ARTICLE OPEN

Check for updates

Towards comprehensive air quality management using low-cost sensors for pollution source apportionment

Dimitrios Bousiotis [1], Gordon Allison[2], David C. S. Beddows[1], Roy M. Harrison [1,3] and Francis D. Pope [1 ✉]

Successful air quality management and control not only requires measurements of air pollution levels. It also requires information on the sources of air pollution, and their relative magnitudes and importance, to plan and enact cost-effective control measures. This paper provides an important breakthrough towards the wider and more comprehensive use of source apportionment via low-cost techniques. Low-cost sensor measurements, along with the statistical methods of Positive Matrix Factorization (PMF) and k-means clustering, were able to successfully pinpoint and quantify the main sources of pollution in three regulatory important sites (a construction site, a quarry and a roadside). The anticipated levels of pollution, which were dependent on meteorological conditions and temporal variations, were assessed. The analysis provides information crucial for successful air quality management and control at a significantly lower cost than previously achieved. The strengths and weaknesses of the methodologies used are illustrated and discussed.

INTRODUCTION

Globally, poor air quality is estimated to cause about 8.8 million premature deaths per year[1], it affects the cardiovascular and respiratory systems, and is also known to cause cancers and affect cognition[2,3]. It is the leading global environmental risk for human health. To tackle this problem and reduce the negative effect of pollution emissions, successful air quality management and control is needed. This requires not only measurements of air pollution levels, it also requires information on the sources of air pollution and their relative importance. Without this critical, targeted information on pollution sources, it is difficult to plan and enact control measures to reduce air pollution. For many years, source apportionment was conducted using research-grade instruments. Many studies have been conducted to develop methodologies for source identification and apportionment which can help in assessing the sources of pollution at a given site. Among the methodologies proposed, the most commonly used ones are Positive Matrix Factorisation (PMF)[4,5] and the k-means clustering[6,7]. The cost and logistics associated with the measurements needed though makes their use scarce and time consuming, and typically limits their usage to the academic literature.

Over the last two decades there has been a revolution in air quality measurement using low-cost sensors (LCS). For PM measurements, these sensors typically use tried and tested methodologies that are used in regulatory-grade instruments, but are miniaturised. These sensors are becoming more dependable and have proven their capabilities in air quality monitoring[8–11] and in several applications for which monitoring costs were previously a limiting factor[12,13]. To achieve lower costs, these LCS do not offer the same level of accuracy and need sophisticated statistical analyses and calibration to provide reliable results[9,14–17].

To date, there have been huge numbers of studies that use LCS for the measurement of air quality but very few studies of their use in source apportionment. Until this point, LCS have been tested for source apportionment in background environments with either less complicated sources from greater distances (which made it easier to distinguish different pollution sources) or in cases where the major sources of pollution were limited in number[18,19]. In some cases LCS data were used with simpler methodologies mainly for pollution source identification[8,20]. Previously, we have shown the use of two statistical techniques (k-means clustering and PMF) that use the PM size distribution data measured by LCS, for source identification and apportionment. While there are great differences in the approach between the two methods, their outcomes complement each other in providing a clearer picture of the sources of pollution and the conditions that affect the extent of their impact[18,21].

This study shows that LCS have great potential to identify local sources of air pollution in complex urban environments. We use k-means clustering and PMF analyses on the PM size distributions measured by low-cost optical particle counters in three distinct locations. Two locations are industrial in character: a construction site and a quarry. The third site is a roadside location. All locations are in central England, UK, and contain activities that require boundary line monitoring. We show that individual sources of PM can be identified and their contribution to overall PM concentrations can be apportioned. This study provides a low-cost methodology for PM source apportionment that will facilitate the pinpointing of PM sources thereby providing the required information for industries and regulatory bodies to reduce PM concentrations and achieve national and international air quality standards.

RESULTS

Results from Curzon Street, Birmingham, UK

The use of the two statistical methods on LCS data provided sufficient insight on both the sources that affect the air quality at

[1]School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. [2]DustScanAQ, Witney, Oxford, UK. [3]Present address: Department of Environmental Sciences / Center of Excellence in Environmental Studies, King Abdulaziz University, PO Box 80203, Jeddah 21589, Saudi Arabia. ✉email: f.pope@bham.ac.uk

Table 1. Conditions and frequency of the clusters from the k-means analysis for all sites.

Curzon Street

Cluster	WS (m s^{-1})	T (ºC)	RH (%)	PM$_1$ (µg m^{-3})	PM$_{2.5}$ (µg m^{-3})	PM$_{10}$ (µg m^{-3})	PM$_1$/PM$_{10}$ (%)	Frequency (%)
CL1	2.2 ± 1.0	12.3 ± 3.4	78.9 ± 10.4	3.2 ± 0.9	19.2 ± 7.2	62.7 ± 47.9	5.1	8.7
CL2	1.1 ± 0.6	10.4 ± 2.5	91.9 ± 6.05	14.7 ± 4.9	23.2 ± 7.9	37.4 ± 17.6	39.2	3.7
CL3	1.8 ± 1.2	12.7 ± 4.5	87.1 ± 11.2	3.3 ± 2.4	8.8 ± 5.7	29.1 ± 36.2	11.5	42.0
CL4	2.0 ± 1.3	12.0 ± 3.7	77.2 ± 13.4	2.0 ± 0.8	11.1 ± 5.8	42.7 ± 47.1	4.8	17.5
CL5	1.9 ± 1.2	11.5 ± 3.1	83.5 ± 11.4	3.5 ± 3.6	10.7 ± 6.3	33.8 ± 25.5	10.2	28.1
Average	1.9 ± 1.2	12.1 ± 3.9	83.8 ± 12.2	3.5 ± 3.5	11.2 ± 7.1	36.1 ± 37.7	9.8	–

Mountsorrel quarry

Cluster	WS (m s^{-1})	T (ºC)	RH (%)	PM$_1$ (µg m^{-3})	PM$_{2.5}$ (µg m^{-3})	PM$_{10}$ (µg m^{-3})	PM$_1$/PM$_{10}$ (%)	Frequency (%)
CL1	0.9 ± 0.7	21.0 ± 4.0	68.5 ± 10.2	1.9 ± 1.0	4.5 ± 2.7	13.4 ± 9.4	14.3	44.8
CL2	0.2 ± 0.3	15.6 ± 1.5	84.4 ± 4.6	9.7 ± 2.5	18.5 ± 5.8	38.9 ± 21.4	25.0	9.24
CL3	0.6 ± 0.7	16.6 ± 2.0	82.0 ± 5.1	5.3 ± 2.0	11.1 ± 4.8	31.9 ± 26.9	16.6	16.8
CL4	0.7 ± 0.7	18.1 ± 3.1	77.5 ± 9.2	3.4 ± 1.5	7.9 ± 3.8	24.2 ± 20.8	14.1	22.2
CL5	0.2 ± 0.3	15.7 ± 1.0	86.1 ± 1.1	22.2 ± 7.3	46.0 ± 16.8	117.7 ± 217.1	18.9	4.31
CL6	1.7 ± 0.8	16.8 ± 2.2	69.5 ± 11.5	2.0 ± 1.0	18.9 ± 6.6	89.4 ± 22.4	3.17	2.62
Average	0.7 ± 0.7	18.7 ± 3.8	75.3 ± 10.7	4.4 ± 4.9	9.8 ± 10.3	27.7 ± 53.3	16.0	–

