at UCCSM

The selection of 4 clusters for this study was based on test running of a range of cluster sizes. This approach in the absence of a priori information re number of clusters provides a solution that offers the most appropriate solution for visualization and interpretation (Carslaw and Ropkins, 2012). For example, in this proof-of-concept, 4 clusters were used as removal of one cluster reduced resolution in the bivariate polar plots and the populations of each cluster are reasonably comparable. The addition of a fifth class or cluster was not used as during testing a fifth cluster was composed of very few data points. These 4 clusters were therefore used to derive information on the proportioning of wind components that contributed to PM pollution at Cape Coast. It has been observed that cluster 4 associated with N winds dominated in contributing to PM concentration. This has been demonstrated using the temporal variation plots for each of the species (see Figures 5b; 6b and 7b). The contribution of each of the clusters have shown that daily averages were not only influenced by cluster 4. Also, there were instances where non-dominating clusters contributed to higher PM levels. For example, for PM_1, from August 30th to October 1st, cluster 3 and 4 dominates with minor contributions from cluster 2 and the least being cluster 1 (Figure 5b). This was observed on daily averages, though the temporal patterns have shown that cluster 3 contributed to levels beyond 20 µg/m^3. Additionally, it was also observed that more than one cluster contributed to PM levels which suggests multiple source domains. The temporal variation (see Figures 5b; 6b and 7b) plots have shown at what degree each of the defined clusters have contributed to PM levels. Some clusters were mainly associated with low PM concentrations. For example, from September 1 to 19, all clusters contributed to high PM_1 levels but cluster 1 contributed to the lowest levels recorded on September 16 and daily average of a little below 5 µg/m^3 (Figure 5b). Cluster 3 though not dominant, contributed to higher daily average (i.e., beyond 20 µg/m^3) followed by cluster 4 which dominates with daily averages of a little below 15 µg/m^3. A similar trend was observed from October 2 to 16 (Figure 5b). This study further demonstrates that cluster analysis if integrated with temporal variation is a tool that is applicable for source identification specifically source attribution based on wind components as evidenced by Carslaw and Beevers, 2013.

With PM_{10}, there were a number of higher daily averages associated with cluster 3; cluster 4 also contributed to the reported high PM_{10} levels (a little above 200 µg/m^3 as shown in Figure 7b). The least contributing cluster to daily levels is cluster 1 with concentrations 10 µg/m^3. These findings agree with similar findings by Zikova *et al*., (2017) using these types of sensors for PM estimation across urban centers. These findings do not only reflect possible multiple sources of PM with some events attributed to higher wind speed specifically for PM_{10} but demonstrate the applicability of low-cost sensor data for source identification (e.g., Mead *et al*., 2013; Zikova *et al*., 2017).

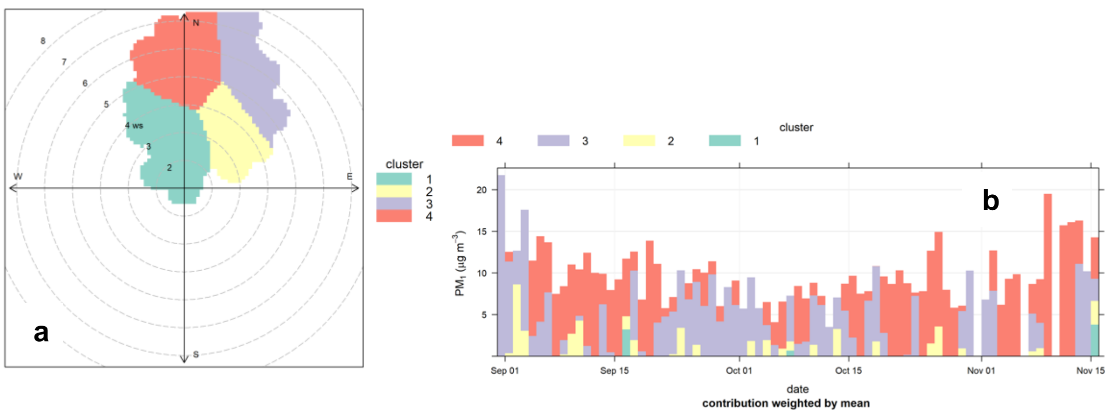


Figure 5: (a) Cluster sources plot and (b) Temporal variation plot of daily PM_1 concentration by the concentration of each cluster at UCCSM

