Assessing the sources of particles at an urban background site using both regulatory instruments and low-cost sensors – A comparative study

Dimitrios Bousiotis[1], Ajit Singh[1], Molly Haugen[3], David C.S. Beddows[1], Sebastián Diez[2], Killian L. Murphy[2], Pete M. Edwards[2], Adam Boies[3], Roy M. Harrison[1] and Francis D. Pope[1]

[1] Division of Environmental Health and Risk Management
School of Geography, Earth and Environmental Sciences
University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

[2] Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, Heslington, York YO10 5DD, United Kingdom

[3] Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, United Kingdom

Abstract

Measurement and source apportionment of atmospheric pollutants is crucial for the assessment of air quality and the implementation of policies for its improvement. In most cases, such measurements use expensive regulatory grade instruments, which makes it difficult to achieve wide spatial coverage. Low-cost sensors may provide a more affordable alternative, but their capability and reliability in separating distinct sources of particles have not been tested extensively yet. The present study examines the ability of a low-cost Optical Particle Counter (OPC) to identify the sources of particles and conditions that affect particle concentrations at an urban background site in Birmingham, UK. To help evaluate the results, the same analysis is performed on data from a regulatory-grade instrument (SMPS) and compared to the outcomes from the OPC analysis. The analysis of the low-cost sensor data manages to separate periods and atmospheric conditions according to the level of pollution at the site. It also successfully identifies a number of sources for the observed particles, which were also identified using the regulatory-grade instruments. The low-cost sensor, due to the particle size range measured (0.35 to 40 μm), performed rather well in differentiating sources of particles with sizes greater than 1 μm, though its ability to distinguish their diurnal variation, as well as to separate sources of smaller particles, at the site was limited. The current level of source identification demonstrated makes the technique useful for background site studies, where larger particles with smaller temporal variations are of significant importance. This study highlights the current capability of low-cost sensors in source identification and differentiation using clustering approaches. Future directions towards particulate matter source apportionment using low cost OPCs are highlighted.

1. Introduction

Particulate matter (PM) plays a dominant role in air quality and is known to cause adverse health effects (Dockery et al., 1993; Pascal et al., 2013; Wu et al., 2016; Zeger et al., 2008). As a result, regulatory limits are set for its concentrations, especially in urban areas (US EPA, 2012; WHO, 2006). For the implementation of such regulations, the identification of the sources of PM is required. To accomplish this, measurements of the concentrations of PM, typically alongside PM composition, in the area of study are conducted. Until recent years these measurements were typically made using regulatory-grade instruments which, while providing high quality data, are rather expensive thereby limiting the number that could be deployed and consequently the spatial resolution of any measurement network. This increases the spatial interpolation uncertainty (Kanaroglou et al., 2005) and can result in inadequate connection between the levels of air pollution exposures and health effects (Holstius et al., 2014), especially in complex urban environments (Harrison, 2017; Mueller et al., 2016). Additionally, many low and middle income countries are unable to invest the large economic assets currently required for source apportionment, even though in many of these countries, the air quality is poor (Ghosh and Parida, 2015; Kan et al., 2009; Petkova et al., 2013; Pope et al., 2018; Singh et al., 2020).

In the past decade, the development of new and cheaper sensors for air quality monitoring has intensified. Many different sensors were introduced measuring either the number concentration or surface area of PM, or the gas phase species (Jovašević-Stojanović et al., 2015; Lewis et al., 2018; Popoola et al., 2018). Overall, the low-cost PM sensors currently offer better comparison with regulatory grade equipment compared to their gas phase counterparts (Lewis et al., 2018). However, many shortcomings have been identified in their application, with the most common being the loss of accuracy in the measurements due to environmental conditions such as relative humidity (RH) variations or high PM concentrations (Castell et al., 2017; Crilley et al., 2018; 2020; Di Antonio et al., 2018; Hagan and Kroll, 2020, Miskell et al., 2017; Zheng et al., 2018). Measurements in ambient conditions also lead to discrepancies with research-grade instruments, which often measure in controlled environments that are air conditioned (U.S. Environmental Protection Agency, 2016). The reproducibility and variability of the outputs from sensors of the same type can also be

problematic (Austin et al., 2015; Sousan et al., 2016; Wang et al., 2015). Therefore, the need for constant and careful calibration is repeatedly highlighted in many studies that evaluate the potential of low cost sensors (Rai et al., 2017; Spinelle et al., 2015, 2017). When these calibration steps are implemented, low-cost sensors have been shown to provide reliable near-real time measurements, maintaining high correlations with research-grade instruments (Kelly et al., 2017; Malings et al., 2020; Sayahi et al., 2019) with the added advantages of the lower cost and portability.

Consequently, low-cost sensors have been successfully deployed in many studies for which the use of more expensive instruments was not feasible. There is a number of applications in low and middle income countries (e.g. Nagendra et al., 2019; Pope et al., 2018), in studies which included mobile measurements within the urban environment (Ionascu et al., 2018; Jerrett et al., 2017; Miskell et al., 2018), or studies of indoor air quality from multiple sites, such as the SKOMOBO project conducted in New Zealand, in which the air quality in schools was assessed (Weyers et al., 2018). The greatest advantage though is likely, as their name implies, their lower cost which made possible the formation of a network of measuring stations (Feinberg et al., 2019; Kotsev et al., 2016; Moltchanov et al., 2015), increasing the spatial resolution and through new data analysis methods improve the mapping of air pollution up to a sub-neighbourhood level (Schneider et al., 2017, Shindler, 2019). Therefore, it is suggested that the development and use of low-cost sensors, either used individually or in conjunction with research-grade instruments (Snyder et al., 2013), have the potential to radically change the conventional approach of both pollution measuring and policy making (Borrego et al., 2018; Kumar et al., 2015; Lagerspetz et al., 2019, Morawska et al., 2018), providing a more effective general public information and enhanced environmental awareness (Penza et al., 2014), even for countries with smaller budgets (Amegah, 2018).

As yet, studying the different sources of particles at a site with the use of data from low-cost sensors has not been widely attempted yet. Pope et al., (2018) managed to identify major pollution sources studying the ratios of PM of different sizes provided by low-cost sensors, while Popoola et al., (2018) using a network of sensors identified the sources of pollution near Heathrow airport in London, UK. Hagan et al., (2019) applying a statistical method (Non-negative Matrix Factorisation) on low-cost sensor data, identified a combustion factor in a three-factor solution in New Delhi, India. The present study investigates the ability of low-cost sensors to provide measurements that can be used to identify the sources of pollution

at a background site in Birmingham, UK, using clustering of particle composition profiles. This method was successfully used in a number of previous studies, though with the use of measurements from research-grade instruments (Beddows et al., 2009, 2015; Von Bismarck-Osten and Weber, 2014; Dall'Osto et al., 2011; 2012; Sabaliauskas et al., 2013). To support the clustering method, chemical composition data from both research-grade and low-cost sensor instruments were used, as well as meteorological data from a closely located measurement station. Apart from attempting the source differentiation with low-cost sensor data, a direct comparison with the results from a similar analysis using research-grade instruments is also conducted to not only validate the results but find the strengths and weaknesses of such an application.

2. Methods

2.1 Location of the site and instruments

The measurement site (fig. 1), characterised as an urban background, is the Birmingham Air Quality Supersite (BAQS) located at the grounds of the University of Birmingham (52.45°N; 1.93°W), about 3 km southwest from the city centre (Alam et al., 2015). In the present study, measurements from the following instruments for the period 24/01/2020 to 12/3/2020 (the date range was chosen to avoid the effect of the lockdown due to COVID-19) were used (Table 1, a picture of the low-cost sensors used at BAQS is found in figure S1):

The Alphasense OPC-N3, which is an optical particle counter, measuring particle number concentrations in the size range between 0.35 to 40 µm at rates up to about 10000 particles per second. As the sample air stream enters the instrument with a sample flow rate of 210 mL m^{-1} (dynamically monitored and corrected by the sensor) it passes through a laser beam (wavelength at 658 nm). The OPC-N3 measures the light scattered by individual particles carried in a sample air stream through a laser beam. These measurements are used to determine the particle size, related to the intensity of light scattered via a calibration based on Mie scattering theory, and particle number concentration. Particle mass loadings (PM_1, $PM_{2.5}$ and PM_{10}) are then calculated from the particle size spectra and concentration data, assuming a particle density and refractive index (default density is 1.65 g/ml and refractive index is 1.5+i0.). Particles of larger size are lost to impaction in the tubing prior to the OPC

and thus are not considered. The OPC is located within the air conditioned station, so measurements represent PM dry mass.

The AethLabs MA200 (microAeth MA200) which provides black carbon (BC) information (0-1 mg BC/m^3). The sample is deposited onto an internal filter, and an IR light (880 nm) is directed through the sample on the filter and into a detector on the other side of the sample. The amount of light absorbed from the sample is proportional to the BC concentration.

Two Naneos Partectors (Naneos Particle Solutions GmbH) which provide the lung deposited surface area metric (LDSA, µm^2/cm^3) in the particle diameter range 10 nm to 10 µm. In general, the provided data is dictated by the particle number concentration and diameter ($Nd^{1.1}$) for both semi-volatile and solid particles. A catalytic stripper (Catalytic Instruments CS015) was used to remove the semi-volatile particles entering one of the two Naneos Partectors. The other Naneos Partector was not subject to the catalytic stripper and therefore measured the surface of all particles. In the present study, apart from the values provided directly from the sensors, the ratio between the measurements of the two Naneos Partectors was also considered according to:

$$LDSA_{ratio} = \frac{LDSA\ after\ the\ catalytic\ stripper}{LDSA\ before\ the\ catalytic\ stripper}$$

This was done to resolve whether such a configuration can also provide information such as the level of pollution or the age of the incoming air masses, as increased concentrations of semi-volatile compounds are usually associated with anthropogenic sources, especially in the urban environment (Harkov, 1989; Mahbub et al., 2011, Schnelle-Kreis et al., 2007, Xu and Zhang, 2011). Thus, a high $LDSA_{ratio}$ is expected to be associated with fresher pollution which usually has a higher content of volatile compounds (i.e., pollution sources at a close distance from the site), while lower ratios are probably associated with either cleaner conditions or more regional and aged pollution with higher concentrations of semi-volatile compounds, usually associated with sources at a greater distance from the measuring site. The specific metric though should be considered with caution, as it can be biased by the absolute surface areas measured.

The sensors monitoring nitrogen dioxide (NO_2) and ozone (O_3) concentrations are part of an Alphasense Box Of Clustered Sensors (BOCS) (Smith et al., 2019), which is a low-power

instrument based on the clustering of multiple low-cost air pollution sensors allocated in two independent circuits to redundantly measure concentrations and other airflow parameters. The air is driven by a pump through the cell (air flow is about 4 L min^{-1}) that hosts the electrochemical sensors (EC) and the nondispersive infrared sensors (NDIR). While the EC sensors redundantly (6 sensors per gas) measure carbon monoxide, NO_2, nitrogen monoxide, oxidizing gases (O_x), the NDIR sensors measure carbon dioxide. EC sensors are based on recording the current generated by redox reactions that occur at the electrode-electrolyte interface in an electrochemical cell composed of three electrodes (working electrode (WE), counter electrode (CE) and reference electrode (RE)). While the gas of interest reacts on the WE surface, the CE completes the redox reaction and the RE ensures that the WE potential remains in the proper range. In the present study, measurements of O_3 (deriving from a linear regression of the values of the six O_x sensors with the measurements from the reference instrument also located at BAQS) and NO_2 were only used from the specific sensor.

The Aethalometer model AE33 by Magee Scientific, collects aerosol particles continuously by drawing the aerosol-laden air stream through a spot on the filter tape. It analyses the aerosol by measuring the transmission of light through one portion of the filter tape containing the sample, versus the transmission through an unloaded portion of the filter tape acting as a reference area. This analysis is done at seven optical wavelengths spanning the range from the near-infrared to the near-ultraviolet. The Aethalometer calculates the instantaneous concentration of optically absorbing aerosols from the rate of change of the attenuation of light transmitted through the particle-laden filter.

