Article

Not peer-reviewed version

Laboratory Comparison of Low-Cost Particulate Matter Sensors to Measure Transient Events of Pollution - Part B - Particle Number Concentrations

Florentin Michel Jacques Bulot *, Hugo Savill Russell, Mohsen Rezaei, Matthew Stanley Johnson, Steven James Ossont, Andrew Kevin Richard Morris, Philip James Basford, Natasha Hazel Celeste Easton, Hazel Louise Mitchell, Gavin Lee Foster, Matthew Loxham, Simon James Cox

Posted Date: 21 April 2023

doi: 10.20944/preprints202304.0701.v1

Keywords: low-cost sensors; particle number concentration; laboratory study; fine particles; particulate matter; air pollution

Article

Laboratory Comparison of Low-Cost Particulate Matter Sensors to Measure Transient Events of Pollution—Part B—Particle Number Concentrations

Florentin Michel Jacques Bulot [1,2,*], Hugo Savill Russell [3,4,5,6], Mohsen Rezaei [6], Matthew Stanley Johnson [4,6], Steven James Ossont [7], Andrew Kevin Richard Morris [8], Philip James Basford [1], Natasha Hazel Celeste Easton [2,9], Hazel Louise Mitchell [1], Gavin Lee Foster [9], Matthew Loxham [2,10,11,12] and Simon James Cox [1,2]

1 Faculty of Engineering and Physical Sciences, University of Southampton, Southampton SO17 1BJ, UK; p.j.basford@soton.ac.uk (P.J.B.); hlm1g16@soton.ac.uk (H.L.M.); s.j.cox@soton.ac.uk (S.J.C.)
2 Southampton Marine and Maritime Institute, University of Southampton, Southampton SO16 7QF, UK; nhcs1g13@soton.ac.uk (N.H.C.E.); m.loxham@soton.ac.uk (M.L.)
3 Danish Big Data Centre for Environment and Health (BERTHA), Aarhus University, DK-4000 Roskilde, Denmark; hugo.russell@envs.au.dk
4 AirScape UK, 88 Baker Street, London, England, W1U 6TQ, Denmark; matthew.johnson@airscape.ai or msj@chem.ku.dk
5 Department of Environmental Science, Atmospheric Measurement, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
6 Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; mrs@bce.au.dk
7 BizData, 278 Collins St, Melbourne, VIC 3000, AU; steven.ossont@bizdata.co.nz
8 National Oceanography Centre, Southampton SO14 3ZH, UK; andmor@noc.ac.uk
9 School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK; gavin.foster@noc.soton.ac.uk
10 Faculty of Medicine, University of Southampton, Southampton SO17 1BJ, UK
11 National Institute for Health Research, Southampton Biomedical Research Centre, Southampton SO16 6YD, UK
12 Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK
* Correspondence: florentin.bulot@centraliens.net

Abstract: Low-cost particulate matter (PM) sensors offer an excellent opportunity to improve our knowledge about this type of pollution. Their size and their cost, which support multi-node network deployment, along with their temporal resolution, enable them to report fine spatio-temporal resolution for a given area. These sensors have known issues across performance metrics. Generally, the literature focuses on the PM mass concentration reported by these sensors but some models of sensors also report particle number concentrations (PNC) segregated into different PM size bins. In this study, 8 units of each Alphasense OPC-R1, Plantower PMS5003 and Sensirion SPS30 have been exposed, under controlled conditions, to short-lived peaks of PM generated using two different combustion sources of PM, exposing the sensors to different particle size distributions to quantify and better understand the low cost sensors performance across a range of relevant environmental ranges. The PNC reported by the sensors were analysed. This study showed that there is added value in directly using the PNC reported by the sensors instead of the mass concentrations, which could aid the efforts to calibrate these sensors to a known accuracy. It demonstrated that all sensors tested here could track the fine temporal variation of PNC, that the Alphasense OPC-R1 could closely follow the variations of size distribution between the two sources of PM, and it showed that particle size distribution and composition are more impactful on sensor measurements than relative humidity.

Keywords: low-cost sensors; particle number concentration; laboratory study; fine particles; particulate matter; air pollution

 doi:10.20944/preprints202304.0701.v1

1. Introduction

Exposure to air pollution is a major cause of environmental morbidity and mortality in the world at present, with Particulate Matter (PM) air pollution being associated with 8.9 million premature deaths per year [1,2]. PM air pollution varies with fine spatio-temporal granularity and can have heterogeneous composition and concentration over a specific area [3]. Current regulatory monitoring networks are based on cumbersome and expensive apparatus that means monitoring with the spatial coverage required to comprehensively understand the spread of air pollution is not feasible. Given the recently substantially reduced WHO exposure limits [4], down to 5 $\mu g/m^3$ as an annual mean, there is an increased need for monitoring. At this lower threshold, local sources can often be the factor causing exceedance, which makes information concerning local levels and sources more important than they have been in the past [5]. The EU is moving towards adapting the more stringent EU standard [6], and voices in the community are saying that the only way to ensure compliance is by using dense networks of low-cost sensors in populated areas [7].

Low-cost PM sensors have been used in the literature and in various projects around the world to determine PM mass concentrations, especially when deployed as networks of sensors to improve the limited spatio-temporal coverage of existing monitoring networks [5]. Considerable research has been conducted to reach a known level of precision and accuracy with some studies achieving the data quality objectives of reference-grade instruments with the proper calibration methods and frequencies [8], at high temporal resolution, providing data that was not previously available to determine population exposure to PM air pollution at a finer level. However, some of these sensors provide not only PM mass concentrations but also Particle Number Concentration (PNC) for different size bins, for example by giving PNC between 0.3–1 µm and PNC between 1–2.5 µm. Although legal limits are based on PM mass concentration, not all PM is equally harmful and other properties of the particles may be more significant in terms of health impact, such as their composition, shape, size, etc. [9]. There is variation in what is reported with some sensors giving a detailed size distribution, while others only outputting PNC with a restricted number of size bins. For instance, the Plantower PMS5003 outputs six different size bins and the Alphasense OPC-R1 outputs 13 size bins. The ability of the sensors to report PNC of different size fractions can be used to identify sources of pollution. Indeed, in Delhi, India, Hagan et al. [10] used the first three size bins of an Alphasense OPC-N2, in conjunction with data on other air pollutants (CO, NO_2, SO_2, O_3) to successfully identify sources of pollution using positive matrix factorisation. Additionally, we previously demonstrated reference-grade improvements to the performances of Plantower PMS5003 and Sensirion SPS30 through calibration methods based on the PNC reported by these sensors [8]. However, it is not clear what confidence to assign to these outcomes, and which methods should be used or not to avoid over-fitting.

There are broadly two types of low-cost PM sensors [11]: (1) volume scattering, or integrating nephelometers, that measure the light scattered by an ensemble of particles; and (2) single particle counters which count individual particles. The two types have different sensitivities to aerosol parameters and environmental factors [12]. However, there is disagreement in the literature about which sensor belongs to which type. There is also concern about whether these low-cost PM sensors can accurately segregate PNC into different size bins [11–15]. Recently, Ouimette et al. [16] conducted a detailed study of the PurpleAir sensors (which is using two Plantower PMS5003) comparing it to a research-grade integrating nephelometer and developed a physical model that showed that the Plantower PMS5003 is a cell-reciprocal nephelometer providing a reliable measurement of the aerosol scattering coefficients for particles between 0.26–0.46 µm. Ouimette et al. [16] is one of the rare studies that focused on sensor-reported PNC. A few laboratory studies have been conducted regarding the size segregation capacity of the sensors [13,15,17,18]. Three of these studies have focused only on sensor-reported mass concentrations, while one has also studied sensor-reported PNC. All of the above studies examined sensor performances with stable concentrations of PM over periods ranging from 5 min to 1 h, depending on the study. They exposed the sensors to PM from a variety of sources and size. Several studies highlighted that low-cost PM sensors are susceptible to a range of environmental

factors, namely particle composition, size distribution and Relative Humidity (RH). However, different studies obtained contrasting results concerning RH which suggests that other factors may be at play that are not accounted for.

Feature selection methods quantify the contribution of individual features (here environmental factors) to the variability of an output variable (here sensor-reported PNC) [19]. They are one of the most popular techniques to improve the explainability of machine learning models [20], which are often used to correct measurements from low-cost sensors. Feature selection methods are divided into three sub-categories of method: filter-based, wrapper-based and embedded methods [21,22]. Filter-based methods class the variables using different metrics such as the Pearson coefficient or the Akaike Information Criteria (AIC). They do not account for possible correlation between variables and are prone to miss patterns [23]. Wrapper-based methods iteratively use supervised learning techniques (e.g., linear model, support vector machine) to classify the variables. They apply algorithms such as recursive feature selection and greedy forward selection. They are generally more accurate than filter-based methods but risk over-fitting and are more computationally intensive [23]. Finally, embedded methods have the reduction of the number of variables embedded in their algorithms, such as Lasso regression, elastic net regression or random forest. They constitute a trade-off between filter and wrapper-based methods [24]. Nonetheless, the features selected are dependent on the methods chosen and the best practice is to use different methods concomitantly and to compare their results [19].

The current study is the second part of a comprehensive experiment that aimed to characterise the response of a range of low-cost PM sensors to transient events of PM pollution. The first part of this study [25] focused on sensor-reported mass concentrations while this current contribution focuses on sensor-reported PNC. Sensors measuring at a 10 s temporal resolution were exposed to short-lived peaks of PM pollution (≈1 min) generated by lighting candles and incense sticks at different RH levels. Using two combustion sources, we can assess the performance of the sensors across different size distributions. Understanding the response of these sensors to short-lived events of PM pollution is important especially if these sensors are to be used indoors, where polluting activities may last only for a few minutes [26], or used outdoors as a network for tracking events of PM pollution as they spread through an area [27]. These data could also be integrated into models to further their granularity through data fusion techniques [28]. Here, we compare eight units for each sensor model Sensirion SPS30, Plantower PMS5003 and Alphasense OPC-R1, for a total of 24 low-cost PM sensors, at 10 s resolution. A TSI OPS 3330 is used as a reference instrument. The aim is to characterise sensor-reported particle size distribution, to determine whether sensor-reported PNC can follow the transient variations of PM observed by the reference instruments and to determine the relative impact of different variables on the performances of the sensors.

The objectives of this study are:

1. to determine whether the sensors can be used at a high temporal resolution to follow trends of PNC
2. to determine whether the different sensor-reported PNC are independent from each other and characterise their accuracy
3. to characterise the capacity of these sensors to capture the size distribution of PM
4. to determine environmental factors impacting the response of the sensors

2. Materials and Methods

2.1. Low-Cost PM Sensors

The low-cost sensors were mounted in the air quality monitors developed in Johnston et al. [29], without their environmental enclosure as can be seen in Figure 1. The absence of enclosure helps reduce potential residual heat build -up in the vicinity of the sensors. The low-cost sensors studied here are the Plantower PMS5003, the Sensirion SPS30 and the Alphasense OPC-R1. These sensors were chosen because they output PNC of the PM measured for different size fractions. Table 1 presents the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

range of the different size bins of these three models of sensors. The Honeywell HPMA115S0 and the Novafitness SDS018 were also measuring during the experiment but the data they produced were not used in this study as they only report PM mass concentrations. All the low-cost sensors tested here are optical measurement devices based on Mie light-scattering. Four air quality monitors were used concomitantly, each containing two of each of the sensor models mentioned above, for a total of eight sensors of each model. In each air quality monitor, the sensors were plugged in via USB to a Raspberry Pi, powered through Power Over Ethernet (PoE), and controlled using Python 3.6 libraries [30–32]. The data recorded by all the sensors were averaged over 10 s for cross comparison purposes. Relative humidity and temperature were measured by each of the four air quality monitors using a Sensirion SHT35 [33] (±1.5 % RH and ±0.1 °C).

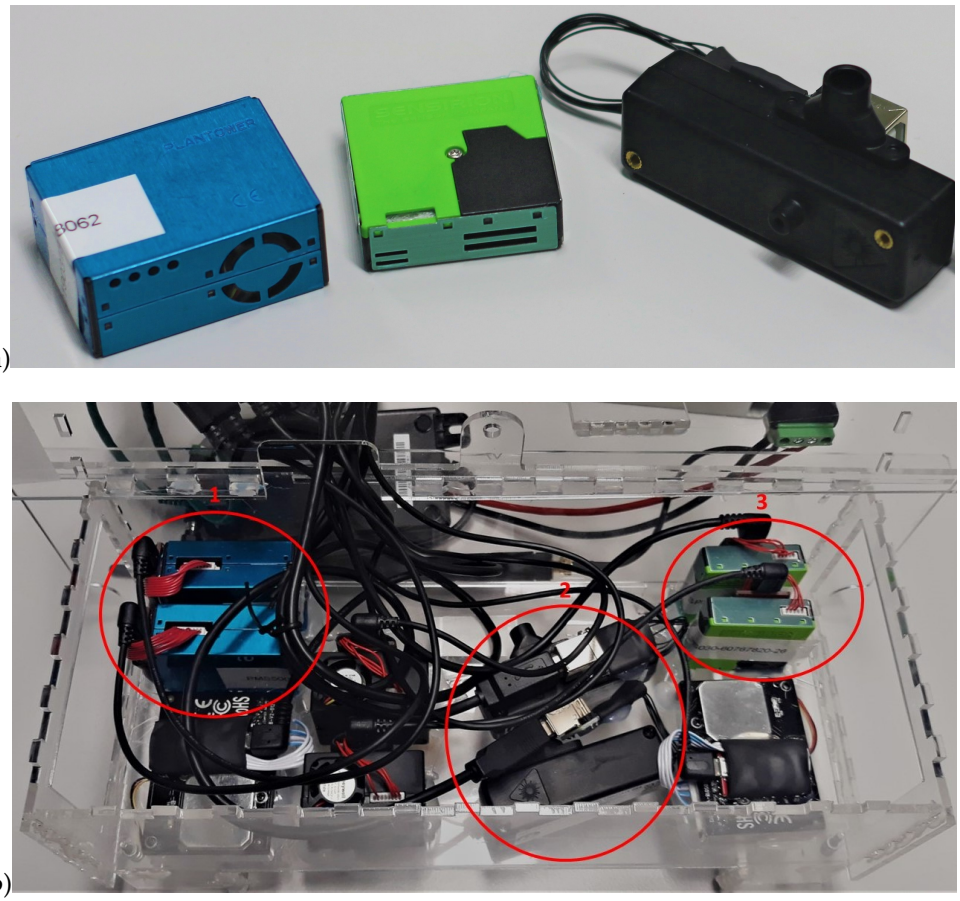


Figure 1. (**a**) Sensors tested from left to right: Plantower PMS5003, Sensirion SPS030, Alphasense OPC-R1, adapted from Bulot et al. [25]. (**b**) Position of the sensors tested within each air quality monitor. From left to right, top to bottom: two Plantower PMS5003 (1), one Novafitness SDS018, two Honeywell HPMA115S0, two Alphasense OPC-R1 (2), two Sensirion SPS30 (3) and one Novafitness SDS018. All the inlets are facing down. Adapted from Bulot et al. [25].