Charlbury roadside

| Cluster	WS (m s^{-1})	T (ºC)	RH (%)	PM$_1$ (µg m^{-3})	PM$_{2.5}$ (µg m^{-3})	PM$_{10}$ (µg m^{-3})	PM$_1$/PM$_{10}$ (%)	Frequency (%)
CL1	2.9 ± 1.1	12.7 ± 1.0	98.6 ± 1.9	21.5 ± 6.3	23.2 ± 5.9	27.4 ± 5.4	78.6	16.9
CL2	2.3 ± 0.7	11.4 ± 1.3	99.0 ± 2.9	24.0 ± 12.4	24.0 ± 11.8	28.5 ± 13.6	84.1	5.23
CL3	3.7 ± 1.6	12.0 ± 2.8	89.1 ± 8.9	6.0 ± 3.4	7.4 ± 3.3	14.8 ± 4.8	40.5	47.7
CL4	3.0 ± 1.5	12.9 ± 2.2	88.0 ± 11.5	19.7 ± 10.2	21.4 ± 10.5	29.5 ± 11.7	66.8	19.8
CL5	4.3 ± 1.3	10.3 ± 2.8	85.3 ± 7.3	3.62 ± 0.6	6.09 ± 0.8	19.3 ± 8.1	18.7	10.5
Average	3.4 ± 1.5	12.1 ± 2.5	90.6 ± 9.4	12.0 ± 9.9	13.6 ± 9.8	21.0 ± 10.0	57.2	–

Curzon Street site, as well as an estimation of the effect of each source. The results of the two methods are found in Supplementary Figs. 5 and 11 (Factors from the PMF analysis are marked as F followed by the factor number, while clusters from the k-means clustering analysis are marked as CL followed by the cluster number). Apart from the Particle Number Size Distributions (PNSD) profiles of the PMF factors, the temporal and wind variation of their G contribution (the relative contribution of each factor on the local atmosphere) of each factor is presented. Additionally, Table 1 shows the average meteorological conditions and PM concentrations of the clusters from the *k*-means analysis.

The effect of the activities from the construction site were found from both methods (CL1, CL3, CL4, CL5 and F2, F4). Specifically, the clustering method separated the effect of the construction site, which was more pronounced for particles of about 10 µm in diameter (Fig. 1). Similarly, using PMF, two hotspots of particle emissions were found, peaking at about 5 µm and having increased concentrations for larger particles (F4), as they presented greater contributions to the PM$_{10}$ concentrations rather than the PM$_{2.5}$ (Table 2). This is consistent with findings from previous studies on construction sites[22,23]. This source also presented a peak at the smallest available particle size, which in most cases points to increased particle concentrations at smaller particle sizes below the lower measurement size limit of the sensor. Looking at the temporal trend of this source, it is directly associated with the earth-moving works performed at the construction site (due to the PNSD profile, this can be associated with emissions from vehicles working at the construction site), though a small contribution of this source is also visible outside the period when such works were undertaken (according to the works diary provided by the constructor, earth moving works

stopped on the 21st of September), which may be associated with resuspension from the same site. This factor presented the greatest contribution of PM$_{10}$ in the area, reaching up to an estimated 60% on working hours (5:00–15:00 on weekdays) (Fig. 2). The effect on smaller PM sizes is significantly smaller (being about 5% in average, though more than double of that on working hours) pointing its association with larger particles associated with earth-moving activities taking place at the construction site. This effect is found to be greatly reduced outside the working hours.

Another source of large particles associated with activities within the area and period of the earth moving activities was also pinpointed (F2), having measured peaks between 1 and 2 µm. This presented a wider area as a source location, though with a similar temporal variation as the F4, and is probably associated with other activities at the construction site. Nevertheless, the possible effect of other external sources occurring at the same time cannot be excluded. The effect of this factor in the PM load at the site was almost negligible for all PM sizes, though its contribution was doubled during working hours. A PNSD profile similar to this was found by Belkacem et al.[24] to be associated with non-exhaust vehicle emissions (i.e. resuspension or tyre and brake wear), which may be the case in the present study as well.

While no significant differences were found in the PNSD using the *k*-means (Supplementary Fig. 9), the PMF analysis showed that the main source of particles <1 µm is to the southeast (Supplementary Fig. 8 and Table 2), where Birmingham city centre is located. The effect of this source is greater during night-time and early morning hours. This may be associated with the daily variation of the boundary layer height (BLH), which is reduced during these hours and tends to increase the effect of any
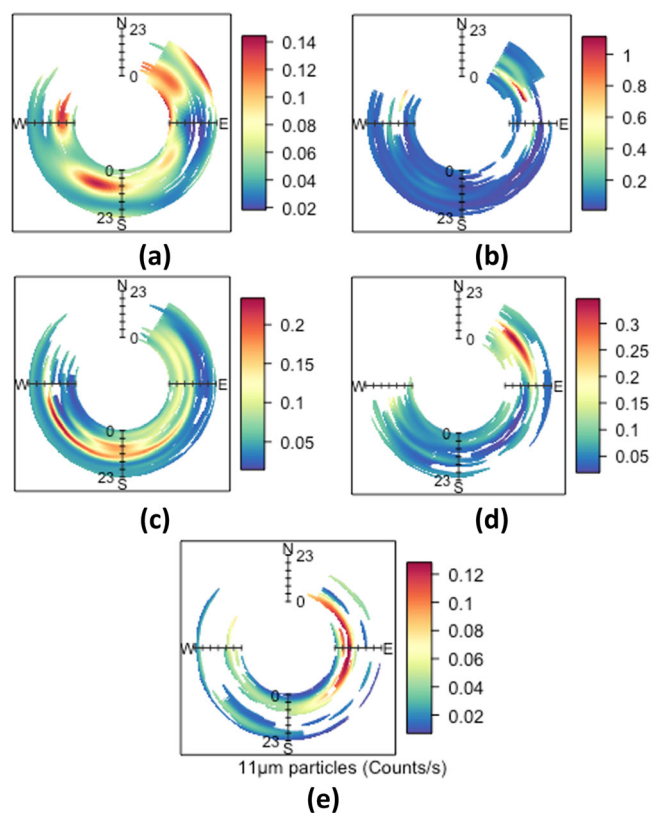
Fig. 1 Variation of 11 μm diameter particles of the clusters in Curzon Street. ZPolar annulus of particles with diameter 11 μm for the clusters from the k-means clustering analysis (Curzon Street). Particles with diameters between 7 and 20 μm are found to be directly associated with the emissions from the works at the construction site. The effect from the works has been found in almost all clusters formed by the k-means analysis showing the significant effect of the specific source regardless of the atmospheric conditions. **a** CL1, **b** CL2, **c** CL3, **d** CL4, **e** CL5.

local pollution source due to reduced mixing[25–27]. This source was pinpointed by both methods (CL2 and F1), and it was associated with the most polluted conditions, with respect to particle mass concentrations, for PM_1.