For the same period data from regulatory-grade instruments were also available. Thus, particle size composition data from a model TSi3082 Scanning Mobility Particle Sizer (SMPS) in the size range 12 – 552 nm, along with PM data for the sizes of 1, 2.5, 4 and 10 μm acquired using a Fidas 200E were used. Additionally, chemical composition data for NO_2, O_3, as well as SO_4^{2-}, NO_3^- and organic content (size range 40 nm to 1 μm) from an Aerodyne Aerosol Chemical Speciation Monitor (ACSM) were also available. Meteorological data (wind speed and direction, temperature, RH and rain level) from the Birmingham Air Quality Supersite were also used in the characterisation of the clusters formed from both methods.

Planetary Boundary Layer (PBL) height data were downloaded from ECMWF's ERA5 (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels - last access 20/3/2021). Back trajectory data calculated using the HYSPLIT model (Draxler and Hess, 1998), were extracted by the NOAA Air Resources Laboratory (https://ready.arl.noaa.gov/READYtransp.php - last access 17/8/2020). Data was processed using the Openair package for R (Carslaw and Ropkins, 2012).

2.2 k-means clustering

In this study, two size spectra are considered, one deriving from the OPC and one from the regulatory-grade SMPS. It is noted that the size spectra from the two instruments only briefly overlap in the size range 350 – 552 nm, with the SMPS mostly measuring smaller particles and the OPC mostly measuring larger particles. For the period studied (24/1/2020 – 12/3/2020), 874 hours of available data (averaged from 10 second intervals - 76% coverage) from the OPC and 732 hours from the SMPS (66% coverage) were exposed to k-means clustering. k-means clustering is a method successfully used in many studies of particle source differentiation (Beddows et al., 2015; Brines et al., 2015, Von Bismarck-Osten and Weber, 2014; Giorio et al., 2015; Wegner et al., 2012) and was proven to have better performance compared to other clustering techniques (Beddows et al., 2009; Salimi et al., 2014), as it was found to produce clusters with the highest similarity between their elements and the highest separation against the other clusters formed (Hennig, 2007). It is a method of vector quantisation which aims to partition observations (x_1, x_2, …, x_n) into k sets, minimising within-clusters variances (squared Euclidean distances) as:

$$arg\ min \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 = arg\ min \sum_{i=1}^{k} |S_i|\ Var\ S_i$$

where S_i are the sets (clusters) formed and μ_i are the centroid point of the cluster (Likas et al., 2003). K-means clustering in this study was performed using the "stats" library for R. The optimal number of clusters was chosen using two metrics, the Dunn Index and the Silhouette width as proposed by Beddows et al., 2009. The Dunn Index provides a measure of the ratio of the minimum cluster separation to the maximum cluster (providing a metric of the compactness and separation of the clusters formed within the space – Pakhira et al., (2004)). The larger the Dunn Index the better separated are the clusters formed. The Silhouette width

is a measure of the similarity of the spectra within each cluster (Rousseeuw, 1987). Both the Dunn index and Silhouette width were calculated using the "fpc" library for R. In the present study the best statistically fitted solution was chosen (the solution for which both metrics maximised), though in source differentiation studies such a solution may not always provide with the best separation of all the available sources. Using the aforementioned statistical tests, a six-cluster solution was independently suggested for both the OPC and SMPS datasets. Though the clustering process could be applied for the FIDAS data, which are comparable in size range, it was not performed in this study.

3. Results

3.1 General conditions, sources of particles and pollution at the site

Being an urban background, the site studied presents relatively low concentrations of most pollutants (the average atmospheric conditions for each cluster formed by both methods is presented in table 2), without the effect of direct sources of pollution, such as traffic. Wind rose and polar plots illustrating the conditions in the period studied are found in figure S2. The main source of pollution lies on the north and northeast sectors, where the city centre is located, as well as in the southern and eastern sectors where a populous residential area is located. As a result, the main sources of NO_2 and BC as well as the smaller sized PM are associated with easterly winds (this though is not reflected in particles observed in the SMPS size range). For the PM_{10} apart from the aforementioned, increased concentrations are also found with southwestern winds likely associated with marine sources. Typical for the UK, the average wind profile for the period consists mainly of western and southwestern winds (McIntosh and Thom, 1969), reducing the effect of the pollution sources in the east of the site. Finally, the secondary pollutants NO_3^- and SO_4^{2-} which are in most cases associated aged pollution and long-distance transport, have less consistent profiles, though they both seem to be mainly associated with southern wind directions. Finally, for the period studied no New Particle Formation events were observed. This is consistent with the general trend in the area as found by Alam et al., (2003) for Birmingham (as well as in a more recent studies by Bousiotis et al., (2019; 2021) at nearby sites in Oxford and London), in which NPF events in Southern

UK are more frequent during the summer months and barely occurring during winter and early spring, mainly due to unfavourable meteorological conditions.

3.2 Clustering of the OPC data

Due to the larger particle sizes measured by the OPC-N3, the differences in the cluster profiles are mainly associated with the particle number concentrations and to a lesser extent on the different peaks, which are less distinct due to the smaller variation found as particle diameter increases. The frequency of the clusters formed, and their diurnal occurrence is shown in figure 2. The average particle size distribution spectra and wind roses for the clusters formed are found in figures S3 and S4.

The six clusters formed from the OPC data are:

OPC.1: A rather polluted group with the highest NO_2 concentrations and average secondary pollutants, PM and LDSA ratio. Its fresher polluted character is further confirmed using the SMPS data which showed higher than average particle concentrations for particles with diameter smaller than 50 nm. This group presents low average temperature, RH, PBL height and slower than average southwestern winds, which is explained, to an extent, by the cluster being slightly more frequent during night-time.

OPC.2: The second group refers mainly to a single midday event on 12/3/2020 (which explains the highest PBL height found) with high-speed southwestern winds, which are associated with lower pollution levels in the area (McGregor and Bamzelis, 1995), high temperature and very low RH. On this day the concentrations of all the pollutants were rather low, though due to the high wind speeds (an increase in the wind speed is observed at the start of the occurrence of this cluster – at 10:00 AM - which affects the particle distribution profile as can be seen in Figure S5) the PM_{10} were close to average (when PM_1 and $PM_{2.5}$ were rather low) indicating the stronger presence of coarser particles, possible of marine origin as shown by the back trajectories, a source with an increasing importance at larger size PM at this area (Harrison et al., 2004; Taiwo et al., 2014). This group presents the highest LDSA ratio, which is in agreement with the low concentrations of the secondary pollutants.

OPC.3: A group occurring mainly during some of the midday periods in January, with the lowest temperature and wind speed averages, as well as the highest average RH, containing both southwestern and southern winds. While the concentrations of the measured pollutants

are close to average, high sulphate and ozone concentrations were found, with the former pointing to air masses with higher concentrations of aged pollutants assisted by the lowest PBL found for this cluster. The LDSA ratio though, was found to be very high despite the higher concentrations of sulphate and nitrate. The near average NO_2 concentrations may point to the effect of a nearby pollution source that may resulted to the increased LDSA ratio found.

OPC.4: A group with low concentrations of NO_2, BC and PM, but close to average secondary pollutants' concentrations. It is associated with close to average temperature, RH, PBL height and wind speed of mainly southwestern directions. It is slightly more frequent during daytime and has lower than average concentrations of particles in the SMPS range.

OPC.5: This group includes the most polluted conditions in the area throughout the day. It is associated with western and southwestern winds of average speed, high temperature and lower than average RH. Most pollutant concentrations, including PM, are rather high while O_3 is low. Similarly, it presents the highest concentrations of particles in all SMPS size ranges which is probably due to the reduced atmospheric mixing in the lowest average PBL height among the OPC clusters. This cluster also includes the more polluted conditions found with north-eastern winds.

OPC.6: A group associated with rather clean conditions, presenting the lowest concentrations of NO_2, BC, NO_3^- and organic content. It is associated with higher than average temperature, PBL height and wind speed and lower than average RH, and has low concentrations of PM_1 and $PM_{2.5}$, while PM_{10} concentration is close to average. Its association with cleaner conditions (lower concentrations of the pollutants with available data) probably explains the highest O_3 concentrations. The fast-moving southwestern air masses, which this group is associated with, are probably of marine origin that have not passed through any significant pollution sources, which can be further suggested by both the low LDSA values and the highest LDSA ratio.

3.3 Clustering of the SMPS data

In the past, a number of studies on the sources of particles were conducted for both the greater area of Birmingham and specifically the site in the University (Harrison et al., 1997; Taiwo, 2016; Yin et al., 2010). As, these studies mainly focused on the chemical composition of coarser particles, to the authors' knowledge this is the first study that uses ultrafine particle

size distribution data to study the sources of particles in Birmingham, UK. The frequency and hourly occurrence of the six clusters formed from the SMPS data is found in figure 3. The average particle size distributions and wind roses for the clusters formed are found in figures S6 and S7.

SMPS.1: This group contains averagely polluted hours and is associated with fresher pollutants (such as NO_2 or NO) and PM, while secondary pollutants such as NH_4^+, NO_3^- and SO_4^{2-} are relatively low. Due to being associated with fresher emissions this group presents higher than average concentrations of particles below 50 nm and a low LDSA ratio. It is associated with average speed southwestern winds (it also includes the small portion of north-eastern winds) and temperature, higher than average RH and occurs more frequently during late night and early morning hours, which explains the low PBL height among the SMPS clusters.

SMPS.2: Similar to the first group, average pollutants' concentrations are found in this group with low concentrations of secondary pollutants. It is associated with slow western and southwestern winds, lower than average temperatures, RH and PBL height and is more frequent during early morning hours. It has the highest concentrations of particles with diameter smaller than 20 nm, but the particle concentrations become relatively smaller as their size increase.

SMPS.3: This is a small group containing very clean night hours mainly in February, with higher than average temperature, lower than average RH, strong western and southwestern winds and a remarkably great PBL height for the time of the day. It has low concentrations of pollutants and PM apart from O_3 (despite the time of day), though PM_{10} concentration is enhanced, probably associating this group with stronger marine origins. The particle concentrations of all size ranges below 500 nm are the lowest among the groups formed and along with the high LDSA ratio are in agreement with the very clean conditions associated with this cluster. This cluster, contrary to all other, presents two peaks: one peaking just below 30 nm and another one just over 100 nm, which indicates that it is probably associated with at least two different sources.

SMPS.4: This group presents near average concentrations of all the pollutants studied. PM_1 average concentration is rather low while PM_{10} is higher than the average. It is associated with average speed southwestern winds, higher average temperature and PBL height and low

RH. It is more frequent during midday and evening hours and it appears to represent the most common conditions in the area, hence having the highest frequency of all clusters.

SMPS.5: This is a unique group associated with southern winds, the side at which the central part of the University resides. This is the most polluted group, probably affected by emissions from the University and the residential area found in that direction assisted by the very low PBL height, with very high concentrations of all the pollutants (apart from O_3), PM and ultrafine particles with available data. The LDSA ratio is very high and this is probably due to the great surface area of the involatile component found. It is associated with very slow wind speeds, low temperature, very high RH and occurred evenly throughout the day, mainly on the first weeks of the campaign when pollution levels were rather high, probably due to increased heating emissions.

SMPS.6: This group presents low concentrations of all pollutants (apart from O_3), PM and ultrafine particles with available data and is associated with western winds with higher than average speed, near average temperatures and PBL height and low RH. It occurred more frequently during evening hours and almost equally frequently throughout the whole study period apart from the first 2 weeks when pollution levels were rather high.