The Plantower PMS5003 reports six size bins called gr03um, gr05um, gr10um, gr25um, gr50um and gr100um, which represent respectively PNC of particles >0.3 µm, >0.5 µm, >1 µm, >2.5 µm, >5 µm and >10 µm. The PNC are reported as particles per 0.1 L of air. According to Sayahi et al. [34], the Plantower PMS5003 has a flow rate of ≈0.1 L/min and a wavelength of 640 ± 10 nm with light polarised at 90° [16].

The Sensirion SPS30 reports size bins called n05, n1, n25, n4 and n10, which represent respectively PNC between 0.3–0.5 µm, between 0.3–1 µm, between 0.3–2.5 µm, between 0.3–4 µm, and between 0.3–10 µm. It utilises a laser beam of 660 nm wavelength and reports PNC as particles/cm^3. The Sensirion SPS30 are calibrated by their manufacturer against a TSI OPS 3330 or a TSI DustTrak DRX

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

8533. The accuracy of the calibration is then verified by the manufacturer using an atomized potassium chloride solution [35]. For the Sensirion SPS30, according to the manufacturer, the particles above 4 µm are not directly measured but determined from the other size bins using a particle distribution profile. The Sensirion SPS30 is certified for UK indicative monitoring and although the sensors used in this study were acquired prior to the certification, private communication with the manufacturer confirmed that there had been no significant changes between the sensors used here and the sensors used for the certification.

The Alphasense OPC-R1 is a single particle counter, which utilises a laser beam at 639 nm of wavelength, which can theoretically count up to 10,000 particles/s or 2500 particles/cm^3 with a maximum coincidence probability of 0.7% at 1000 particles/cm^3. The PNC are output as particles/cm^3 into 16 different size bins (see Table 1) and can measure particles between 0.35–12.4 µm. It has a flow rate of 0.24 L/min. The Alphasense OPC-R1 was calibrated by their manufacturer using monodisperse Polystyrene Sphere (PLS) particles against an Alphasense OPC-R1, which itself had previously been calibrated against a TSI OPS 3330 [36].

Table 1. Size bins of the Plantower PMS5003, Sensirion SPS30, Alphasense OPC-R1 and TSI OPS 3330.

Sensor	Size Bins Range (µm)
PMS5003	0.3–10; 0.5–10; 1–10; 2.5–10; 5–10; >10
SPS30	0.3–0.5; 0.3–1; 0.3–2.5; 0.3–4; 0.3–10
OPC-R1	0.4–0.7; 0.7–1.1; 1.1–1.5; 1.5–1.9; 1.9–2.4; 2.4–3; 3–4; 4–5; 5–6; 6–7; 7–8; 8–9; 9–10; 10–11; 11–12; 12–12.4
OPS 3330	0.3–0.4; 0.4–0.5; 0.5–0.6; 0.6–0.7; 0.7–0.9; 0.9–1.1; 1.1–1.4; 1.4–1.7; 1.7–2.2; 2.2–2.7; 2.7–3.3; 3.3–4.2; 4.2–5.2; 5.2–6.5; 6.5–8.0; 8.0–10

2.2. Reference Instruments

2.2.1. TSI OPS 3330

The TSI Optical Particle Sizer (OPS) 3330 (TSI Inc. Shoreview, MN, USA) is an optical instrument based on light scattering which reports PNC divided into 16 size bins (see Table 1), reported in particles/cm^3, from 0.3–10 µm. It uses a laser beam at 660 nm and a flow rate of 1 L/min. It is calibrated by its manufacturer for size using Polystyrene Sphere (PLS) [37]. In this study, the TSI OPS 3330 was used as a reference instrument and set to measure every 10 s. The size bin of the TSI OPS 3330 were re-calculated to match the size bins of the different models of sensors to enable comparison. The method for redistributing the bin size is described below in the Data analysis subsection.

2.2.2. Aerasense Nanotracer

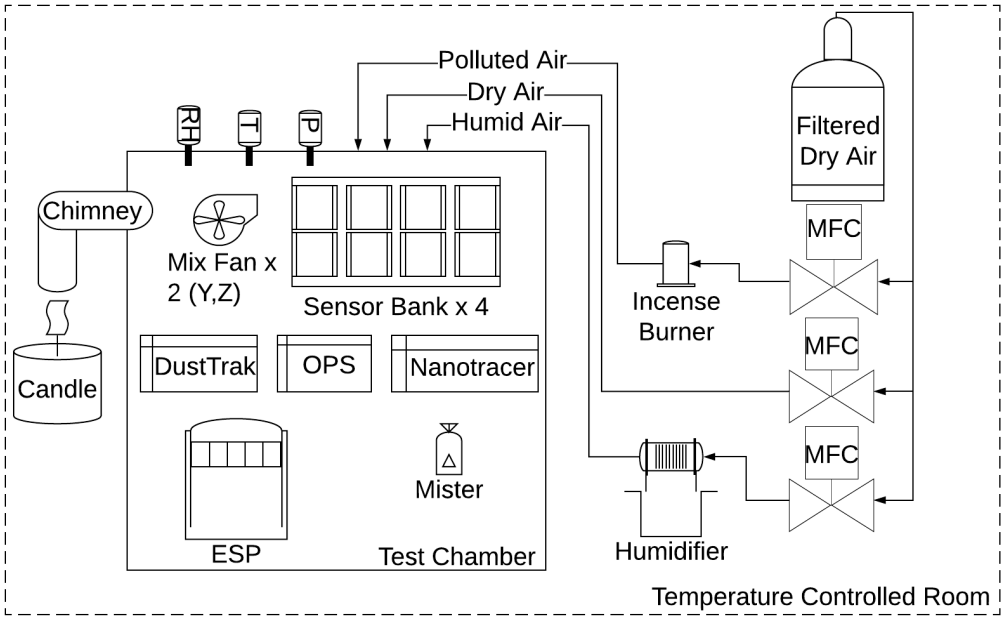
The Aerasense Nanotracer (Oxility BV, Best, Netherlands) counts particles between 10–300 nm based on diffusion charging. It is measuring below the advertised cut-off size of the low-cost sensors. It was calibrated by its manufacturer using KNO_3 polydisperse particles and has a flow rate of 0.3–0.4 L/min. It is used because some of the sensors may be able to measure below 0.3 µm.

2.3. Experimental Conditions

The experimental conditions and set-up are the same as in Bulot et al. [25], the relevant elements are summarised here and the experimental set-up is described in Figure 2. Candle and incense smoke were used as the two different combustion sources, enabling testing at different particle size distributions. Candle smoke, here produced by smouldering, contains mostly particles between 0.02–0.1 µm with particles size peaks between 0.3–0.5 µm in terms of PNC [38,39]. For incense smoke, particles are mostly comprised between 0.05–0.7 µm with a peak at 0.2 µm, in terms of PNC [40]. Five sets of experiments were conducted. For each set of experiment, several peaks of candle smoke were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

generated, followed by a longer concentration of candle smoke, then the air was cleaned of particles using the Electrostatic Precipitator (ESP), then a series of peaks of incense smoke were generated followed by a stable concentration of incense smoke. The peaks of PM lasted around 1 min and had a targeted concentration of 20–50 µg/m^3 as measured in real-time by a DustTrak DRX 8533 Desktop (TSI Inc.). During each experiment, the RH was set at different targets: 54, 69, 72, 76, 79% RH. RH was controlled by a mist generator. The experimental conditions are further described in Bulot et al. [25].



(**a**)

(**b**)

Figure 2. (**a**) Schematic showing the arrangement of the test chamber and supporting equipment, adapted from Bulot et al. [25]. (**b**) Image showing the air quality boxes located in the test chamber, adapted from Bulot et al. [25]. Mass Flow Controllers (MFC).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

The particle size distribution from the TSI OPS 3330 is available in Supplementary Figure A13 and shows that for candle-generated PM, there is less than 10 particles/cm^3 above 5 µm and for incense-generated PM above 2.5 µm. Given the low values of PNC above 2.5 µm, only the size bins of the sensors having a lower cut size <2.5 µm will be considered during this study.

2.4. Data Analysis

2.4.1. Feature Selection Methods

As detailed in the introduction of this paper, the outcome of feature selection methods depend on the methods chosen and it is best practice to use different methods concomitantly and to compare their results. In this paper, we tested three methods: Ridge regression, Boruta, Recursive Feature Elimination (RFE) with Support Vector Machines (SVM). Ridge regression feature selection is an extension of linear model with a penalisation term on the residual of the sums of squares using L_2 norm [41]. Ridge regression is an embedded method, as the selection of features is part of its algorithms. Boruta and RFE are both wrapper methods. Boruta is a wrapper based on random forest, it starts by adding shuffled copies of the existing variables to the dataset, called shadow variables. It then trains a random forest on this dataset and measures the variable importance (using the variable importance measure built into the random forest) and compares the importance of the initial variables to the importance of the shadow variables. Variables that obtained a significantly lower score than the higher score of the shadow variables are removed. It then reiterates the process [42]. RFE was initially developed to enable SVM to perform feature selection [43]. It trains a SVM model, computes a ranking criterion for all the variables considered (the weights of the SVM) and then removes the feature with the smallest ranking criterion. Filter-based methods have not been tested here as they generally do not allow for complex interactions between the variables considered. It is important to note that the scores obtained across the different methods cannot be compared, only the relative scores of each variable within a method can be compared.

In this study, the following variables were used for feature selection to predict each size bin of each sensor: the source of PM (candle or incense), the number of particles <0.3 µm (measured by the Nanotracer), the number of particles between 0.3–0.8 µm (measured by the TSI OPS 3330), the number of particles between 0.3–10 µm (measured by the TSI OPS 3330) and RH. The feature selection methods are performed on the data from the five different levels of RH and using both source of PM, aggregated by model of sensor.

2.4.2. Lognormal Size Distribution

Particle size distribution is presented using lognormal distributions of the normalised concentrations calculated using the formula, for each size bins of the instrument considered:

$$\frac{dN}{dlog(D_p)} = \frac{dN}{log(D_{p,u}) - log(D_{p,l})} \tag{1}$$

with dN the PNC, $D_{p,u}$ the diameter of the upper boundary of the size bin and $D_{p,l}$ of the lower boundary of the size bin.

2.4.3. Redistribution of OPS Size Bins

The sensors and the TSI OPS 3330 measure the particle distribution using different numbers of size bins or different cut-off diameters. In this study, we re-calculated the size bins of the TSI OPS 3330 to match the size bins of each sensor tested. Overlapping size bins fractions are computed with the formulas used by Di Antonio et al. [44] and described in Figure 3 on a simple example. For this example, the equivalent TSI OPS 3330 size bin b_{eq}^{ops} is defined by:

$$b_{eq}^{ops} = b_0^{ops} * f_{low} + b_1^{ops} + b_2^{ops} * f_{upp} \tag{2}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

with

$$f_{low} = \frac{b_{upp} - b^{ops}_{0,low}}{b^{ops}_{0,upp} - b^{ops}_{0,low}} \quad (3)$$

and

$$f_{upp} = \frac{b^{ops}_{2,upp} - b_{low}}{b^{ops}_{2,upp} - b^{ops}_{2,low}} \quad (4)$$

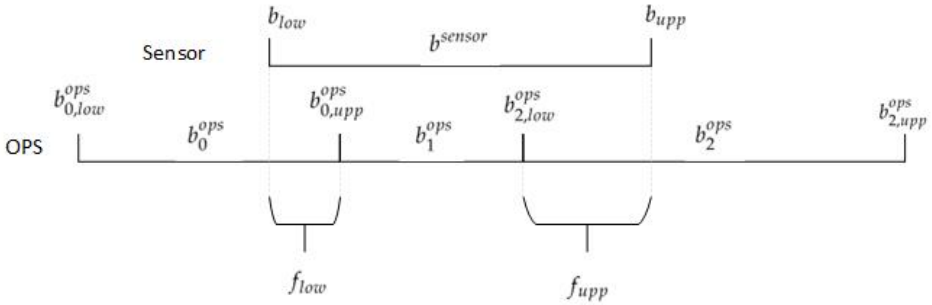


Figure 3. Principle of the redistribution of the size bins of the TSI OPS 3330 to match the size bins of each model of sensor with b^{sensor} the sensor size bin, b_{low} and b_{upp} the lower and upper cut size of the sensor size bin, b^{ops}_i the i corresponding size bins of the TSI OPS 3330, $b^{ops}_{i,low}$ and $b^{ops}_{i,upp}$ the lower and upper cut sizes of the i^{th} size bin of the TSI OPS 3330, f_{low} and f_{upp}, the lower and upper fractions of the size bins corresponding to b_{low} and b_{upp}.