Finally, the PMF also identified two more constant sources of particles (F3 and F5), which did not present a clear spatio-temporal variations. These are probably associated with back-ground emissions related to activities from the urban environment or the greater area, which are not associated to the construction site according to their other characteristics. Among these two, F5 appeared to have a more significant contribution on larger PM (both $PM_{2.5}$ and PM_{10}), which may associate it with sources of particles from marine origins. These two factors did not present any variation during working hours further suggesting their background character.

The use of the clustering provides insight to the real-world conditions occurring at the site. Thus, while the separation of the sources that affect the air quality at the site is not as distinct as with the PMF, this method helps in better understanding the combination of the sources as well as their effect with different conditions. The small wind variation for the period studied resulted in rather homogenised clusters, with wind direction originating in the direction of the construction sites. All clusters carry the effect of the source to the southeast (the city centre and probably the railroad also located on that side), which affects the site throughout the day for the whole measuring period, though its effect is enhanced during the night hours, as mentioned earlier.

While not clearly visible when considering PM measurements, the activities of the construction site are more visible when different particle size ranges are considered. It is found that for particles above 5 μm and above, two hotspots are observed that are similar to that in the PMF analysis (Fig. 2). The clustering method managed to further quantify the effect of these two hotspots (hence the large number of clusters associated with the works on the construction site), showing that the one on the northeast contributed a greater number of larger particles, not only having a larger PM_{10} concentration, but a smaller PM_1 to PM_{10} ratio as well (Table 1). The latter though may be the effect of the sources of smaller particles found at the south side. This information though cannot be clarified without measurements of particles at smaller sizes than that the optical particle counter (OPC) provides, or without chemical composition data that were not available for this study.

Results from the Mountsorrel quarry in Leicestershire, UK

At the quarry in Leicestershire, UK, both methods (Fig. 3 and Supplementary Figs. 12–15 and 17–19) identified the background particle composition profiles (the particle composition profiles occurring when no significant local source affects the measuring site) and their variation with the conditions these occur (CL1, CL3, CL4 and F1). Additionally, they managed to identify three important sources of particles, one being associated directly with the works in the quarry (F2). This was separated into two parts, one being the crushing area of the quarry, located about 500 m southwest from the measuring site. Due to the great distance from the measuring site, its effect is important with stronger winds from that direction (CL6). Such conditions bring particles predominantly in the size range of 1.5 to 5 μm to the site and can increase the concentration of $PM_{2.5}$ and PM_{10} up to three times compared to the average conditions when downwind (though this only occurred for 3% of the measurement period) (Table 1). This effect is confirmed by the PMF as well. Using the estimated PM concentrations for the factor F2, which is the factor directly associated with the work at the quarry, F2 is responsible for about 17% and 10% of the PM_{10} and $PM_{2.5}$ concentrations at the site respectively (Table 2). As with the Curzon Street site, work on the construction site had a greater effect on the PM_{10} concentrations compared to $PM_{2.5}$. These values are almost doubled with southwesterly winds (32% and 18% respectively), making the quarry an important source of such PM (the first factor though remains the greatest contributor of PM at the site), especially on working days (Fig. 3).

The second important source is at the area directly to the south of the measurement site located at a smaller distance compared to the crushing area, which was grouped together with the crushing part of the quarry. At this location other works take place which seem to contribute to a significant increase in the concentration of particles of smaller sizes, between 0.5 and 1.5 μm.

Additionally, the PMF managed to specify a source associated with marine origin at the site (F3), being responsible for a significant portion of the PM_{10} and $PM_{2.5}$ at the site. This factor presented the same PNSD profile as the marine factors in the other areas of study, with two peaks at sizes <1 and 2 μm. While this factor was separated from the one associated with the quarry its effect is more significant with southwesterly winds (Supple-mentary Fig. 16). The PM concentrations of this factor are probably enhanced by the emissions from the quary as the incoming air masses of marine origin pass from the quarry area. Still, its contribution to the PM is smaller both in general and specifically with SW winds (about 14% and 19% respectively) compared to the factor that was directly assigned to the quarry.

Finally, an undefined source of particles which is probably located very close to the measuring site (as it is associated with rather low wind speeds) probably to the north and mainly during

Table 2. Estimation of the PM_1, $PM_{2.5}$ and PM_{10} contribution (%) of each PMF factor.

Curzon Street			
Factor	PM$_1$ (%)	PM$_{2.5}$ (%)	PM$_{10}$ (%)
F1	55.2	15.8	Negligible
F2	0.2	0.2	Negligible
F3	8.1	20.0	Negligible
F4	1.2	5.0	33.6
F5	27.1	50.0	38.5
Unexplained	8.3	9.0	27.8
Mountsorrel quarry			
F1	70.2	68.6	56.6
F2	3.5	8.86	16.0
F3	12.4	13.4	12.0
F4	0.7	Negligible	0.2
Unexplained	13.2	9.2	16.2
Charlbury roadside			
F1	74.5	75.0	65.2
F2	15.7	11.4	Negligible
F3	3.0	8.1	19.2
Unexplained	6.8	5.5	15.6

