



HAL
open science

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.

► To cite this version:

Baptiste Languille, Valérie Gros, Nicolas Bonnaire, Clément Pommier, Cécile Honoré, et al.. A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science. *Science of the Total Environment*, 2020, 708, pp.134698. 10.1016/j.scitotenv.2019.134698 . hal-02861698

HAL Id: hal-02861698

<https://hal.science/hal-02861698>

Submitted on 11 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

3
4 **Baptiste Languille^{1*}, Valérie Gros¹, Nicolas Bonnaire¹, Clément Pommier¹,**
5 **Cécile Honoré^{2†}, Christophe Debert², Laurent Gauvin², Salim Srairi³, Isabella**
6 **Annesi-Maesano⁴, Basile Chaix⁵, Karine Zeitouni⁶**

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

9 *²Airparif, Paris, France.*

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

11 *³CEREMA, Trappes-en-Yvelines, France.*

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

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

15 *⁶Université de Versailles Saint-Quentin, Versailles, France.*

16 **Abstract**

17 The field of small air quality sensors is of growing interest within the scientific community,
18 especially because this new technology is liable to improve air pollutant monitoring as well as be
19 used for personal exposure quantification. Amongst the myriad existing devices, the
20 performances are highly variable; this is why the sensors must be rigorously assessed before
21 deployment, according to the intended use. This study is included in the Polluscope project; its
22 purpose is to quantify personal exposure to air pollutants by using portable sensors. This paper
23 designs and applies a methodology for the evaluation of portable air quality sensors to eight
24 devices measuring PM, BC, NO₂ and O₃. The dedicated testing protocol includes static ambient
25 air measurements compared with reference instruments, controlled chamber and mobility tests, as
26 well as reproducibility evaluation. Three sensors (AE51, Cairclip and Canarin) were retained to
27 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

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

33

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

36

37

38 **1 INTRODUCTION**

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

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

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

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

62 Conversely, the main drawback of these devices is their questionable accuracy. In the field of
63 small air quality sensors, several kinds of studies were published, such as sensor development
64 (Hu *et al.*, 2016; Mead *et al.*, 2013; Peng *et al.*, 2013), sensor assessment (Lin *et al.*, 2015;
65 Burkart *et al.*, 2010; Sousan *et al.*, 2016), exploratory measures of personal exposure (Velasco *et*
66 *al.*, 2016; Velasco and Tan, 2016; Hu *et al.*, 2014) and full-scale project involving large field
67 campaigns (Mead *et al.*, 2013; Castell *et al.*, 2015, 2017; Schneider *et al.*, 2017; Hasenfratz *et al.*,
68 2015).

69 Although previous projects were interesting, some limitations invite researchers to keep on
70 improving the methods to quantify personal exposure. First, it is highlighted that most of the
71 small sensors suffer from lower precision than reference instruments, which shows the
72 importance of a robust assessment prior to launching field campaigns. The report by Lewis *et al.*
73 (2018) (not yet published during the experiments presented below) gave the state of the art of the
74 low-cost air quality sensors. However, not all studies pushed the sensor characterization as far as
75 it should be. Second, personal exposure is about measuring air quality as close as possible to the
76 inhaled air, which is why it is of primary importance to give the sensor units directly to people.
77 But some studies used sensor units on static measurements or attached to vehicles (Velasco and
78 Tan, 2016; Castell *et al.*, 2017; Deville Cavellin *et al.*, 2016; Duvall *et al.*, 2016; Fishbain *et al.*,
79 2017; Gao *et al.*, 2015; Holstius *et al.*, 2014), which is not as relevant as asking people to
80 personally carry the sensor units. Finally, personal exposure is relevant in health impact studies
81 for which it is interesting to have several measured pollutants. This last consideration points out
82 single pollutant measurement studies as a limitation.

83 Île-de-France is the Paris region, it is the most densely populated in France but with great
84 disparities between Paris centre and remote places. This region is also characterized as an
85 important atmospheric pollutant emission area. More than one million inhabitants are exposed to
86 nitrogen dioxide concentration exceeding the *limit value*¹ (annual mean of 40 µg m⁻³) and 85 %
87 of the population is exposed to PM_{2.5} levels above the *long-term objective*¹ (annual mean of
88 10 µg m⁻³) (Airparif, 2017, 2018, 2019).