2.4.4. Software and Data

The data was analysed using R 4.2.2 (R Foundation for Statistical Computing, Vienna, Austria) [45]. The underlying dataset is openly available at https://doi.org/10.5281/zenodo.7808620, and the code used for the analysis and to generate the tables and graphs of this study is openly available at https://doi.org/10.5281/zenodo.7808794. Boruta feature selection was conducted using the Boruta package [46]. RFE-SVM was conducted using the Caret package [47] using a 10 times repeated cross-validation and based on SVM radial [48]. Ridge was performed using the packages caret and glmnet [49] with a 10 times repeated cross-validation, and a grid search to optimise the penalisation coefficient *lambda* between 0 and 1.

3. Results

3.1. Correlation between the Different Size Bins

Tables 2–4 present the correlations obtained by the different models of sensors across the different size bins they report. For comparison, the same has been done for the equivalent size bins calculated from the TSI OPS 3330 readings and are presented between brackets in the same tables. This gives a baseline for the levels of correlation to expect in the actual size distribution of the particles measured. If the difference between correlation between the sensor bins and the correlation of the TSI OPS 3330 is > 0.15, we consider that the two considered bins of the sensors are not truly independent. If the difference in correlation is < 0.15, the bins of the sensors are considered independent.

For the Plantower PMS5003, the first bin gr03um is not independent from the other three size bins. gr05um, gr10um and gr25um are independent from each other. For the Alphasense OPC-R1, all the bins are independent from each other. For the Sensirion SPS30, the very high correlations obtained between this sensor and the TSI OPS 3330 makes it hard to conclude on the actual independence of the bins sizes and it must be noted that each successive size bin of the Sensirion SPS30 encompasses the following one (i.e., n1 is part of n05), this, together with the fact that the number of particles decreases

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

with increasing size means that the correlation observed here may be due to a mathematical artefact rather than to an actual description of the accuracy and precision of the Sensirion SPS30.

Table 2. Correlation and linear model between the different Particle Number Concentration (PNC) size bins reported by the Plantower PMS5003 during the period of the study. The numbers between brackets represent the correlation between equivalent size bins of the TSI OPS 3330. Cells are shaded in red if the difference between the correlation of the sensor and the TSI OPS 3330 is greater than 0.15. If the sensor bins accurately measured the size distribution, they should obtain a similar correlation to the TSI OPS 3330.

PMS5003		**Range**	**gr05um 0.5–10 µm**	**gr10um 1–10 µm**	**gr25um 2.5–10 µm**
R^2	gr03um	0.3–10 µm	0.998 (0.61)	0.83 (0.2461)	0.46(0.08)
R^2	gr05um	0.5–10 µm		0.85(0.771)	0.49(0.45)
R^2	gr10um	1–10 µm			0.75(0.85)

Table 3. Correlation and linear model between the different Particle Number Concentration (PNC) size bins reported by the Sensirion SPS30 during the period of the study. The numbers between brackets represent the correlation between equivalent size bins of the TSI OPS 3330. Cells are shaded in red if the difference between the correlation of the sensor and the TSI OPS 3330 is greater than 0.15. If the sensor bins accurately measured the size distribution, they should obtain a similar correlation to the TSI OPS 3330.

SPS30		**Range**	**n1 0.3–1 µm**	**n25 0.32.5 µm**
R^2	n05	0.3–0.5 µm	0.99(0.998)	0.98 (0.97)
R^2	n1	0.3–1 µm		0.997 (0.997)

Table 4. Correlation and linear model between the different Particle Number Concentration (PNC) size bins reported by the Alphasense OPC-R1 during the period of the study. The numbers between brackets represent the correlation between equivalent size bins of the TSI OPS 3330. Cells are shaded in red if the difference between the correlation of the sensor and the TSI OPS 3330 is greater than 0.15. If the sensor bins accurately measured the size distribution, they should obtain a similar correlation to the TSI OPS 3330.

OPCR1		**Range**	**Bin1 0.7–1.1 µm**	**Bin2 1.1–1.5 µm**	**Bin3 1.5–1.9 µm**	**Bin4 1.9–2.4 µm**
R^2	Bin0	0.35–0.7 µm	0.305 (0.426)	0.189 (0.173)	0.142 (0.091)	0.116 (0.055)
R^2	Bin1	0.7–1.1 µm		0.784 (0.802)	0.612 (0.637)	0.530 (0.528)
R^2	Bin2	1.1–1.5 µm			0.934 (0.956)	0.8765 (0.892)
R^2	Bin3	1.5–1.9 µm				0.956 (0.982)

3.2. Time Series of the Experiments

The time series presented in Figures 4–6 are focused on the experiment performed at 69% RH for brevity and the time series for the other experiments are available in Supplementary Figures A1–A12. They show similar results to Experiment 2. The first seven peaks correspond to the generation of peaks of candle-generated PM followed by stable concentrations of candle-generated PM, then a series of six peaks of incense-generated PM and a stable concentration of incense-generated PM. For this section, the bins sizes of the OPS have been converted to the size bins of each individual sensor. The y-axis follows a logarithmic scale.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

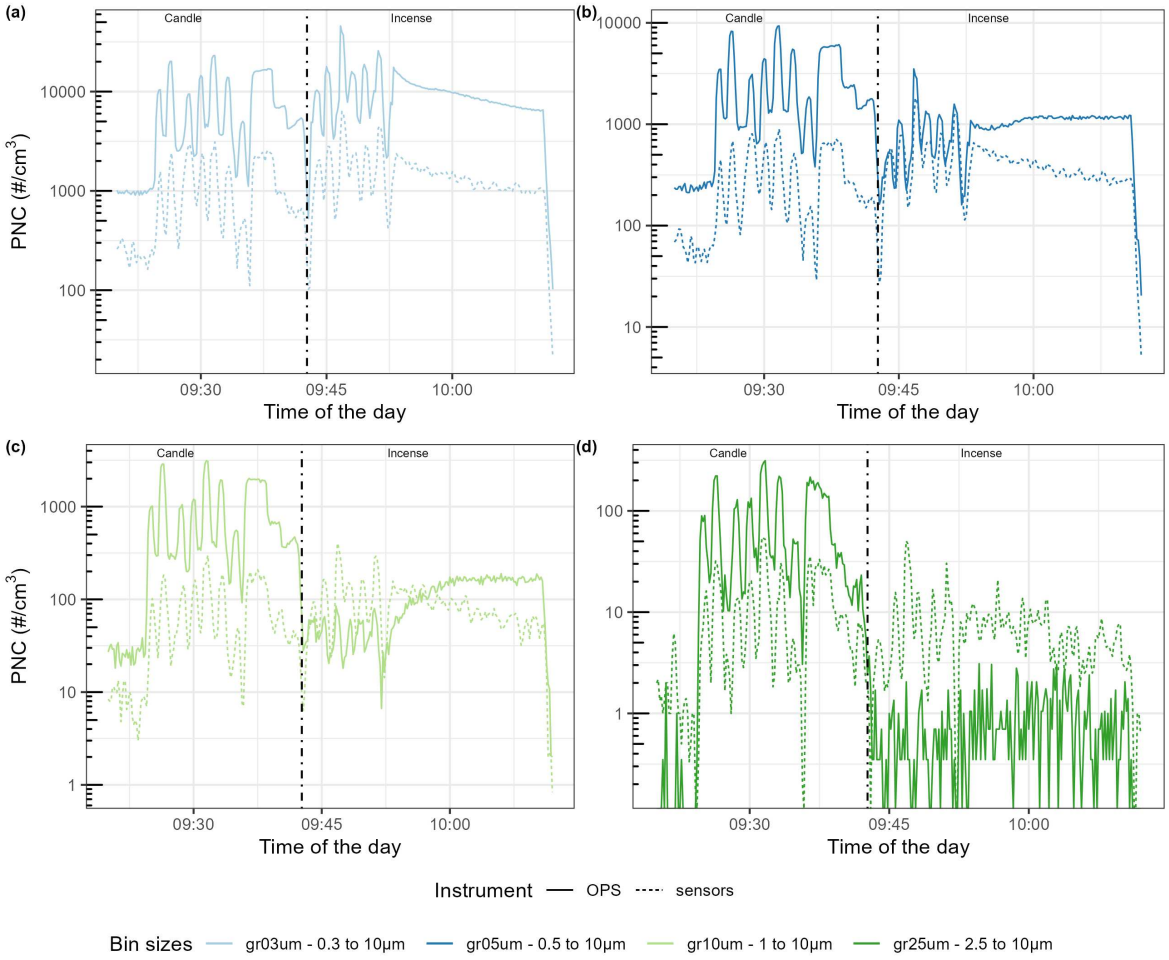


Figure 4. Time-series of the Particle Number Concentration (PNC) size bins of the Plantower PMS5003 compared with the size bins computed from the OPS size bins. On the left of the dotted line, PM were generated using candle, on the right using incense. The different categories are, from left to right and top to bottom, (**a**) gr03um (between 0.3–10 µm), (**b**) gr05um (between 0.5–10 µm), (**c**) gr10um (between 1–10 µm, (**d**) gr25um (between 2.5–10 µm).

For the three models of sensors, the time series of the different size bins closely follow the variation of the size bins of the TSI OPS 3330, for both sources of particles. For the Plantower PMS5003, for candle-generated PM, the magnitude of sensor-reported PNC is 10 times lower than the magnitude of the TSI OPS 3330, for all the bins of this sensor model. For incense-generated PM, the first bin (0.3–10 µm) is also 10 times lower than the magnitude of the TSI OPS 3330, but the second bin (0.5–10 µm) obtained the same magnitude than the TSI OPS 3330 for peaks, but underestimated PNC for the stable concentration of incense-generated PM. This bin also presents a downward slope for stable concentrations of incense-generated PM, which does not match the TSI OPS 3330 measurements. The third bin (1–10 µm), for incense-generated PM overestimated PNC for peak concentrations, but for the stable concentration, it starts by over-reporting before then under-reporting. For the fourth bin, for incense-generated PM, the sensor over-reports PNC but it is difficult to compare the measurements with the measurements of the TSI OPS 3330 given the high variance of the measurements from the TSI OPS 3330. For the Sensirion SPS30, the magnitude is 100 times lower for all the bins and little difference is observed between candle- and incense-generated PM. For the Alphasense OPC-R1, the magnitude of the first bin (from 0.35–0.7 µm) is about 10 times lower, this bin also presents a lot of variability that is not present in the measurement made by the TSI OPS 3330. For the other size bins of this sensor, the magnitude is similar to the TSI OPS 3330: for the size bin 0.7–1.1 µm, the Alphasense

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

OPC-R1 slightly over-reports PNC while for the three remaining size bins, it slightly under-reports PNC compared to the TSI OPS 3330.

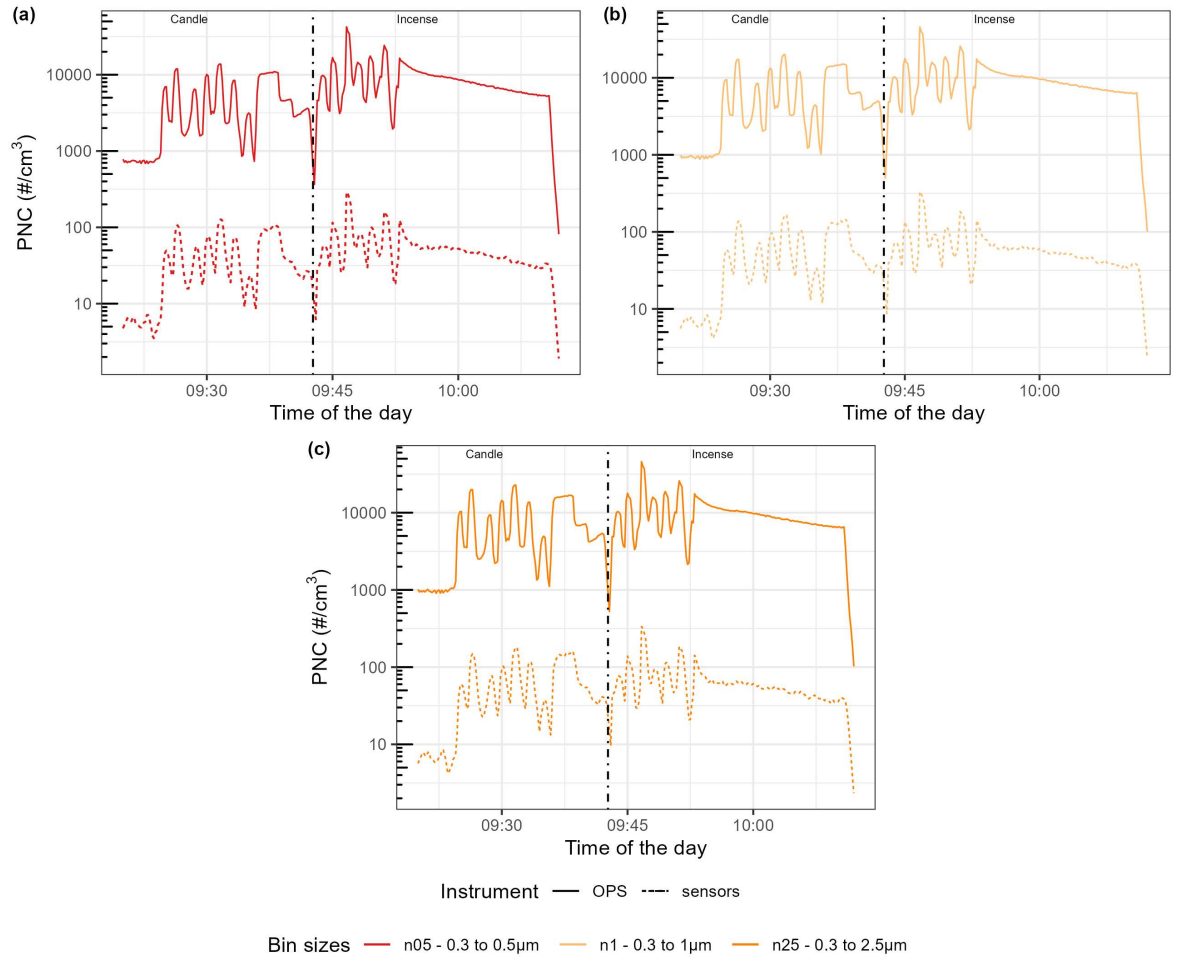


Figure 5. Time-series of the Particle Number Concentration (PNC) size bins of the Sensirion SPS30 compared with the size bins computed from the OPS size bins. On the left of the dotted line, PM were generated using candle, on the right using incense. The different categories are, from left to right and top to bottom, (**a**) n05 (between 0.3–0.5 µm), (**b**) n1 (between 0.3–1 µm), and (**c**) n25 (between 0.3–2.5 µm).