3.4 Direct comparison between the methods

Due to the difference in the size ranges measured by the SMPS and OPC instruments, it is evident that a direct comparison between the two methods would provide mixed results as some clusters found using the SMPS data are not detectable with the OPC, and vice versa. The particle size range that is common between the two instruments lies at about 350 – 550 nm. Therefore, many particle sources associated with particles in the size range below the minimum detectable size of the OPC are not expected to be found using its data and vice versa. At a background site though, many of the sources of smaller sized particles play a less important role as they are usually associated with fresher emissions, which are not common to such sites.

The clustering process attempts to separate the particle size distributions into groups with as similar spectral profiles as possible, while being as different to the other groups as possible. As expected, the SMPS is more capable in separating different cluster profiles at the size range

smaller than 500 nm, a size range in which the cluster profiles (using the SMPS data) formed by the groups from the OPC are almost uniform (fig. 4). This shows the limitation of the OPC data to distinguish ultrafine particle variations and thus it does not provide insight for the sources of particles within this size range. On the other hand, the OPC performs much better in identifying different sources when considering larger particles in the range between 1 – 10 μm, for which it manages to clearly distinguish variations between the groups formed (fig. 5). The clusters formed using the OPC data appear to be better associated with different sources of PM_1 (fig. 6), compared to those deriving from the SMPS data (fig. S8), as distinct "hot spots" of PM_1 are more clearly defined on the polar plots from the OPC compared to the less clear and mainly associated with calm (or almost calm) conditions from the SMPS (providing no separation among possible sources of PM_1).

Table 3 contains the cluster relationships between the two methods, while Table S1 contains the conditions observed when pairs of clusters from both methods are considered. The OPC.2 and OPC.3 clusters appear infrequently, and it would be nonsensical to directly associate them with SMPS groups, as they appear under very specific conditions, that either are not detected or are not identified as separate cases by the SMPS. As a result, they will be separately studied later in this study.

The OPC.1 was mainly associated with SMPS.4 and SMPS.6 and to a lesser extend with SMPS.1. OPC.1 has a somehow higher frequency during night times and it shares many of these hours with groups SMPS.4 and SMPS.6, while with SMPS.1 it mainly shares early morning hours. It includes the more polluted portion of the rather clean SMPS.6 and a portion with lower PM_{10} (though not much difference from average pollutants' concentrations) from the more polluted SMPS.4. It is interesting that the variation between the subgroups (in relation to SMPS clusters) of the OPC.1 is very small for the NO_2 concentrations, a pollutant for which its variations are not expected to be directly "visible" at the size range of the OPC as it is mainly associated to fresher emissions. No great variation was found for the wind direction in the subgroups of OPC.1, though it includes the lower temperature and higher RH conditions of the SMPS clusters it is associated with. The OPC.1 includes the relatively clean part of the more polluted SMPS.1 and the more polluted portion of the cleaner SMPS.6. While this does not provide a clear connection between the OPC and SMPS results, it shows that there is consistency in the results provided by the former in identifying particle sources of specific qualities.

Similarly, OPC.4 was mainly associated with SMPS.4 and SMPS.6. As the OPC.4 occurs under cleaner conditions, it includes the less polluted hours of both the SMPS clusters it is mainly associated with, though the concentrations of the secondary pollutants such as NO_3^- and SO_4^{2-} are closer to the average. The OPC.4 is associated with the cleaner portion of the aforementioned SMPS clusters with higher average temperature and RH though with variable wind speeds.

OPC.5 represents a polluted group of hours associated mainly with SMPS.4, SMPS.5 and SMPS.6. Being a group of hours associated with higher concentrations of pollutants, it includes the more polluted portions of SMPS.4 and SMPS.6 with average meteorological conditions, though lower wind speeds. It also coincides with the largest portion of SMPS.5, mainly in the sixth week when the temperature was the lowest, including the portion with the higher concentrations organic content and NO_3^-. SMPS.5 is the group that is associated with southern wind directions, a side from which a source of secondary pollutants (NO_3^-, SO_4^{2-}, NH_4^+), organic content and particles of diameter greater than 100 nm occurs. The OPC.5 is associated with the part of SMPS.5 which is more burdened from secondary pollutants, hence very large concentrations are observed for them.

Finally, OPC.6 is mainly associated with SMPS.2, SMPS.4 and SMPS.6. Being a cleaner group of hours, it includes the portion of these SMPS clusters with lower pollutant concentrations but higher PM_{10} concentrations (though with lower PM_1 concentrations). These rather clean conditions, along with the western and southwestern high-speed winds in average and the large PM_{10} concentrations, further enhance the possible marine character of this cluster. Due to the size range of these particles such variation is not clearly identified by the SMPS, resulting to them not being clearly separated when its data is considered.

The weekly contribution of each cluster group from the analysis of either dataset is found in Figure 7 and the conditions on each week studied in Table S2. It is evident that the variation from the SMPS is greater than that of the OPC, as the latter is less affected by the diurnal variations. It is apparent that it is easier to comprehend the clusters' variation in association with the levels of pollution in the site (the more polluted weeks have a greater portion of SMPS.1 and SMPS.5), while for those with lower concentrations of pollutants the SMPS.4 and SMPS.6 are more enhanced. These variations are harder to distinguish using the OPC data, as they are less apparent in the size range measured by the sensor. To further understand the possible sources using the latter, information from other instrument which provide chemical

composition data are needed, though it is still hard to pinpoint exact sources, due to the OPC's weakness in explaining distinct particle sources within the day.

3.5 Case studies

OPC.2

OPC.2 occurs mainly on a single day in March (12th) with higher than average temperature and strong western winds. This was the cluster with the lowest concentrations of NH_4^+, NO_3^- (about an order of magnitude compared to average conditions) and SO_4^{2-}, rather low concentrations of NO_2, BC and high O_3, which is probably the result of the strong winds and the very high PBL height assisting in the removal of the pollutants from the site. Using the SMPS data, this group of hours seems to follow the trends of BC, associating it with SMPS.6 for low, SMPS.1 and SMPS.2 for medium and SMPS.4 for higher concentrations of BC. This cluster has very low PM_1 and $PM_{2.5}$ and near average PM_{10} concentrations, probably associating it with marine sources (due to the high wind speed). Due to this, it is not clearly separated using the SMPS data, which does so for the hours of this group according to the level of fresher pollutants, the variation of which is smaller in this type of environments. This cluster seems to be the result of the change in the wind profile which greatly affected the coarser particles at the site (figure S5).

OPC.3

The third cluster formed using the OPC data, was a rather small group of hours in late January (25,27 and 28th), with the lowest average temperature, wind speed and PBL height compared to the rest of the clusters. The wind direction profile for this group contains both western and southern winds, the latter being associated with high concentrations of pollutants (as found by the study of the SMPS data). The majority of the hours in this group (65%) were characterised as freshly polluted when using the SMPS data, mainly associated with SMPS.2. Unfortunately, data of NO_2, BC, O_3 and PM for this group were very scarce from regulatory-grade instruments (due to instrument error – the results provided in table 2 for the OPC.3 are only from 2 hours of data that were available from the regulatory grade instrument). The ACSM data, which were available for the hours of this cluster pointed to marginally lower

than average values of organic content, nitrate and ammonium, while the sulphate concentrations were rather high. Using the low-cost sensor data, it is found that this group has the highest BC (data from this low-cost sensor is not included), and involatile component of LDSA. This group also had the highest average particle concentration in the size range of the OPC, which is in agreement with the highest PM concentrations in all ranges (PM_1, $PM_{2.5}$, PM_{10}), and is probably the result of the low wind speed and PBL height. As this is not visible from the SMPS, the cluster associated with this group has nearly average particle concentrations in the SMPS particle ranges. This group was not distinctively detected by the SMPS due to presenting variation in larger sized particles, which can be one of the weaknesses of studying the sources of such particles using SMPS data alone. The OPC.3 appears to contain the more polluted slow-moving portion of SMPS.2 with enhanced SO_4^{2-}, BC and PM concentrations.

SMPS.3

The third cluster from the analysis of SMPS data presented a unique profile with two peaks, one below 30 nm and one a bit over 100 nm. This unique group was associated with very clean conditions, with very low concentrations for all the pollutants with available data (apart from O_3), as well as low particle concentrations for all the ranges in the SMPS and OPC range as well as PM_1 and $PM_{2.5}$. The concentrations of PM_{10} and SO_4^{2-} were somehow higher but still lower than the average in the area for the period of the study. This group is associated with high average temperature and wind speed and rather low RH, with wind directions being mainly southwestern and western. This group occurred solely at night hours during a number of relatively warm nights mainly in February and to a lesser extend in March. Even with very low particle concentrations (as found by both the SMPS and OPC) the presence of two separate peaks in the size range of the ultrafine particles is indicative of more than one simultaneous source. Due to these sources of particles occurring at the ultrafine particle range, the OPC was not able to distinguish this special condition and grouped the hours of this cluster to a number of clusters (mainly OPC.5 and to a lesser extend OPC.1 and OPC.6), occurring either during night-time or throughout the day. The inability of the OPC to distinguish complicated conditions in the ultrafine range is a weakness of the OPC that should be considered when such conditions are anticipated.

4. Discussion

As the SMPS measures smaller particle sizes and with better accuracy, compared to the OPC, it managed to better separate the different sources of fresher pollution with the main differentiating factor being the time of the day, for which the variability of such sources is more prominent. The differences in NO_2 concentrations, which are mainly associated with fresher emissions are more distinct between the groups and using this data better separation of very clean (SMPS.3) and very polluted conditions from a distinct source (SMPS.5) was achieved, while the other groups described mostly average conditions with lesser variability (as expected in this range at a background site). Additionally, using the SMPS data it is possible to distinguish multiple sources of ultrafine particles (SMPS.3), as they can appear as multiple peaks within the SMPS spectra. This is not possible using the OPC data as the size range measured by this instrument cannot identify such cases.

Contrary to the SMPS, using the OPC data provided less distinct separation of fresher emissions (as expected due to the lack of data of small sized particles). Additionally, the OPC data is less sensitive to diurnal variations due to the range of particles covered, which are in a size range that does not vary significantly through the day but between days. This results in the less distinct diurnal variations found between the groups formed. The analysis of the OPC data though managed to adequately separate conditions and/or sources associated with larger particles, such as aged pollution (for which it also managed to separate a small time-window with very strong sulphate presence – OPC.3) which has the greatest contribution in the particle chemical composition for the study area (Harrison et al., 2003; Taiwo, 2016; Yin et al., 2010), RH variations or air masses of marine origin. To an extent, this might be due to the number of clusters chosen as there is a possibility that a larger number of clusters from the SMPS may separate sources of larger particles better, though with the risk of also separating similar sources. Additionally, the pollution levels of the clusters formed directly follow the trends of the PBL height in the area, a variation captured by both instruments, showing the importance of this variable in the air quality of an area.

To sum up, the study of SMPS data with k-means clustering is far superior at separating complex pollution sources within urban environments in which the variation of very small

particles is crucial for identifying particle and emission sources. This advantage of the SMPS will not be overcome even with a denser measuring network of OPCs that could be acquired for the same cost of the SMPS. However, clustering of the OPC data can provide useful information to assess the sources of air pollution at background sites in which the direct (smaller) particle sources are few. The method managed to find sources of greater pollution associated with higher concentrations of particles of greater sizes (which are mainly associated with aged pollution though), showing that the footprint of pollution in the ultrafine particle range can have a detectable effect in coarser particle distributions as well. While not as precise as the SMPS, a denser network of such instruments in background sites can be more beneficial and more cost efficient in studying multiple pollution sources or "hot spots" within the urban environment.

The current inability of low-cost PM sensors in measuring particle size spectra at small sizes (<300 nm) is the greatest drawback in their application for separating particle sources, since much information is contained in these smaller sizes. OPCs using shorter wavelength light sources and hence smaller particle detection could be beneficial here. Also, there are several low-cost sensors that provide insight for the surface area or the total number of particles in the ultrafine particle size range (such as the LDSA sensor used in this study). The combined use of the OPC with these instruments, along with sophisticated statistical techniques, may provide possibilities for more precise source differentiation than shown in the present study. It is noted that while clustering of particle number size distributions is one approach in the study of the source assessment of particles, other alternative methods, such as the Positive Matrix Factorisation (PMF), may also provide useful results.

