A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science

Baptiste Languille, Valérie Gros, Nicolas Bonnaire, Clément Pommier, Cécile Honoré, Christophe Debert, Laurent Gauvin, Salim Srairi, Isabella Annesi-Maesano, Basile Chaix, et al.

A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science

Baptiste Languille[1*], Valérie Gros[1], Nicolas Bonnaire[1], Clément Pommier[1], Cécile Honoré[2†], Christophe Debert[2], Laurent Gauvin[2], Salim Srairi[3], Isabella Annesi-Maesano[4], Basile Chaix[5], Karine Zeitouni[6]

[1]Laboratoire des sciences du climat et de l'environnement CNRS-CEA-UVSQ, IPSL, Gif-Sur-Yvette, France.

[2]Airparif, Paris, France.

[†]Now at: Mairie de Paris, Direction de la voirie et des déplacements, Paris, France

[3]CEREMA, Trappes-en-Yvelines, France.

[4]EPAR, IPLESP, INSERM et Sorbonne université, Faculté de médecine Saint-Antoine, Paris, France.

[5]INSERM, Sorbonne université, institut Pierre Louis d'épidémiologie et de santé publique IPLESP, Nemesis team, Paris, France.

[6]Université de Versailles Saint-Quentin, Versailles, France.

Abstract

The field of small air quality sensors is of growing interest within the scientific community, especially because this new technology is liable to improve air pollutant monitoring as well as be used for personal exposure quantification. Amongst the myriad existing devices, the performances are highly variable; this is why the sensors must be rigorously assessed before deployment, according to the intended use. This study is included in the Polluscope project; its purpose is to quantify personal exposure to air pollutants by using portable sensors. This paper designs and applies a methodology for the evaluation of portable air quality sensors to eight devices measuring PM, BC, NO_2 and O_3. The dedicated testing protocol includes static ambient air measurements compared with reference instruments, controlled chamber and mobility tests, as well as reproducibility evaluation. Three sensors (AE51, Cairclip and Canarin) were retained to be used for the field campaigns. The reliability of their performances were robustly quantified by

[*] Corresponding author. Tel: +331 69 08 32 63

E-mail address: baptiste.languille@lsce.ipsl.fr

using several metrics. These three devices (for a total of 36 units) were deployed to be worn by volunteers for a week. The results show the ability of sensors to discriminate between different environments (i.e., cooking, commuting or in an office). This work demonstrates, first, the ability of the three selected sensors to deliver data reliable enough to enable personal exposure estimations, and second, the robustness of this testing methodology.

Keywords: Paris region; Black carbon; Nitrogen oxides; Particulate matter; Personal exposure, Mobile measurements.

1 INTRODUCTION

Atmospheric pollution is a well-identified threat to health (WHO, 2003; IARC, 2013; Sante Publique France, 2016; European Environment Agency, 2017). WHO (2014) indicates that 94% of world population is exposed to levels of air pollution that are hazardous. This highlights that monitoring pollution accurately is very important to understand better the phenomenon and to suggest solutions for its mitigation.

So far, monitoring networks do not enable the precise measurement of personal exposure to pollution, defined as the pollutant concentration inhaled by people over a period of time. The first reason is the interpolation between stations. Even with a great number of stations, interpolation is still required to quantify personal exposure, and that leads to errors. Second, the air sampling height (set by the legislation between 1.5 and 3 meters above ground level) is often above the average height of our respiratory system, which also induces a difference with the inhaled air. Furthermore, monitoring stations maps and daily reports are only based on outdoor measurements, but people spend most of their time indoors (Klepeis *et al.*, 2001), where high concentrations of air pollutants may exist (Adgate *et al.*, 2004). This is why it is of primary importance to measure also indoor concentrations and to take into account the different environments where people live to quantify their personal exposure.

The solution to quantifying personal exposure could be portable air pollutant sensors. The field of small sensors is constantly improving (Borghi *et al.*, 2017) thanks to the progress of available technologies. This type of sensor presents two main advantages over classical measurements. First, the sensor units are small and thus easily worn all day long by people, which could enable robust 24/7 personal exposure measurements (including indoor air measurements). Second, some

of these sensor units are relatively low cost, which allows large numbers of units to be purchased and enables simultaneous monitoring of a large number of places.

Conversely, the main drawback of these devices is their questionable accuracy. In the field of small air quality sensors, several kinds of studies were published, such as sensor development (Hu *et al.*, 2016; Mead et al., 2013; Peng *et al.*, 2013), sensor assessment (Lin *et al.*, 2015; Burkart *et al.*, 2010; Sousan *et al.*, 2016), exploratory measures of personal exposure (Velasco *et al.*, 2016; Velasco and Tan, 2016; Hu *et al.*, 2014) and full-scale project involving large field campaigns (Mead *et al.*, 2013; Castell *et al.*, 2015, 2017; Schneider *et al.*, 2017; Hasenfratz *et al.*, 2015).

Although previous projects were interesting, some limitations invite researchers to keep on improving the methods to quantify personal exposure. First, it is highlighted that most of the small sensors suffer from lower precision than reference instruments, which shows the importance of a robust assessment prior to launching field campaigns. The report by Lewis et al. (2018) (not yet published during the experiments presented below) gave the state of the art of the low-cost air quality sensors. However, not all studies pushed the sensor characterization as far as it should be. Second, personal exposure is about measuring air quality as close as possible to the inhaled air, which is why it is of primary importance to give the sensor units directly to people. But some studies used sensor units on static measurements or attached to vehicles (Velasco and Tan, 2016; Castell *et al*., 2017; Deville Cavellin *et al.*, 2016; Duvall *et al.*, 2016; Fishbain *et al.*, 2017; Gao *et al.*, 2015; Holstius *et al.*, 2014), which is not as relevant as asking people to personally carry the sensor units. Finally, personal exposure is relevant in health impact studies for which it is interesting to have several measured pollutants. This last consideration points out single pollutant measurement studies as a limitation.

Île-de-France is the Paris region, it is the most densely populated in France but with great disparities between Paris centre and remote places. This region is also characterized as an important atmospheric pollutant emission area. More than one million inhabitants are exposed to nitrogen dioxide concentration exceeding the *limit value*[1] (annual mean of 40 µg m^{-3}) and 85 % of the population is exposed to $PM_{2.5}$ levels above the *long-term objective*[1] (annual mean of 10 µg m^{-3}) (Airparif, 2017, 2018, 2019).

The project Polluscope funded by the French National Research Agency addresses precisely the previously mentioned issues by asking volunteers to carry portable geolocalised sensor units all day long during one week in order to quantify personal exposure to several pollutants. This project is characterized by multidisciplinary objectives (large field campaigns with many volunteers, a cloud platform for data processing, a big data analysis, an epidemiological study, a deep data processing, etc.), but the first step was to conduct a robust sensor selection and assessment.

Polluscope is also defined by the diversity of the studied pollutants as this project will monitor PM (PM_{10}, $PM_{2.5}$ and PM_1), black carbon (BC) and nitrogen dioxide (NO_2). The choice of the monitored pollutants was made according to their impact on health as well as the exceedances experienced in Île-de-France for each pollutant. Indeed, Airparif, the French agency for air quality monitoring, states that some pollutants (PM_{10}, $PM_{2.5}$, NO_2, O_3, benzene) still exceed the limits. As these pollutants (as well as BC, which is not regulated) have a positive deleterious

[1] Definitions of *limit value* and *long-term objective* are deeply detailed in Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe

impact on health (Schwartz *et al.*, 1996; WHO, 2012, 2003; Liu and Peng, 2018; IARC, 2013; Niranjan and Thakur, 2017; Janssen *et al.*, 2011) they are of great interest for the project.

Within Polluscope, the purposes of this paper consist of:

- Selecting the sensors in accordance with the project expectations (measurement performance, size, price, etc.); and
- Assessing the selected sensors to determine precisely their ability to be used in the project.

2 METHODS

2.1 Sensor selection protocol

2.1.1 Pre-selection

Among the main expected sensors specifications (listed in Table 1, measurement ranges were set in accordance with the maximum measured hourly levels in Île-de-France by Airparif from 1990 to 2017 (O_3: 170 ppb, NO_2: 270 ppb and PM_{10}: 660 µg m^{-3}). BC hourly mean values up to tens of micrograms per cubic meter were measured close to major traffic axes. As one of the objectives of this work is to monitor the exposure in different environments (inside, outside, in a car, etc.) throughout the day, a fine time resolution (below 5 min) was needed. Furthermore, some practical requirements were added due to the need to have sensors able to measure all day long, whatever the season, and which are easy to carry every day of the week. This is why a long enough battery life was necessary, as well as being light weight and having a large working temperature range. In addition to these fundamental specifications, the sensors had to present a detection limit, a precision and an accuracy in accordance with the project goals. Moreover, close

attention had to be paid to other possible issues, such as interference from other pollutants, long-term drift, lack of data storage capacity, etc.

Tens of sensors measuring the pollutants of interest existed, but very few matched our expectations. First, the sensor units must be commercially available; nevertheless, beta versions of sensor units were considered in order to give a chance to almost commercialized units. Furthermore, the considered units had to be built for mobile measurements.

According to these specifications, a pre-selection was made, mostly based on a bibliographic survey. About fifty articles were studied; they are listed in Table S1 in the appendix.

At the end of the pre-selection stage, eight sensors fulfilling or mostly fulfilling the requirements were chosen to be tested. Unfortunately, some devices gave aberrant values, this is why no VOC results are presented in this work even if it was a pollutant of interest.