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

96 Polluscope is also defined by the diversity of the studied pollutants as this project will monitor
97 PM (PM₁₀, PM_{2.5} and PM₁), black carbon (BC) and nitrogen dioxide (NO₂). The choice of the
98 monitored pollutants was made according to their impact on health as well as the exceedances
99 experienced in Île-de-France for each pollutant. Indeed, Airparif, the French agency for air
100 quality monitoring, states that some pollutants (PM₁₀, PM_{2.5}, NO₂, O₃, benzene) still exceed the
101 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

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

104 Within Polluscope, the purposes of this paper consist of:

105 •Selecting the sensors in accordance with the project expectations (measurement
106 performance, size, price, etc.); and

107 •Assessing the selected sensors to determine precisely their ability to be used in the
108 project.

109 **2 METHODS**

110 **2.1 Sensor selection protocol**

111 **2.1.1 Pre-selection**

112 Among the main expected sensors specifications (listed in Table 1, measurement ranges were
113 set in accordance with the maximum measured hourly levels in Île-de-France by Airparif from
114 1990 to 2017 (O₃: 170 ppb, NO₂: 270 ppb and PM₁₀: 660 µg m⁻³). BC hourly mean values up to
115 tens of micrograms per cubic meter were measured close to major traffic axes. As one of the
116 objectives of this work is to monitor the exposure in different environments (inside, outside, in a
117 car, etc.) throughout the day, a fine time resolution (below 5 min) was needed. Furthermore, some
118 practical requirements were added due to the need to have sensors able to measure all day long,
119 whatever the season, and which are easy to carry every day of the week. This is why a long
120 enough battery life was necessary, as well as being light weight and having a large working
121 temperature range. In addition to these fundamental specifications, the sensors had to present a
122 detection limit, a precision and an accuracy in accordance with the project goals. Moreover, close

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

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

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

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

134 **2.1.2 Testing strategy**

135 The first step consisted of testing the sensors in static measurements, for up to 15 days, in
136 ambient air, close to fixed reference instruments. These measurements took place at *Site*
137 *instrumental de recherche par télédétection atmosphérique* (SIRTA²) which is an atmospheric
138 observatory belonging to the research infrastructure ACTRIS³. For particulate matter, the
139 reference instruments were Fidas (Palas), TEOM 1405F for PM₁₀ (Thermo Scientific), TEOM
140 1400 for PM_{2.5} and PM₁ (Thermo Electron) and Aethalometer AE33-7 for BC (Magee Scientific).
141 For gas monitoring, the reference instruments were T200UP for nitric oxides and T400 for ozone
142 (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/>

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

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

175 **2.2 Selected sensors assessment**

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

179 **2.2.1 Reproducibility tests**

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

185 **2.2.2 Controlled chamber tests**

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

189 (from approximately 40 % to 80 % relative humidity, the entire cycle lasted one hour, two cycles
190 were conducted) as well as NO₂ concentration (from 0 to 200 ppb). These are the typical orders
191 of magnitude the sensors will have to deal with due to environmental changes during the coming
192 campaigns. NO₂ variation is of course useful to characterize sensors measuring NO₂, but this test
193 was also relevant for other gas sensors, as some of these devices can suffer from NO₂
194 interference.

195 **2.2.3 Feasibility campaign**

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

202 **2.2.4 Final sensor selection**

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

208 **3 RESULTS**

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

211 **3.1 Static ambient air tests**

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

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

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

236 NO₂. The best one was sensor B, with an IPI of 0.76 and a RMSE of 5 ppb. Sensor C gave poor
237 results and even aberrant values highlighted by very low correlation coefficients (below 0.15).
238 Sensor D has a fair correlation coefficient (higher than 0.5) but suffers from a large RMSE
239 (37 ppb), poor match score (0.24) and quite significant data loss (presence of 67 %). Particulate
240 matter was measured by three sensors: sensor B measured PM₁₀ and the sensors E and F
241 measured PM₁₀, PM_{2.5} and PM₁. Sensor F gave the best results for all the PM sizes, with an IPI of
242 0.64, 0.80 and 0.78 for PM₁₀, PM_{2.5} and PM₁, respectively. The others gave a much lower IPI.
243 The major advantage of sensor F is the data availability, which does not suffer from data loss. Its
244 match score is acceptable for the three PM sizes (always larger than 0.6) although this parameter
245 is lower than 0.43 for the others. Even if its RMSE is large, sensor F gives a relevant
246 approximation of the PM concentration.

247 This first static ambient test led us to rule out sensor E, which gave aberrant values for PM₁₀,
248 as well as sensor D because of its non-satisfactory results. Sensor B was kept despite the
249 mediocre results for O₃ (IPI of 0.46) and PM₁₀ (IPI of 0.40), thanks to its multi-pollutant
250 measuring ability and because the producing firm should improve the sensor before the next
251 testing step. Sensor C gave here poor results, but the authors were aware that good results had
252 been obtained with this device, and it was suggested that these unsatisfactory results could be due
253 to an out-of-date electrochemical sensing cell or an inappropriate storage, which would lower the
254 sensor performance. This is why new units of this sensor were purchased for the following testing
255 steps.