3.3. Features Selection

Tables 5–7 present the importance of the variables (source; PNC 0.01–0.3 µm; PNC 0.3–0.8 µm; PNC 0.3–10 µm; RH) for the different bins sizes of the sensors, computed by using the three methods described in the Methods section: Boruta, Ridge and RFE-SVM. High scores denote the relevance of the variable to explain the size bin considered. Each method computes their score differently, and the values obtained should not be compared between the methods.

For the Plantower PMS5003, RH was consistently given a score of zero or close except for its bin representing the largest particle size (2.5–10 µm) where its score was still low compared to the other variables considered. For the two first bins, the PNC for particles 0.01–0.3 µm was given the highest scores for the Boruta and Ridge method and a high score for the RFE-SVM method. The two remaining bins were given much lower scores for this variable for the three methods. The source of the PM was given relatively low scores for the first two bins, scores close to zero for the third bin but high scores for Boruta and Ridge for the last bin and a score close to zero for RFE-SVM for the last bin. PNC 0.3–0.8 µm and 0.3–10 µm were given high scores for all methods and all bins.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

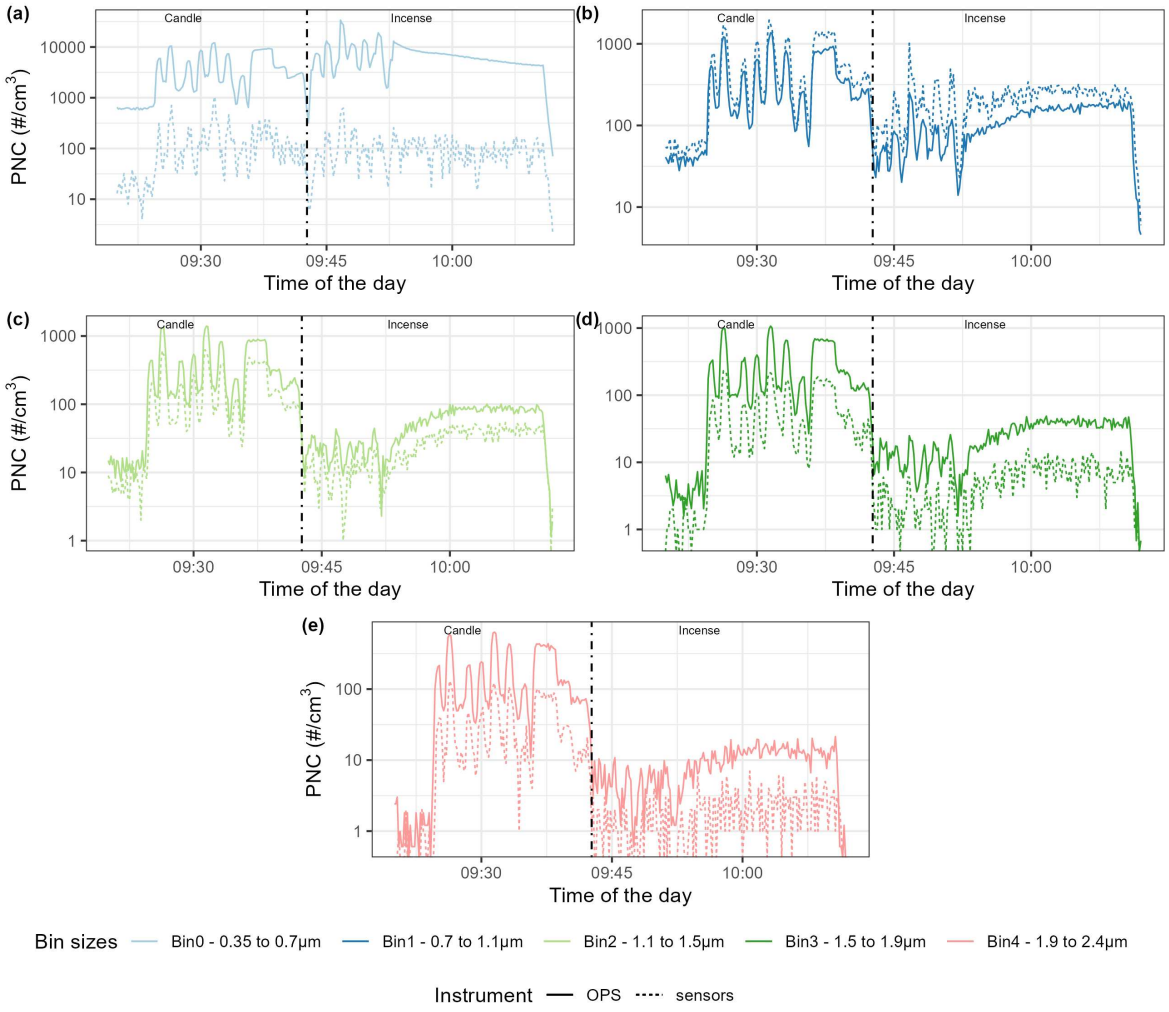


Figure 6. Time-series of the Particle Number Concentration (PNC) size bins of the Alphasense OPC-R1 compared with the size bins computed from the OPS size bins. On the left of the dotted line, PM were generated using candle, on the right using incense. The different categories are, from left to right and top to bottom, (**a**) Bin0 (between 0.4–0.7 µm), (**b**) Bin1 (between 0.7–1.1 µm), (**c**) Bin2 (between 1.1–1.5 µm), (**d**) Bin3 (between 1.5–1.9 µm, and (e) Bin4 (between 1.9–2.4 µm).

For the Sensirion SPS30, similarly, RH was given scores of zero or close to zero on its three size bins for all three methods. Source was given similar scores apart for Ridge on the first size bin, with a score of 22, which is small compared to the score given for PNC 0.3–0.8 µm and 0.3–10 µm for this method and size bin. PNC 0.3–0.8 µm and 0.3–10 µm were given the highest scores for all methods and all bins. PNC for particles 0.01–0.3 µm was given a score close to zero on the three bins for Ridge selection and a relatively low score for the other two methods.

For the Alphasense OPC-R1, RH was given low scores, of zero or close to zero for most bins and most methods except for Bin1 for Boruta and Ridge for which it was given moderate scores. Source was given the highest score for Bin2 to Bin4 for the three methods, the highest score for Bin1 for Boruta and Ridge and a moderate score for RFE-SVM. The first bin was given lower scores for source. PNC for particles 0.01–0.3 µm was given relatively low scores for all methods for all size bins. 0.3–0.8 µm and 0.3–10 µm was given moderate to high scores on all bins for the three methods considered.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

Table 5. Scores of the features selection methods for the different size bins of the Plantower PMS5003 computed using (a) Boruta, (b) Ridge, (c) RFE-SVM. High scores denote the relevance of the variable to explain the size bin. Each method computes the score differently and the values obtained should not be compared between the methods. PNC 0.01–0.3 µm is measured by the Nanotracer and the other two size fractions are measured by the TSI OPS 3330.

PMS5003	**gr03um 0.3–10µm**			**gr05um 0.5–10µm**			**gr10um 1–10µm**			**gr25um 2.5–10µm**		
Method	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Source	17	31	14	17	32	13	7	10	1	39	71	7
PNC 0.01–0.3 µm	47	100	49	46	100	47	15	20	21	15	0	5
PNC 0.3–0.8 µm	27	78	64	30	87	66	29	87	79	28	42	48
PNC 0.3–10 µm	28	68	61	29	76	63	35	100	80	36	100	53
RH	9	0	0	8	0	0	7	0	0	15	18	2

Table 6. Scores of the features selection methods for the different size bins of the Sensirion SPS30 computed using (a) Boruta, (b) Ridge, (c) RFE-SVM. High scores denote the relevance of the variable to explain the size bin. Each method computes the score differently and the values obtained should not be compared between the methods.

SPS30	**n05 0.3–0.5 µm**			**n1 0.3–1 µm**			**n25 0.3–2.5 µm**		
Method	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Source	6	22	4	6	7	1	7	2	1
PNC 0.01–0.3 µm	17	2	22	15	4	19	15	2	18
PNC 0.3–0.8 µm	31	100	79	30	100	75	29	96	71
PNC 0.3–10 µm	32	84	79	35	96	76	35	100	74
RH	10	0	0	12	0	0	13	0	1

Table 7. Scores of the features selection methods for the different size bins of the Alphasense OPC-R1 computed using (a) Boruta, (b) Ridge, (c) RFE-SVM. High scores denote the relevance of the variable to explain the size bin. Each method computes the score differently and the values obtained should not be compared between the methods.

OPC-R1	**Bin0 0.35–0.7 µm**			**Bin1 0.7–1.1 µm**			**Bin2 1.1–1.5 µm**			**Bin3 1.5–1.9 µm**			**Bin4 1.92.4 µm**		
Method	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
Source	17	22	10	35	100	16	74	100	35	69	100	36	69	100	35
PNC 0.01–0.3 µm	12	3	23	10	1	3	14	15	5	13	12	5	13	10	5
PNC 0.3–0.8 µm	35	100	98	20	0	32	23	33	15	23	34	13	21	35	13
PNC 0.3–10 µm	29	85	96	21	81	34	33	95	16	32	95	15	32	96	14
RH	5	0	0	15	46	7	8	0	0	5	0	0	4	0	0

3.4. Particle Size Distribution

Figure 7 presents the size distribution measured by the sensors and the TSI OPS 3330 during stable concentrations of candle and incense-generated PM. The data are averaged per sensor model and is average over the five sets of experiments. The different size distributions recorded by the TSI OPS 3330 in each experiment are presented in Supplementary Figure A13. To facilitate the visualisation, the data from the Sensirion SPS30 has between multiplied by 100 and the data from the Plantower PMS5003 by 10.

For the TSI OPS 3330, the incense-generated PM shows a relatively steeper decrease of PNC with increasing sizes and higher PNC with candle-generated PM for particles >0.5 µm than for incense-generated PM.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

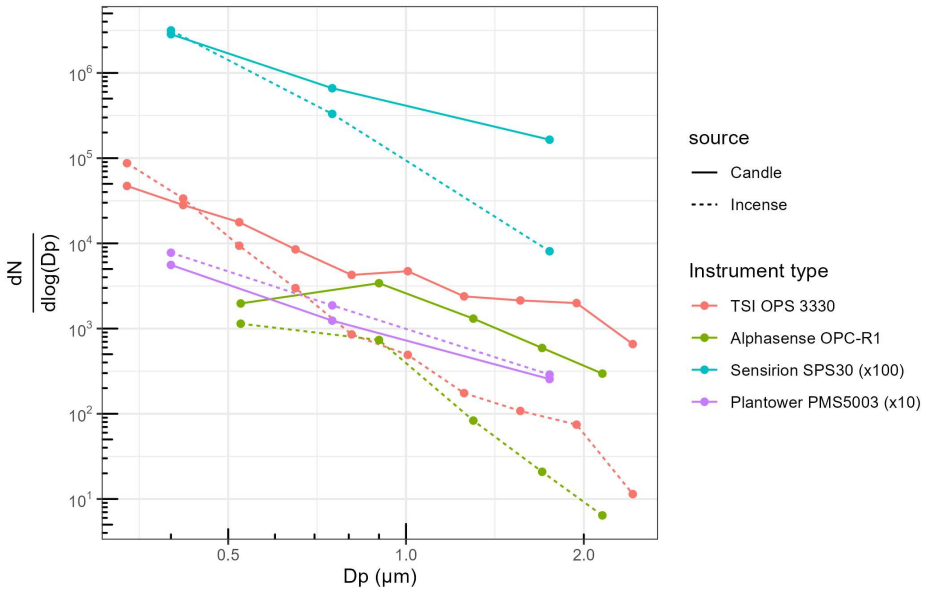


Figure 7. Particle size distribution reported by the sensors and the OPS TSI for stable concentrations of incense and candle-generated PM, aggregated over the five experiments and per sensor model. For each sensor model, the data presented is the average per size bin. The dots are on the midpoint of the size bins. To facilitate the visualisation, the data from the Sensirion SPS30 has between multiplied by 100 and the data from the Plantower PMS5003 by 10. The dots represent the actual datapoints used to construct the plot.

The Plantower PMS5003 reports almost the same distribution for candle and incense-generated PM and surprisingly reports higher numbers for incense-generated PM than for candle-generated PM. For candle-generated PM, the magnitude is off by a factor 10. For incense-generated PM, the Plantower PMS5003 over-reports PNC for particles $\approx <0.6$ µm and under-reports for particles of wider diameter. The Sensirion SPS30 presents the same steeper decrease than the TSI OPS 3330 between candle and incense-generated PM and detects fewer particles >0.7 µm for incense-generated PM than for candle-generated PM. As in the time series, the magnitude of this sensors differs by a factor of 100. For the Alphasense OPC-R1, the steeper decrease between incense and candle-generated PM is also present. The magnitude of the first size bin is much lower for the sensor than for the TSI OPS 3330 for both sources of PM but the other sizes follow each other with almost similar magnitude.

The readings of the individual sensors are available in Appendix Figures A14–A16 and limited variability is observed between units of the same model of sensor for each size bins although the Alphasense OPC-R1 demonstrated a higher variability for its first size bin.