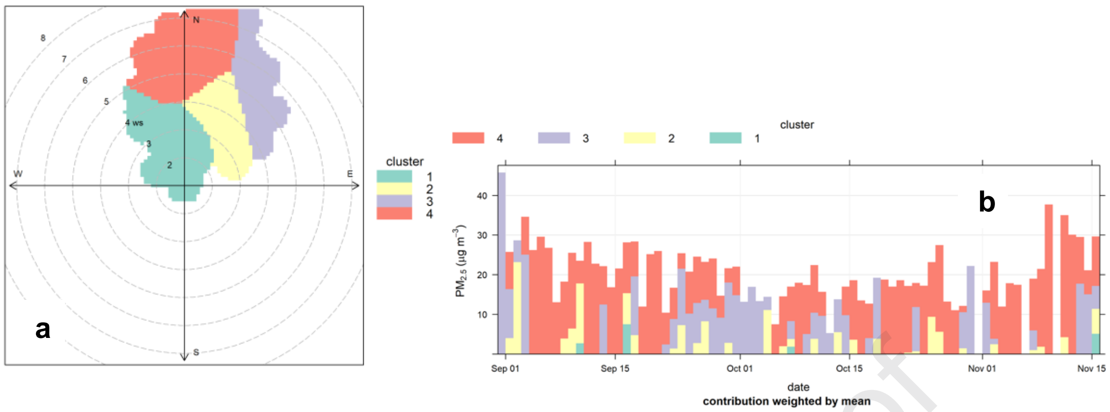


Figure 6: (a) 4 cluster plot and (b) Temporal variation plot of daily $PM_{2.5}$ concentration by the concentration of each cluster at UCCSM

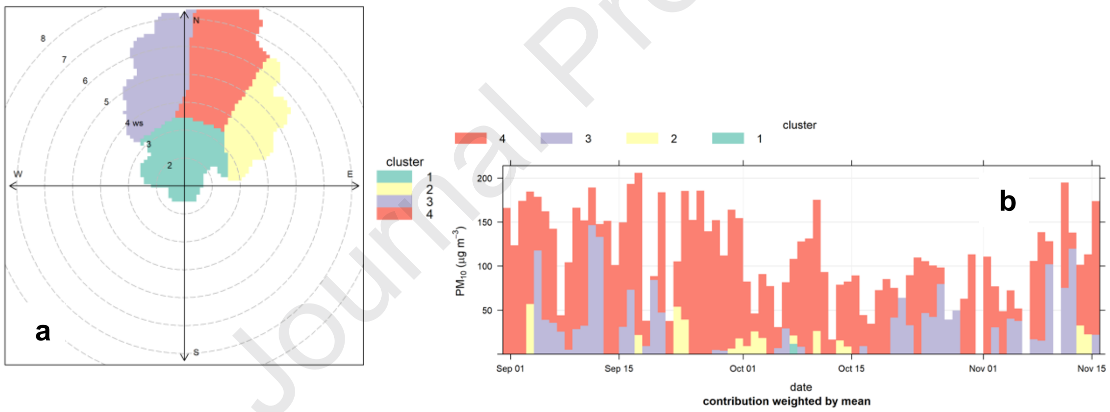


Figure 7: (a) 4 cluster plot and (b) Temporal variation plot of daily PM_{10} concentration by the concentration of each cluster at UCCSM