5. Conclusions

The present study investigates the capabilities of a low-cost OPC sensor for source differentiation at an urban background site in Birmingham, UK. It is used alongside a regulatory-grade SMPS instrument, which has previously been used successfully for source differentiation. The clustering approach identified optimal solutions of six clusters for both the SMPS and OPC data. There were similarities between the SMPS and OPC solutions, which provide insights into periods of low and high pollution. However, large differences were also

observed. A more distinct separation of direct emission sources was achieved using the SMPS data, which identified sources with time windows that correlated with extreme NO_2 concentrations (either high or low), as well as periods with more complex sources. The OPC was able to distinguish time periods with greater variation of super micron sized particle sources (e.g. marine sources). There seems to be a clearer distinction of the diurnal variability of sources using the SMPS data, while the OPC seems to be able to only distinguish the variability within periods of days rather than hours, as found by the less variable diurnal and weekly variation. This though might not be a great drawback when considering background sites, as this variability is smaller in such environments which are mainly affected by regional pollution, while the local emissions are less and more distinct. Low-cost sensors can be a reliable alternative for source identification studies in environments with less complex sources, which present smaller alterations within the span of the day. Still, such instruments cannot be used for scientific analyses which require greater precision. Their application will probably be adequate when studying the sources of particles with a more regional character (e.g. marine sources) rather than direct and variable sources (e.g. traffic or cooking emissions) and can provide enough information for the air quality levels, sources and conditions these are anticipated from. Such studies may include the analysis of mineral dust events resulting from either anthropogenic activities or meteorological events (e.g. dust storms), bioaerosol events in forested areas and other sources which affect mainly the composition of coarser particles.

This study demonstrates that single low-cost sensor PM units can provide sensible source differentiation of large sized PM pollution sources. This allows for the prospect of source apportionment via networks of low-cost sensors in the near future, thereby allowing triangulation of sources. The development of more sophisticated low-cost sensors in conjunction with their low cost ensures the prospect of the application of a denser measurement network, making better air quality monitoring and control feasible in the near future. This though, requires more similar studies which can further elucidate the strengths and weaknesses of those sensors compared to the regulatory-grade ones, as they develop.

Author Contributions

The study was conceived and planned by FDP who also contributed to the final manuscript, and DB who also carried out the analysis and prepared the first draft. AS, MH, DCSB and SD have provided with the data for the analysis. DCSB provided help with the analysis of the data. RMH provided advice on the analysis. PME and AB contributed to the final manuscript.

Competing Interests

The authors have no conflict of interests.

Acknowledgements

The work is funded by NERC (NE/T001879/1) and EPSRC (EP/T030100/1). We thank the OSCA team at the Birmingham Air Quality Supersite (BAQS), funded by NERC (NE/T001909/1), for help in data collection for the regulatory grade instruments. We thank Lee Chapman for access to his meteorological data set used in the analysis.

References

Alam, A., Shi, J. P. and Harrison, R. M.: Observations of new particle formation in urban air, J. Geophys. Res. Atmos., 108(D3), 4093, doi:10.1029/2001JD001417, 2003.

Alam, M. S., Keyte, I. J., Yin, J., Stark, C., Jones, A. M. and Harrison, R. M.: Diurnal variability of polycyclic aromatic compound (PAC) concentrations: Relationship with meteorological conditions and inferred sources, Atmos. Environ., 122, 427–438, doi:10.1016/j.atmosenv.2015.09.050, 2015.

Alphasense User Manual for OPC-N3 Optical Particle Counter, Alphasense Ltd., 072-0502, pp. 32, Essex, UK, 2019.

Amegah, A. K.: Proliferation of low-cost sensors. What prospects for air pollution epidemiologic research in Sub-Saharan Africa?, Environ. Pollut., 241, 1132–1137, doi:10.1016/j.envpol.2018.06.044, 2018.

Austin, E., Novosselov, I., Seto, E. and Yost, M. G.: Laboratory evaluation of the Shinyei PPD42NS low-cost particulate matter sensor, PLoS One, 10(9), 1–17, doi:10.1371/journal.pone.0137789, 2015.

Beddows, D. C. S., Dall'Osto, M. and Harrison, R. M.: Cluster analysis of rural, urban, and curbside atmospheric particle size data, Environ. Sci. Technol., 43(13), 4694–4700, doi:10.1021/es803121t, 2009.

Beddows, D. C. S., Harrison, R. M., Green, D. C. and Fuller, G. W.: Receptor modelling of both particle composition and size distribution from a background site in London, UK, Atmos. Chem. Phys., 15(17), 10107–10125, doi:10.5194/acp-15-10107-2015, 2015.

Von Bismarck-Osten, C. and Weber, S.: A uniform classification of aerosol signature size distributions based on regression-guided and observational cluster analysis, Atmos. Environ., 89, 346–357, doi:10.1016/j.atmosenv.2014.02.050, 2014.

Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Salvato, M., Smith, P., André, N., Gérard, P., Francis, L. A., Castell, N., Schneider, P., Viana, M., Minguillón, M. C., Reimringer, W., Otjes, R. P., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S. and Penza, M.: Assessment of air quality microsensors versus reference methods: The EuNetAir Joint Exercise – Part II, Atmos. Environ., 193, 127–142, doi:10.1016/j.atmosenv.2018.08.028, 2018.

Bousiotis, D., Brean, J., Pope, F.D., Dall'Osto, M., Querol, X., Alastuey, A., Perez, N., Petäjä, T., Massling, A., Nøjgaard, J.K., Nordstrøm, C., Kouvarakis, G., Vratolis, S., Eleftheriadis, K., Niemi, J.V., Portin, H., Wiedensohler, A., Weinhold, K., Merkel, M., Tuch, T. and Harrison, R.M.: The effect of meteorological conditions and atmospheric composition in the occurrence and development of new particle formation (NPF) events in Europe, Atmos. Chem. Phys., 21(5), 3345 - 3370, https://doi.org/10.5194/acp-21-3345-2021, 2021.

Bousiotis, D., Dall'Osto, M., Beddows, D. C. S., Pope, F. D., Harrison, R. M. and Harrison, C. R. M.: Analysis of new particle formation (NPF) events at nearby rural , urban background and urban roadside sites, 5679–5694, 2019.

Brines, M., Dall'Osto, M., Beddows, D.C.S., Harrison, R.M., Gomez-Moreno, F., Nuñez, L., Artiñano, B., Costabile, F., Gobbi, G.P., Salimi, F., Morawska, L., Sioutas, C. and Querol, X.: Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities, Atmos. Chem. Phys., 15, 5929 - 5945, https://doi.org/10.5194/acp-15-5929-2015, 2015.

Carslaw, D. C. and Ropkins, K.: openair — An R package for air quality data analysis, Environ. Model. Softw., 27–28, 52–61, doi:10.1016/j.envsoft.2011.09.008, 2012.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D. and Bartonova, A.: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, Environ. Int., 99, 293–302, doi:10.1016/j.envint.2016.12.007, 2017.

672 Crilley, L. R., Singh, A., Kramer, L. J., Shaw, M. D., Alam, M. S., Apte, J. S., Bloss, W. J.,
673 Hildebrandt Ruiz, L., Fu, P., Fu, W., Gani, S., Gatari, M., Ilyinskaya, E., Lewis, A. C., Ng'ang'a,
674 D., Sun, Y., Whitty, R. C. W., Yue, S., Young, S. and Pope, F. D.: Effect of aerosol composition
675 on the performance of low-cost optical particle counter correction factors, Atmos. Meas.
676 Tech., 13(3), 1181–1193, doi:10.5194/amt-13-1181-2020, 2020.
677
678 Crilley, L.R., Shaw, M., Pound, R., Kramer, L.J., Price, R., Young, S., Lewis, A.C. and Pope F.D.:
679 Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient
680 monitoring, Atmos. Meas. Tech., 11(2), 709 - 720, https://doi.org/10.5194/amt-11-709-
681 2018, 2018.
682
683 Dall'Osto, M., Beddows, D.C.S, Pey, J., Rodriguez, S., Alastuey, A., Harrison, R.M., and
684 Querol, X.: Urban aerosol size distributions over the Mediterannean city of Barcelona, NE
685 Spain, Atmos. Chem. Phys., 12, 10693 - 10707, https://doi.org/10.5194/acp-12-10693-2012,
686 2012.
687
688 Dall'Osto, M., Monahan, C., Greaney, R., Beddows, D. C. S., Harrison, R. M., Ceburnis, D. and
689 O'Dowd, C. D.: A statistical analysis of North East Atlantic (submicron) aerosol size
690 distributions, Atmos. Chem. Phys., 11(24), 12567–12578, doi:10.5194/acp-11-12567-2011,
691 2011.
692
693 Di Antonio, A., Popoola, O. A. M., Ouyang, B., Saffell, J and Jones, R. L.: Developing a relative
694 humidity correction for low-cost sensors measuring ambient particulate matter, Sensors
695 (Switzerland), 18(9). https://doi.org/10.3390/s18092790, 2018.
696
697 Dockery, D. W., Pope III, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., G., F. B. and E., S.
698 F.: An association between air pollution and mortality in six U.S. cities, N. Engl. J. Med.,
699 329(24), 1753–1759, 1993.
700
701 Draxler, R. R. and Hess, G. D.: An Overview of the HYSPLIT_4 Modelling System for
702 Trajectories, Dispersion, and Deposition, Aust. Meteorol. Mag., 47(January), 295–308, 1998.
703 Feinberg, S. N., Williams, R., Hagler, G., Low, J., Smith, L., Brown, R., Garver, D., Davis, M.,

Morton, M., Schaefer, J. and Campbell, J.: Examining spatiotemporal variability of urban particulate matter and application of high-time resolution data from a network of low-cost air pollution sensors, Atmos. Environ., 213(May), 579–584, doi:10.1016/j.atmosenv.2019.06.026, 2019.

Ghosh, D. and Parida, P.: Air Pollution and India: Current Scenario, Int. J. Curr. Res., 7(11), 22194–22196, 2015.

Giorio, C., Tapparo, A., Dallosto, M., Beddows, D. C. S., Esser-Gietl, J. K., Healy, R. M. and Harrison, R. M.: Local and regional components of aerosol in a heavily trafficked street canyon in central London derived from PMF and cluster analysis of single-particle ATOFMS spectra, Environ. Sci. Technol., 49(6), 3330–3340, doi:10.1021/es506249z, 2015.

Hagan, D.H., Kroll, J.H., Assessing the accuracy of low-cost optical particle sensors using a physics-based approcach, Atmos. Meas. Tech., 13, 6343 - 6355, 2020.

Hagan, D. H., Gani, S., Bhandari, S., Patel, K., Habib, G., Apte, J. S., Hildebrandt Ruiz, L. and Kroll, J. H.: Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India, Environ. Sci. Technol. Lett., 6(8), 467–472, doi:10.1021/acs.estlett.9b00393, 2019.

Harrison, R. M.: Urban atmospheric chemistry: a very special case for study, npj Clim. Atmos. Sci., 1(1), 5, doi:10.1038/s41612-017-0010-8, 2017.

Harrison, R. M., Deacon, A. R., Jones, M. R. and Appleby, R. S.: Sources and processes affection concentrations of PM10 and PM2.5 particulate matter in Birmingham (U.K.), Atmos. Environ., 31(24), 4103–4117, doi:10.1016/S1352-2310(97)00296-3, 1997.

Harrison, R. M., Jones, A. M. and Lawrence, R. G.: A pragmatic mass closure model for airborne particulate matter at urban background and roadside sites, Atmos. Environ., 37(35), 4927–4933, doi:10.1016/j.atmosenv.2003.08.025, 2003.