2.1.2 Testing strategy

The first step consisted of testing the sensors in static measurements, for up to 15 days, in ambient air, close to fixed reference instruments. These measurements took place at *Site instrumental de recherche par télédétection atmospherique* (SIRTA[2]) which is an atmospheric observatory belonging to the research insfrastructure ACTRIS[3]. For particulate matter, the reference instruments were Fidas (Palas), TEOM 1405F for PM_{10} (Thermo Scientific), TEOM 1400 for $PM_{2.5}$ and PM_1 (Thermo Electron) and Aethalometer AE33-7 for BC (Magee Scientific). For gas monitoring, the reference instruments were T200UP for nitric oxides and T400 for ozone (Teledyne). The time step was 1 minute for every instrument, except for TEOM, which was 15

[2] http://sirta.ipsl.fr

[3] https://www.actris.eu/

minutes. The reference instruments' inlets were situated on the roof of the building (approximately 4 meters above ground level). Sensor units were fastened under a steel shelter (next to the reference instruments building) in order to be protected from rain; they sampled the air at around 2.5 meters above ground level. The units were powered all day long. This first step consisting of static measurements compared with reference instruments aimed to be a first assessment of sensor accuracy. For this kind of test, sensors are usually assessed compared with reference instruments by using root mean squared error (RMSE) and Pearson correlation R^2. However, these two indicators do not totally assess the capability of the sensor units to be used in field campaigns. First, RMSE is too sensitive to brief discrepancies between the reference instrument and the sensor. For instance, if two time series are very similar most of the time, but a sudden and significant difference occurs, the RMSE could be large. The Pearson correlation is a good indicator, but only for the linear domain. However, for extreme values, sensors could measure outside their linear domain. Therefore, other correlation coefficients dealing with non-linear domains are needed. This is why a tool developed by Fishbain *et al.* (2017) that quantitatively assesses the sensors was used. The algorithm is called SET for sensor evaluation toolbox, and as it is well described in the paper, only a brief explanation is given here. The SET adds six parameters to the two above-mentioned (RMSE and Pearson correlation): the Kendall and Spearman correlation (assessing the non-linear correlation), the presence (quantifying the data loss), the source analysis (not used in the present work because not relevant in this context), the match score (a metric assessing the ranking order similarities) and the lower frequencies energy (LFE, estimating the sensor ability to properly capture the signal variability). Then, an integrated performance index (IPI) is calculated from the eight parameters; this index goes from 0 to 1 and the higher the IPI, the better is the sensor.

The second step consisted of mobility/reproducibility tests. These tests were done to assess the ability of the sensors to measure not only in static positions but also in motion. To assess the reproducibility of the measurements, a group of three persons were equipped, each with one unit of the selected sensors. The measurements lasted one working day and took place by following a route previously set, as done in scripted exposure studies (Jarjour *et al.*, 2013). The itinerary was chosen to pass through different environments (public bus, tramway, metro, office and restaurant) in downtown and suburbs of Paris. The route also went close to Airparif monitoring stations (www.airparif.fr), so as to get some "reference points" throughout the day. This itinerary is plotted in Fig. 1.

2.2 Selected sensors assessment

Once the selection tests were over and the sensors were selected, it was of primary importance to conduct a robust assessment of these instruments to be used in the field campaigns. The different steps of this assessment are described below.

2.2.1 Reproducibility tests

Fifteen sensor units were purchased for the "real" campaigns, all of them had to measure in the same way, which is why reproducibility, accuracy and precision had to be properly checked. From June 26th to July 2nd 2018, all fifteen units of each selected sensor were tested simultaneously in static measurements co-located with reference instruments just like during the selection step.

2.2.2 Controlled chamber tests

The second step's goal was to assess the sensors' sensitivity to humidity, reactivity and interferences to pollutant level changes as well as the sensors' signal to noise ratio under constant conditions. These tests were conducted in a controlled chamber where the humidity is controlled

(from approximately 40 % to 80 % relative humidity, the entire cycle lasted one hour, two cycles were conducted) as well as NO_2 concentration (from 0 to 200 ppb). These are the typical orders of magnitude the sensors will have to deal with due to environmental changes during the coming campaigns. NO_2 variation is of course useful to characterize sensors measuring NO_2, but this test was also relevant for other gas sensors, as some of these devices can suffer from NO_2 interference.

2.2.3 Feasibility campaign

Finally, a feasibility campaign was performed, which consisted of a rehearsal or a practice for the coming "real" campaigns. In other words, the selected sensor units were worn by collaborators all day long for a week, in the exact same conditions that would occur during the real campaigns. The goal here was to check for any issues of any kind (battery life, inlet sampling, data flow, data processing, etc.) This exercise also allowed us to validate the final campaign protocol.

2.2.4 Final sensor selection

The authors have decided to anonymize the non-selected sensors in order not to discredit devices that did not give good results during our experiments but which may have been improved since then. The final selected sensors are presented in detail in this work: the AE51, the Cairclip and the Canarin; below called, respectively, sensor A, sensor C and sensor F (technical details about measurement principle are shown in appendix A).

3 RESULTS

Eight sensors were tested during the selection tests. The main specifications are listed in Table 2.

3.1 Static ambient air tests

For static ambient air tests the results of six of them are presented (the sensors G and F showed aberrant values). An example of the time series is given for the sensors A, B and C in Fig. 2 and gives a preliminary assessment of the sensors' reliability. As shown in Fig. 2, the results from several days of continuous measurements of sensors versus reference instruments are heterogeneous among the different sensors and thus difficult to assess. The first time series exhibits a BC sensor (black line), which gave results very close to the reference instrument (grey line). For this sensor, the results were satisfying: the two lines are almost always overlapping. This first basic tool (studying the time series) identified sensor A as being in agreement with the project expectations.

However, the results were not always that unambiguous, and some sensors gave medium results like the nitrogen dioxide sensors presented in the second time series (Fig. 2). For these two devices, it is difficult to assess the performance of the sensors by only using the time series. Furthermore, the difference with the reference instrument and the correlation are not the only characteristics to focus on, but also the medium term shifting, the lack of data, dynamics, etc. are important. This is why the SET tool (presented in the methods section) is relevant.

In Table 3, the integrated performance index (IPI) and the other SET results are presented for the six sensors used in this work. The measurements time bases and the dates of the considered period of time are also given here. BC measurements were done only by sensor A. The satisfying performance of this sensor demonstrated with the time series is corroborated by the SET evaluation, with a very good IPI of 0.91, which is due to the high results for every single parameter. Ozone was only measured by sensor B. For this pollutant, the IPI is mediocre with a value of 0.46. This is explained by the non-negligible data loss: the presence is 0.75, which means that one value out of four is missing. Furthermore, the RMSE is high (15 ppb) compared with the mean value of 8 ppb, and even the match score (0.3) is poor. Three sensors measured

NO_2. The best one was sensor B, with an IPI of 0.76 and a RMSE of 5 ppb. Sensor C gave poor results and even aberrant values highlighted by very low correlation coefficients (below 0.15). Sensor D has a fair correlation coefficient (higher than 0.5) but suffers from a large RMSE (37 ppb), poor match score (0.24) and quite significant data loss (presence of 67 %). Particulate matter was measured by three sensors: sensor B measured PM_{10} and the sensors E and F measured PM_{10}, $PM_{2.5}$ and PM_1. Sensor F gave the best results for all the PM sizes, with an IPI of 0.64, 0.80 and 0.78 for PM_{10}, $PM_{2.5}$ and PM_1, respectively. The others gave a much lower IPI. The major advantage of sensor F is the data availability, which does not suffer from data loss. Its match score is acceptable for the three PM sizes (always larger than 0.6) although this parameter is lower than 0.43 for the others. Even if its RMSE is large, sensor F gives a relevant approximation of the PM concentration.

This first static ambient test led us to rule out sensor E, which gave aberrant values for PM_{10}, as well as sensor D because of its non-satisfactory results. Sensor B was kept despite the mediocre results for O_3 (IPI of 0.46) and PM_{10} (IPI of 0.40), thanks to its multi-pollutant measuring ability and because the producing firm should improve the sensor before the next testing step. Sensor C gave here poor results, but the authors were aware that good results had been obtained with this device, and it was suggested that these unsatisfactory results could be due to an out-of-date electrochemical sensing cell or an inappropriate storage, which would lower the sensor performance. This is why new units of this sensor were purchased for the following testing steps.

3.2 Mobility/reproducibility test

This test involved the following sensors: sensor B, sensor C and sensor F. The BC sensor A was not involved in this mobility tests because studies (Ezani *et al.*, 2018; Lin *et al.*, 2017) have already demonstrated its ability to perform mobile measurements. Fig. 3 and Fig. 4 (and Fig. S1

260 in Appendix) show, for the three units of each sensor, the entire time series for every sensor of
261 each pollutant and the comparison with the Airparif stations when the route goes by these stations.
262 Fig. 3 refers to the NO_2 sensors. The whole-day time series shows that sensors C present a better
263 reproducibility between the units than sensors B. Secondly, the sensors C show a better dynamic
264 response whereas sensors B present averaged values. Moreover, this figure demonstrates the
265 sensitivity of the sensors C to monitor the environmental changes. Three specific environments
266 are pointed out in the time series: "Opéra" is an Airparif monitoring station classified as traffic
267 influenced, "Restaurant" refers to the lunch break, which took place in a cafeteria and "Bus"
268 stands for bus travel. Sensor C was able to identify different levels associated with different
269 environments: NO_2 was high (around 50 ppb) during the time spent close to the Opéra traffic
270 location, low (about 10 ppb) in the restaurant (indoors) and presented strong variations during bus
271 travel. Furthermore, these sensors quickly detected the environment changes. On the contrary,
272 sensors B were very slow to monitor pollutant level variations, and were unable to properly
273 discriminate environmental changes. Below the main time series, graphs allow us to estimate the
274 accuracy of the sensors against Airparif stations, which can be considered as reference
275 measurements. Except for the "Paris centre" station, the three sensors C were in the right range;
276 the variations were also well monitored, especially for the "Célestins" and "Opéra" stations. The
277 sensors B never showed clear variation, and a significant difference existed between the three
278 units (up to tens of ppb for the "RN2" station). Fig. 4 shows the mobility test results for PM_{10}
279 sensors B and F. The time series depict poor consistency between the units, especially for sensor
280 B, for which a gap of up to 50 $\mu g\, m^{-3}$ was observed. The comparison with the Airparif stations
281 shows that sensors B gave poor results: for the stations "Paris centre" and "Bobigny", the sensor
282 values were significantly different from the reference; for the "Opéra" station, results were fair (a
283 shift of about 10 $\mu g\, m^{-3}$ appears); and for the "RN2" station, the difference between the three
284 units was substantial (about 30 $\mu g\, m^{-3}$). Results were better for the sensors F: the sensors were

almost always in the range of the reference station, with an inter-sensor difference not larger than 20 µg m^{-3}. Nevertheless, sensors F results were poor for the "Bobigny" station, for which the difference with the reference went up to 30 µg m^{-3}.Three units of each sensor (B, C and F) were tested during this mobility steps. Sensors C and F presented the best results.