4. Discussion

The particle size distribution showed that Plantower PMS5003 did not capture the difference in size distribution between candle and incense smoke. Incense smoke had clearly fewer particles >0.5 µm than candle smoke, according to the TSI OPS 3330, however, while this was not captured by the Plantower PMS5003, it was captured by the Sensirion SPS30 to a certain extent, and more clearly by the Alphasense OPC-R1. The size distribution captured by the latter sensor is close to the size distribution measured by the TSI OPS 3330, except for its first size bin (0.35–0.7 µm). Ouimette et al. [16] found that the Plantower PMS5003 behaved as an integrating nephelometer, reporting correctly the aerosol scattering coefficient for particles between 0.26–0.46 µm. Using the global database of PurpleAir sensors, they also showed that the Plantower PMS5003 obtained similar shape of size distribution in different sites outdoors around the world, this being partly attributed to the fact that aerosol scattering coefficient of outdoor PM is generally constant. Similarly, He et al. [15] showed that the different size bins of the Plantower PMS5003 had cut-off diameters between 0.1–0.7 µm, when

testing the sensors with ammonium sulfate and sodium chloride. Together, these support the claim that the size distribution is computed by an algorithm rather than actually measured for each size bins. Tryner et al. [13] studied the size segregation of the mass concentration reported by the Plantower PMS5003, and the Sensirion SPS30 in an environmental chamber and exposed eight of each of these sensor models to stable concentrations of PM between 10–1000 $\mu g/m^3$ lasting around 45 min each and generated using ammonium sulfate, Arizona road dust, NIST urban PM, wood smoke and oil mist and PLS of different diameters (0.1, 0.27, 0.72 and 2 µm). They found that the Plantower PMS5003 obtained similar shape of size distribution for all the diameters of PLS and for the different sources of pollution. However, Kuula et al. [17] exposed Plantower PMS5003 and Sensirion SPS30 sensors, amongst other sensor models, to monodisperse particles of diameters between 0.45–9.8 µm and they showed that while the Plantower PMS5003 misclassified the size of the particles, it was still producing two different signals, one for particles 0.3–2.5 µm and one for particles 2.5–10 µm. Similarly, Zamora et al. [18] exposed the sensor to PLS of 0.081, 0.3, 0.8, 1.1, 2.5 and 4.8 µm and while the Plantower PMS5003 misclassified and misreported the size distribution of the particle measured, it showed some differences in the size distribution it reported between the different diameters tested. These suggest that the algorithm used by the Plantower PMS5003 may include a second measurement to compute the size distribution it reports. This is further supported by: (1) the between size bins correlations obtained here by the sensor; and (2) the Plantower PMS5003 differences between the scores obtained for PNC $<$0.3 µm for its first two bins and the two remaining bins.

The Sensirion SPS30 captured some of the variations in the size distribution between incense and candle-generated PM. Tryner et al. [13] obtained two different sizes, using sensor-reported PM mass concentrations, for PLS particles of 0.1 and 0.27 µm and of 0.72 and 2 µm, conversely to the Plantower PMS5003. Nonetheless, in their study, the Sensirion SPS30 did not agree with the Aerodynamic Particle Sizer Spectrometer that was used as a reference instrument. Kuula et al. [17], again using sensor-reported PM mass concentrations, suggested that this sensor was able to differentiate two different size ranges, 0.3–0.9 µm and 0.7–1.3 µm, with a valid detection range for PM_1 mass concentration. Given the fact that the size bins reported by this sensor are inherently correlated (the size bins are 0.3–0.5 , 0.3–1 , and 0.3–2.5 µm so the first size bin mostly drives the value they report), it was not possible to differentiate between them in this study. Therefore, the results of this study cannot corroborate or invalidate the results obtained by the above mentioned studies.

The Alphasense OPC-R1 showed clear differences between its size bins, which generally followed the correlation obtained by the TSI OPS 3330, for both PM sources. In our 2020 study [25], which analyses the same set of experiments but focuses on sensor-reported mass concentrations, the Alphasense OPC-R1 obtained lower correlation coefficients between the mass concentration of the sensors and a DustTrak DRX 8533 for incense than for candle-generated PM. This can have several explanations: (1) the algorithm used by the Alphasense OPC-R1 to convert PNC to mass concentration is based on factors, which differ according to specific properties of the particles measured; (2) there are some differences in the measurement taken by the TSI OPS 3330 and the DustTrak DRX 8533, which is unlikely as they are based on the same technology and are made by the same manufacturer.

For the three models of sensors, the variables that impacted their readings the most were the variables linked to the particle size distribution, followed by source and with RH having less impact on the readings. This corroborates the suggestion in our earlier paper et al. [25] that the performances of the sensors are primarily impacted by the size distribution of the particles, and secondly the source of those particles. This also means that at RH < 79%, this variable does not need to be corrected. This would need to be verified on different sources of PM, especially with sources having a different size distribution and refractive index than candle and incense. Jayaratne et al. [50] showed that the Plantower PMS1003 and a DustTrak DRX8530 started to over-report PM mass concentration for RH>75%. Tryner et al. [13] also found an impact of RH on the readings of the Plantower PMS5003 and the Sensirion SPS30 for RH > 75-80%.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

We found here that the accuracy of the readings of the first two size bin of the Plantower PMS5003 were impacted by the PNC <0.3 µm. This is similar to the results obtained by both He et al. [15], who developed a transfer-function based model that predicted that the sensor will output a signal for particles with diameter <0.3 µm, and by Ouimette et al. [16] whose physical model of the Plantower PMS5003 as an integrating nephelometer, based on the Mie theory, also predicted that the sensor would be able to measure particles <0.3 µm, in direct proportion to their contribution to the aerosol scattering coefficient. The Alphasense OPC-R1 and the Sensirion SPS30 readings were not impacted by the PNC in that size range. While it is quite clear that the Alphasense OPC-R1 is an Optical Particle Counter (OPC), this, along with the differences observed earlier on the sensitivity to particle size distribution between the Plantower PMS5003 and the Sensirion SPS30, may suggest that the Sensirion SPS30 is not an integrated nephelometer and/or measures and interprets the PNC differently from the Plantower PMS5003. Tryner et al. [13] also suggested that these two sensors had a different method for measuring or interpreting light-scattering data.

Although the three models of sensors were able to capture the temporal variations of the PNC, as reported by the TSI OPS 3330, the only sensor that reliably reported the particle size distribution of the aerosols was the Alphasense OPC-R1.

5. Conclusions

In this study, eight sensors of each of three models, Alphasense OPC-R1, Plantower PMS5003 and Sensirion SPS30, for a total of 24 sensors, were studied at a 10 s resolution and exposed to short-lived events of PM pollution, generated from two combustion sources having a different size distribution profile, at varying levels of RH.

The time-series obtained revealed that the sensors were able to closely follow PNC variation measured by the reference TSI OPS 3330, for both sources of PM, but for the Plantower PMS5003 and the Sensirion SPS30 PNC measurements recorded were respectively 10 and 100 times lower than the measurements of the TSI OPS 3330. The magnitude was correct for the Alphasense OPC-R1, except for its first bin, between 0.35–0.7 µm.

Regarding the independence of the size bin reported by the sensors and their accuracy, the Plantower PMS5003 reports two independent signals; the Alphasense OPC-R1 reports independent signal for each of its size bins; for the Sensirion SPS30 the inherent correlations of its size bins did not allow to verify their independence. The analysis conducted suggested that the Plantower PMS5003 and the Sensirion SPS30 have a different method for measuring or interpreting light-scattering data and reporting the determined PNC.

For capturing the particle size distribution, the Plantower PMS5003 showed no difference between incense and candle-generated PM, while the two other sensors recorded differences that were also recorded by the TSI OPS 3330. The Alphasense OPC-R1 measured values that were close to the data reported by the TSI OPS 3330, except for its first bin (0.35–0.7 µm).

The analysis of the features selection revealed that the sensors are more susceptible to the composition of the particles and their size distribution than to RH at the levels of humidity considered. We therefore recommend that a RH correction is not required below 75-79%. PNC between 0.01–0.3 µm impacted the first size bin of the Plantower PMS5003 supporting the fact that this sensor is an integrating nephelometer.

For studies requiring more detailed knowledge of the particle size distribution of the aerosol measured, the Alphasense OPC-R1 should be preferred, although our previous study [25] also showed that this sensor was less suited to report PM mass concentration for the two sources of PM used here. If a more general image of the particle size distribution is sufficient, the Sensirion SPS30 should be considered. It is not clear from the results of this study if the Plantower PMS5003 can be used in this scenario but it can be used to measure the general trends of PNC. This work shows that there is added value in directly using the PNC instead of PM mass concentration, as the size bins provide some level of information about the particle size distribution, especially in the case of the Alphasense OPC-R1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

This differential information collected by the PNC size bins can be used to improve the calibration models develop to calibrate the sensors to standard performances and provide extra granularity with regard to source profiling.

Author Contributions: Conceptualization, F.M.J.B., H.S.R., M.R., M.S.J., S.J.O., A.K.R.M., P.J.B., N.H.C.E., H.L.M., G.L.F., M.L. and S.J.C.; Data curation, F.M.J.B.; Formal analysis, F.M.J.B. and M.L.; Funding acquisition, F.M.J.B., M.S.J., S.J.O., A.K.R.M., G.L.F., M.L. and S.J.C.; Investigation, F.M.J.B., H.S.R., M.R. and M.S.J.; Methodology, F.M.J.B., H.S.R., M.R., M.S.J., S.J.O., A.K.R.M., P.J.B., N.H.C.E., H.L.M., G.L.F., M.L. and S.J.C.; Project administration, M.S.J.; Resources, F.M.J.B., H.S.R., M.R., M.S.J. and A.K.R.M.; Software, F.M.J.B., S.J.O. and P.J.B.; Supervision, M.S.J., S.J.O., A.K.R.M., G.L.F., M.L. and S.J.C.; Visualization, F.M.J.B.; Writing—original draft, F.M.J.B.; Writing—review & editing, F.M.J.B., H.S.R., M.R., M.S.J., S.J.O., A.K.R.M., P.J.B., N.H.C.E., H.L.M., G.L.F., M.L. and S.J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Next Generation of Unmanned Systems Centre for Doctoral Training supported by the Natural Environmental Research Council grant number [NE/N012070/1]; the Leverhulme Trust through the Southampton Marine and Maritime Institute; Engineering and Physical Sciences Research Council UK grant [EP/T517859/1]. Dr Matthew Loxham is supported by a BBSRC Future Leader Fellowship, grant number [BB/P011365/1], and an NIHR Southampton Biomedical Research Centre Senior Fellowship [BB/V004573/1]. Hugo S. Russell was supported by Airscape, Aarhus University Graduate School of Science and Technology (GSST), and BERTHA - the Danish Big Data Centre for Environment and Health funded by the Novo Nordisk Foundation Challenge Programme (grant NNF17OC0027864). The test chamber at the University of Copenhagen is supported by ACTRIS-DK. The APC was funded by the Engineering and Physical Sciences Research Council.

Data Availability Statement: The underlying dataset are openly available at https://doi.org/10.5281/zenodo.7808620, and the code used for the analysis and to generate the tables and graphs of this study is openly available at https://doi.org/10.5281/zenodo.7808794.

Conflicts of Interest: M.J. and H.R. receive part of their salary from Airscape, London, UK, which builds and sells low cost sensor nodes. No Airscape equipment was used in this study.

Appendix A. Time Series

Appendix A.1. RH = 54%

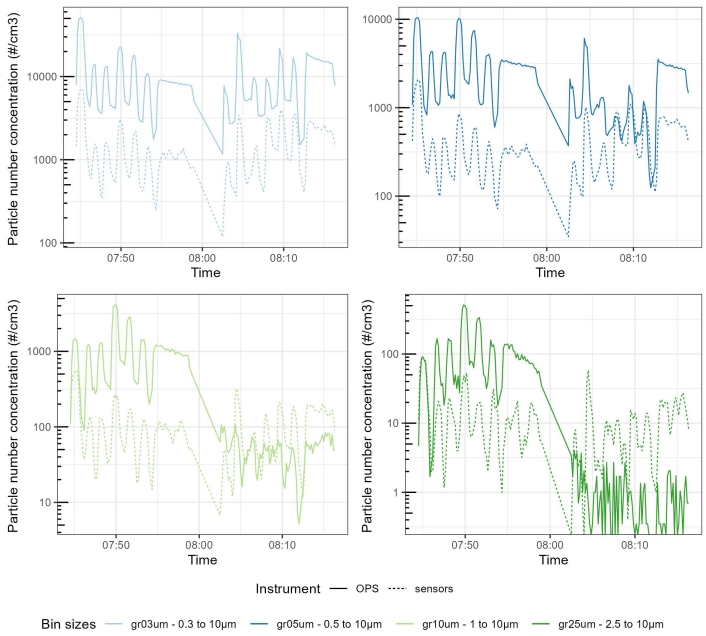


Figure A1. Time-series of the Particle Number Concentration (PNC) size bins of the Plantower PMS5003 compared with the size bins computed from the OPS size bins for RH=54%. The different categories are gr03um (between 0.3–10 µg/m^3), gr05um (between 0.5–10 µg/m^3), gr10um (between 1–10 µg/m^3, gr25um (between 25–10 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

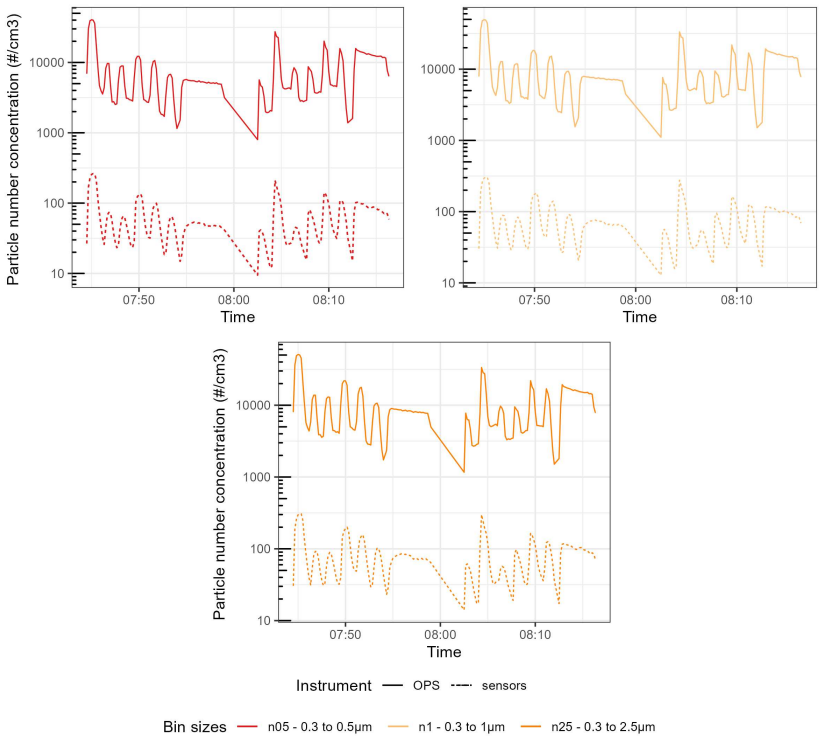


Figure A2. Time-series of the Particle Number Concentration (PNC) size bins of the Sensirion SPS30 compared with the size bins computed from the OPS size bins for RH=54%. The different categories are n05 (between 0.3–0.5 µg/m^3), n1 (between 0.3–1 µg/m^3), and n25 (between 0.3–2.5 µg/m^3).