4 Conclusions

In this paper, we have demonstrated the efficacy of using bivariate polar plots and cluster analysis to characterize PM sources at Cape Coast with high-resolution (1-minute) data from LCS. The applicability of low-cost devices for these types of studies is important for emission source identification in resource-constrained environments i.e., areas with sporadic or entirely absent conventional monitoring. This study also demonstrates the advantages of using LCS to provide initial, relative concentration data for tracking air pollution sources. While most of the recent studies on LCS focused on developing data correction mechanisms which are dependent on reference-grade data, this is not always achievable in technologically lagging countries with limited air quality monitoring capabilities. This study demonstrated the usefulness of LCS data using complex methodologies to identify PM sources which do not necessarily require such advanced calibration methodology since relative concentrations are more important in distinguishing point sources and clusters. Therefore, LCS is a useful tool for source feature extraction, particularly in environments with limited to non-existent regulatory air quality monitoring stations. Overall, the analysis as presented in this paper is reproducible provided basic protocols in the analysis methodologies are followed.

findings presented here show that LCS can provide useable data at an appropriate resolution for source feature extraction in urban environments such as those found in Ghana and wider SSA when combined with representative local wind data. This is of particular importance in resource-constrained settings such as those encountered in Ghana and the wider SSA region where urban areas are often characterized by complex and varying sources of atmospheric pollution. Though further studies are recommended especially for validating the reported data in this study, the protocols used in this preliminary study are reproducible for source identification with high-resolution data from low-cost sensors in Ghana and wider SSA. These are useful low-cost approaches that regulatory agencies, for example, the Ghana Environmental Protection Agency can use for source characterization, and informing location-specific mitigation strategies for real-world benefits.

Acknowledgements

The authors would like to thank the Ghana Education Trust Fund (GETFund) for funding this research, Ghana Environmental Protection Agency (GhEPA), Ghana for supporting the deployment of sensors and Ms Rosine A Atidepe, Mrs Ann Poku, Mrs Zakiyah, Dr Kofi Amegah, Dr Pallavi Pant and Mr John Terry Moladza for their support.

References

Baron, R., Saffell, J., 2017. Amperometric Gas Sensors as a Low-cost Emerging Technology Platform for AQ Monitoring Applications: A Review, ACS Sensors 2:1553-1566

Bousiotis, D., Beddows, D. C. S., Singh, A., Haugen, M., Diez, S., Edwards, P. M., Boies, A., Harrison, R. M., and Pope, F. D. (2022). A study on the performance of low-cost sensors for source apportionment at an urban background site, Atmos. Meas. Tech., 15, 4047–4061, https://doi.org/10.5194/amt-15-4047-2022

C. Gameli Hodoli, F. Coulon, M.I. Mead (2020). Applicability of factory calibrated optical particle counters for high-density air quality monitoring networks in Ghana,

Heliyon, Volume 6, Issue 6, https://doi.org/10.1016/j.heliyon.2020.e04206

Carl Malings, Rebecca Tanzer, Aliaksei Hauryliuk, Provat K. Saha, Allen L. Robinson, Albert A. Presto & R Subramanian., 2019. Fine particle mass monitoring with low-cost sensors: Corrections and long-term performance evaluation, Aerosol Science and Technology, DOI: 10.1080/02786826.2019.1623863

Carslaw DC, Beevers SD, 2013. Characterising and understanding emission sources using bivariate polar plots and k-means clustering. Environmental Modelling & Software, 40, 325-329

Carslaw DC, Ropkins K (2012). "Openair — An R package for air quality data analysis." Environmental Modelling & Software, 27–28(0), 52–61. ISSN 1364-8152, doi: 10.1016/j.envsoft.2011.09.008

Carslaw, D. C., S. D. Beevers, SD., Ropkins K and 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". In: Atmospheric Environment 40.28, pp. 5424–5434 (cit. on p. 125)

Carslaw, D.C. (2015). The Openair manual — open-source tools for analysing air pollution data. Manual for version 1.1-4, King's College London.