3.3 Selected sensors

The BC sensor A gave satisfying results and was therefore selected for the next step. The sensors C gave suitable results during the mobility tests, which confirms the hypothesis that a deteriorated unit was used during the static measurement tests. This sensor gave better results than sensor B for NO_2. The PM sensor F gave more accurate results than sensor B. For this reason and because of the poor results for NO_2 and O_3, sensor B was excluded.

The retained sensors were sensors A (BC), C (NO_2) and F (PM_{10}, $PM_{2.5}$ and PM_1). The next section deals with the assessment of their capabilities. Several units of each of the three sensors were purchased: 15 units for the sensors C and F, and 6 units for the sensor A (due to its high cost).

3.4 Reproducibility tests in static measurements

An overview of the results is presented in Fig. 5. From the top to the bottom, are presented the BC, NO_2 and $PM_{2.5}$ measurements. The reference instrument is plotted in black and the sensors in coloured lines. Generally, the sensors closely followed the reference instrument trend even if some discrepancies were observed. The BC sensors A were very accurate, despite some noisy periods. On the whole, the NO_2 sensors C overestimated the concentration, this is certainly due to the low NO_2 ambient concentration compared with the limit of detection (20 ppb according to the manufacturer's specifications). This behaviour was already observed by Duvall *et al.* (2016). Even if these sensors were still able to monitor the global variability in these conditions, another measurement campaign was conducted. The objective was to submit the sensors to NO_2 ambient

levels higher than the devices' detection limit. From August 28th to September 4th, fifteen NO_2 sensors C were used to measure conditions at the Airparif station close to a major road (Paris ring road). The results overview is presented in Appendix (Fig. S2), the rest of this work is based on the results from this experiment. Concerning the PM sensors F, they were both quite close to the reference and very reproducible to one another, except for one unit plotted in light green, which presented erratic values. For the $PM_{2.5}$, the mean RMSE is 6 µg m^{-3}, which is fairly low compared with the measured concentration in mobility in Paris and its suburbs (often higher than 40 µg m^{-3}).

The SET results for the BC sensors are presented in details in Table S2 in appendix. The IPI is high (around 0.8) for all the sensor units. These sensors did not suffer from data loss at all (the presence parameter is always almost 100 %). The reproducibility between the units can be quantified by the measuring range, defined as the average of the difference across the units between the maximum and the minimum for each measurement date. For these BC sensor units, the measurement range is 616 ng m^{-3}, which is not negligible but below the mean concentration value.

The SET results for the fifteen NO_2 sensor units (Table S3 in appendix) are homogeneous, the IPI spans from 0.75 to 0.79. The mean measured concentration was above 40 ppb, which is higher than the limit of detection. This leads to good correlation coefficients (above 0.76 for the mean Pearson coefficient) and reasonable RMSE compared with the ambient levels (mean RMSE is 14 ppb). Moreover, data loss was very uncommon as shown by the presence parameter.

The PM SET results are presented in Tables S4, S5 and S6 in appendix for the three PM sizes. Regarding $PM_{2.5}$ sensor units, the IPI is above 0.68, except for the sensor unit F7, which was defective (plotted in light blue in Fig. 5). The presence is higher than 0.72 (except for F6 at 0.62): the sensor units F data loss was low. The correlation coefficients were never below 0.6, which

shows good agreement with the reference. The mean measuring ranges for the three classes of PM are 16 µg m^{-3}, 19 µg m^{-3} and 20 µg m^{-3} for PM_1, $PM_{2.5}$ and PM_{10}, respectively. These measuring ranges are not negligible but the mean RMSE values are lower: 4 µg m^{-3}, 6 µg m^{-3} and 14 µg m^{-3} for PM_1, $PM_{2.5}$ and PM_{10}, respectively. This means that a discrepancy existed between the units, but the general agreement with the true value is acceptable.

3.5 Chamber controlled test

Fig. 6 shows the results of this test. The top chart represents the controlled parameters monitored: humidity and NO_2 concentration. The three others are the sensors results.

The sensor F values were always zero: it was not affected at all by humidity changes (at least in the absence of PM). The BC sensor A was clearly affected by humidity variations as it showed BC concentration variations at the same time as humidity ones. When the humidity was decreasing, a positive artefact of up to 250 ng m^{-3} was observed. Conversely, a negative artefact of 150 ng m^{-3} was reached when the humidity decreased to 40 % RH. The mean value and the standard deviation of this sensor over the constant humidity period were respectively 15 ng m^{-3} and 14 ng m^{-3}, which is low compared with ambient levels. Two NO_2 sensor units C were measuring in the chamber. They presented a sensitivity to humidity changes with the same pattern as the BC sensor. The positive artefacts went up to 66 ppb and the negative artefact was constrained at zero (there is no negative value). We have also to note that the two sensor units gave results close to one another, with a RMSE of 7 ppb. Over constant humidity and a NO_2 concentration of 25 ppb, their mean values were 21 ppb and 26 ppb with a standard deviation around 1 ppb.

The second test conducted in the controlled chamber was the NO_2 variation (Fig. S3 in Appendix). This experiment consisted of a succession of one-hour steps of 0 ppb, 50 ppb,

100 ppb, 200 ppb, 0 ppb and a final longer 50 ppb stage over several hours. Inside the chamber, the relative humidity was set at 60 % and there is no particulate matter.

The ability of the sensors to monitor quick concentration changes was demonstrated here as the two sensor units reacted at the same time as the reference instrument. The sensor units were able to monitor increases (up to 200 ppb) and decreases down to 0 ppb. However, a gap can be observed between the sensor units and the reference, the associated RMSE is 11 ppb and 15 ppb. For the final longer step at 50 ppb, the RMSE stood at 9 ppb and 2 ppb.

To conclude for the controlled chamber tests, although sensor F presented no artefacts due to humidity changes (with a zero concentration of PM), the sensors A and C were sensitive to humidity. During the following campaigns, this will have to be taken into account, a correction or an invalidation protocol may be needed.

3.6 Feasibility campaign

The feasibility campaign was conducted with fifteen volunteers from Monday 18th to Friday 22nd June 2018. One sensor unit C and one sensor unit F were given to each volunteer and six sensor units A were shared between the participants. This campaign was conceived as a proof of the Polluscope concept, thus, only a limited analysis of the results was done.

Globally, the campaign was a success: all the sensor units were worn for the whole week. The data availability (the time resolution was one minute) is 66 % for sensors A and 69 % for sensors F, which can be considered as a satisfying result. The data loss was due to minor problems and routine maintenance (filter change, turning on and off, powering, etc.). However, the data availability for the sensors C only reaches 41 %. This was caused by storage memory erasure when the sensor ran out of power. The coming campaigns protocols will prevent this issue. Generally, the data availability was slightly lower than during the previous tests, this was due to

379 the campaign environment and the fact that the sensor units were operated by volunteers without
380 expert skills.

381 Fig. 7 shows the results of the three sensor units worn for the whole week by a volunteer. Four
382 kinds of environments are pointed out: “indoor” for the time spent inside, “polluted indoor” for
383 emitting activities conducted indoors (cafeteria, smoking or cooking, for instance), “commuting”
384 journeys (whatever the travel mode) and “outdoor” for the time spent outside any building. The
385 indoor environment is the more frequent environment, nevertheless, the spikes were usually
386 observed during commuting or in polluted indoor environments. The major BC and NO_2 peaks
387 occurred most frequently during commuting trips. Inversely, the highest particulate values were
388 measured during “polluted indoor” episodes. An example of contrasted environments
389 (commuting, indoors and tobacco smoke in indoor environment) measurements is presented in
390 Appendix, Fig. S4. During the campaign, artefacts due to quick environmental change (studied in
391 the controlled chamber in Section 3.5) were observed; this is more detailed in Appendix, Fig. S5).

392 This feasibility campaign demonstrated the capability of the Polluscope protocol to conduct a
393 campaign lasting a whole week with volunteers. The results from the sensors enable us to
394 discriminate several emitting activities; a preliminary estimation of the personal exposure is thus
395 available.

396 # 4 DISCUSSION

397 The first stages of Polluscope (the selection and assessment of the sensors) have been
398 conducted. The AE51 (BC), the Cairclip (NO_2) and the Canarin (PM_{10}, $PM_{2.5}$ and PM_1) have
399 been selected and assessed. The feasibility campaign demonstrated that these three sensors are
400 reliable enough to be used for full-scale campaigns involving volunteers from the general public.
401 Their ability to discriminate different environments (commuting trips, polluted or clean indoor
402 environment, etc.) has been proven.

For the static measurement assessments, we used the SET algorithm designed by Fishbain *et al.* (2017), available as an open source resource[4]. In their article, they presented results from 25 AQMesh NO_2 sensor units that had taken measurements for about three months in static positions. The mean related IPI is 0.58, this is very close to our mean Cairclip IPI (from the reproducibility test) of 0.54; even if the Cairclip is designed for mobile measurements whereas the AQMesh system is designed for static monitoring (i.e. expecting to have a better performance than a portable device). Knowing the successful deployment of the AQMesh sensor units, this result demonstrates the reliability of the Cairclip. Fishbain *et al.* (2017) also used the SET with PM sensors (DC1700 Dylos and GeoTech), the mean resulting IPI is 0.63. The $PM_{2.5}$ Canarin sensor used in our study gave significantly better results with a mean IPI of 0.73.

The SET algorithm was also used by Broday and the Citi-Sense Project Collaborators (2017). They presented unpublished results from about three months of ambient air measurements of six PM_{10} sensor units located in Ostrava, Czech Republic. The mean IPI is 0.72, which is very close to the result from our study (0.73).

Due to its recent release, SET algorithm results have not been published in other articles yet. To the best of our knowledge, our study is the first to apply the SET evaluation to the AE51 and the Cairclip sensors. However, these two devices have been largely used and several results have been published, some of the more relevant for our study are discussed below.