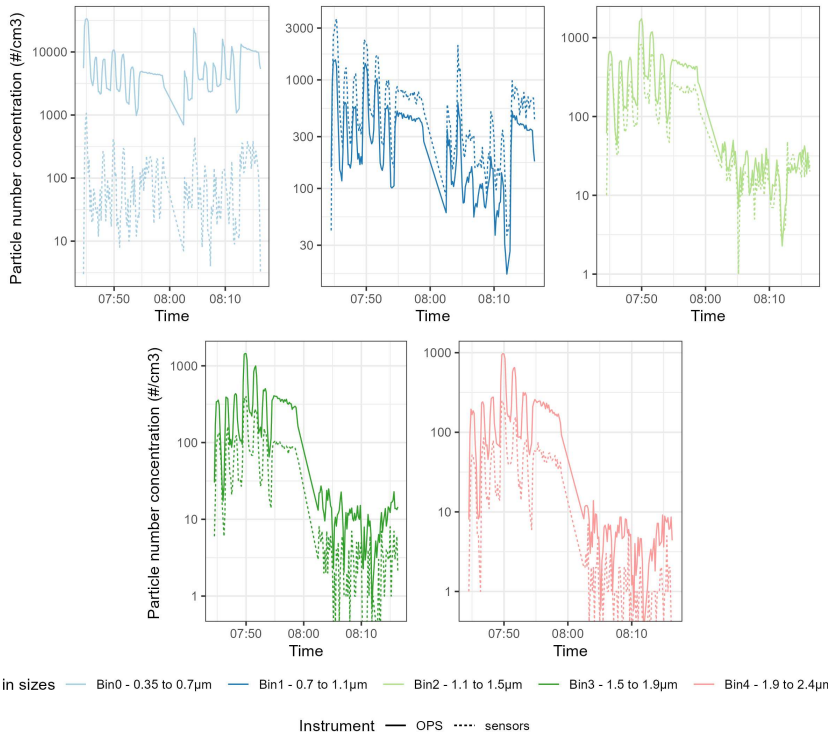


Figure A3. Time-series of the Particle Number Concentration (PNC) size bins of the Alphasense OPC-R1 compared with the size bins computed from the OPS size bins for RH=54%. The different categories are Bin0 (between 0.4–0.7 µg/m^3), Bin1 (between 0.7–1.1 µg/m^3), Bin2 (between 1.1–1.5 µg/m^3), Bin3 (between 1.5–1.9 µg/m^3, and Bin4 (between 1.9–2.4 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

Appendix A.2. RH = 72%

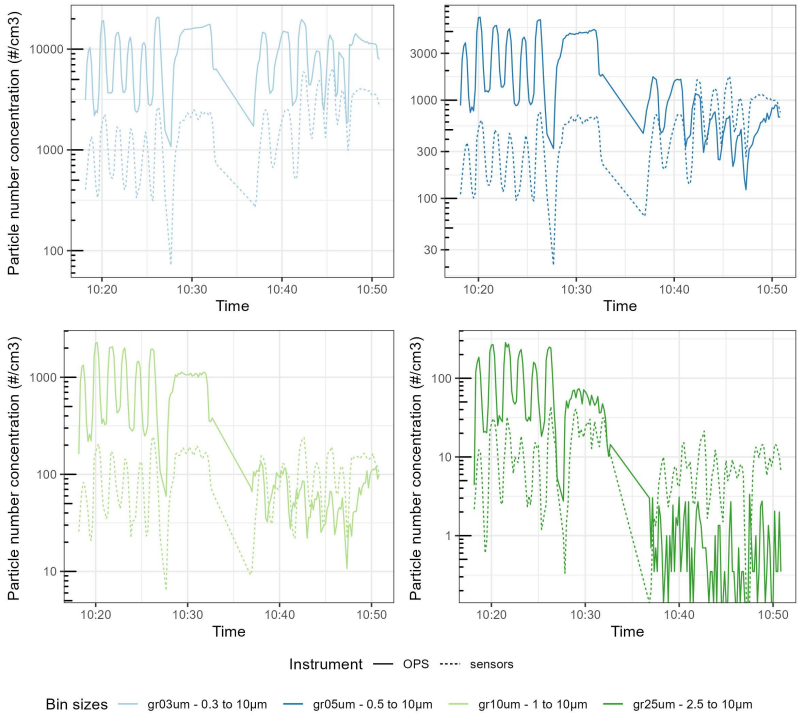


Figure A4. Time-series of the Particle Number Concentration (PNC) size bins of the Plantower PMS5003 compared with the size bins computed from the OPS size bins for RH=72%. The different categories are gr03um (between 0.3–10 µg/m^3), gr05um (between 0.5–10 µg/m^3), gr10um (between 1–10 µg/m^3, gr25um (between 25–10 µg/m^3).

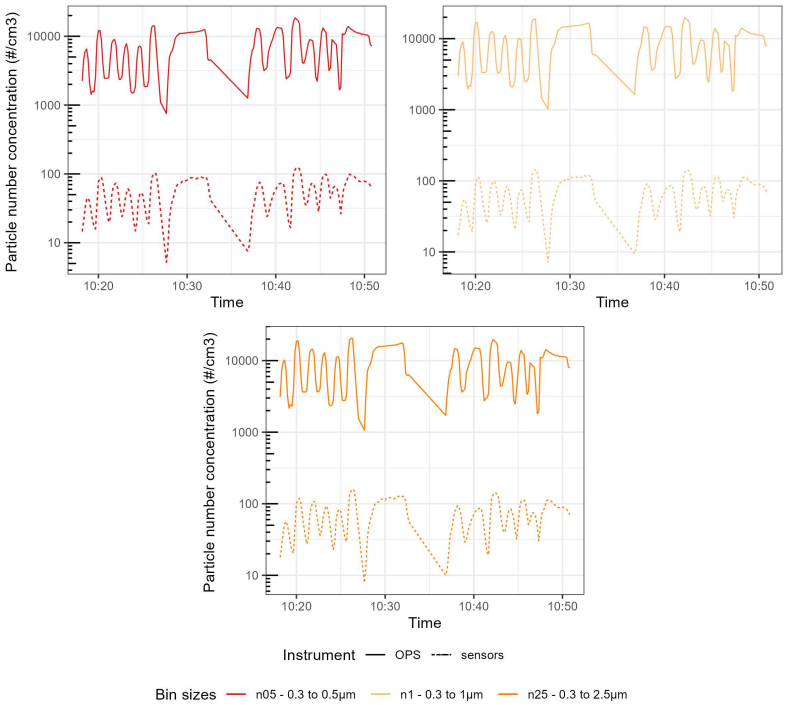


Figure A5. Time-series of the Particle Number Concentration (PNC) size bins of the Sensirion SPS30 compared with the size bins computed from the OPS size bins for RH=72%. The different categories are n05 (between 0.3–0.5 µg/m^3), n1 (between 0.3–1 µg/m^3), and n25 (between 0.3–2.5 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

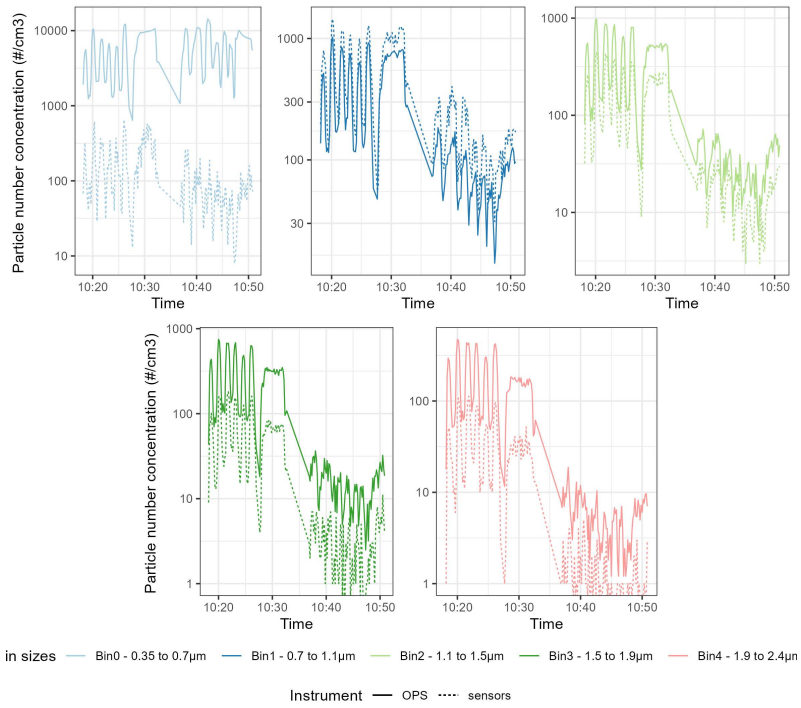


Figure A6. Time-series of the Particle Number Concentration (PNC) size bins of the Alphasense OPC-R1 compared with the size bins computed from the OPS size bins for RH=72%. The different categories are Bin0 (between 0.4–0.7 µg/m^3), Bin1 (between 0.7–1.1 µg/m^3), Bin2 (between 1.1–1.5 µg/m^3), Bin3 (between 1.5–1.9 µg/m^3, and Bin4 (between 1.9–2.4 µg/m^3).

Appendix A.3. RH = 76%

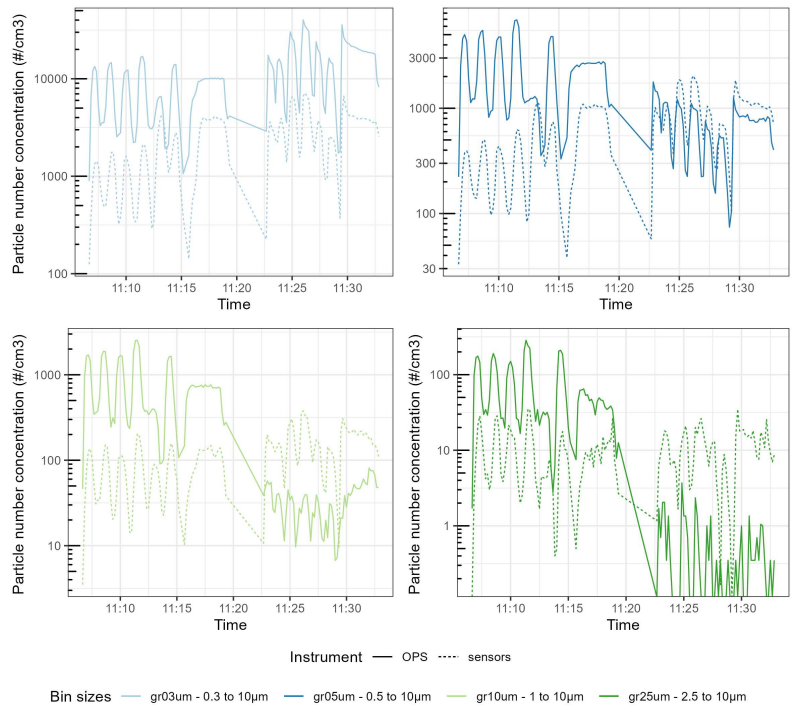


Figure A7. Time-series of the Particle Number Concentration (PNC) size bins of the Plantower PMS5003 compared with the size bins computed from the OPS size bins for RH=76%. The different categories are gr03um (between 0.3–10 µg/m^3), gr05um (between 0.5–10 µg/m^3), gr10um (between 1–10 µg/m^3, gr25um (between 25–10 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

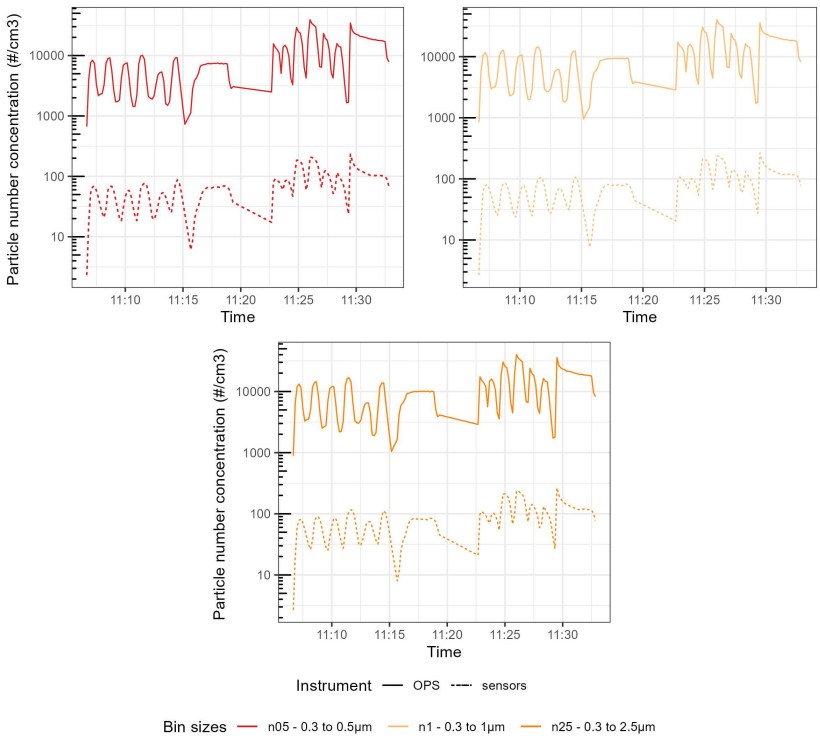


Figure A8. Time-series of the Particle Number Concentration (PNC) size bins of the Sensirion SPS30 compared with the size bins computed from the OPS size bins for RH=76%. The different categories are n05 (between 0.3–0.5 µg/m^3), n1 (between 0.3–1 µg/m^3), and n25 (between 0.3–2.5 µg/m^3).