Castell, Nuria, Dauge, Franck R., Schneider, Philipp, Vogt, Matthias, Lerner, Uri, Fishbain, Barak, Broday, David, Bartonova, Alena, 2017. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? Environment International, 99, 293-302

E. Manoli, D. Voutsa, C. Samara, 2002. Chemical characterization and source identification/apportionment of fine and coarse air particles in Thessaloniki, Greece Atmos. Environ., 36 (6), 949-961

Eric S. Coker, A. Kofi Amegah, Ernest Mwebaze, Joel Ssematimba, Engineer Bainomugisha (2021). A land use regression model using machine learning and locally developed low cost particulate matter sensors in Uganda, Environmental Research, Volume 199, 111352, https://doi.org/10.1016/j.envres.2021.111352

Garima Raheja, Kokou Sabi, Hèzouwè Sonla, Eric Kokou Gbedjangni, Celeste M. McFarlane, Collins Gameli Hodoli, and Daniel M. Westervelt (2022). A Network of Field-Calibrated Low-Cost Sensor Measurements of $PM_{2.5}$ in Lomé, Togo, Over One to Two Years *ACS Earth and Space Chemistry 6* (4), 1011-1021 DOI: 10.1021/acsearthspacechem.1c00391

Hartigan, J.A. (1975). Clustering Algorithms, Wiley & Sons, New York.

Hastie, T. J. and R. Tibshirani (1990). Generalized additive models. London: Chapman and Hall (cit. on p. 126)

Health Effects Institute. 2019. State of Global Air 2019. A Special Report on Global Exposure to Air Pollution and its Disease Burden. Boston, MA: Health Effects Institute.

Hinds, W. C., 999. Aerosol Technology: Properties, Behaviour, and Measurement of Airborne Particles, 2nd Ed., New York: Wiley-Interscience.

Javier Fernández-López, Klaus Schliep., 2019. rWind v1.1.3: Download, edit and include wind data in ecological and evolutionary analysis. Ecography 42: 804–810.

Jiang, Q., Kresin, F., Bregt, A.K., Kooistra, L., Pareschi, E., Van Putten, E., Volten, H., Wesseling, J., 2016. Citizen sensing for improved urban environmental monitoring. J. Sens

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., 2016. Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. Atmos. Meas. Tech. 9, 5281–5292.

Jovašević-Stojanović, Milena, Bartonova, Alena, Topalović, Dušan, Lazović, Ivan, Pokrić, Boris, Ristovski, Zoran, 2015. On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. Environmental Pollution, 206, 696-704

Kim, K.H., Lee, S.-B., Woo, S.H., Bae, G.-N., 2014. NOx profile around a signalized intersection of busy roadway. Atmos. Environ. 97, 144e154

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytouh, M., Di Sabatino, S., Bell, M., Norfordk, L., Britter, R., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75, 199-205

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., Pozzer, A., 2015. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature 525, 367–371. https://doi.org/10.1038/nature15371.

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., and White, A., 2016. Evaluating the performance of low-cost chemical sensors for air pollution research, Faraday Discuss., 189, 85–103, https://doi.org/10.1039/c5fd00201j

M. Statheropoulos, N. Vassiliadis, A. Pappa, 1998. Principal component and canonical correlation analysis for examining air pollution and meteorological data Atmos. Environ., 32 (6), 1087-1095

McFarlane, C., Isevulambire, P.K., Lumbuenamo, R.S., Ndinga, A.M.E., Dhammapala, R., Jin, X., McNeill, V.F., Malings, C., Subramanian, R., Westervelt, D.M. (2021). First Measurements of Ambient $PM_{2.5}$ in Kinshasa, Democratic Republic of Congo and Brazzaville, Republic of Congo Using Field-calibrated Low-cost Sensors. Aerosol Air Qual. Res. 21, 200619. https://doi.org/10.4209/aaqr.200619

Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203. doi:10.1016/j.atmosenv.2012.11.060

NOAA, 2019 available at https://www.ncdc.noaa.gov/data-access/model-data/model datasets/global-forcast-system-gfs [accessed, 13/09/2019