4.1 Static comparison

Lin *et al.* (2017) compared AE51 with reference instruments and found a good mean correlation of 0.77. Viana *et al.* (2015) conducted a study involving six AE51 and a reference

[4] fishbain.net.technion.ac.il

station in static measurements, the correlation coefficient was above 0.75. In our study, the mean correlation coefficient was 0.80. This higher agreement may be due to higher inlet flow (150 mL min^{-1} compared with 100 mL min^{-1} in the Viana *et al.* (2015) study) or coarser time resolution (1 min in our study and 1 sec in the Lin *et al.* (2017) study).

Several other studies have pointed to the good results of this black carbon sensor in agreement with our results (Cai *et al.*, 2014; Gillespie *et al.*, 2017; Velasco and Tan, 2016).

The recent low-cost sensors review by the World Metrological Organization (Lewis *et al.*, 2018) described several performance evaluation programs as the work supported by the United States Environment Protection Agency (EPA)[5]; they described an air sensor toolbox where the main performances of tens of sensors were gathered and compared. For the Cairclip NO_2 sensor, the EPA and Jiao *et al.* (2016) state that a correlation coefficient between 0.42 and 0.76 was obtained with reference instruments. Our mean Pearson correlation was 0.76, which is in the high part of the EPA range. The additional information given by the SET algorithm in our study is the good match score of 0.64 and the absence of data loss (the presence parameter reaches almost 100 %). Another example is the study conducted by Spinelle *et al.* (2015). They found a correlation of up to 0.75 for the Cairclip sensor units. This result is both in the EPA range and close to our result of 0.76.

Due to its new release, only a few research works including Canarin have been conducted. For instance, Tse *et al.*, 2018B presented a project based on static measurements from four Canarin units. Some tests were conducted in Bologna, Italia, and PM_{10} maps have been produced. These works were preliminary and the most accomplished article about Canarin sensor is certainly the one conducted by Tse *et al.*, 2018A. They deployed nine Canarin units in a library inducted at the

[5]https://www.epa.gov/air-sensor-toolbox

UNESCO world heritage list. The sensors were measuring 24/7 for months (from Summer 2017 to Spring 2018), which enabled to show that a clear diurnal pattern occurred with higher levels during night time. On a longer period of time it was the winter season which experienced more pollution. This protocol also permitted to quantify that 56 % of the time, the $PM_{2.5}$ air pollution level was low according to the EPA standards (below 12 $\mu g\, m^{-3}$). The coming improvements announced in these three articles suggested a wider use of the Canarin in a near future. Overall, these works underlined the promising capabilities of this sensor. The present paper confirmed this first evaluation and went a step further (larger amount of units, mobile measurements, etc.) to prove the ability to use the Canarin to equip volunteers for the personal exposure quantification.

4.2 Mobile measurements

It is usually more difficult to robustly assess sensor accuracy in mobile measurements as the reference instruments are unlikely to be usable in motion. A metric that can be used is the agreement between several units of portable sensors (previously assessed – or not – in static measurements versus a reference instrument). This provides information on the reproducibility and thus on the reliability of the mobile device. Another possibility is to compare the sensor measurements with static stations considered as a reference if the mobility route goes close to this kind of monitoring site.

The Ezani *et al.* (2018) study was based on mobile measurements performed with two AE51 units. There were no reference instruments but the correlation between the two units was good: 0.92. Lin *et al.* (2017) performed mobile AE51 measurements. Comparison was possible thanks to 17 transient immobile periods (of less than one hour) nearby reference stations. The AE51 unit showed an interquartile range agreement with the reference instrument of 82 %. High-resolution mapping is possible with the AE51, as in the study of den Bossche *et al.* (2015) where sensor units were mounted on bikes in Antwerp, Belgium. A 50-meter resolution was obtained with an

uncertainty of 25 %. Pant *et al.* (2017) performed a study aiming at quantifying personal exposure to BC in New Delhi, India. The AE51 were given to volunteers and environments (commuting, cooking, etc.) were distinguished.

Few studies have been published on the Cairclip sensor being used in mobile measurements, especially compared with the abundant literature related to the AE51. The recent work by Chambers *et al.* (2018) found no consistent relationship between NO_2 concentrations and health parameters. The authors state that the Cairclip was able to appropriately monitor personal exposure and a clear diurnal cycle was observed but no more validation data was provided. The study by Reid (2015) is based on the qualification of Cairclip sensors. Mobile measurements were conducted with two Cairclip units in different environments: public transport, outdoor and indoor. The sensors monitored interesting variability, especially close to traffic.

Lastly, Aguiari *et al.*, 2018 introduced a possible use of the Canarin by attaching them to bikes.

4.3 Methodology discussion

Our work has revealed that the three selected sensors are appropriate for personal exposure assessment. Beyond that first result, the Polluscope selection and assessment methodology was also an outcome of this study.

It is now well known, even in the emerging field of small air quality sensors that a complete sensor assessment is of primary importance to obtain reliable data. Some studies were only based on laboratory experiments (Manikonda *et al.*, 2016; Ng *et al.*, 2018), but in-the-field calibration was identified as necessary to properly assess the sensors' capabilities (Castell *et al.*, 2017; Schneider *et al.*, 2017) as the results can be substantially different from laboratory-controlled environments. For instance, during our tests, the Cairclip showed better results during the controlled chamber tests. The ambient air tests were very useful to reveal that the Cairclip had difficulties in measuring low ambient concentrations.

As seen in the section concerning the reproducibility tests, non-negligible differences were observed between units of the same sensor. This highlights the importance of testing several units at a time. In this study, we conducted tests with 6 AE51, 15 Cairclip and 15 Canarin. For some of the previously published studies, the small number of tested units was a limitation, for example, Lin *et al.* (2017) (two AE51 and only one for the mobility tests), Ezani *et al.* (2018) (two AE51), Duvall *et al.* (2016) (two Cairclip).

4.4 Conclusions

No remote sensor is perfect, and the three selected ones are the result of compromises and each have strengths and weaknesses. The AE51 is accurate, its IPI (above 0.8) is higher than all other sensors. This BC sensor is also reliable and easy to use with very little data loss. But it is sensitive to humidity, which leads to some artefacts when quick environmental changes occur, and its high price is also a weakness because fewer units can be purchased. The Cairclip is very light and thus easy to carry all day long, but the storage memory is erased if the sensor unit runs out of battery, thus it is consequently more demanding for the operator. Moreover, even if these sensors demonstrated their ability to perform reliable measurements in mobile measurement (see related section), the detection limit (20 ppb) is not appropriate for low NO_2 levels. The Canarin is able to send data via Wi-Fi and has a high storage capacity (several weeks of measurements), which is useful when the data sending is not possible. Its robustness in mobility is also an important advantage: the Canarin was the sensor that presented the highest data availability during the feasibility campaign (69 %). This sensor showed satisfying results for the $PM_{2.5}$ measurements (IPI of 0.7) but substantially lower for PM_{10} (IPI of 0.4). Finally, the weight of the sensor unit (it is quite heavy) is a drawback.

Even if the three selected sensors have some weaknesses, their ability to be used in mobile measurements has been demonstrated. For the coming campaigns, attention will be given to their drawbacks.

To conclude, the Polluscope project is one of the few studies that has conducted an in-depth sensor assessment including the most important following steps:

- Several kinds of tests were performed. Ambient air static measurement tests against reference instruments was the first assessment of the sensors in real atmosphere with natural meteorological parameters variability (temperature, humidity, wind speed and direction, etc.) The laboratory tests were of primary importance to quantify the sensors; responses to rapid atmospheric changes (i.e., humidity, pollutants levels). Mobile measurements were necessary as the project goal is to use the sensors for personal exposure, to be worn by volunteers.
- A large number of units of each sensor was tested to quantify the reproducibility and to eliminate problems arising from a single deficient unit.
- A robust multi-metric static measurement assessment with the SET algorithm was conducted in order to be as rigorous as possible in the assessment.

The next stage of the Polluscope project will be the full-scale campaigns involving fifteen volunteers each week during the six weeks per studied season. These campaigns will take place over two years and will involve 160 people. Even if the size of the project is already consistent, a valuable perspective would be to recruit more volunteers over a larger area in order to increase the representativeness of the study.

ACKNOWLEDGEMENTS

This project was funded by the French agency: *Agence nationale de la recherche* (ANR – http://www.agencenationale-recherche.fr/Projet-ANR-15-CE22-0018). Part of the equipment was funded by iDEX Paris-Saclay, in the framework of the project ACE-ICSEN. Measurements from the SIRTA station were performed within the ACTRIS research infrastructure under the H2020 grant agreement 654109. Additional support from CEA and CNRS are acknowledged.

546

547 # REFERENCES

548 Adgate, J.L., Church, T.R., Ryan, A.D., Ramachandran, G., Fredrickson, A.L., Stock, T.H.,
549 Morandi, M.T. and Sexton, K. (2004). Outdoor, indoor, and personal exposure to VOCs in
550 children. *Environ. Health Perspect.* 112(14): 1386–1392.

551 Aguiari, D., Delnevo, G., Monti, L., Ghini, V., Mirri, S., Salomoni, P., Pau, G., Im, M., Tse, R.,
552 Ekpanyapong, M., and Battistini, R. (2018). Canarin ii : Designing a smart e-bike eco-system.
553 *15th IEEE Annual Consumer Communications & Networking Conference (CCNC)* 1-6

554 Airparif – La qualité de l'air dans les enceintes du métro ou du rer,
555 https://www.airparif.asso.fr/pollution/air-interieur-metro, Accessed: 22 November 2018.

556 Airparif (2017). Bilan de la qualité de l'air 2016 – surveillance et information en Île-de-France.

557 Airparif (2018). Bilan de la qualité de l'air 2017 – surveillance et information en Île-de-France.

558 Airparif (2019). Bilan de la qualité de l'air 2018 – surveillance et information en Île-de-France.

559 Borghi, F., Spinazzè, A., Rovelli, S., Campagnolo, D., Del Buono, L., Cattaneo, A., and Cavallo,
560 D.M. (2017). Miniaturized monitors for assessment of exposure to air pollutants: A review. *Int.*
561 *J. Environ. Res. Public Health* 14(8): 909.