736 Harrison, R. M., Jones, A. M. and Lawrence, R. G.: Major component composition of PM10
737 and PM2.5 from roadside and urban background sites, Atmos. Environ., 38(27), 4531–4538,
738 doi:10.1016/j.atmosenv.2004.05.022, 2004.
739
740 Holstius, D. M., Pillarisetti, A., Smith, K. R. and Seto, E.: Field calibrations of a low-cost aerosol
741 sensor at a regulatory monitoring site in California, Atmos. Meas. Tech., 7(4), 1121–1131,
742 doi:10.5194/amt-7-1121-2014, 2014.
743
744 Ionascu, M. E., Gruicin, I. and Marcu, M.: Laboratory evaluation and calibration of low-cost
745 sensors for air quality measurement, SACI 2018 - IEEE 12th Int. Symp. Appl. Comput. Intell.
746 Informatics, Proc., 395–400, doi:10.1109/SACI.2018.8440974, 2018.
747
748 Jerrett, M., Donaire-Gonzalez, D., Popoola, O., Jones, R., Cohen, R. C., Almanza, E., de Nazelle,
749 A., Mead, I., Carrasco-Turigas, G., Cole-Hunter, T., Triguero-Mas, M., Seto, E. and
750 Nieuwenhuijsen, M.: Validating novel air pollution sensors to improve exposure estimates for
751 epidemiological analyses and citizen science, Environ. Res., 158(April), 286–294,
752 doi:10.1016/j.envres.2017.04.023, 2017.
753
754 Jovašević-Stojanović, M., Bartonova, A., Topalović, D., Lazović, I., Pokrić, B. and Ristovski, Z.:
755 On the use of small and cheaper sensors and devices for indicative citizen-based monitoring
756 of respirable particulate matter, Environ. Pollut., 206, 696–704,
757 doi:10.1016/j.envpol.2015.08.035, 2015.
758
759 Kan, H., Chen, B. and Hong, C.: Health impact of outdoor air pollution in China: Current
760 knowledge and future research needs, Environ. Health Perspect., 117(5), 12737,
761 doi:10.1289/ehp.12737, 2009.
762
763 Kanaroglou, P. S., Jerrett, M., Morrison, J., Beckerman, B., Arain, M. A., Gilbert, N. L. and
764 Brook, J. R.: Establishing an air pollution monitoring network for intra-urban population
765 exposure assessment: A location-allocation approach, Atmos. Environ., 39(13), 2399–2409,
766 doi:10.1016/j.atmosenv.2004.06.049, 2005.
767

768 Kelly, K. E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R. and
769 Butterfield, A.: Ambient and laboratory evaluation of a low-cost particulate matter sensor,
770 Environ. Pollut., 221, 491–500, doi:10.1016/j.envpol.2016.12.039, 2017.
771
772 Kotsev, A., Schade, S., Craglia, M., Gerboles, M., Spinelle, L. and Signorini, M.: Next generation
773 air quality platform: Openness and interoperability for the internet of things, Sensors
774 (Switzerland), 16(3), doi:10.3390/s16030403, 2016.
775
776 Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M.,
777 Norford, L. and Britter, R.: The rise of low-cost sensing for managing air pollution in cities,
778 Environ. Int., 75, 199–205, doi:10.1016/j.envint.2014.11.019, 2015.
779
780 Lagerspetz, E., Motlagh, N. H., Arbayani Zaidan, M., Fung, P. L., Mineraud, J., Varjonen, S.,
781 Siekkinen, M., Nurmi, P., Matsumi, Y., Tarkoma, S. and Hussein, T.: MegaSense: Feasibility of
782 Low-Cost Sensors for Pollution Hot-spot Detection, IEEE Int. Conf. Ind. Informatics, 2019-July,
783 1083–1090, doi:10.1109/INDIN41052.2019.8971963, 2019.
784
785 Lewis, A. C., von Schneidemesser, E., Peltier, R. E., Lung, C., Jones, R., Zellweger, C., Karppinen,
786 A., Penza, M., Dye, T., Hüglin, C., Ning, Z., Leigh, R., Hagan, D. H., Laurent, O. and Carmichael,
787 G.: Low-cost sensors for the measurement of atmospheric composition: overview of topic and
788 future applications. [online] Available from:
789 http://www.wmo.int/pages/prog/arep/gaw/documents/Draft_low_cost_sensors.pdf, 2018.
790
791 Likas, A., Vlassis, N. and Verbeek, J.J.: The global k-means clustering algorithm, Pattern
792 Recognition, 36(2), 451-461, https://doi.org/10.1016/S0031-3203(02)00060-2, 2003.
793
794 Mahbub, P., Ayoko, G.A., Goonetilleke, A., Egodawatta, P.: Analysis of the build-up of semi
795 and non volatile organic compounds on urban roads, Water Res. 45(9), 2835 - 2844, doi:
796 10.1016/j.watres.2011.02.033, 2011.
797
798 Malings, C., Tanzer, R., Hauryliuk, A., Saha, P. K., Robinson, A. L., Presto, A. A. and
799 Subramanian, R.: Fine particle mass monitoring with low-cost sensors: Corrections and long-
800 term performance evaluation, Aerosol Sci. Technol., 54(2), 160–174,

801 doi:10.1080/02786826.2019.1623863, 2020.
802
803 McGregor, G.R., Bamzelis, D.: Synoptic typing and its application to the investigation of
804 weather air pollution relationships, Birmingham, United Kingdom, Theoretical and Applied
805 Climatology, 51, 223 - 236, 1995.
806
807 Miskell, G., Salmond, J. and Williams, D. E.: Low-cost sensors and crowd-sourced data:
808 Observations of siting impacts on a network of air-quality instruments, Sci. Total Environ.,
809 575, 1119–1129, doi:10.1016/j.scitotenv.2016.09.177, 2017.
810
811 Miskell, G., Salmond, J. A. and Williams, D. E.: Use of a handheld low-cost sensor to explore
812 the effect of urban design features on local-scale spatial and temporal air quality variability,
813 Sci. Total Environ., 619–620, 480–490, doi:10.1016/j.scitotenv.2017.11.024, 2018.
814
815 Moltchanov, S., Levy, I., Etzion, Y., Lerner, U., Broday, D. M. and Fishbain, B.: On the feasibility
816 of measuring urban air pollution by wireless distributed sensor networks, Sci. Total Environ.,
817 502, 537–547, doi:10.1016/j.scitotenv.2014.09.059, 2015.
818
819 Morawska, L., Thai, P.K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai,
820 F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G.S.W., Jayaratne, R., Kumar, M., Lau, A.K.H.,
821 Louie, P.K.K., Mazaheri, M., Ning, Z., Motta, N., Mullins, B., Rahman, M.M., Ristovski, Z.,
822 Shafiei, M., Tjondronegoro, D., Westerdahl, D. and Williams, R.: Applications of low-
823 costsensing technologies for a quality air monitoring and exposure assessment: How far have
824 they gone?, Environment International, 116, 286 - 299,
825 https://doi.org/10.1016/j.envint.2018.04.018, 2018.
826
827 Mueller, M. D., Hasenfratz, D., Saukh, O., Fierz, M. and Hueglin, C.: Statistical modelling of
828 particle number concentration in Zurich at high spatio-temporal resolution utilizing data from
829 a mobile sensor network, Atmos. Environ., 126, 171–181,
830 doi:10.1016/j.atmosenv.2015.11.033, 2016.
831
832 Nagendra, S., Reddy Yasa, P., Narayana, M., Khadirnaikar, S. and Pooja Rani: Mobile

monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots, Sustain. Cities Soc., 44(September 2018), 520–535, doi:10.1016/j.scs.2018.10.006, 2019.

Pakhira, M.K., Bandyopadhyay, S. and Maulik, U.: Validity index of crisp and fuzzy clusters, Pattern Recognition, 37(3), 487-501, https://doi.org/10.1016/j.patcog.2003.06.005, 2004.

Pascal, M., Corso, M., Chanel, O., Declercq, C., Badaloni, C., Cesaroni, G., Henschel, S., Meister, K., Haluza, D., Martin-Olmedo, P. and Medina, S.: Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project, Sci. Total Environ., 449(2007105), 390–400, doi:10.1016/j.scitotenv.2013.01.077, 2013.

Penza, M., Suriano, D., Villani, M. G., Spinelle, L. and Gerboles, M.: Towards air quality indices in smart cities by calibrated low-cost sensors applied to networks, in Proceedings of IEEE Sensors, vol. 2014-Decem, pp. 2012–2017., 2014.

Petkova, E. P., Jack, D. W., Volavka-Close, N. H. and Kinney, P. L.: Particulate matter pollution in African cities, Air Qual. Atmos. Heal., 6(3), 603–614, doi:10.1007/s11869-013-0199-6, 2013.

Pope, F. D., Gatari, M., Ng'ang'a, D., Poynter, A. and Blake, R.: Airborne particulate matter monitoring in Kenya using calibrated low cost sensors, Atmos. Chem. Phys. Discuss., 1–31, doi:10.5194/acp-2018-327, 2018.

Popoola, O. A. M., Carruthers, D., Lad, C., Bright, V. B., Mead, M. I., Stettler, M. E. J., Saffell, J. R. and Jones, R. L.: Use of networks of low cost air quality sensors to quantify air quality in urban settings, Atmos. Environ., 194(February), 58–70, doi:10.1016/j.atmosenv.2018.09.030, 2018.

Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A. and Rickerby, D.: End-user perspective of low-cost sensors for outdoor air pollution monitoring, Sci. Total Environ., 607–608, 691–705, doi:10.1016/j.scitotenv.2017.06.266, 2017.

Rousseeuw, P.J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics, 20, 53-65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Sabaliauskas, K., Jeong, C.-H. H., Yao, X., Jun, Y.-S. S. and Evans, G.: Cluster analysis of roadside ultrafine particle size distributions, Atmos. Environ., 70(0), 64–74, doi:http://dx.doi.org/10.1016/j.atmosenv.2012.12.025, 2013.

Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., Clifford, S. and Morawska, L.: Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment, Atmos. Chem. Phys., 14(21), 11883–11892, doi:10.5194/acp-14-11883-2014, 2014.

Sayahi, T., Butterfield, A. and Kelly, K. E.: Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors, Environ. Pollut., 245, 932–940, doi:10.1016/j.envpol.2018.11.065, 2019.

Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A. and Bartonova, A.: Mapping urban air quality in near real-time using observations from low-cost sensors and model information, Environ. Int., 106(May), 234–247, doi:10.1016/j.envint.2017.05.005, 2017.

Shindler, L., Development of a low-cost sensing platform for air quality monitoring: Application in the city of Rome, Environmental Technology, 42:4, 618 - 631, doi:10.1080/09593330.2019.1640290, 2019.

Schnelle-Kreis, J., Sklorz, M., Orasche, J., Stölzel, M., Peters, A. and Zimmermann, R.: Semi volatile organic compounds in ambient PM2.5. Seasonal trends and daily resolved source contributions, Environ. Sci. Technol., 41(11), 3821–3828, doi:10.1021/es060666e, 2007.

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. and Preuss, P. W.: The changing paradigm of air pollution monitoring, Environ. Sci. Technol., 47(20), 11369–11377, doi:10.1021/es4022602, 2013.

Sousan, S., Koehler, K., Thomas, G., Park, J. H., Hillman, M., Halterman, A. and Peters, T. M.: Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols, Aerosol Sci. Technol., 50(5), 462–473, doi:10.1080/02786826.2016.1162901, 2016.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M. and Bonavitacola, F.: Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide, Sensors Actuators, B Chem., 215, 249–257, doi:10.1016/j.snb.2015.03.031, 2015.

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M. and Bonavitacola, F.: Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2, Sensors Actuators, B Chem., 238, 706–715, doi:10.1016/j.snb.2016.07.036, 2017.

Taiwo, A. M.: Source apportionment of urban background particulate matter in Birmingham, United Kingdom using a mass closure model, Aerosol Air Qual. Res., 16(5), 1244–1252, doi:10.4209/aaqr.2015.09.0537, 2016.