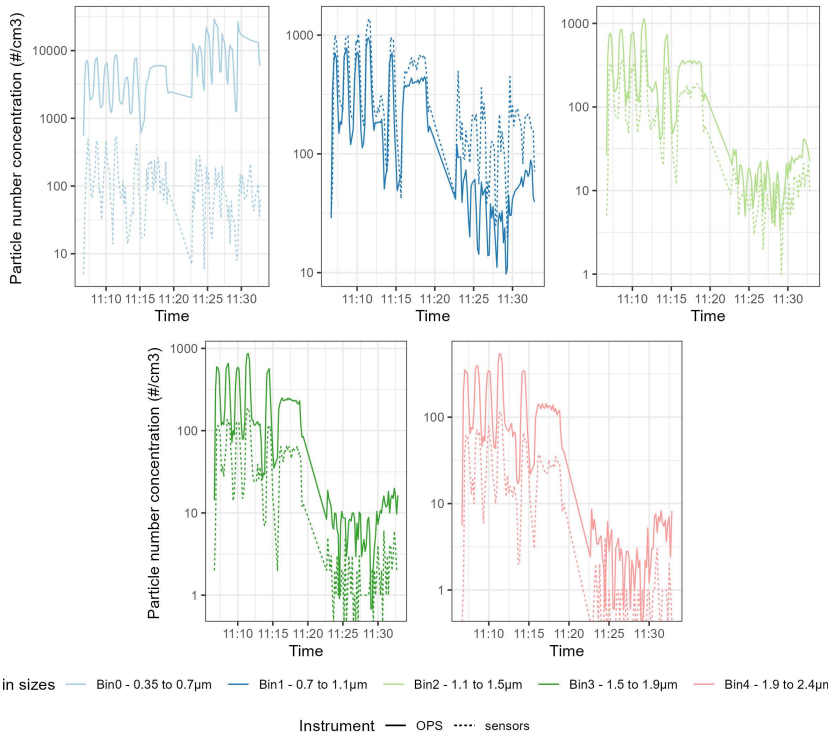


Figure A9. Time-series of the Particle Number Concentration (PNC) size bins of the Alphasense OPC-R1 compared with the size bins computed from the OPS size bins for RH=76%. The different categories are Bin0 (between 0.4–0.7 µg/m^3), Bin1 (between 0.7–1.1 µg/m^3), Bin2 (between 1.1–1.5 µg/m^3), Bin3 (between 1.5–1.9 µg/m^3, and Bin4 (between 1.9–2.4 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

Appendix A.4. RH = 79%

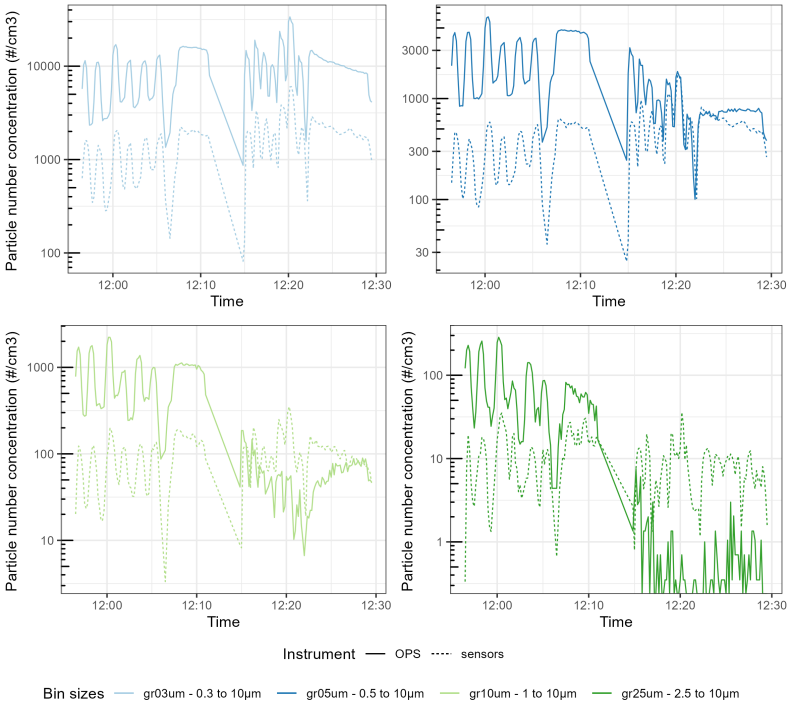


Figure A10. Time-series of the Particle Number Concentration (PNC) size bins of the Plantower PMS5003 compared with the size bins computed from the OPS size bins for RH=79%. The different categories are gr03um (between 0.3–10 µg/m^3), gr05um (between 0.5–10 µg/m^3), gr10um (between 1–10 µg/m^3, gr25um (between 25–10 µg/m^3).

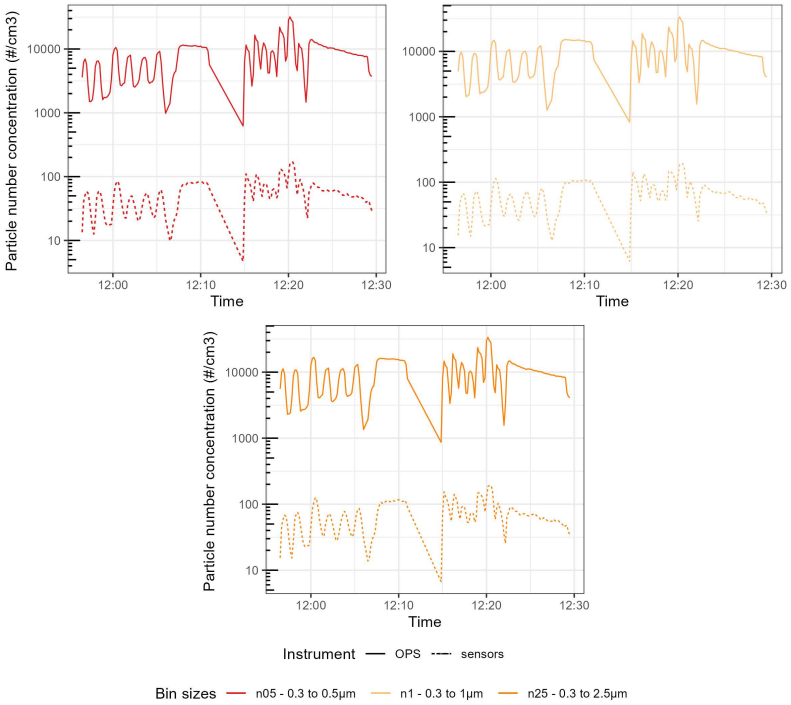


Figure A11. Time-series of the Particle Number Concentration (PNC) size bins of the Sensirion SPS30 compared with the size bins computed from the OPS size bins for RH=79%. The different categories are n05 (between 0.3–0.5 µg/m^3), n1 (between 0.3–1 µg/m^3), and n25 (between 0.3–2.5 µg/m^3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

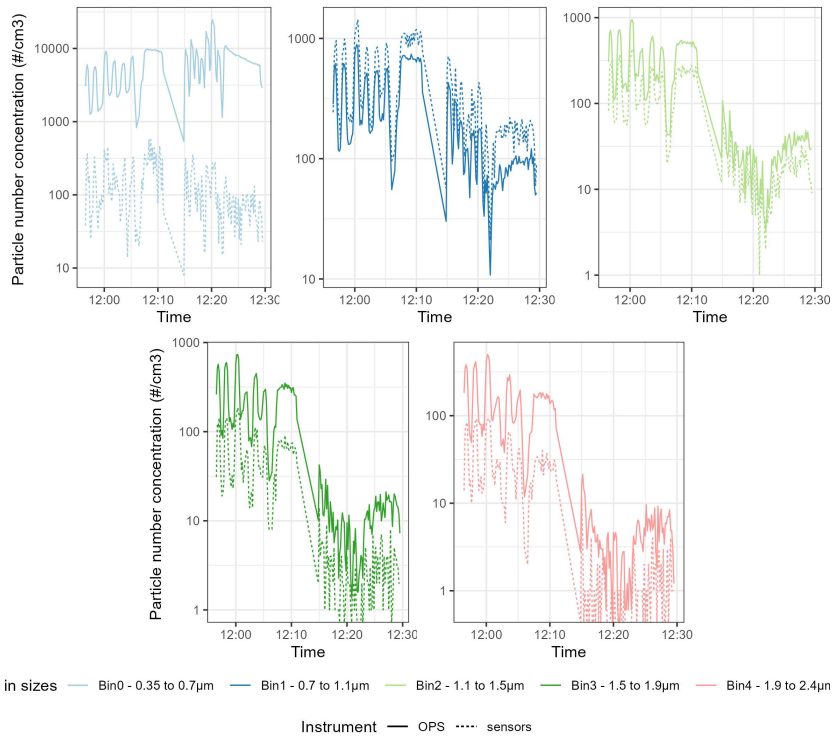


Figure A12. Time-series of the Particle Number Concentration (PNC) size bins of the Alphasense OPC-R1 compared with the size bins computed from the OPS size bins for RH=79%. The different categories are Bin0 (between 0.4–0.7 µg/m^3), Bin1 (between 0.7–1.1 µg/m^3), Bin2 (between 1.1–1.5 µg/m^3), Bin3 (between 1.5–1.9 µg/m^3, and Bin4 (between 1.9–2.4 µg/m^3).

Appendix B. Particle Size Distribution

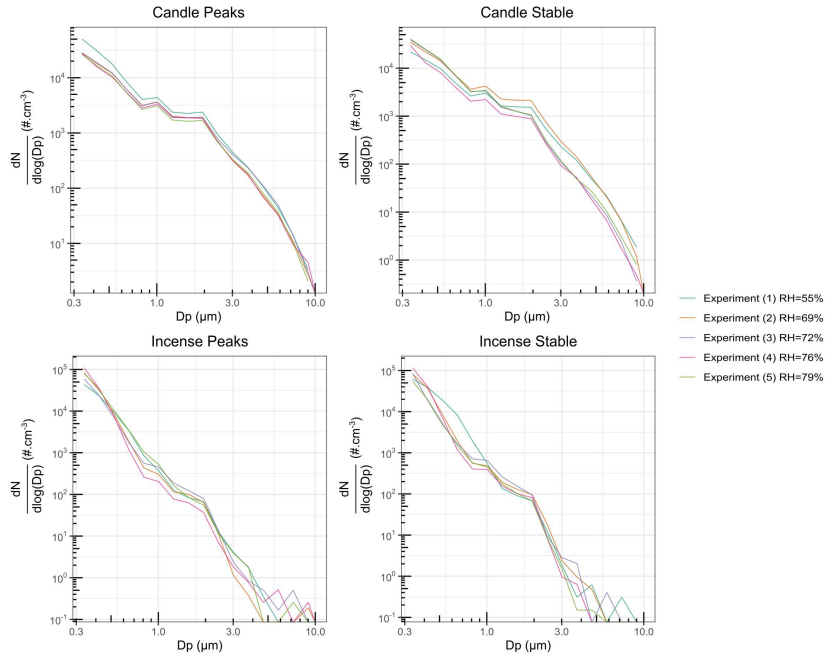


Figure A13. Size distribution measured from the OPS during the different experiments for peaks and stable concentrations of incense and candle smoke. The axis are in a logarithmic scale. N is the number of particles in # 1/cm, D_p is the mean diameter of the particles. Adapted from Bulot et al. [25].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 April 2023 doi:10.20944/preprints202304.0701.v1

Appendix B.1. Particle Size Distribution per Sensor

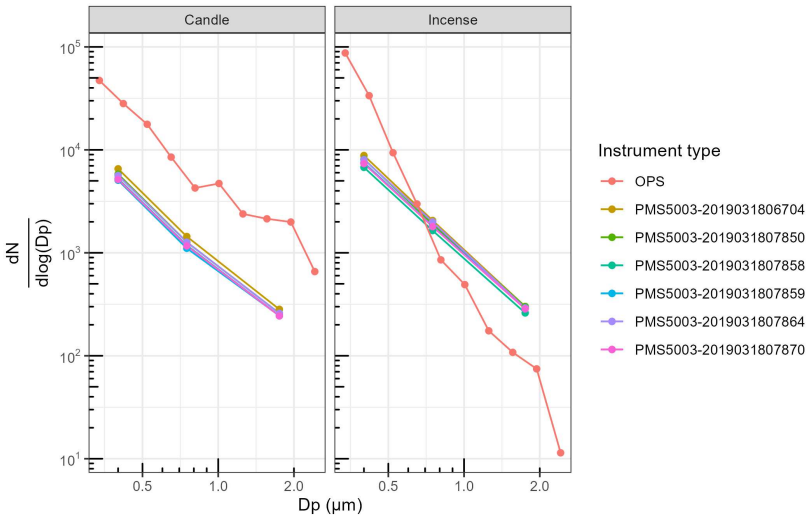


Figure A14. Particle size distribution reported by the PMS5003 and the OPS TSI for stable concentrations of incense and candle. The dots represent the actual datapoints used. To facilitate the visualisation, the data from the Plantower PMS5003 were multiplied by a factor 10.