562 den Bossche, J.V., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J. and Baets, B.D. (2015).
563 Mobile monitoring for mapping spatial variation in urban air quality: Development and
564 validation of a methodology based on an extensive dataset. *Atmos. Environ.* 105: 148–161.

565 Broday, D.M. and the Citi-Sense Project Collaborators (2017). Wireless distributed
566 environmental sensor networks for air pollution measurement: the promise and the current
567 reality. *Sensors* 17(10): 2263.

568 Burkart, J., Steiner, G., Reischl, G., Moshammer, H., Neuberger, M. and Hitzenberger, R. (2010).

569 Characterizing the performance of two optical particle counters (grimm opc1.108 and opc1.109)
570 under urban aerosol conditions. *J. Aerosol Sci.* 41(10): 953–962.

571 Cai, J., Yan, B., Ross, J.M., Zhang, D., Kinney, P.L., Perzanowski, M.S., Jung, K., Miller, R.L.
572 and Chillrud, S.N. (2014). Validation of microaeth® as a black carbon monitor for fixed-site
573 measurement and optimization for personal exposure characterization. *Aerosol Air Qual. Res.*
574 14(1): 1–9.

575 Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D. and
576 Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality
577 monitoring and exposure estimates? *Environ. Int.* 99: 293–302.

578 Castell, N., Kobernus, M., Liu, H.-Y., Schneider, P., Lahoz, W., Berre, A.J. and Noll, J. (2015).
579 Mobile technologies and services for environmental monitoring: The Citi-sense-MOB
580 approach. *Urban Clim.* 14(Part 3): 370–382.

581 Chambers, L., Finch, J., Edwards, K., Jeanjean, A., Leigh, R. and Gonem, S. (2018). Effects of
582 personal air pollution exposure on asthma symptoms, lung function and airway inflammation.
583 *Clin. Exp. Allergy*, 48(7): 798–805.

584 Deville Cavellin, L., Weichenthal, S., Tack, R., Ragettli, M.S., Smargiassi, A. and Hatzopoulou,
585 M. (2016). Investigating the use of portable air pollution sensors to capture the spatial
586 variability of traffic-related air pollution. *Environ. Sci. Technol.* 50(1): 313–320.

587 Duvall, R.M., Long, R.W., Beaver, M.R., Kronmiller, K.G., Wheeler, M.L. and Szykman, J.J.
588 (2016). Performance evaluation and community application of low-cost sensors for ozone and
589 nitrogen dioxide. *Sensors* 16(10): 1698.

590 European Environment Agency (2017). Air quality in Europe 2017 report. European
591 Environment Agency.

592 Ezani, E., Masey, N., Gillespie, J., Beattie, T.K., Shipton, Z.K. and Beverland, I.J. (2018).
593 Measurement of diesel combustion-related air pollution downwind of an experimental
594 unconventional natural gas operations site. *Atmos. Environ.* 189: 30–40.

595 Fishbain, B., Lerner, U., Castell, N., Cole-Hunter, T., Popoola, O., Broday, D.M., Iñiguez, T.M.,
596 Nieuwenhuijsen, M., Jovasevic-Stojanovic, M., Topalovic, D., Jones, R.L., Galea, K.S., Etzion,
597 Y., Kizel, F., Golumbic, Y.N., Baram-Tsabari, A., Yacobi, T., Drahler, D., Robinson, J.A.,
598 Kocman, D., Horvat, M., Svecova, V., Arpaci, A. and Bartonova, A. (2017). An evaluation
599 tool kit of air quality micro-sensing units. *Sci. Total Environ.* 575: 639–648.

600 Gao, M., Cao, J. and Seto, E. (2015). A distributed network of low-cost continuous reading
601 sensors to measure spatiotemporal variations of $PM_{2.5}$ in Xi'an, China. *Environ. Pollut.* 199:
602 56–65.

603 Gillespie, J., Masey, N., Heal, M.R., Hamilton, S. and Beverland, I.J. (2017). Estimation of
604 spatial patterns of urban air pollution over a four-week period from repeated 5-min
605 measurements. *Atmos. Environ.* 150: 295–302.

606 Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J. and Thiele, L.
607 (2015). Deriving high-resolution urban air pollution maps using mobile sensor nodes.
608 *Pervasive Mob. Comput.* 16(Part B): 268–285.

609 Hinds, W.C. and Bellin, P. (1988). Effect of facial-seal leaks on protection provided by half-mask
610 respirator. *Appl. Ind. Hyg.* 3(5): 158–164.

611 Holstius, D.M., Pillarisetti, A., Smith, K.R. and Seto, E. (2014). Field calibrations of a low-cost
612 aerosol sensor at a regulatory monitoring site in California. *Atmos. Meas. Tech.* 7(4): 1121–
613 1131.

614 Hu, K., Sivaraman, V., Luxan, B.G. and Rahman, A. (2016). Design and evaluation of a

metropolitan air pollution sensing system. *IEEE Sens. J.* 16(5): 1448–1459.

Hu, K., Wang, Y., Rahman, A. and Sivaraman, V. (2014). Personalising pollution exposure estimates using wearable activity sensors. In *2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)* (pp. 1–6).

IARC (2013). Air pollution and cancer. IARC Scientific Publications.

Janssen, N.A., Hoek, G., Simic-Lawson, M., Fischer, P., van Bree, L., ten Brink, H., Keuken, M., Atkinson, R.W., Anderson, H.R., Brunekreef, B. and Cassee, F.R. (2011). Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM_{10} and $PM_{2.5}$. *Environ. Health Perspect.* 119(12): 1691–1699.

Jarjour, S., Jerrett, M., Westerdahl, D., de Nazelle, A., Hanning, C., Daly, L., Lipsitt, J. and Balmes, J. (2013). Cyclist route choice, traffic-related air pollution, and lung function: a scripted exposure study. *Environmental Health* 12(1): 14.

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S. and Buckley, K. (2016). Community air sensor network (Cairsense) project: evaluation of low-cost sensor performance in a suburban environment in the south-eastern United States. *Atmos. Meas. Tech.* 9(11): 5281–5292.

Klepeis, N.E., Nelson, W.C., Ott, W.R., Robinson, J.P., Tsang, A.M., Switzer, P., Behar, J.V., Hern, S.C. and Engelmann, W.H. (2001). The national human activity pattern survey (NHAPS): A resource for assessing exposure to environmental pollutants. *J. Exposure Anal. Environ. Epidemiol.* 11(3): 231–252.

Lewis, A.C., von Schneidemesser, E. and Peltier, R.E. (2018). Low-cost sensors for the measurement of atmospheric composition: overview of topic and future applications. *World*

Meteorological Organization, 1215.

Lin, C., Gillespie, J., Schuder, M., Duberstein, W., Beverland, I. and Heal, M. (2015). Evaluation and calibration of aeroqual series 500 portable gas sensors for accurate measurement of ambient ozone and nitrogen dioxide. *Atmos. Environ.* 100: 111–116.

Lin, C., Masey, N., Wu, H., Jackson, M., Carruthers, D.J., Reis, S., Doherty, R.M., Beverland, I.J. and Heal, M.R. (2017). Practical field calibration of portable monitors for mobile measurements of multiple air pollutants. *Atmosphere* 8(12): 231.

Liu, J.C. and Peng, R.D. (2018). Health effect of mixtures of ozone, nitrogen dioxide, and fine particulates in 85 US counties. *Air Qual. Atmos. Health*. 11(3): 311–332.

Manikonda, A., Zikova, N., Hopke, P.K. and Ferro, A.R. (2016). Laboratory assessment of low-cost PM monitors. *J. Aerosol Sci.* 102: 29–40.

Mead, M., Popoola, O., Stewart, G., Landsho, P., Calleja, M., Hayes, M., Baldovi, J., McLeod, M., Hodgson, T., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J. and Jones, R. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmos. Environ.* 70: 186–203.

Ng, C.-L., Kai, F.-M., Tee, M.-H., Tan, N. and Hemond, H.F. (2018). A prototype sensor for in situ sensing of fine particulate matter and volatile organic compounds. *Sensors* 18(1): 265.

Niranjan, R. and Thakur, A.K. (2017). The toxicological mechanisms of environmental soot (black carbon) and carbon black: focus on oxidative stress and inflammatory pathways. *Front. Immunol.* 8: 763.

Pant, P., Habib, G., Marshall, J.D. and Peltier, R.E. (2017). $PM_{2.5}$ exposure in highly polluted cities: a case study from New Delhi, India. *Environ. Res.* 156: 167–174.

Peng, I.H., Chu, Y.Y., Kong, C.Y. and Su, Y.S. (2013). Implementation of indoor VOC air

661 pollution monitoring system with sensor network. In *2013 Seventh International Conference*
662 *on Complex, Intelligent, and Software Intensive Systems* (pp. 639–643).

663 Reid, B. (2015). A quantitative analysis of the Cairclip O_3/NO_2 sensor. Thesis, University of
664 Alberta.

665 Sante Publique France (2016). Impacts sanitaires de la pollution de l'air en France : nouvelles
666 données et perspectives. Sante Publique France.

667 Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A. and Bartonova, A. (2017).
668 Mapping urban air quality in near real-time using observations from low-cost sensors and
669 model information. *Environ. Int.* 106: 234–247.

670 Schwartz, J., Dockery, D.W. and Neas, L.M. (1996). Is daily mortality associated specifically
671 with fine particles? *J. Air Waste Manage. Assoc.* 46(10): 927–939.

672 Sousan, S., Koehler, K., Thomas, G., Park, J.H., Hillman, M., Halterman, A. and Peters, T.M.
673 (2016). Inter-comparison of low-cost sensors for measuring the mass concentration of
674 occupational aerosols. *Aerosol Sci. Technol.* 50(5): 462–473.

675 Spinelle, L., Gerboles, M., Kok, G., Persijn, S. and Sauerwald, T. (2017). Review of portable and
676 low-cost sensors for the ambient air monitoring of benzene and other volatile organic
677 compounds. *Sensors* 17(7): 1520.