Taiwo, A. M., Beddows, D. C. S., Shi, Z. and Harrison, R. M.: Mass and number size distributions of particulate matter components: Comparison of an industrial site and an urban background site, Sci. Total Environ., 475, 29–38, doi:10.1016/j.scitotenv.2013.12.076, 2014.

U.S. Environmental Protection Agency: Quality Assurance Guidance Document 2.12, , 105 [online] Available from: https://www3.epa.gov/ttnamti1/files/ambient/pm25/qa/m212.pdf, 2016.

United States Environmental Protection Agency: The National Ambient Air Quality Standards for Particle Matter: Revised Air Quality Standards for Particle Pollution and Updates to the Air Quality Index (AQI), Environ. Prot. Agency, 1–5 [online] Available from: http://www.epa.gov/pm/2012/decfsstandards.pdf, 2012.

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J. and Biswas, P.: Laboratory Evaluation and

929 Calibration of Three Low-Cost Particle Sensors for Particulate Matter Measurement, Aerosol
930 Sci. Technol., 49(11), 1063–1077, doi:10.1080/02786826.2015.1100710, 2015.
931
932 Wegner, T., Hussein, T., Hämeri, K., Vesala, T., Kulmala, M. and Weber, S.: Properties of
933 aerosol signature size distributions in the urban environment as derived by cluster analysis,
934 Atmos. Environ., 61, 350–360, doi:10.1016/j.atmosenv.2012.07.048, 2012.
935
936 Weyers, R., Jang-Jaccard, J., Moses, A., Wang, Y., Boulic, M., Chitty, C., Phipps, R. and
937 Cunningham, C.: Low-cost Indoor Air Quality (IAQ) Platform for Healthier Classrooms in New
938 Zealand: Engineering Issues, Proc. - 2017 4th Asia-Pacific World Congr. Comput. Sci. Eng.
939 APWC CSE 2017, (December), 208–215, doi:10.1109/APWConCSE.2017.00045, 2018.
940
941 World Health Organization: Air quality guidelines for particulate matter, ozone, nitrogen
942 dioxide and sulfur dioxide - Global update 2005., 2006.
943
944 Wu, S., Ni, Y., Li, H., Pan, L., Yang, D., Baccarelli, A. A., Deng, F., Chen, Y., Shima, M. and Guo,
945 X.: Short-term exposure to high ambient air pollution increases airway inflammation and
946 respiratory symptoms in chronic obstructive pulmonary disease patients in Beijing, China,
947 Environ. Int., 94, 76–82, doi:10.1016/j.envint.2016.05.004, 2016.
948
949 Xu, Y., Zhang, J.S.: Understanding SVOCs, ASHRAE Journal, 53 (12), 121 - 125, 2011.
950
951 Yin, J., Harrison, R. M., Chen, Q., Rutter, A. and Schauer, J. J.: Source apportionment of fine
952 particles at urban background and rural sites in the UK atmosphere, Atmos. Environ., 44(6),
953 841–851, doi:10.1016/j.atmosenv.2009.11.026, 2010.
954
955 Zeger, S. L., Dominici, F., McDermott, A. and Samet, J. M.: Mortality in the medicare
956 population and Chronic exposure to fine Particulate air pollution in urban centers (2000-
957 2005), Environ. Health Perspect., 116(12), 1614–1619, doi:10.1289/ehp.11449, 2008.
958
959 Zheng, T., Bergin, M. H., Johnson, K. K., Tripathi, S. N., Shirodkar, S., Landis, M. S., Sutaria, R.
960 and Carlson, D. E.: Field evaluation of low-cost particulate matter sensors in high-and low-

961 concentration environments, Atmos. Meas. Tech., 11(8), 4823–4846, doi:10.5194/amt-11-
962 4823-2018, 2018.
963
964

965 **TABLE LEGENDS**

966

967 **Table 1:** List of the measuring instruments used in the present study.

968

969 **Table 2:** Average atmospheric conditions for the clusters formed by both methods.

970

971 **Table 3:** Simultaneous occurrences of the clusters formed by both the OPC and SMPS.
972

973 **FIGURE LEGENDS**

974

975 **Figure 1:** Map of the location of the Birmingham Air Quality Supersite (BAQS) site in the
976 U.K. (Map by ©HERE).

977

978 **Figure 2:** Frequency and diurnal variation of the clusters formed by the OPC data.

979

980 **Figure 3:** Frequency and diurnal variation of the clusters formed by the SMPS data.

981

982 **Figure 4:** Particle contributions in the range 12 – 550 nm (using the SMPS data), for the
983 clusters formed using the OPC data (top) and the SMPS data (bottom).

984

985 **Figure 5:** Particle contributions in the range up to 10 μm (using the FIDAS data), for the
986 clusters formed using the OPC data (top) and the SMPS data (bottom).

987

988 **Figure 6:** Polar plots for the PM_1 (μg m^{-3}) for the clusters formed by the OPC data.

989

990 **Figure 7:** Weekly contribution (week number refers to week of year 2020) of the clusters
991 formed by the OPC (top) and SMPS (bottom).

992

993 **Table 1: List of the measuring instruments used in the present study.**

994

Monitoring	Model	Manufacturer	Regulatory grade	Approximate cost (£)
NO_2	NO2-B43F	Alphasense	No	250
O_x	Ox-B43I	Alphasense	No	160
Black Carbon	MA200	Aethlabs	No	5,700
Lung Deposited Surface Area		Naneos	No	8,500
OPC	OPC-N3	Alphasense	No	250
SMPS	TSi3082	TSi	Yes	80,000
ACSM	Quad - ACSM	Aerodyne	Yes	170,000
PM	Fidas 200E	Palas	Yes	25,000
NO_2	T500U	Teledyne	Yes	15,000
Black Carbon	AE33 Aethalometer	Magee Scientific	Yes	25,000
O_3	49i	Thermo	Yes	3,000

995

996

Table 2: Average atmospheric conditions for the clusters formed by both methods.

	NO_2 (ppb)	BC (ng m^{-3})	PM_1 (µg m^{-3})	$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})	O_3 (ppb)	Organic content (µg m^{-3})	SO_4^{2-} (µg m^{-3})	NO_3^- (µg m^{-3})	LDSA ratio	RH (%)	WS (m s^{-1})	T (°C)	PBL height (m)
OPC.1	18.6±13.9	555±630	4.32±4.08	6.53±4.62	9.97±5.81	31.9±9.81	0.254±0.231	4.12E-02±5.42E-02	8.90E-02±1.15E-01	0.443	83.9±13.1	4.16±2.50	5.20±3.11	852±568
OPC.2	9.64±1.90	233±32.8	2.56±0.72	5.61±1.58	10.7±2.97	38.6±1.34	0.142±0.082	2.98E-02±5.67E-02	1.64E-02±5.53E-03	0.847	65.1±10.5	7.1±1.01	7.16±1.53	1622±264
OPC.3	13.1±8.20	278±153	2.95±0.78	5.80±1.98	9.70±2.69	37.6±6.79	0.241±0.254	6.73E-02±6.25E-02	8.41E-02±1.54E-01	0.830	91.8±8.73	3.47±1.11	4.60±1.95	732±312
OPC.4	11.5±7.15	281±191	2.51±1.55	4.84±3.20	8.33±5.35	36.5±5.17	0.192±0.235	4.53E-02±6.62E-02	1.08E-01±2.53E-01	0.536	83.5±11.5	4.37±2.09	6.26±2.73	930±430
OPC.5	18.3±16.3	659±879	6.27±6.56	9.10±7.18	13.3±8.37	31.5±11.9	0.338±0.558	4.10E-02±6.49E-02	1.31E-01±2.62E-01	0.417	82.6±11.5	4.38±2.50	6.68±3.31	835±485
OPC.6	8.58±6.72	197±155	2.85±1.12	5.96±2.51	10.3±4.30	40.0±4.69	0.116±0.152	3.50E-02±5.08E-02	3.50E-02±1.18E-01	0.588	81.2±12.3	4.87±2.07	6.42±2.89	1135±408
Average	**15.9±13.7**	**498±673**	**4.53±4.93**	**7.11±5.61**	**11.0±6.94**	**33.6±9.95**	**0.252±0.403**	**4.19E-02±6.05E-02**	**1.00E-01±2.08E-01**	**0.499**	**83.1±12.3**	**4.37±2.37**	**6.05±3.11**	**901±504**

	NO_2 (ppb)	BC (ng m^{-3})	PM_1 (µg m^{-3})	$PM_{2.5}$ (µg m^{-3})	PM_{10} (µg m^{-3})	O_3 (ppb)	Organic content (µg m^{-3})	SO_4^{2-} (µg m^{-3})	NO_3^- (µg m^{-3})	LDSA ratio	RH (%)	WS (m s^{-1})	T (°C)	PBL height (m)
SMPS.1	16.0±14.9	485±852	3.35±2.64	5.70±3.89	9.52±6.05	32.2±10.3	0.215±0.300	3.06E-02±4.80E-02	5.47E-02±7.76E-02	0.331	85.1±10.7	4.1±2.70	5.53±3.06	771±558
SMPS.2	16.8±12.0	406±539	2.70±1.57	5.11±2.33	8.91±3.75	32.9±8.10	0.132±0.156	2.53E-02±4.11E-02	2.56E-02±4.31E-02	0.501	83.2±9.71	3.74±1.67	4.64±2.86	831±441
SMPS.3	4.38±2.91	88.1±62.2	2.64±1.62	5.57±3.62	9.26±5.87	41.6±3.24	0.062±0.063	3.74E-02±5.75E-02	2.07E-02±7.15E-02	0.555	80.1±8.93	7.19±2.48	7.43±2.72	1378±290
SMPS.4	14.3±12.3	452±592	3.77±2.56	6.71±3.75	11.1±5.67	35.6±9.32	0.249±0.306	4.68E-02±6.27E-02	8.12E-02±1.53E-01	0.499	79.4±13.9	4.74±2.38	6.97±2.62	1022±540
SMPS.5	29.8±17,2	1389±838	17.95±7.89	21.1±8.08	25.1±7.95	16.1±10.6	1.066±0.562	1.41E-01±7.58E-02	5.74E-01±3.60E-01	0.833	93.9±7.49	2.6±1.63	4.90±2.94	454±330
SMPS.6	13.2±10.8	340±395	2.68±1.58	5.23±3.12	9.12±5.42	36.0±6.54	0.164±0.189	2.93E-02±4.31E-02	3.86E-02±7.17E-02	0.467	81.0±12.7	4.73±2.11	6.1±3.11	1092±426
Average	**15.1±13.2**	**460±649**	**4.12±4.72**	**6.78±5.48**	**10.8±6.90**	**33.8±9.84**	**0.280±0.403**	**4.61E-02±6.40E-02**	**1.07E-02±2.23E-01**	**0.499**	**82.8±12.4**	**4.41±2.42**	**5.95±2.99**	**929±517**

Table 3: Simultaneous occurrences of the clusters formed by both the OPC and SMPS.

OPC/SMPS	SMPS.1	SMPS.2	SMPS.3	SMPS.4	SMPS.5	SMPS.6	Total OPC
OPC.1	48	30	9	71	13	66	237
OPC.2	1	3		5		3	12
OPC.3		15		2	4	2	23
OPC.4	25	27	6	52	19	50	179
OPC.5	24	26	17	39	40	38	184
OPC.6	7	25	9	28	3	25	97
Total SMPS	105	126	41	197	79	184	732

Figure 1: Map of the location of the Birmingham Air Quality Supersite (BAQS) site in the U.K. (Map by ©HERE).