256 **3.2 Mobility/reproducibility test**

257 This test involved the following sensors: sensor B, sensor C and sensor F. The BC sensor A
258 was not involved in this mobility tests because studies (Ezani *et al.*, 2018; Lin *et al.*, 2017) have
259 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 NO₂ 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: NO₂ 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 PM₁₀
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 µg 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 µg m⁻³ appears); and for the “RN2” station, the difference between the three
284 units was substantial (about 30 µg m⁻³). Results were better for the sensors F: the sensors were

285 almost always in the range of the reference station, with an inter-sensor difference not larger than
286 $20 \mu\text{g m}^{-3}$. Nevertheless, sensors F results were poor for the “Bobigny” station, for which the
287 difference with the reference went up to $30 \mu\text{g m}^{-3}$. Three units of each sensor (B, C and F) were
288 tested during this mobility steps. Sensors C and F presented the best results.

289 **3.3 Selected sensors**

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

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

299 **3.4 Reproducibility tests in static measurements**

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

309 levels higher than the devices' detection limit. From August 28th to September 4th, fifteen NO₂
310 sensors C were used to measure conditions at the Airparif station close to a major road (Paris ring
311 road). The results overview is presented in Appendix (Fig. S2), the rest of this work is based on
312 the results from this experiment. Concerning the PM sensors F, they were both quite close to the
313 reference and very reproducible to one another, except for one unit plotted in light green, which
314 presented erratic values. For the PM_{2.5}, the mean RMSE is 6 µg m⁻³, which is fairly low
315 compared with the measured concentration in mobility in Paris and its suburbs (often higher than
316 40 µg m⁻³).

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

324 The SET results for the fifteen NO₂ sensor units (Table S3 in appendix) are homogeneous, the
325 IPI spans from 0.75 to 0.79. The mean measured concentration was above 40 ppb, which is
326 higher than the limit of detection. This leads to good correlation coefficients (above 0.76 for the
327 mean Pearson coefficient) and reasonable RMSE compared with the ambient levels (mean RMSE
328 is 14 ppb). Moreover, data loss was very uncommon as shown by the presence parameter.

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

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

338 **3.5 Chamber controlled test**

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

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

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

356 100 ppb, 200 ppb, 0 ppb and a final longer 50 ppb stage over several hours. Inside the chamber,
357 the relative humidity was set at 60 % and there is no particulate matter.

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

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

367 **3.6 Feasibility campaign**

368 The feasibility campaign was conducted with fifteen volunteers from Monday 18th to Friday
369 22nd June 2018. One sensor unit C and one sensor unit F were given to each volunteer and six
370 sensor units A were shared between the participants. This campaign was conceived as a proof of
371 the Polluscope concept, thus, only a limited analysis of the results was done.

372 Globally, the campaign was a success: all the sensor units were worn for the whole week. The
373 data availability (the time resolution was one minute) is 66 % for sensors A and 69 % for sensors
374 F, which can be considered as a satisfying result. The data loss was due to minor problems and
375 routine maintenance (filter change, turning on and off, powering, etc.). However, the data
376 availability for the sensors C only reaches 41 %. This was caused by storage memory erasure
377 when the sensor ran out of power. The coming campaigns protocols will prevent this issue.
378 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 NO₂ 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 (NO₂) and the Canarin (PM₁₀, PM_{2.5} and PM₁) 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.

403 For the static measurement assessments, we used the SET algorithm designed by Fishbain *et al.*
404 (2017), available as an open source resource⁴. In their article, they presented results from 25
405 AQMesh NO₂ sensor units that had taken measurements for about three months in static positions.
406 The mean related IPI is 0.58, this is very close to our mean Cairclip IPI (from the reproducibility
407 test) of 0.54; even if the Cairclip is designed for mobile measurements whereas the AQMesh
408 system is designed for static monitoring (i.e. expecting to have a better performance than a
409 portable device). Knowing the successful deployment of the AQMesh sensor units, this result
410 demonstrates the reliability of the Cairclip. Fishbain *et al.* (2017) also used the SET with PM
411 sensors (DC1700 Dylos and GeoTech), the mean resulting IPI is 0.63. The PM_{2.5} Canarin sensor
412 used in our study gave significantly better results with a mean IPI of 0.73.