678 Tse, R., Aguiari, D., Chou, K.S., Tang, S.K., Giusto, D., and Pau, G. (2018A) Monitoring cultural
679 heritage buildings via low-cost edge computing/sensing platforms: The biblioteca joanina de
680 coimbra case study. *4th International Conference on Smart Objects and Technologies for*
681 *Social Good, Goodtechs* 148-152.

682 Tse, R., Monti, L., Prandi, C., Aguiari, D., Pau, G., and Salomoni, P. On assessing the accuracy
683 of air pollution models exploiting a strategic sensors deployment (2018B). *4th EAI*

International Conference on Smart Objects and Technologies for Social Good, Goodtechs 55-58.

Velasco, A., Ferrero, R., Gandino, F., Montrucchio, B. and Rebaudengo, M. (2016). A mobile and low-cost system for environmental monitoring: a case study. *Sensors* 16: 710.

Velasco, E. and Tan, S.H. (2016). Particles exposure while sitting at bus stops of hot and humid Singapore. *Atmos. Environ.* 142: 251–263.

Viana, M., Rivas, I., Reche, C., Fonseca, A., Perez, N., Querol, X., Alastuey, A., Alvarez Pedrerol, M. and Sunyer, J. (2015). Field comparison of portable and stationary instruments for outdoor urban air exposure assessments. *Atmos. Environ.* 123(Part A): 220–228.

World Health Organization (2014). Seven million premature deaths annually linked to air pollution. http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/. Accessed: 22 November 2018.

World Health Organization (2012). Health effects of black carbon. World Health Organization.

World Health Organization (2003). Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. World Health Organization, Regional Office for Europe.

Table titles

Table 1. Tested sensor specifications.

Table 2. Main expected specifications of the sensors.

Table 3. SET results for every sensor, the units are "ppb" for gases, ng m^{-3} for BC and µg m^{-3} for PM. Pollutant sensor: measured pollutant and sensor's name; *M*: mean concentration; Match: match score; RMSE: root mean squared error; *r*: Pearson correlation coefficient; *t*: Kendall correlation coefficient; *S*: Spearman correlation coefficient; Pres: presence parameter; LFE: low frequencies energy parameter; IPI: SET integrated performance index; SensTB: sensor time base: RefTB: reference time base; Start time: measurement's beginning; End time: measurement's ending.

Table 1. Tested sensor specifications.

Expected specifications	
Measurement range	O_3 = [0;250] ppb NO_2 = [0;500] ppb BC = [0;50000] ng m^{-3} PM_{10} and $PM_{2.5}$ = [0;1000] µg m^{-3} VOC = depending on the sensor specificity ability
Time step	below 5 min
Battery life	12 hours as a minimum
Temperature range	[–10;40] °C
Weight (total to be worn)	below 2 kg
Detection limit, precision and accuracy	To be specified by the manufacturer

17 **Table 2.** Main expected specifications of the sensors.

Sensor	Price (€)	Weight (g)	Measured pollutant
A	5000–10000	200–400	BC
B	2000–5000	600–800	PM_{10}, NO_2, O_3, VOC
C	500–1000	0–100	NO_2
D	1000–2000	400–600	NO_2
E	2000–5000	600–800	PM_{10}, $PM_{2.5}$, PM_1
F	500–1000	600–800	PM_{10}, $PM_{2.5}$, PM_1
G	1000–2000	100–200	VOC
H	2000–5000	400–600	CO, NO, NO_2, O_3, PM_{10}

18

19

Table 3. SET results for every sensor, the units are "ppb" for gases, ng m^{-3} for BC and µg m^{-3} for PM. Pollutant sensor: measured pollutant and sensor's name; *M*: mean concentration; Match: match score; RMSE: root mean squared error; *r*: Pearson correlation coefficient; *t*: Kendall correlation coefficient; *S*: Spearman correlation coefficient; Pres: presence parameter; LFE: low frequencies energy parameter; IPI: SET integrated performance index; SensTB: sensor time base: RefTB: reference time base; Start time: measurement's beginning; End time: measurement's ending.

Pollutant sensor	*M*	Match	RMSE	*r*	*t*	*S*	Pres	LFE	IPI	SensTB (min)	RefTB (min)	Start time (TU)	End time (TU)
BC A	1077	0.88	268	0.98	0.85	0.96	0.97	1.00	0.91	5	1	23/01/2017 07:55	06/02/2017 15:05
O_3 B	8	0.30	15	0.70	0.60	0.80	0.75	1.00	0.46	1	1	23/01/2017 07:43	03/02/2017 08:44
NO_2 B	11	0.46	5	0.94	0.72	0.89	0.78	1.00	0.76	1	1	23/01/2017 07:43	03/02/2017 08:44
NO_2 C	20	0.35	13	0.04	0.08	0.12	1.00	1.00	0.42	1	1	20/01/2017 14:00	27/01/2017 10:00
NO_2 D	47	0.24	37	0.54	0.54	0.73	0.67	1.00	0.56	1	1	23/01/2017 13:56	01/02/2017 10:09
PM_{10} B	104	0.43	112	0.18	0.20	0.26	0.81	0.99	0.40	1	15	23/01/2017 07:45	03/02/2017 08:45
PM_{10} E	535	0.37	1819	0.06	0.38	0.52	0.69	0.89	0.07	1	15	23/01/2017 08:30	03/02/2017 16:00
PM_{10} F	21	0.63	16	0.84	0.33	0.46	1.00	0.99	0.64	1.5	15	26/01/2017 12:30	06/02/2017 15:15
$PM_{2.5}$ E	136	0.43	185	0.45	0.52	0.72	0.69	0.99	0.49	1	15	23/01/2017 08:30	03/02/2017 16:00
$PM_{2.5}$ F	18	0.76	10	0.91	0.66	0.82	1.00	0.99	0.80	1.5	15	26/01/2017 12:30	06/02/2017 15:15
PM_1 E	43	0.43	33	0.76	0.62	0.82	0.69	1.00	0.65	1	15	23/01/2017 08:30	03/02/2017 16:00
PM_1 F	13	0.77	8	0.88	0.66	0.82	1.00	0.99	0.78	1.5	15	26/01/2017 12:30	06/02/2017 15:15

Figure Captions

Fig. 1. Routes for the mobility tests (plotted in red and blue). Sources: actualitix, and OpenStreetMap (modified).

Fig. 2. BC and NO_2 time series in static measurements.

Fig. 3. Mobility tests for NO_2 sensors. The three sensors C are plotted in shades of green, the three sensors B in shades of orange, the reference Airparif stations are plotted in black.

Fig. 4. Mobility tests for PM_{10} sensors. The three sensors F are plotted in shades of blue, the three sensors B in shades of orange, the reference Airparif stations are plotted in black.

Fig. 5. Reproducibility tests for sensors A, C and F with reference instruments.

Fig. 6. Chamber controlled tests for BC sensor A, PM sensor F and NO_2 sensor C.

Fig. 7. Mobility test results.

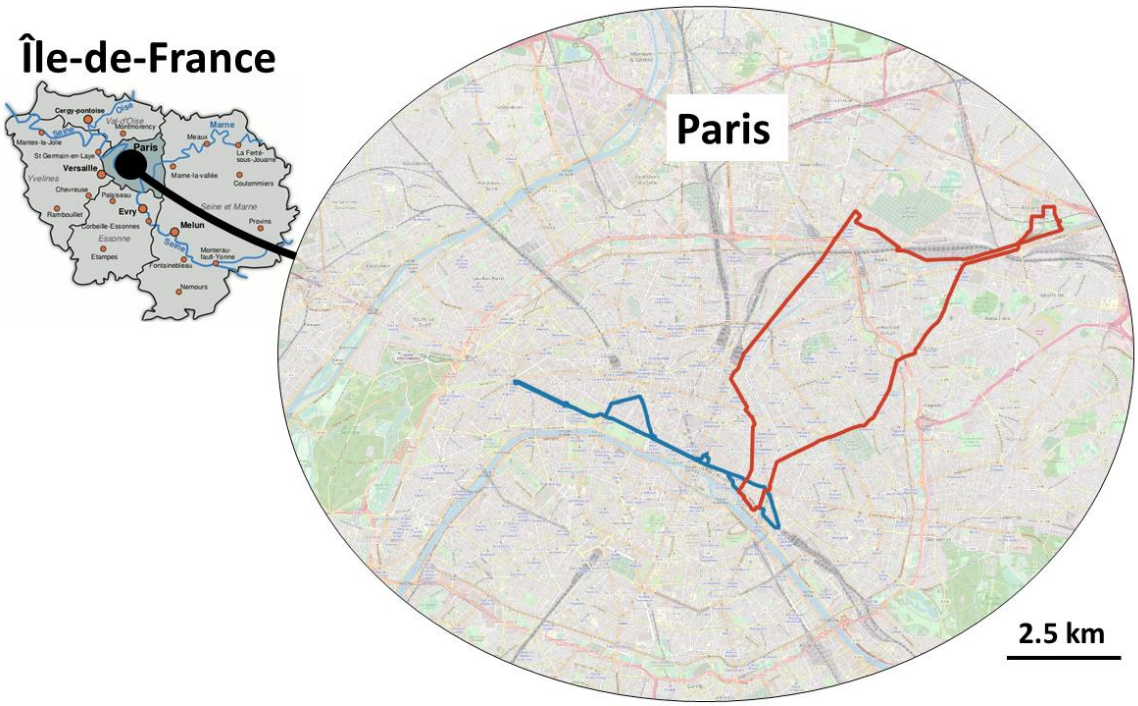
Île-de-France
Paris
Versailles
Melun
Evry
Meaux
Marne
Seine et Marne
Yvelines
Essonne
Paris
2.5 km

Fig. 1.