The unexplained variation includes additional sources not described by the factors deriving from the analysis.
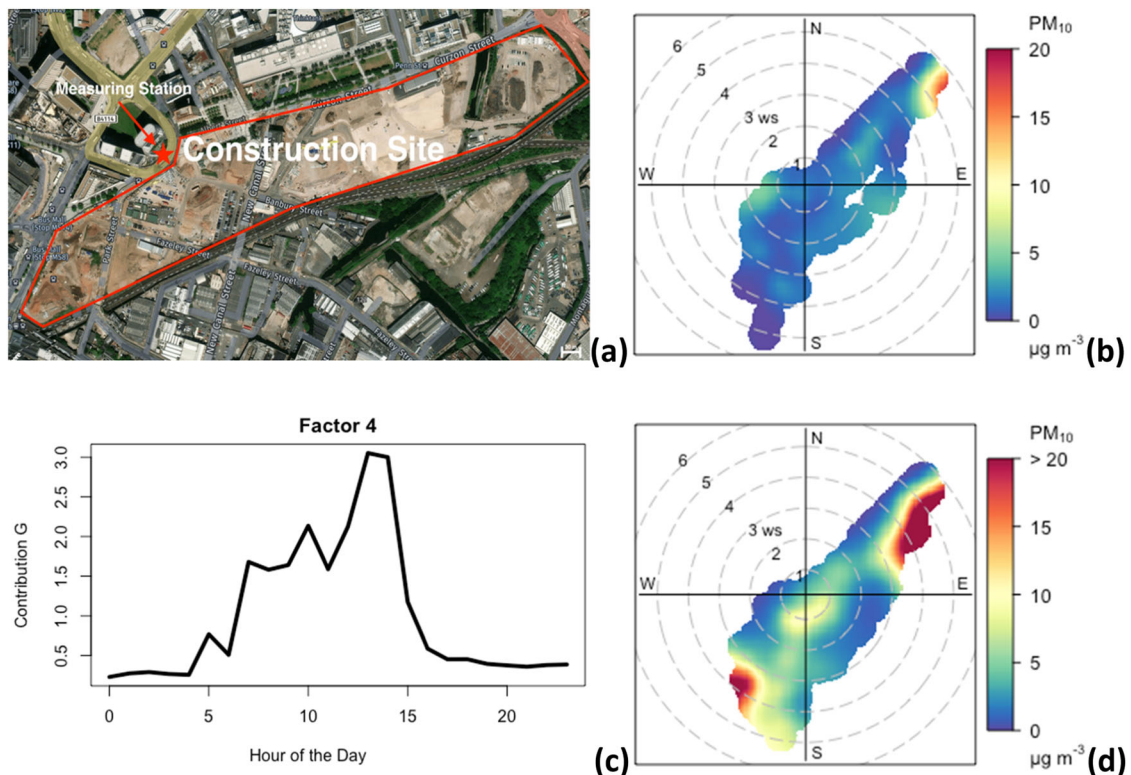
Fig. 2 Characteristics of the factor associated with the industrial site in Curzon Street. Map of the Curzon Street construction site and the location of the measuring station (image by ©Google) (**a**). Diurnal variation of the G contribution (the relative contribution of the factor compared to the average contribution of the factor which is equal to 1) of the factor mainly associated with the works on the construction site (**c**). The polar plots on the right show the estimated PM_{10} concentrations attributed to the factor on non-working (**b**) and working hours (**d**). Working hours are considered between 5:00 and 15:00.
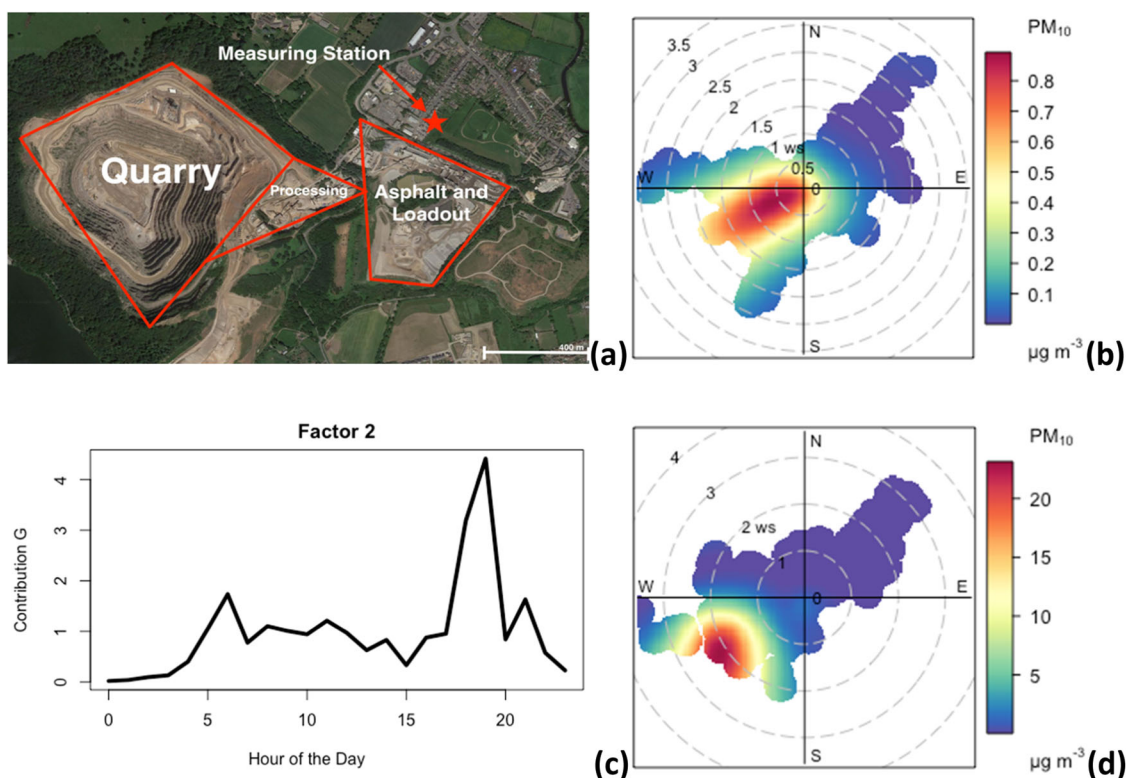
Fig. 3 Characteristics of the factor associated with the works at the Mountsorrel quarry. Map of the Mountsorrel quarry and the location of the measuring station (image by ©Google) (**a**). Diurnal variation of the G contribution (the relative contribution of the factor compared to the average contribution of the factor which is equal to 1) of the factor mainly associated with the quarry and the additional works area (**c**). The polar plots on the right show the estimated PM_{10} concentrations attributed to the factor on non-working (**b**) and working hours (**d**). Please note the difference in the scale. Working hours are considered between 6:00 and 17:00 on weekdays and 6:00 to 12:00 on Saturdays (according to the schedule provided by the quarry operator).

the night-time was also found. While with such low wind speeds it is hard to infer the source of a location with good precision, there was a persistent connection of the specific source with wind directions from the northern sector (CL5 and F4). The PMF analysis separated this source from the emissions from the nearby town to the north (though this does not exclude the possible connection to it, as wind speeds are lower during the night times), and the estimated effect of this factor was found to be rather low on the PM concentrations at the site.

From the results of the analyses, the quarry remains a notable but not the most important source of particles at the site during its working hours (according to the clustering analysis its effect exceeded the updated daily limit set by the World Health Organisation[3] only for 12% for $PM_{2.5}$, and 3% for the PM_{10} of the monitoring period), though the highest PM concentrations were observed with almost calm conditions and are probably associated with other sources (mainly the nearby town to the north). Its effect is separated into two parts at the site, and it is probably due to different works that take place in these areas which have a different effect on the PNSD. Specifically for the southwestern site (the crushing area), its effect is more pronounced when downwind and specifically with higher than average wind speeds, while its effect is a lot smaller with lower wind speeds, still increasing the PM concentrations but with average values considerably smaller than double those of the average conditions. Using the PMF the factor associated with the works at the quarry presented similar results, exceeding the limits set by WHO on only 5% of the hours of the measuring period spread in several days within the measuring period (using the estimated PM concentrations), though this effect is slightly increased when the effect of the marine factor which appears to

be affected by emissions from the quarry is considered. The consistency of the results between the two methods is highly encouraging as it shows that both methods can provide meaningful and self-consistent results for air quality studies.

Results from the roadside site in Charlbury, UK

This case study investigated a location, a provincial roadside site in Charlbury UK, predominantly impacted by traffic emissions rather than industrial emissions. Once again, both k-means and PMF methods were applied on the dataset and their results are summarised in Fig. 4 and Supplementary Figs. 20–26. It should be noted that, to the authors' knowledge, roadside measurements using LCS were never analysed in this manner before, as many particles derived from traffic are mostly smaller in size than those measured by them.