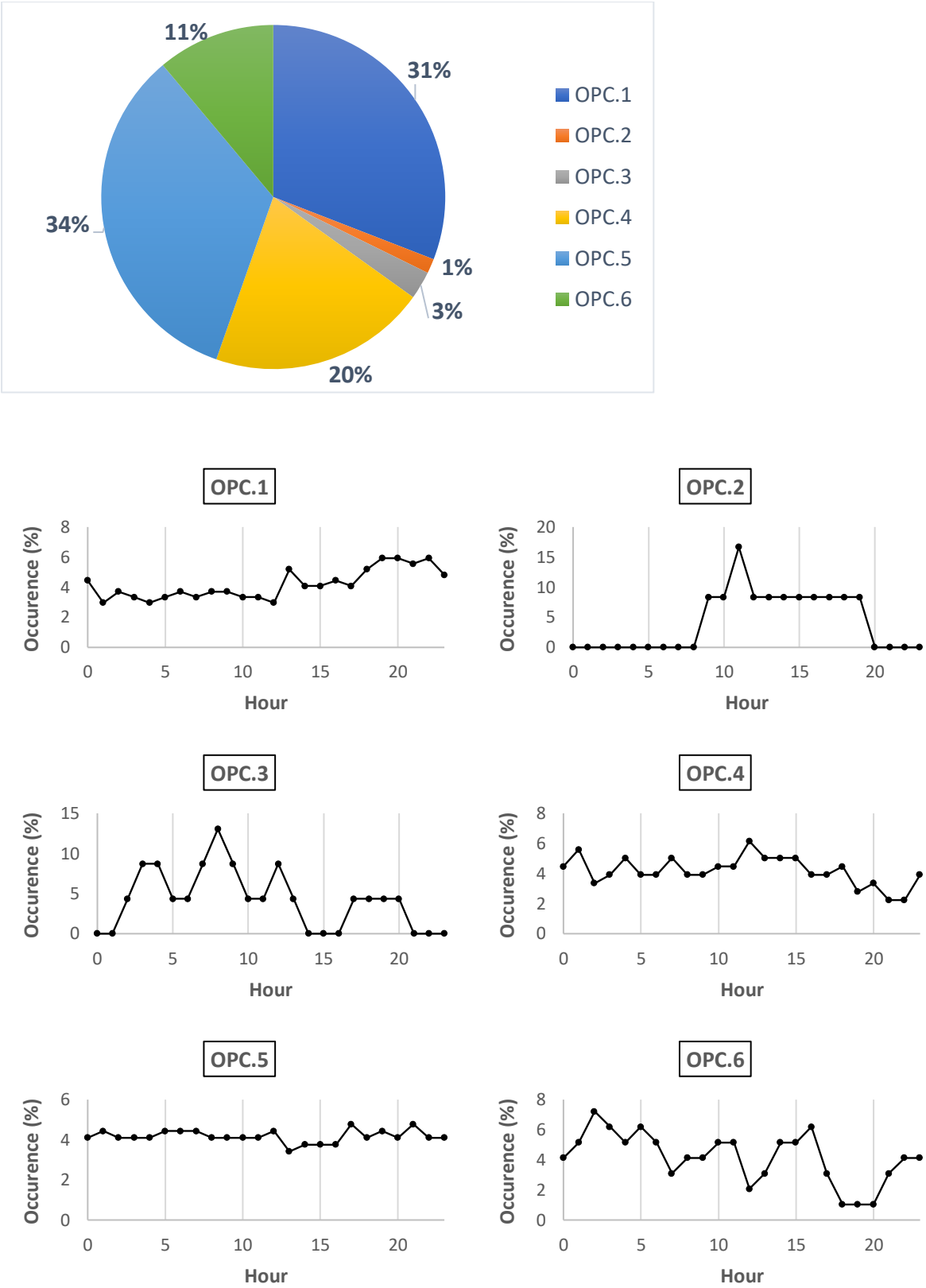
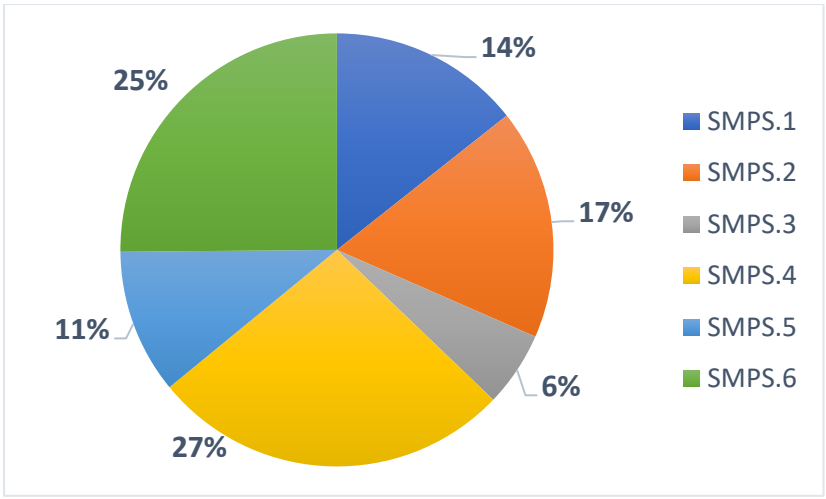


Figure 2: Frequency and diurnal variation of the clusters formed by the OPC data.



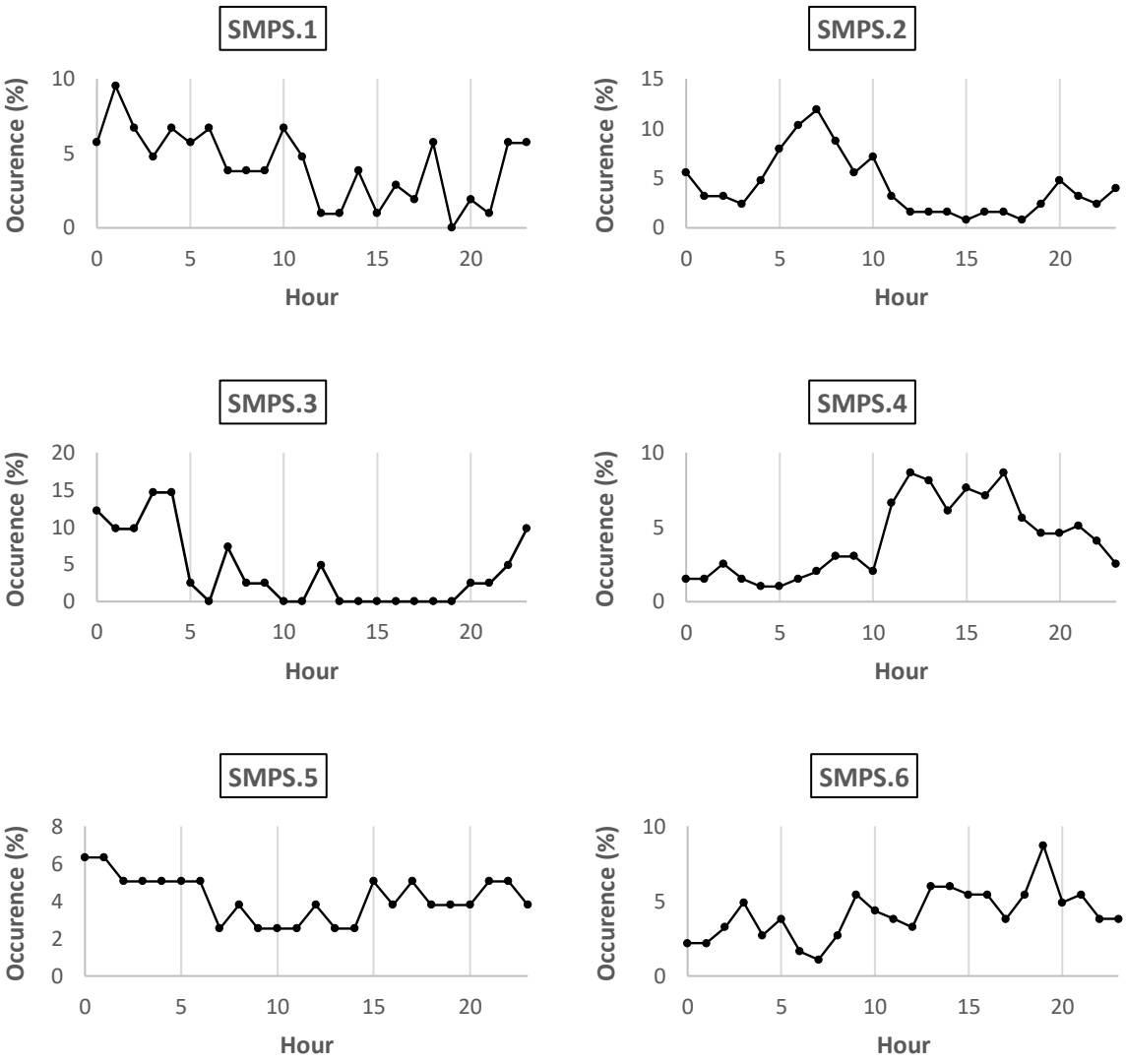


Figure 3: Frequency and diurnal variation of the clusters formed by the SMPS data.

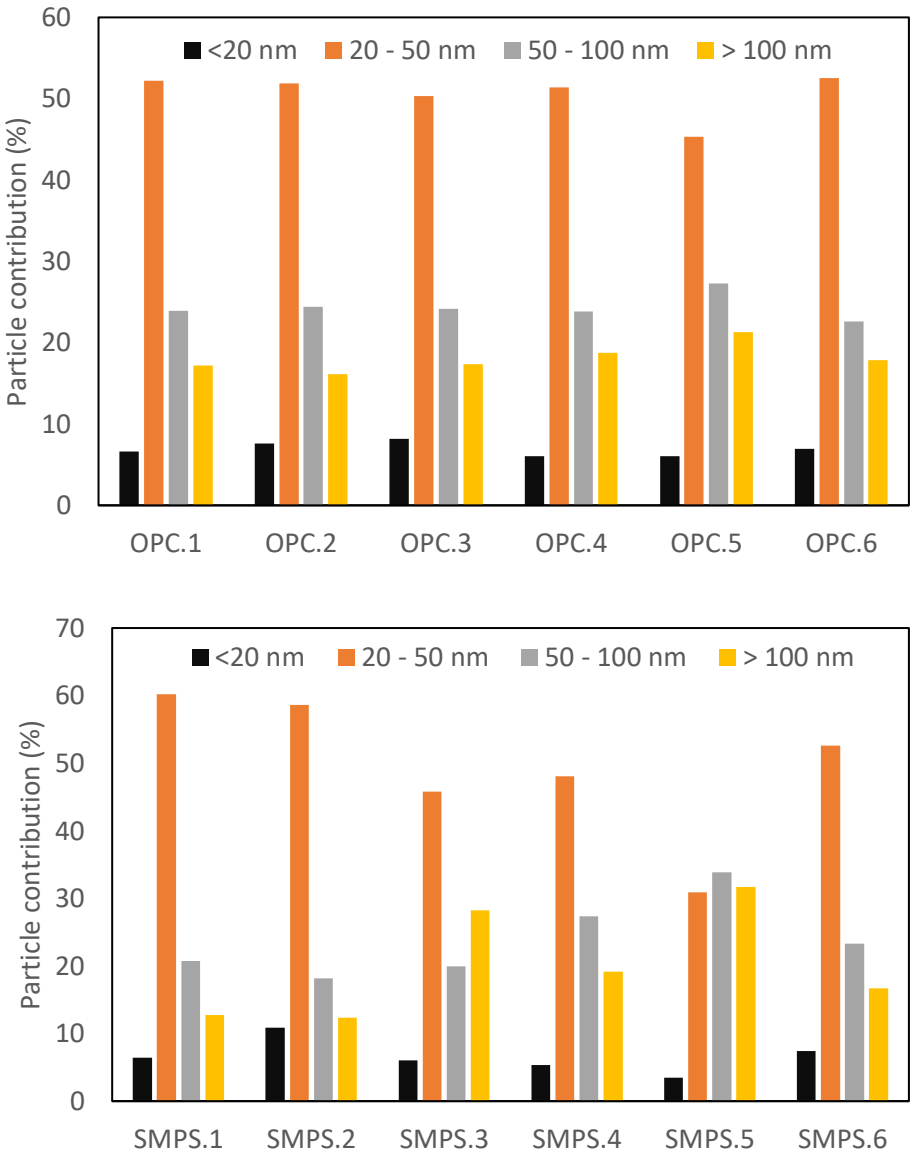


Figure 4: Particle contributions in the range 12 – 550 nm (using the SMPS data), for the clusters formed using the OPC data (top) and the SMPS data (bottom).