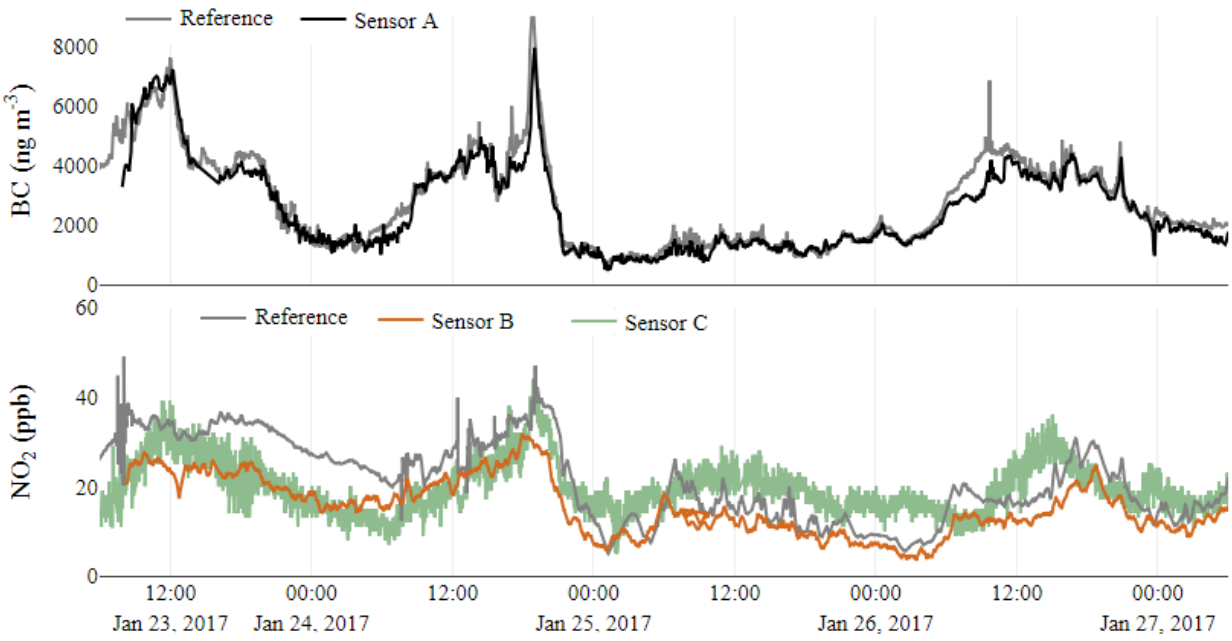
Reference
Sensor A
8000
6000
4000
2000
0
BC (ng m-3)
Reference
Sensor B
Sensor C
60
40
20
0
NO2 (ppb)
12:00
00:00
12:00
00:00
12:00
00:00
12:00
00:00
Jan 23, 2017
Jan 24, 2017
Jan 25, 2017
Jan 26, 2017
Jan 27, 2017

Fig. 2.

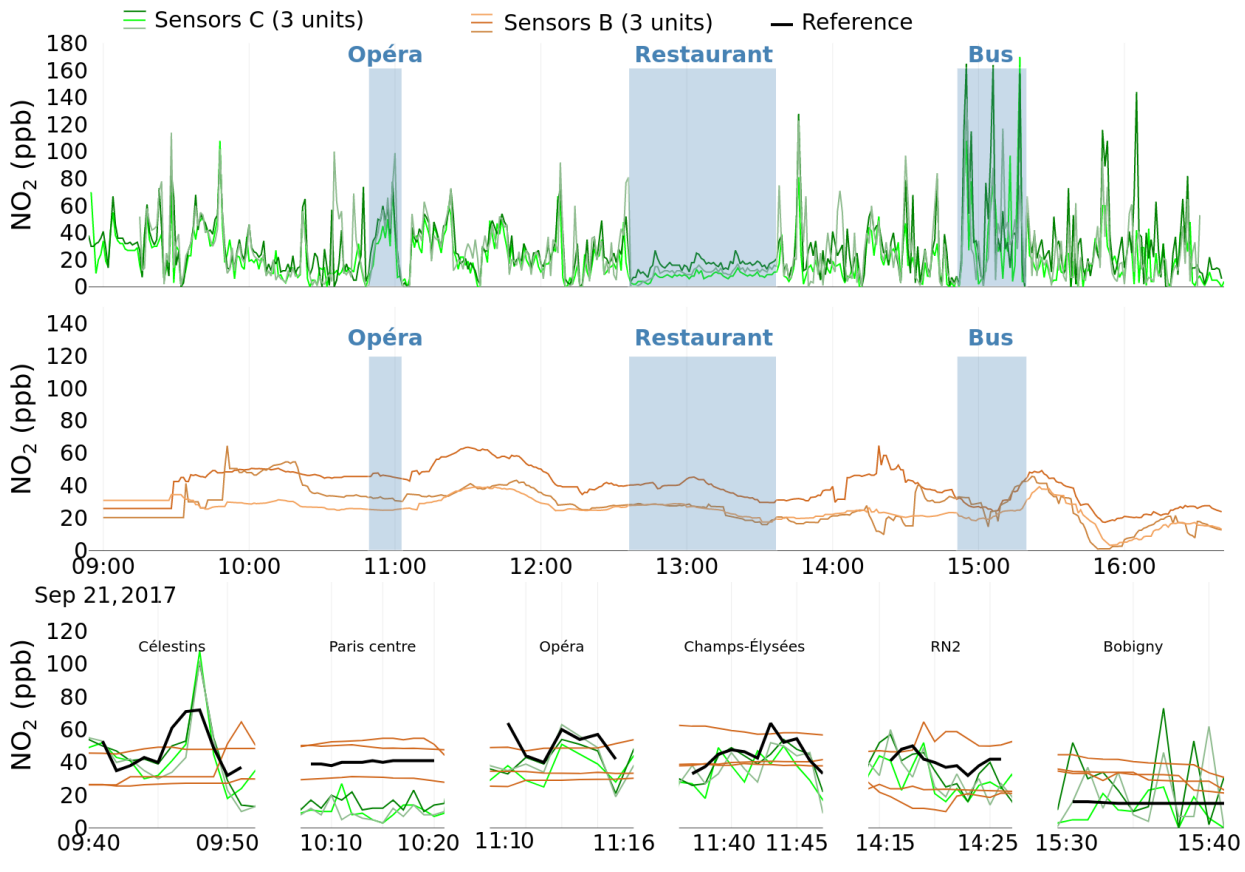
Sensors C (3 units)
Sensors B (3 units)
Reference
Opéra
Restaurant
Bus
NO2 (ppb)
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
Sep 21, 2017
Célestins
Paris centre
Opéra
Champs-Élysées
RN2
Bobigny
09:40
09:50
10:10
10:20
11:10
11:16
11:40
11:45
14:15
14:25
15:30
15:40

Fig. 3.

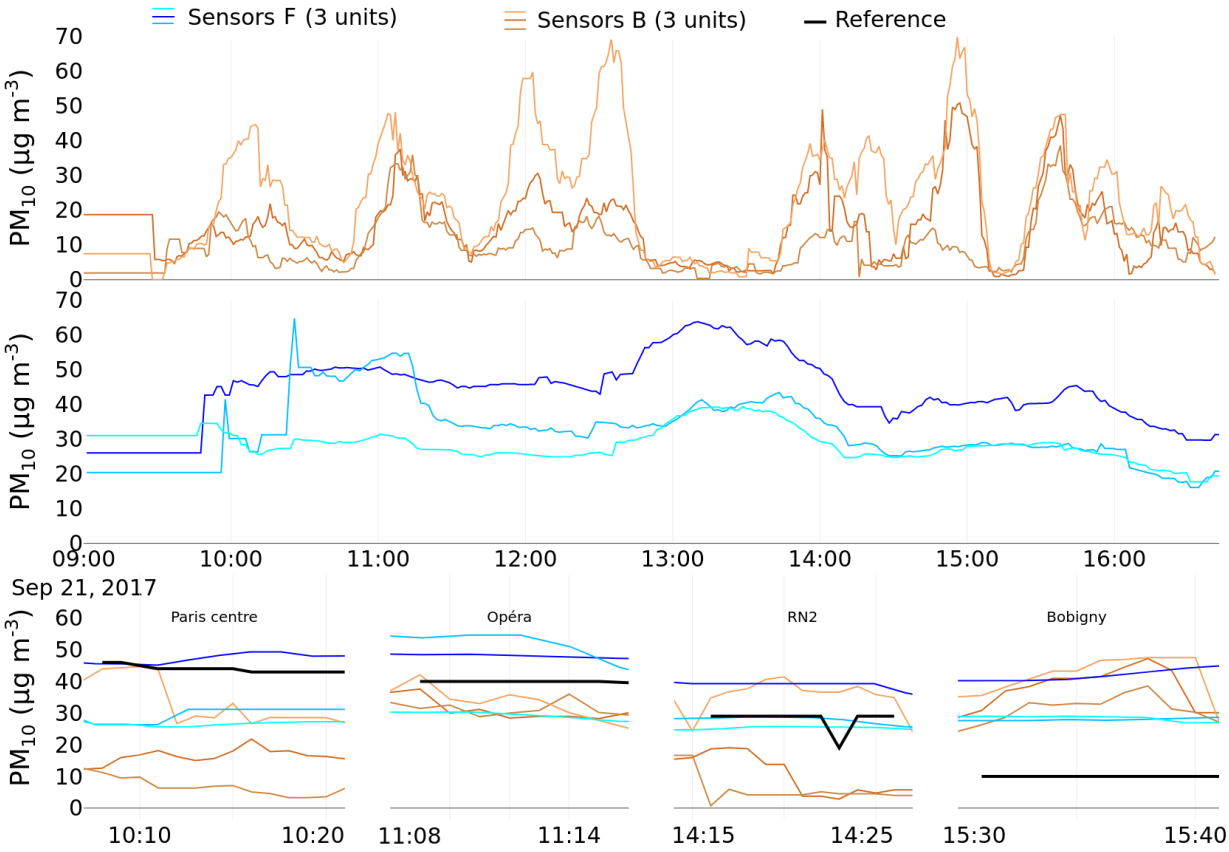
Sensors F (3 units)
Sensors B (3 units)
Reference
PM10 (µg m-3)
09:00
10:00
11:00
12:00
13:00
14:00
15:00
16:00
Sep 21, 2017
Paris centre
Opéra
RN2
Bobigny
10:10
10:20
11:08
11:14
14:15
14:25
15:30
15:40

Fig. 4.