In this case, the clustering method managed to identify 5 unique conditions (Table 1) that lead to different particle profiles, separating night and day (CL1 and CL3), polluted (CL1, CL2 and CL4) and clean conditions (CL3) as well as separating a period when strong wind was blowing parallel to the road (CL5). The most polluted conditions were found when wind was blowing from the south (CL2), a side which coincides with the main area of Charlbury, without this excluding the effect of the traffic from the nearby road (junction). Contrary to this, particle concentrations were rather low when the wind was blowing parallel to the axis of the road from the north regardless of this occurring on a normal weekday (CL5). CL3 and CL4 were the most common clusters (about 70% of time combined), occurring almost every day (Supplementary Table 1) with higher frequency during midday and the rush hours respectively. CL3 is not
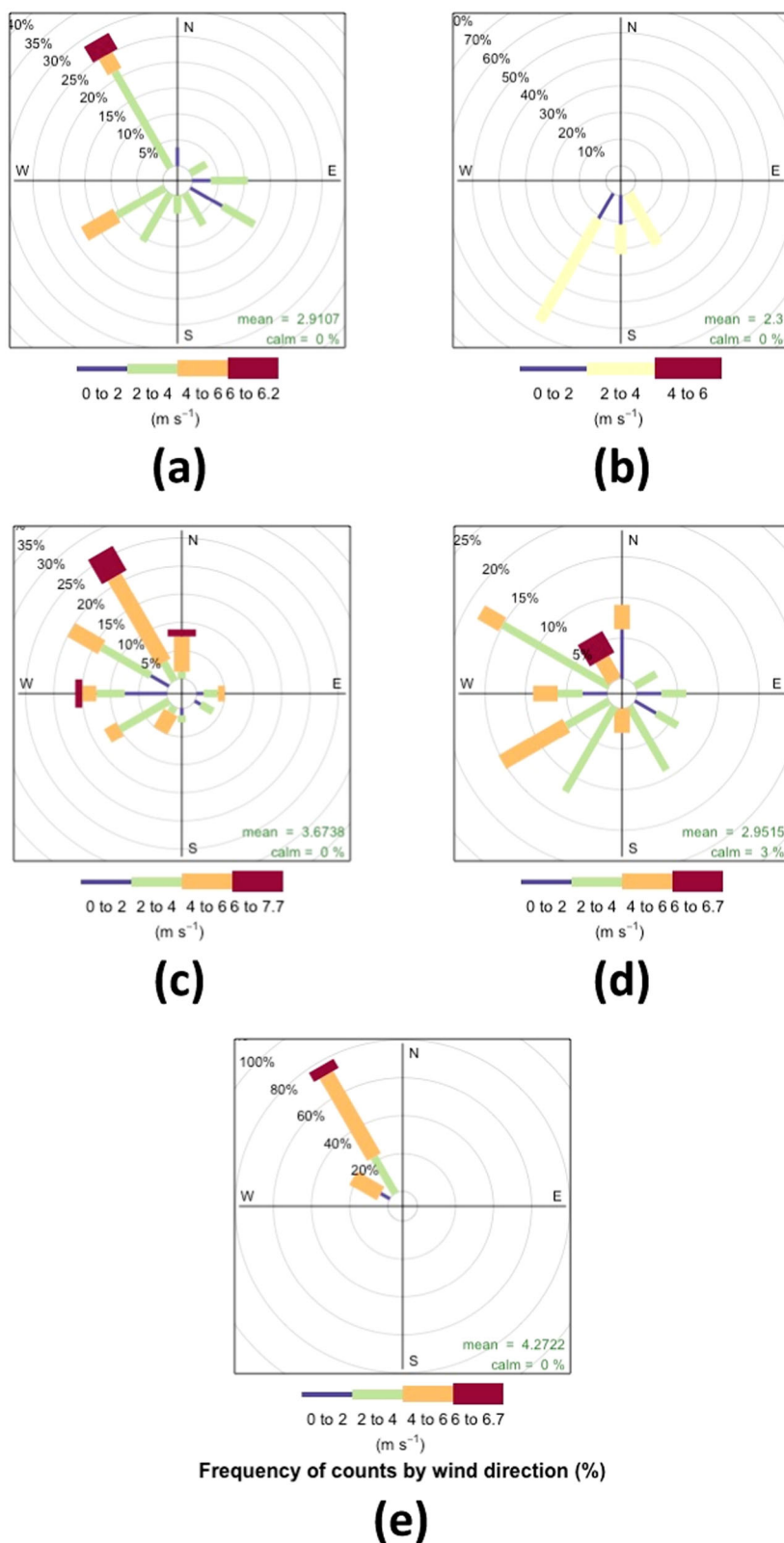
Fig. 4 Wind profiles of the clusters at the Charlbury roadside site. Windroses of the five clusters identified via the *k*-means clustering analysis generated for the Charlbury roadside location. **a** CL1, **b** CL2, **c** CL3, **d** CL4, **e** CL5.

greatly affected by the nearby traffic having low PM concentrations and PM_1 to PM_{10} ratio, while CL4 is, as expected, greatly affected by traffic having the opposite characteristics (a high PM_1 to PM_{10} ratio indicates a greater relative content of fine

particles which, especially for such an environment, are probably associated with traffic).

The PMF separated three distinct particle profiles at the site. The main source that causes the greatest increase of particles at the

site (F1), has increased contributions as particle size decreases. This factor is probably associated with the traffic from the nearby road as increased contributions of this factor were found with easterly wind directions, though the effect of the part of the town at that side cannot be excluded. This factor presents a rather balanced diurnal variation and is associated with the majority of the PM load for all size ranges, as found by the estimated PM concentrations (Table 2).

Another factor with similar PNSD profile but significantly lower contribution to the PM concentrations at the site was also found (F2), though this one presents its highest contributions with southerly winds. As mentioned earlier, while the town centre of Charlbury is located at that side, the possible effect of the nearby junction should not be overlooked. This factor presents a more distinct diurnal variation with greater contributions during the night and early morning hours, which may associate the variation of this factor with that of the BLH. While it has a similar PNSD profile with F1, using the estimated PM concentrations, this factor has a significantly greater effect on the $PM_{2.5}$ and PM_1 concentrations rather than the PM_{10}.

Finally, one more source of larger particles (F3), which seems to have a more regional character and presented a more balanced contribution profile with wind directions from almost all sides. This factor presents a two-peak PNSD profile, with the peaks being at the same particle sizes as the ones assigned to the marine factor in the other study areas. This factor does not have any distinct wind direction or speed hotspots, though it is the only factor with its estimated PM concentrations increasing with increased wind speeds. This factor is associated with more than 20% of the estimated PM_{10} concentrations at the site though its contribution reduces for smaller PM sizes. The very high PM_{10} contribution found for this factor makes it unlikely for it to be solely attributed to marine sources. Solutions with higher number of factors were attempted, though they had a very small effect on the variation of this factor.

On the roadside site, the PMF when used on PNSDs and PM data was able to identify and separate rather distinct particle profiles, ones though that could not be confidently attributed to specific sources. This is probably due to the small measuring period, as greater datasets tend to smoothen and clarify the results of such analyses. The clustering method on the other hand provided a clearer picture of the conditions at the site according to the combination of particle sources, meteorological conditions, and time of the day (which is associated with different anticipated traffic densities). The uncertainty of the results from the PMF analysis in the specific site shows the need for either additional information (such as of the atmospheric chemical composition) or longer measurement periods for confident association of factors with sources. This will likely be the case for any analysis which includes sources that are associated with particles below the measuring limit of the LCS.