Peng Shi, Pin-Hua Xie, Min Qin, Fu-Qi Si Ke Dou and Ke Du., 2014. Cluster Analysis for Daily Patterns of SO2 and NO2 Measured by the DOAS System in Xiamen, Aerosol and Air Quality Research, 14: 1455–1465, doi: 10.4209/aaqr.2013.05.0160

Petkova, E.P., Jack, D.W., Volavka-Close, N.H., Kinney, P.L., 2013. Particulate matter pollution in African cities. Air Qual. Atmosphere Health 6, 603–614. doi:10.1007/s11869-013 0199-6

R Subramanian, Kagabo, A. S. ., Baharane, V., Guhirwa, S., Sindayigaya, C. ., Malings, C., Williams, N. J., Kalisa, E., Li, H., Adams, P., Robinson, A. L., DeWitt, H. L. ., Gasore, J., & Jaramillo, P. (2020). Air pollution in Kigali, Rwanda: spatial and temporal variability, source contributions, and the impact of car-free Sundays. *Clean Air Journal*, *30*(2). https://doi.org/10.17159/caj/2020/30/2.8023

Schwela, D., 2012a. Review of urban air quality in Sub-Saharan Africa region - air quality profile of SSA countries (No. 67794). The World Bank

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. 47, 11369-11377

Stuart K. Grange, Alastair C. Lewis, David C. Carslaw, 2016. Source apportionment advances using polar plots of bivariate correlation and regression statistics, Atmos. Environ., 145, 128-134

Technology Platform for AQ Monitoring Applications: A Review, ACS Sensors 2:1553-1566

Wood, S. N., 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC (cit. on pp. 126, 278).

Yang Wang, Jiayu Li, He Jing, Qiang Zhang, Jingkun Jiang & Pratim Biswas (2015) Laboratory Evaluation and Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement, Aerosol Science and Technology, 49:11, 1063-1077, DOI: 10.1080/02786826.2015.1100710

Yu, K.N., Cheung, Y.P., Cheung, T., Henry, R.C., 2004. Identifying the impact of large urban airports on local air quality by nonparametric regression. Atmos. Environ., 38 (27), 4501–4507.

Zikova N, Hopke PK and Ferro AR (2017). Evaluation of new low-cost particle monitors for $PM_{2.5}$ concentrations measurements. *J. Aerosol Sci* 105: 24–34.

Key highlights

i. Low-cost sensor data is suitable for source feature extraction.
ii. Low-cost sensors are applicable for complementing air quality capabilities in Africa.
iii. Cluster analysis and bivariate sectorial plots provide great insights into source apportionment studies.
iv. Meteorology plays a key role in aerosol pollution.
v. Localized PM pollution.

AUTHORSHIP STATEMENT

Manuscript title: Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana**.**

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *Environmental Pollution*.

Authorship contributions

Please indicate the specific contributions made by each author. The name of each author must appear at least once in each of the three categories below.

Category 1

Conception and design of study: Collins Gameli Hodoli, Iq Mead, Frederic Coulon;

acquisition of data: Collins Gameli Hodoli;

analysis and/or interpretation of data: Collins Gameli Hodoli, Iq Mead, Frederic Coulon.

Category 2

Drafting the manuscript: Collins Gameli Hodoli, Iq Mead, Frederic Coulon;

revising the manuscript critically for important intellectual content: Collins Gameli Hodoli, Iq Mead,

Frederic Coulon.

Category 3

Approval of the version of the manuscript to be published (the names of all authors must be listed):

Collins Gameli Hodoli, Iq Mead, Frederic Coulon

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

This statement is signed by all the authors (*a photocopy of this form may be used if there are more than 10 authors***):**

Author's name (typed)	Author's signature	Date

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana

Gameli Hodoli, Collins

2022-11-28

Gameli Hodoli C, Coulon F, Mead MI. (2023) Source identification with high-temporal resolution data from low-cost sensors using bivariate polar plots in urban areas of Ghana, Environmental Pollution, Volume 317, January 2023, Article number 120448

https://doi.org/10.1016/j.envpol.2022.120448