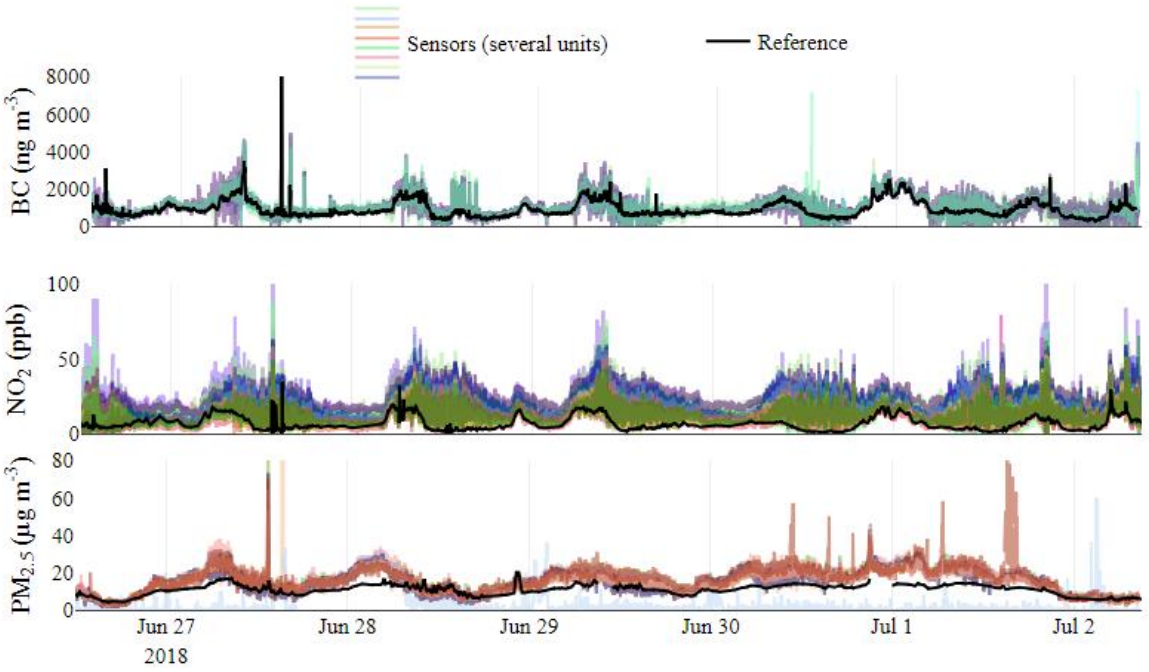
Sensors (several units)
Reference
BC (ng m-3)
8000
6000
4000
2000
0
NO2 (ppb)
100
50
0
PM2.5 (µg m-3)
80
60
40
20
0
Jun 27
2018
Jun 28
Jun 29
Jun 30
Jul 1
Jul 2

Fig.5.

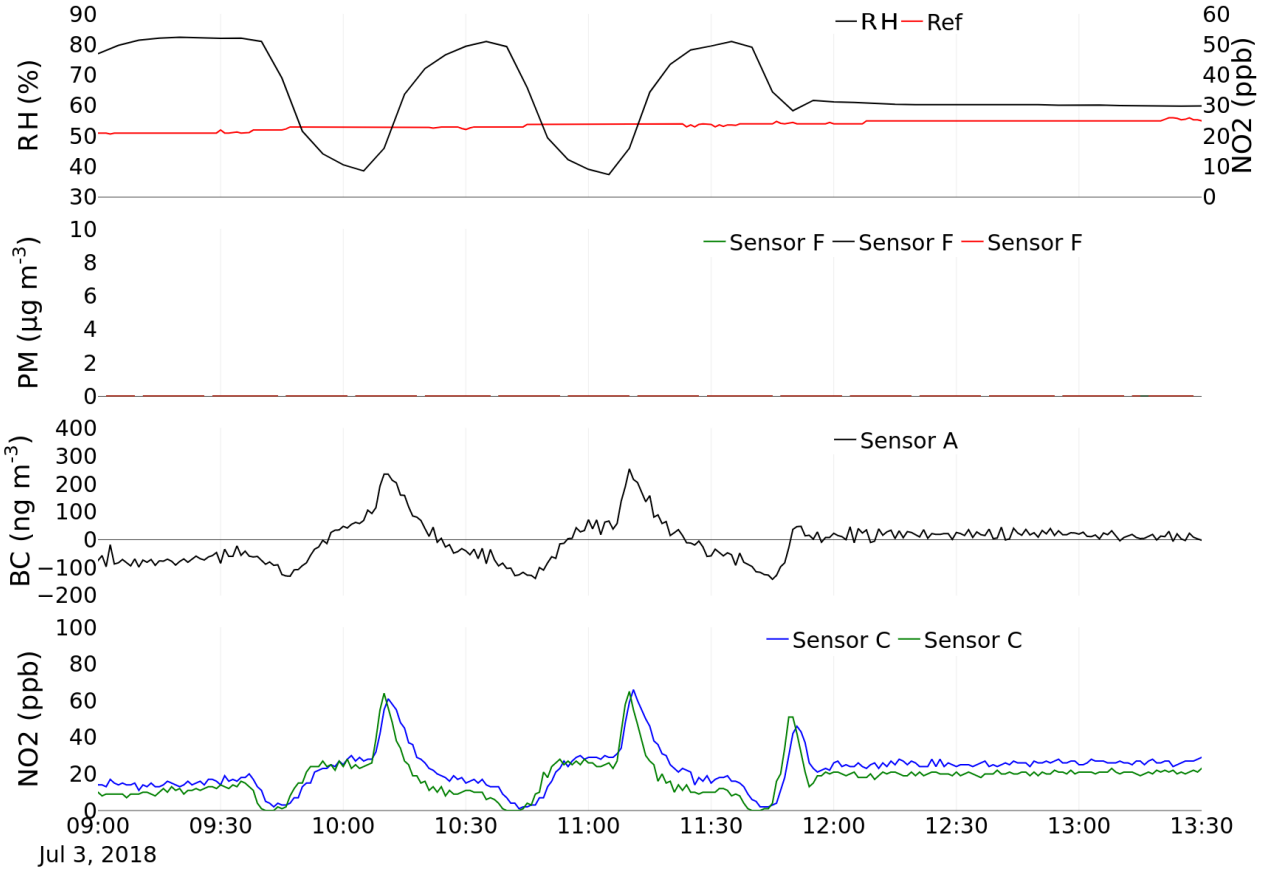
RH
Ref
RH (%)
NO2 (ppb)
Sensor F
Sensor F
Sensor F
PM (µg m-3)
Sensor A
BC (ng m-3)
Sensor C
Sensor C
NO2 (ppb)
09:00
09:30
10:00
10:30
11:00
11:30
12:00
12:30
13:00
13:30
Jul 3, 2018

Fig. 6.

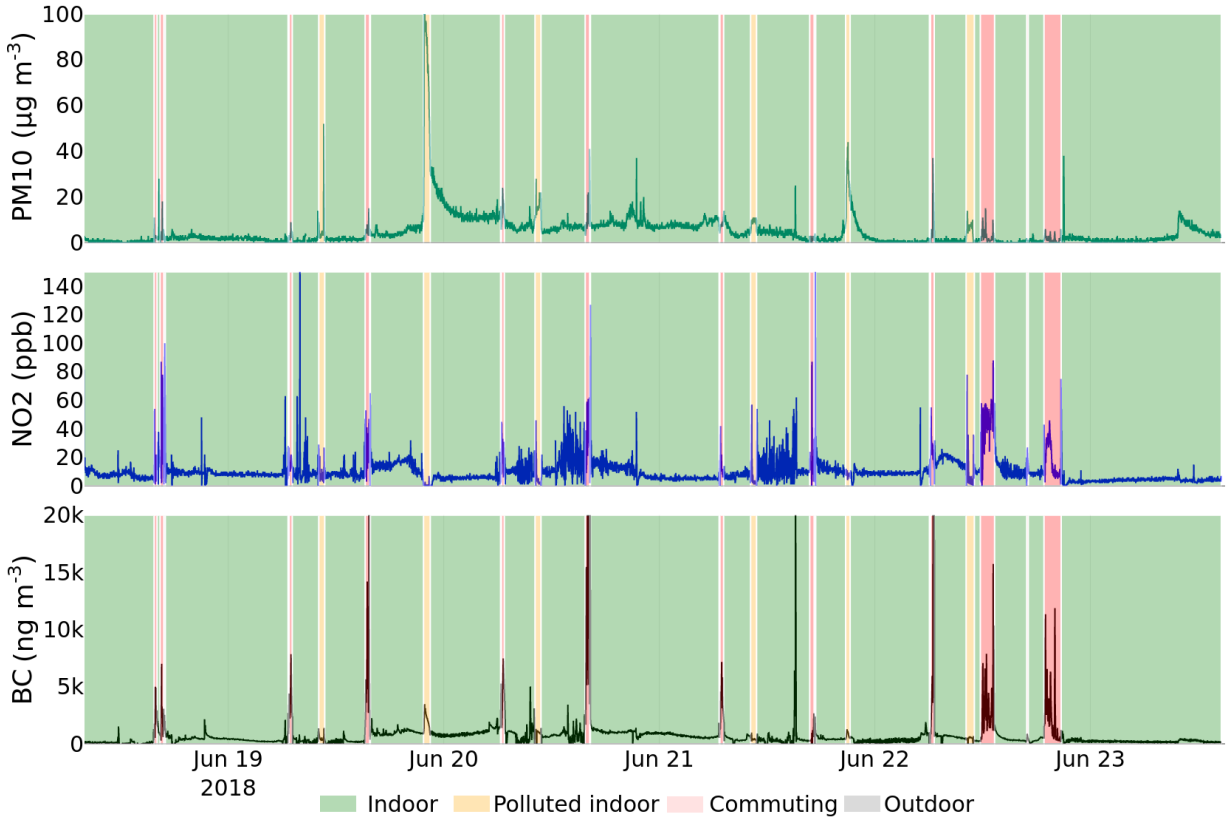
PM10 (µg m-3)
100
80
60
40
20
0
NO2 (ppb)
140
120
100
80
60
40
20
0
BC (ng m-3)
20k
15k
10k
5k
0
Jun 19
2018
Jun 20
Jun 21
Jun 22
Jun 23
Indoor
Polluted indoor
Commuting
Outdoor

Fig. 7.