DISCUSSION

The low-cost methodology developed in this paper was able to identify the main pollution sources affecting air quality at the three case study sites. Low-cost PM size distribution data, provided by the OPC sensor, gives sufficient information for the k-means clustering and PMF analyses to identify and apportion various pollution sources. Furthermore, it was possible to estimate the contributions of the identified sources to overall PM mass concentrations in the regulatory $PM_{2.5}$ and PM_{10} size ranges.

The three case studies provided complex urban environments in which to assess and apportion air pollution sources. The approach managed to effectively separate the effect of sources in the surrounding area, find the periods and conditions that affect their variability, and quantify the effect of the pollution sources depending on real conditions (weekday, meteorology etc.). The

k-means clustering provides a clear image of the air quality at each site according to changes in the general air quality patterns, making it a reliable method for identifying the effect of near-constant sources, even with limited datasets. The PMF approach can identify and apportion distinct sources and estimate their contribution to total $PM_{2.5}$ and PM_{10} mass concentrations.

The low-cost approach was particularly good at identifying and apportioning air pollution sources that have super-micron-sized PM contributions. The lack of information, from the OPCs, at particle diameters <0.38 μm makes identification of most regional sources more difficult, because their PNSDs of regional sources typically have characteristic peak PNSD features smaller than the OPC size cut-off. Hence, most regional sources were typically homogenised into a single source with little information about their origin. These sources appeared as a single blurred factor or cluster, with the PNSD profile peaking at the smallest particle size bin available. The low-cost approach will also likely miss local combustion sources, which are typically characterised by small particles. This omission will be important at roadside sites and other locations under heavy influence of internal combustion engines, as pointed by our study.

In all the areas studied, the PMF approach was able to identify one distinct regional source that was attributed to a marine factor. This source had a PNSD profile with super micron features and was observed across all three case study sites as well as a previous study investigating an urban background site, described in Bousiotis et al.[14] (Fig. 5). This ability to identify distinct sources by PNSD profile is highly encouraging because it reduces the number of factors that need to be distinguished using the PMF analysis, as well as indicating that similar sources show similar profiles regardless of the local environment. This can be very useful in simplifying the source apportionment process, making such applications easier in the future and its wider use feasible, without the need of expert personel.

Currently, PM regulations are focused on $PM_{2.5}$ and PM_{10}. While the methods presented here may not be able to distinguish all the PM sources that affect air quality at a particular site, they can separate and approximately quantify the predominant ones, which in most cases contribute the most in the deterioration of the local air quality. The approach cannot assess total PM number concentrations, a metric that is becoming increasingly important from a regulatory perspective, which is dominated by particles of smaller diameter than that measured by the OPC. Going forward, low-cost particle number counters used in combination with the low-cost OPCs might offer good potential in this direction.

While the previous works[18,21] focused characterising the capabilities of the methodologies of LCS source identification and apportionment, the present study shows that the methodology can be useful for industrial applications. Additionally, compared to the previous work, this study highlights the ability of PMF to provide quantitative results of the effect of different industrial activities on the air quality in the surrounding area. This is crucial for the assessment of the environmental footprint of such activities which was limited until now due to the high cost that came with such analyses. Additionally, the k-means showed the effect of the meteorological and temporal variations in air quality level which can be useful for applications such as urban planning or environmental alerts.

There is great potential for low-cost source identification and apportionment. By achieving local source apportionment in a lower cost manner, this study makes clear that PM source apportionment could be used more widely for regulatory and compliance purposes. By pinpointing individual pollution sources, the new methodology provides significant prospects to reduce overall PM concentrations through the targeting of specific sources, for example, nuisance dust from construction. Primarily, the technique should allow for greater use of source apportionment by significantly reducing the associated
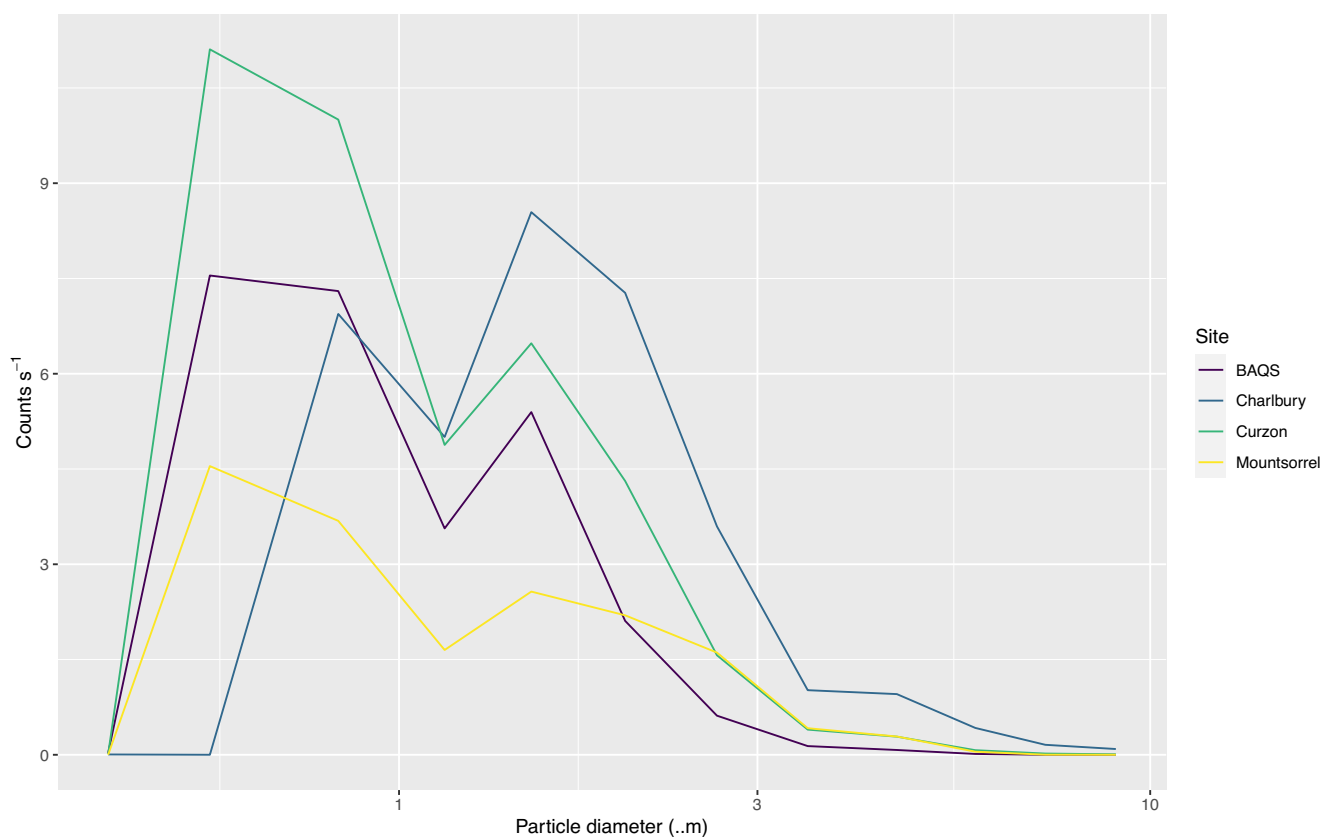
Fig. 5   The marine factor as identified at the studied sites. The marine factor was identified at the sites of the study and BAQS. A factor associated with marine sources has been identified at all the sites in this study as well as in a similar study done for the BAQS[17]. At all sites, it presents a unimodal PNSD profile with peaks at about 750 nm and 1.5 μm. Regardless of the difference in the particle counts (BAQS site result is multiplied by 10) which is affected by the local conditions, uniformity of the result is expected as all the sites are greatly affected by air masses originating from the Atlantic Ocean, which is the most common origin for incoming air masses in the U.K.[39].