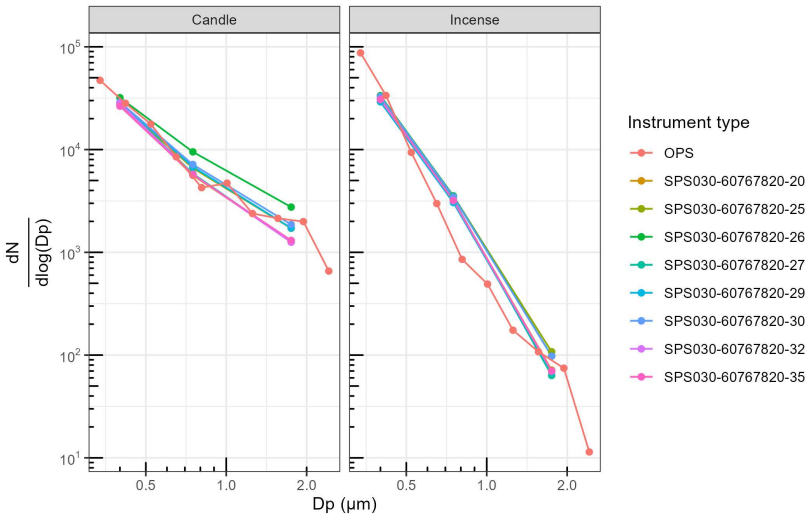


Figure A15. Particle size distribution reported by the SPS30 and the OPS TSI for stable concentrations of incense and candle. The dots represent the actual datapoints used. To facilitate the visualisation, the data from the Sensirion SPS30 were multiplied by a factor 100.

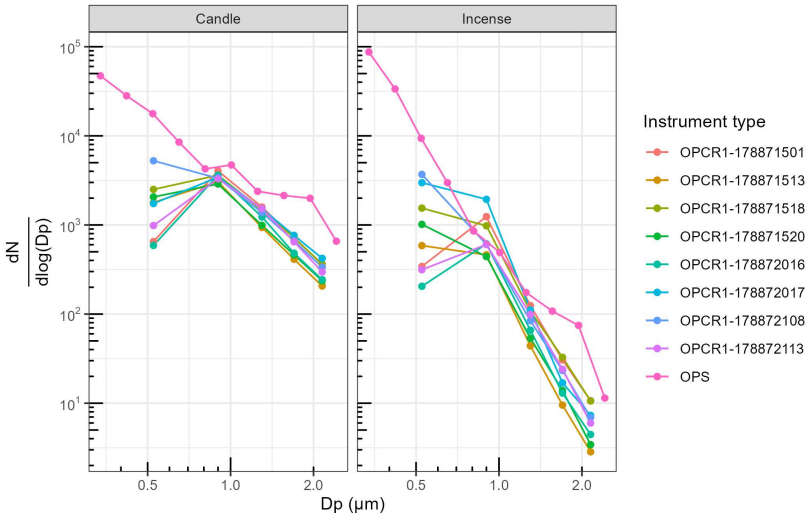


Figure A16. Particle size distribution reported by the OPCR1 and the OPS TSI for stable concentrations of incense and candle. The dots represent the actual datapoints used.

References

1. Landrigan, P.J.; Fuller, R.; Acosta, N.J.; Adeyi, O.; Arnold, R.; Basu, N.; Baldé, A.B.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J.I.; et al. The Lancet Commission on pollution and health. *Lancet* **2017**, *391*, 462–512. doi:10.1016/S0140-6736(17)32345-0.
2. Burnett, R.; Chen, H.; Fann, N.; Hubbell, B.; Pope, C.A.; Frostad, J.; Lim, S.S.; Kan, H.; Walker, K.D.; Thurston, G.D.; et al. Global estimates of mortality associated with long- term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9592–9597. doi:10.1073/pnas.1803222115.
3. Kim, E.; Hopke, P.K.; Pinto, J.P.; Wilson, W.E. Spatial variability of fine particle mass, components, and source contributions during the Regional Air Pollution Study in St. Louis. *Environ. Sci. Technol.* **2005**, *39*, 4172–4179. doi:10.1021/es049824x.
4. Organization, W.H. *WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization: 2021;
5. Frederickson, L.B.; Sidaraviciute, R.; Schmidt, J.A.; Hertel, O.; Johnson, M.S. Are dense networks of low-cost nodes really useful for monitoring air pollution? A case study in Staffordshire. *Atmos. Chem. Phys.* **2022**, *22*, 13949–13965. doi:10.5194/acp-22-13949-2022.
6. Questions and Answers on New Air Quality Rules. Available online: https://ec.europa.eu/commission/presscorner/detail/en/qanda_22_6348 (accessed on 5 April 2023).
7. Kuula, J.; Timonen, H.; Niemi, J.V.; Manninen, H.E.; Rönkkö, T.; Hussein, T.; Fung, P.L.; Tarkoma, S.; Laakso, M.; Saukko, E.; et al. Opinion: Insights into updating Ambient Air Quality Directive 2008/50/EC. *Atmos. Chem. Phys.* **2022**, *22*, 4801–4808. doi:10.5194/acp-22-4801-2022.
8. Bulot, F. Systematic Studies of Commodity Particulate Matter Air Pollution Sensors. Ph.D. Thesis, University of Southampton, 2022.
9. Kim, K.H.; Kabir, E.; Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015**, *74*, 136–143. doi:10.1016/j.envint.2014.10.005.
10. Hagan, D.H.; Gani, S.; Bhandari, S.; Patel, K.; Habib, G.; Apte, J.S.; Hildebrandt Ruiz, L.; Kroll, J.H. Inferring Aerosol Sources from Low-Cost Air Quality Sensor Measurements: A Case Study in Delhi, India. *Environ. Sci. Technol. Lett.* **2019**, *6*, 467–472. doi:10.1021/acs.estlett.9b00393.
11. Morawska, L.; Thai, P.K.; Liu, X.; Asumadu-Sakyi, A.; Ayoko, G.; Bartonova, A.; Bedini, A.; Chai, F.; Christensen, B.; Dunbabin, M.; et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **2018**, *116*, 286–299. doi:10.1016/j.envint.2018.04.018.
12. Hagan, D.H.; Kroll, J.H. Assessing the accuracy of low-cost optical particle sensors using a physics-based approach. *Atmos. Meas. Tech.* **2020**, *13*, 6343–6355. doi:10.5194/amt-13-6343-2020.

13. Tryner, J.; Mehaffy, J.; Miller-Lionberg, D.; Volckens, J. Effects of aerosol type and simulated aging on performance of low-cost PM sensors. *J. Aerosol Sci.* **2020**, *150*, 105654. doi:10.1016/j.jaerosci.2020.105654.
14. Alfano, B.; Barretta, L.; Giudice, A.D.; De Vito, S.; Francia, G.D.; Esposito, E.; Formisano, F.; Massera, E.; Miglietta, M.L.; Polichetti, T. A review of low-cost particulate matter sensors from the developers' perspectives. *Sensors* **2020**, *20*, 6819. doi:10.3390/s20236819.
15. He, M.; Kuerbanjiang, N.; Dhaniyala, S. Performance characteristics of the low-cost Plantower PMS optical sensor. *Aerosol Sci. Technol.* **2020**, *54*, 232–241. doi:10.1080/02786826.2019.1696015.
16. Ouimette, J.R.; Malm, W.C.; Schichtel, B.A.; Sheridan, P.J.; Andrews, E.; Ogren, J.A.; Arnott, W.P. Evaluating the PurpleAir monitor as an aerosol light scattering instrument. *Atmos. Meas. Tech.* **2022**, *15*, 655–676. doi:10.5194/amt-15-655-2022.
17. Kuula, J.; Mäkelä, T.; Aurela, M.; Teinilä, K.; Varjonen, S.; González, Ó.; Timonen, H. Laboratory evaluation of particle-size selectivity of optical low-cost particulate matter sensors. *Atmos. Meas. Tech.* **2020**, *13*, 2413–2423. doi:10.5194/amt-13-2413-2020.
18. Levy Zamora, M.; Xiong, F.; Gentner, D.; Kerkez, B.; Kohrman-Glaser, J.; Koehler, K. Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environ. Sci. Technol.* **2019**, *53*, 838–849. doi:10.1021/acs.est.8b05174.
19. Saarela, M.; Jauhiainen, S. Comparison of feature importance measures as explanations for classification models. *SN Appl. Sci.* **2021**, *3*, 272. doi:10.1007/s42452-021-04148-9.
20. Bhatt, U.; Xiang, A.; Sharma, S.; Weller, A.; Taly, A.; Jia, Y.; Ghosh, J.; Puri, R.; Moura, J.M.F.; Eckersley, P. Explainable Machine Learning in Deployment. *arXiv* **2020**, arXiv:1909.06342. doi:10.48550/arXiv.1909.06342.
21. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. *J. King Saud Univ. - Comput. Inf. Sci.* **2022**, *34*, 1060–1073. doi:10.1016/j.jksuci.2019.06.012.
22. Kumar, V. Feature Selection: A literature Review. *Smart Comput. Rev.* **2014**, *4*. doi:10.6029/smartcr.2014.03.007.
23. Venkatesh, B.; Anuradha, J. A Review of Feature Selection and Its Methods. *Cybern. Inf. Technol.* **2019**, *19*, 3–26. doi:10.2478/cait-2019-0001.
24. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature Selection: A Data Perspective. *ACM Comput. Surv.* **2018**, *50*, 1–45. doi:10.1145/3136625.
25. Bulot, F.M.J.; Russell, H.S.; Rezaei, M.; Johnson, M.S.; Johnston Ossont, S.J.; Morris, A.K.R.; Basford, P.J.; Easton, N.H.C.; Foster, G.L.; Loxham, M.; et al. Laboratory comparison of low-cost particulate matter sensors to measure transient events of pollution. *Sensors* **2020**, *20*, 2219. doi:10.3390/s20082219.
26. Cowell, N.; Chapman, L.; Bloss, W.; Srivastava, D.; Bartington, S.; Singh, A. Particulate matter in a lockdown home: Evaluation, calibration, results and health risk from an IoT enabled low-cost sensor network for residential air quality monitoring. *Environ. Sci. Atmos.* **2023**, *3*, 65–84. doi:10.1039/D2EA00124A.
27. Gupta, P.; Doraiswamy, P.; Levy, R.; Pikelnaya, O.; Maibach, J.; Feenstra, B.; Polidori, A.; Kiros, F.; Mills, K.C. Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth* **2018**, *2*, 172–181. doi:10.1029/2018GH000136.
28. Gressent, A.; Malherbe, L.; Colette, A.; Rollin, H.; Scimia, R. Data fusion for air quality mapping using low-cost sensor observations: Feasibility and added-value. *Environ. Int.* **2020**, *143*, 105965. doi:10.1016/j.envint.2020.105965.
29. Johnston, S.J.; Basford, P.J.; Bulot, F.M.J.; Apetroaie-Cristea, M.; Easton, N.H.C.; Davenport, C.; Foster, G.L.; Loxham, M.; Morris, A.K.R.; Cox, S.J. City scale particulate matter monitoring using LoRaWAN based air quality IoT devices. *Sensors* **2019**, *19*, 209. doi:10.3390/s19010209.
30. Bulot, F.M.J.; Basford, P.J.; Johnston, S.J.; Cox, S.J. FEEprojects/sensirion-sps030: Fix and improve test.py. 2019. doi:10.5281/zenodo.3543603.
31. Bulot, F.M.J.; Basford, P.J.; Johnston, S.J.; Cox, S.J. FEEprojects/plantower: Timestamp now included. 2019. doi:10.5281/zenodo.2579227.
32. Hagan, D.H.; Bulot, F.; Tolmie, A.; Rach8cww.; Will.; Badger, T.G.; Trochim, J.; Smith, A. FlorentinBulotUoS/py-opc: Implementation of the OPC-R1. 2019. doi:10.5281/ZENODO.3571966.
33. Sensirion. Datasheet SHT3x-DIS Humidity and Temperature Sensor.
34. Sayahi, T.; Butterfield, A.; Kelly, K.E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **2019**, *245*, 932–940. doi:10.1016/j.envpol.2018.11.065.
35. Sensirion. Datasheet SPS30-Particulate Matter Sensor for Air Quality Monitoring and Control.

36. Alphasense User Manual OPC-R1 Optical Particle Counter, 2019.
37. Han, H.S.; Sreenath, A.; Birkeland, N.T.; Chancellor, G.J. Performance of a High Resolution Optical Particle Spectrometer. In Proceedings of the EAC2011, 2011.
38. Klosterköther, A.; Kurtenbach, R.; Wiesen, P.; Kleffmann, J. Determination of the emission indices for NO, NO2, HONO, HCHO, CO, and particles emitted from candles. *Indoor Air* **2021**, *31*, 116–127. doi:10.1111/ina.12714.
39. Zai, S.; Zhen, H.; Jia-song, W. Studies on the size distribution, number and mass emission factors of candle particles characterized by modes of burning. *J. Aerosol Sci.* **2006**, *37*, 1484–1496. doi:10.1016/j.jaerosci.2006.05.001.
40. Li, W.; Hopke, P.K. Initial size distributions and hygroscopicity of indoor combustion aerosol particles. *Aerosol Sci. Technol.* **1993**, *19*, 305–316. doi:10.1080/02786829308959638.
41. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, 2nd ed.; Springer Series in Statistics; Springer: New York, NY, USA, 2009.
42. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the **Boruta** Package. *J. Stat. Softw.* **2010**, *36*. doi:10.18637/jss.v036.i11.
43. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **2002**, *46*, 389–422. doi:10.1023/A:1012487302797.
44. Di Antonio, A.; Popoola, O.A.; Ouyang, B.; Saffell, J.; Jones, R.L. Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter. *Sensors* **2018**, *18*, 2790. doi:10.3390/s18092790.
45. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
46. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13.
47. Kuhn, M. *caret: Classification and Regression Training*; R package version 6.0-93; 2022.
48. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. kernlab—An S4 Package for Kernel Methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20. doi:10.18637/jss.v011.i09.
49. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22. doi:10.18637/jss.v033.i01.
50. Jayaratne, R.; Liu, X.; Thai, P.; Dunbabin, M.; Morawska, L. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmos. Meas. Tech.* **2018**, *11*, 4883–4890. doi:10.5194/amt-11-4883-2018.