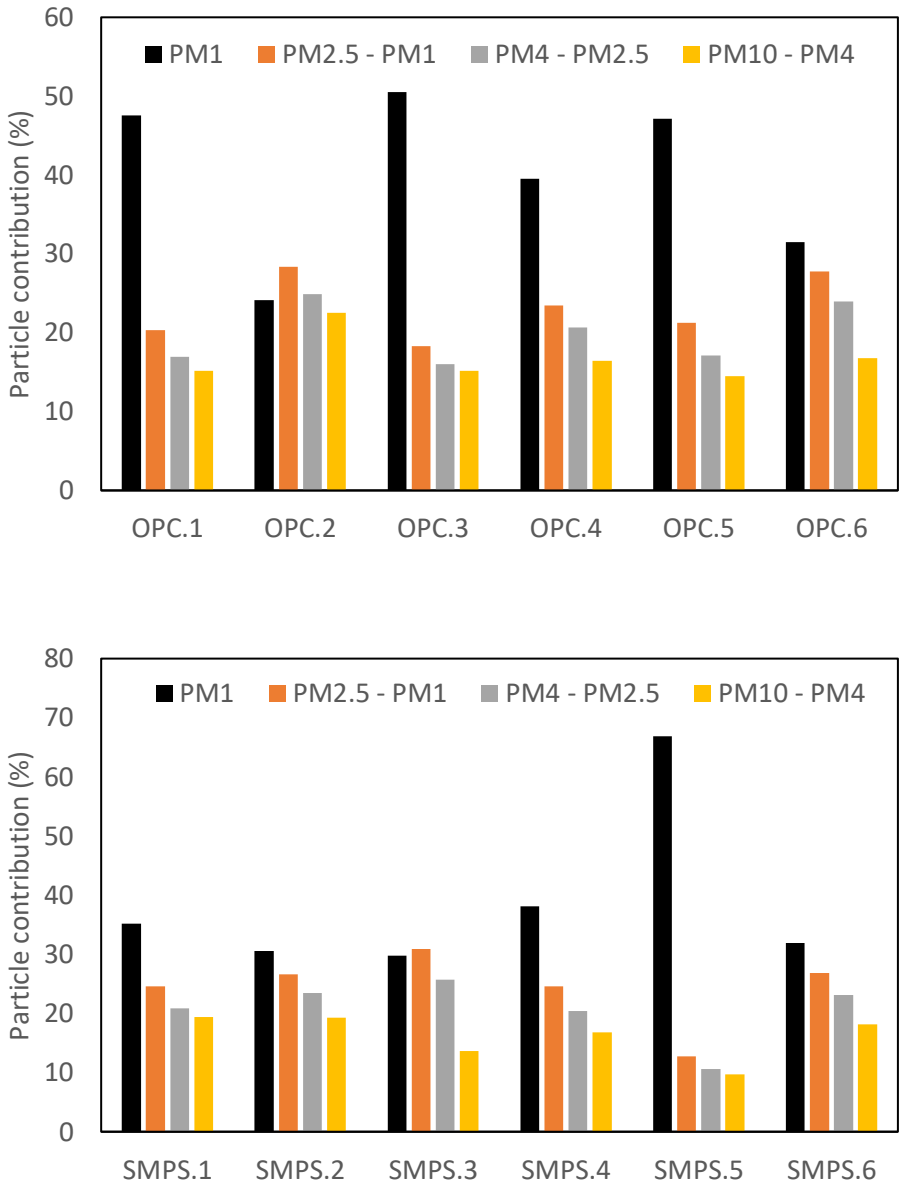


Figure 5: Particle contributions up to 10 µm (using the FIDAS data), for the clusters formed using the OPC data (top) and the SMPS data (bottom).

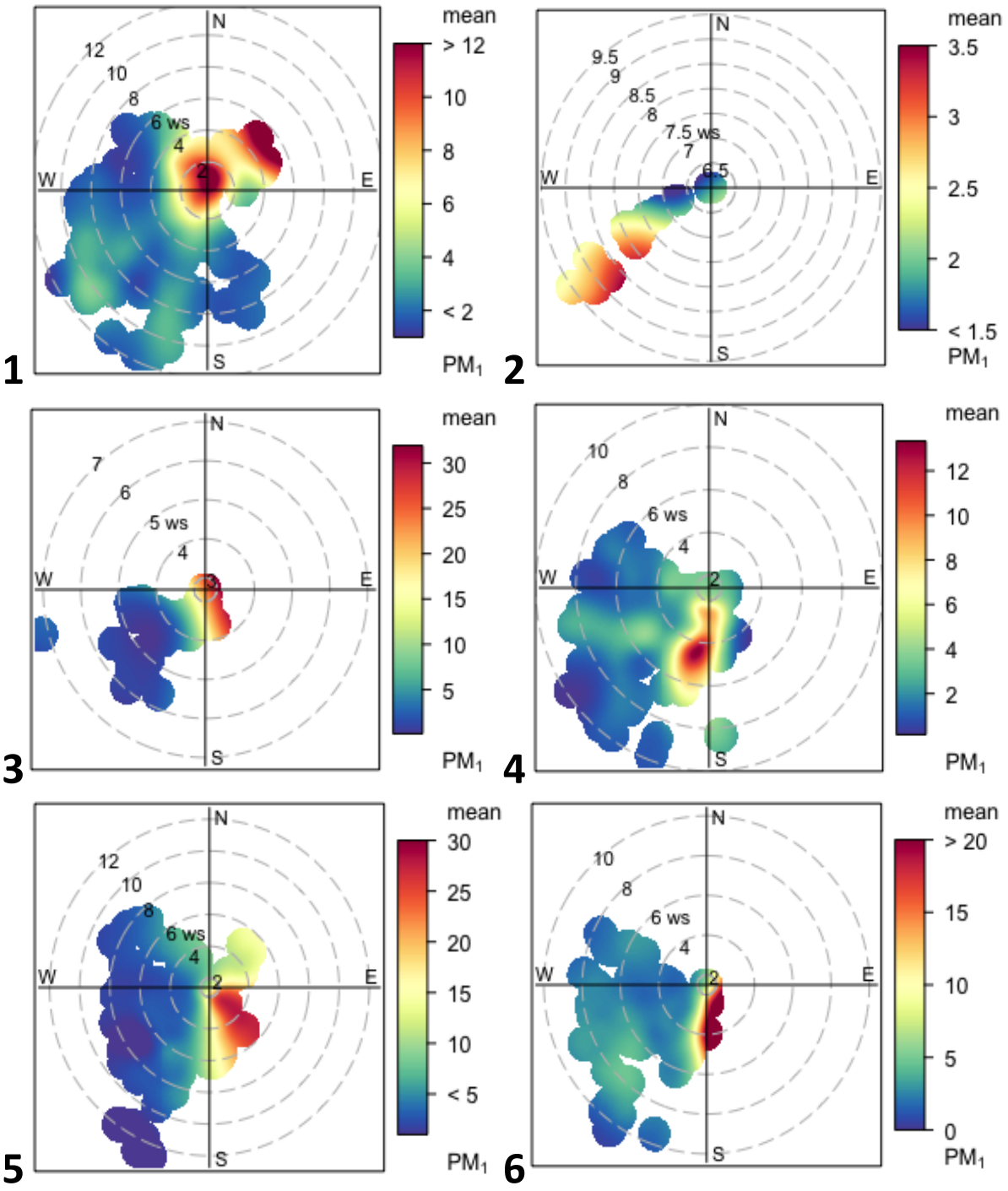


Figure 6: Polar plots for the PM_1 (µg m^{-3}) for the clusters formed by the OPC data.

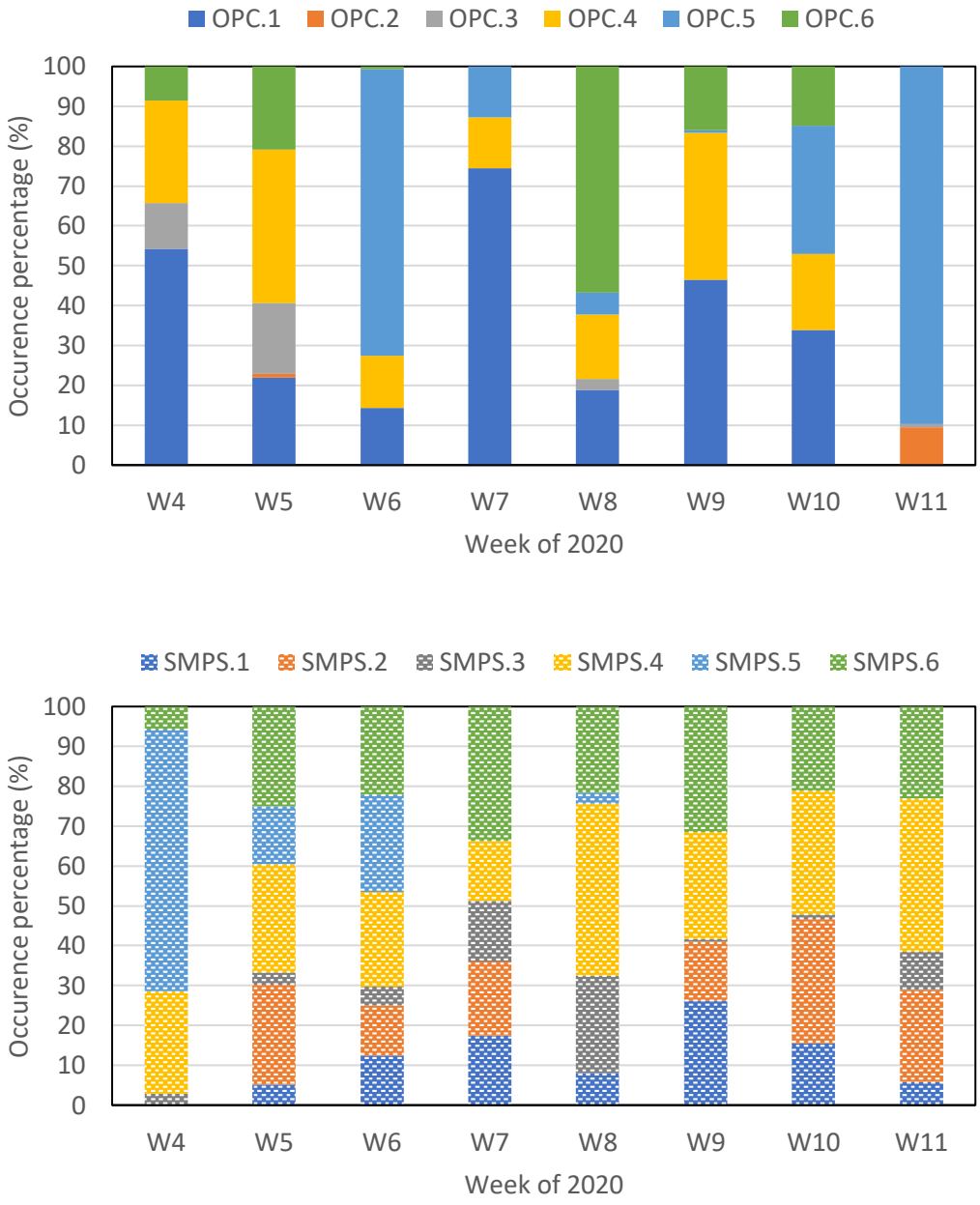


Figure 7: Weekly contribution (week number refers to week of year 2020) of the clusters formed by the OPC (top) and SMPS (bottom).