economic and logistical costs. Already, we believe that low-cost source apportionment can supplement regulatory grade source apportionment. Further field testing is now needed to see if low-cost source apportionment can replace regulatory grade source apportionment and in what situations this is appropriate. The low-cost technique should be particularly beneficial in boundary line source apportionment of point source polluters. The use of sensor arrays across urban areas in combination with the techniques described in this paper opens up the prospect of source triangulation to pinpoint precisely the location of emitters.

METHODS
Aerosol counts and sizing
For the PM size distribution measurements at all sites, Alphasense OPC-N3 sensors (http://www.aqmd.gov/aq-spec/product/alphasense) were used, with each sensor costing approximately GBP250. The sensor is a small-sized Optical Particle Counter (OPC), with dimensions of $75 \times 63.5 \times 60$ mm and weight under 105 g. The sensors measure particle number concentration in the size range of 0.35–40 μm in 24 bins, with a maximum particle count of 10,000 per second. The count and classification of the particles is achieved by measuring the scattered light by individual particles as they pass through a laser beam within the sensor. It has a minimum sampling time resolution of <1 s, though in the present study measurements were averaged into hourly periods. At the industrial sites, in a distance of <2 m there were regulatory instruments used to monitor the dust emissions and air quality either by the respective operator or

DustScanAQ. In both cases the OPC was directly calibrated against the measurements from the regulatory instrument.

Measuring sites
Three distinct sites were used in this study, all of which are located within central England, UK, see Supplementary Fig. 1. The site in Curzon Street, Birmingham (52.48°N; 1.89°W), is located close to the construction works of a major new railway station and north of the Birmingham city centre. The construction site lies in a broad arc from the monitoring station, that extends from the south to the northeast of the monitoring station. Particle number size distribution data were collected for the period between 2/9/2020 to 26/10/2020.

The second site is the Mountsorrel quarry in Leicestershire (52.73°N; 1.16°W) and is located south of the nearby town. It is one of the largest quarries in Europe and comprises of the crushing site and several other facilities. Data were available for a 4-week period between 1st and 28th of July 2021 at a location northeast of the crushing area of the quarry and about 100 m north of the other work areas within the quarry site.

The third site in this study is a roadside location on Market Street (51.87°N; 1.48°W), a provincial road in Charlbury, UK. Though Charlbury is a rather small town (population of about 3000), the road is a relatively busy one traversing the centre of the town. The measuring period is the week between 7/10/2021 and 14/10/21.

The heights of all measurements were at ~2 m above ground level. For all the sites, particle number concentrations were measured at a 1 min resolution, then mean averaged to hourly

intervals. Time periods were chosen such as no missing or below zero values were found in the datasets.

Meteorological data

For Curzon Street, data from the meteorological station at the University of Birmingham were used (the average wind rose for the measurement period is found in Supplementary Fig. 2). While the University of Birmingham meteorological station is located about 6–7 km away from the measuring site, the data provided are not biased by the local urban topography at Curzon Street, thus providing a better representation of the regional conditions in the greater area.

For the Mountsorrel quarry, met data was available from a met station located a couple of hundred metres to the west of the measuring station (the average wind rose for the period is found in Supplementary Fig. 3). Wind speed and direction measurements were made at 5 m height for this site.

Meteorological data were not available on the measurement site at Market Street in Charlbury, thus data from the nearest met station in Little Rissington (14 km west from the site) were used (the average wind rose for the period is found in Supplementary Fig. 4). No data that may identify individuals were collected for this study.

k-means clustering

The k-means clustering is a widely used source identification method, used in many studies in the past either on regulatory grade instrument[6,28,29] or LCS[21] data. It performs better than other clustering methods, as it produces clusters with greater consistency between their elements and better separation from the elements of other clusters[30,31]. k-means is a variable reduction method which partitions the observations into k sets, forming groups with the minimum possible variance (squared Euclidean distance) of the elements of each cluster[32]. The choice of the optimal number of clusters was informed by testing the different solutions with two metrics, the Dunn Index[33] and the Silhouette width[34] (Supplementary Fig. 27), as proposed by Beddows et al.[30]. When in some cases the choice of the best solution was not clear, the solution among the candidates that better described the conditions at each site was chosen.

Positive Matrix Factorization

The PMF is a multivariate data analysis method, commonly used for source apportionment in air quality studies, with numerous applications[5,6,35]. It describes the relationships between species measurements using a least-squares technique[36]. If X is the matrix of measured observed data with known experimental uncertainty (u), the method solves the $X = GF + E$ bilinear matrix problem, where F is the matrix of the factors (sources), G is the factors' contribution and E is the matrix of residuals. F and G are determined so that the Euclidean norm of E divided by u is minimised, and as the solution is constrained the elements of F and G are required to be non-negative[37]. As the PMF is a descriptive model with no objective criterion for the optimal number of factors[38], the solution that better described the conditions at each site was chosen.

Calculation of the estimated PM concentrations

The estimation of the PM contribution for factor (i) is calculated as:

$$PM_{est_{(i)}} = F_{PM_{(i)}} \times G_{(i)} \qquad (1)$$

where the elements of $F_{(i)}$ (provided as a result by the PMF method) represent the mean concentration of each variable, in this case the PM concentration, in each factor (i) (when the elements of $G_{(i)}$ are normalised so that their mean values are equal to one[37]). $G_{(i)}$ is the relative contribution of the factor (i) in the local atmosphere at a given time which is averaged to 1 for the period of the measurement. This also takes into account the non-explained variance from all the factors combined, i.e. the variable's variation not explained by the factors of the solution. While this method does not provide an accurate value for each variable (all variables of a factor are considered to fluctuate according to a single G contribution, attributed to its corresponding factor for the given period), it can be used as a reliable estimation of their variation. As each factor (source) is associated with a specific PNSD profile for which variation is mainly expected on their intensity rather than the particle concentration proportions within the factors, the estimations are expected to present a relatively accurate contribution of each factor on the PM concentrations.

REFERENCES

1. Lelieveld, J. et al. Loss of life expectancy from air pollution compared to other risk factors: a worldwide perspective. *Cardiovasc. Res* **116**, 1910–1917 (2020).
2. Politis, M., Pilinis, C. & Lekkas, T. D. Ultrafine particles (UFP) and health effects. Dangerous. Like no other PM? Review and analysis. *Glob. Nest J.* **10**, 439–452 (2008).
3. Kim, K. H., Kabir, E. & Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **74**, 136–143 (2015).
4. Pokorná, P., Hovorka, J. & Hopke, P. K. Elemental composition and source identification of very fine aerosol particles in a European air pollution hot-spot. *Atmos. Pollut. Res.* **7**, 671–679 (2016).
5. Harrison, R. M., Beddows, D. C. S. & Dall'Osto, M. PMF analysis of wide-range particle size spectra collected on a major highway. *Environ. Sci. Technol.* **45**, 5522–5528 (2011).
6. Beddows, D. C. S., Harrison, R. M., Green, D. C. & Fuller, G. W. Receptor modelling of both particle composition and size distribution from a background site in London, UK. *Atmos. Chem. Phys.* **15**, 10107–10125 (2015).
7. Dall'Osto, M. et al. Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain. *Atmos. Chem. Phys.* **12**, 10693–10707 (2012).
8. Pope, F. D., Gatari, M., Ng'ang'a, D., Poynter, A. & Blake, R. Airborne particulate matter monitoring in Kenya using calibrated low cost sensors. *Atmos. Chem. Phys. Discuss.* https://doi.org/10.5194/acp-2018-327 (2018).
9. Crilley, L. R. et al. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmos. Meas. Tech.* **11**, 709–720 (2018).
10. Stamp, S. et al. A critical evaluation of the dynamic nature of indoor-outdoor air quality ratios. *Atmos. Environ.* **273**, 118955 (2022).
11. Popoola, O. A. M. et al. Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmos. Environ.* **194**, 58–70 (2018).
12. Weyers, R. et al. Low-cost Indoor Air Quality (IAQ) platfo rm for healthier classrooms in new zealand: engineering issues. *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE).* p. 208–215 (2018).
13. Ionascu, M. E., Gruicin, I. & Marcu, M. Laboratory evaluation and calibration of low-cost sensors for air quality measurement. *2018 IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI).* p. 395–400 (2018).
14. Crilley, L. R. et al. Effect of aerosol composition on the performance of low-cost optical particle counter correction factors. *Atmos. Meas. Tech.* **13**, 1181–1193 (2020).
15. Hagan, D. & Kroll, J. Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. *Atmos. Meas. Tech. Discuss.* https://doi.org/10.5194/amt-2020-188 (2020).

16. Sousan, S. et al. Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols. *Aerosol Sci. Technol.* **50**, 462–473 (2016).

17. Masic, A. et al. Evaluation of optical particulate matter sensors under realistic conditions of strong and mild urban pollution. *Atmos. Meas. Tech.* **13**, 6427–6443 (2020).

18. Bousiotis, D. et al. A study on the performance of low-cost sensors for source apportionment at an urban background site. *Atmos. Meas. Tech.* https://doi.org/10.5194/amt-15-4047-2022 (2022).

19. Hagan, D. H. et al. Inferring aerosol sources from low-cost air quality sensor measurements: a case study in Delhi, India. *Environ. Sci. Technol. Lett.* **6**, 467–472 (2019).

20. Owoade, O. K. et al. Spatial-temporal variation and local source identification of air pollutants in a semi-urban settlement in nigeria using low-cost sensors. *Aerosol Air Qual. Res.* **21**, 1–18 (2021).

21. Bousiotis, D. et al. Assessing the sources of particles at an urban background site using both regulatory instruments and low-cost sensors – a comparative study. *Atmos. Meas. Tech.* **14**, 4139–4155 (2021).

22. Azarmi, F., Kumar, P., Marsh, D. & Fuller, G. Assessment of the long-term impacts of PM10 and PM2.5 particles from construction works on surrounding areas. *Environ. Sci. Process. Impacts* **18**, 208–221 (2016).

23. Muleski, G. E., Cowherd, C. & Kinsey, J. S. Particulate emissions from construction activities. *J. Air Waste Manag. Assoc.* **55**, 772–783 (2005).

24. Belkacem, I., Helali, A., Khardi, S. & Slimi, K. Investigations on vehicle non-exhaust particle emissions: real-time measurements. *Int. J. Environ. Sci. Technol.* https://doi.org/10.1007/s13762-022-03955-w (2022).

25. Petäjä, T. et al. Enhanced air pollution via aerosol-boundary layer feedback in China. *Sci. Rep.* **6**, 1–6 (2016).

26. Wang, Z., Huang, X. & Ding, A. Dome effect of black carbon and its key influencing factors: A one-dimensional modelling study. *Atmos. Chem. Phys.* **18**, 2821–2834 (2018).

27. Ding, A. J. et al. Intense atmospheric pollution modifies weather: a case of mixed biomass burning with fossil fuel combustion pollution in eastern China. *Atmos. Chem. Phys.* **13**, 10545–10554 (2013).

28. Brines, M. et al. Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities. *Atmos. Chem. Phys.* **15**, 5929–5945 (2015).

29. von Bismarck-Osten, C., Birmili, W., Ketzel, M. & Weber, S. Statistical modelling of aerosol particle number size distributions in urban and rural environments - a multi-site study. *Urban Clim.* **11**, 51–66 (2015).

30. Beddows, D. C. S. et al. Cluster analysis of rural, urban, and curbside atmospheric particle size data. *Environ. Sci. Technol.* **43**, 4694–4700 (2009).

31. Salimi, F. et al. Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment. *Atmos. Chem. Phys.* **14**, 11883–11892 (2014).

32. Likas, A., Vlassis, N. & Verbeek, J. The global k-means clustering algorithm. *Pattern Recognit.* **36**, 451–461 (2003). J.

33. Pakhira, M. K., Bandyopadhyay, S. & Maulik, U. Validity index for crisp and fuzzy clusters. *Pattern Recognit.* **37**, 487–501 (2004).

34. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).

35. Hopke, P. K. Review of receptor modeling methods for source apportionment. *J. Air Waste Manag. Assoc.* **66**, 237–259 (2016).

36. Reff, A., Eberly, S. I. & Bhave, P. V. Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. *J. Air Waste Manag. Assoc.* **57**, 146–154 (2007).

37. Paatero, P. & Tapper, U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**, 111–126 (1994).

38. Paatero, P., Hopke, P. K., Song, X. H. & Ramadan, Z. Understanding and controlling rotations in factor analytic models. *Chemom. Intell. Lab. Syst.* **60**, 253–264 (2002).

39. McIntosh, D. H. & Thom, A. S. *Essentials of Meteorology* (Wykeham Publications, 1969).

AUTHOR CONTRIBUTIONS

The study was conceived and planned by FDP who also contributed to the final manuscript and DB who carried out the analyses and wrote the first draft of the manuscript. GA provided the data for the analysis. RMH and DCSB provided feedback on the analysis. All authors reviewed and edited the final paper.

COMPETING INTERESTS

The authors declare no competing non-financial interests but the following competing financial interests. D.B., D.C.S.B., R.M.H. and F.D.P. are employed by the University of Birmingham, and GA is employed by DustScan Limited, which hope to commercialise the approach discussed in the paper. Author Roy Harrison is co-Editor-in-Chief for npj Climate and Atmospheric Science and was not involved in the journal's review of, or decisions related to, this manuscript.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00424-0.

Correspondence and requests for materials should be addressed to Francis D. Pope.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.