413 The SET algorithm was also used by Broday and the Citi-Sense Project Collaborators (2017).
414 They presented unpublished results from about three months of ambient air measurements of six
415 PM₁₀ sensor units located in Ostrava, Czech Republic. The mean IPI is 0.72, which is very close
416 to the result from our study (0.73).

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

421 **4.1 Static comparison**

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

⁴ fishbain.net.technion.ac.il

424 station in static measurements, the correlation coefficient was above 0.75. In our study, the mean
425 correlation coefficient was 0.80. This higher agreement may be due to higher inlet flow
426 (150 mL min⁻¹ compared with 100 mL min⁻¹ in the Viana *et al.* (2015) study) or coarser time
427 resolution (1 min in our study and 1 sec in the Lin *et al.* (2017) study).

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

430 The recent low-cost sensors review by the World Metrological Organization (Lewis *et al.*,
431 2018) described several performance evaluation programs as the work supported by the United
432 States Environment Protection Agency (EPA)⁵; they described an air sensor toolbox where the
433 main performances of tens of sensors were gathered and compared. For the Cairclip NO₂ sensor,
434 the EPA and Jiao *et al.* (2016) state that a correlation coefficient between 0.42 and 0.76 was
435 obtained with reference instruments. Our mean Pearson correlation was 0.76, which is in the high
436 part of the EPA range. The additional information given by the SET algorithm in our study is the
437 good match score of 0.64 and the absence of data loss (the presence parameter reaches almost
438 100 %). Another example is the study conducted by Spinelle *et al.* (2015). They found a
439 correlation of up to 0.75 for the Cairclip sensor units. This result is both in the EPA range and
440 close to our result of 0.76.

441 Due to its new release, only a few research works including Canarin have been conducted. For
442 instance, Tse *et al.*, 2018B presented a project based on static measurements from four Canarin
443 units. Some tests were conducted in Bologna, Italia, and PM₁₀ maps have been produced. These
444 works were preliminary and the most accomplished article about Canarin sensor is certainly the
445 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>

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

455 **4.2 Mobile measurements**

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

463 The Ezani *et al.* (2018) study was based on mobile measurements performed with two AE51
464 units. There were no reference instruments but the correlation between the two units was good:
465 0.92. Lin *et al.* (2017) performed mobile AE51 measurements. Comparison was possible thanks
466 to 17 transient immobile periods (of less than one hour) nearby reference stations. The AE51 unit
467 showed an interquartile range agreement with the reference instrument of 82 %. High-resolution
468 mapping is possible with the AE51, as in the study of den Bossche *et al.* (2015) where sensor
469 units were mounted on bikes in Antwerp, Belgium. A 50-meter resolution was obtained with an

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

473 Few studies have been published on the Cairclip sensor being used in mobile measurements,
474 especially compared with the abundant literature related to the AE51. The recent work by
475 Chambers *et al.* (2018) found no consistent relationship between NO₂ concentrations and health
476 parameters. The authors state that the Cairclip was able to appropriately monitor personal
477 exposure and a clear diurnal cycle was observed but no more validation data was provided. The
478 study by Reid (2015) is based on the qualification of Cairclip sensors. Mobile measurements
479 were conducted with two Cairclip units in different environments: public transport, outdoor and
480 indoor. The sensors monitored interesting variability, especially close to traffic.

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

482 **4.3 Methodology discussion**

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

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

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

500 **4.4 Conclusions**

501 No remote sensor is perfect, and the three selected ones are the result of compromises and each
502 have strengths and weaknesses. The AE51 is accurate, its IPI (above 0.8) is higher than all other
503 sensors. This BC sensor is also reliable and easy to use with very little data loss. But it is
504 sensitive to humidity, which leads to some artefacts when quick environmental changes occur,
505 and its high price is also a weakness because fewer units can be purchased. The Cairclip is very
506 light and thus easy to carry all day long, but the storage memory is erased if the sensor unit runs
507 out of battery, thus it is consequently more demanding for the operator. Moreover, even if these
508 sensors demonstrated their ability to perform reliable measurements in mobile measurement (see
509 related section), the detection limit (20 ppb) is not appropriate for low NO₂ levels. The Canarin is
510 able to send data via Wi-Fi and has a high storage capacity (several weeks of measurements),
511 which is useful when the data sending is not possible. Its robustness in mobility is also an
512 important advantage: the Canarin was the sensor that presented the highest data availability
513 during the feasibility campaign (69 %). This sensor showed satisfying results for the PM_{2.5}
514 measurements (IPI of 0.7) but substantially lower for PM₁₀ (IPI of 0.4). Finally, the weight of the
515 sensor unit (it is quite heavy) is a drawback.

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

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

521 •Several kinds of tests were performed. Ambient air static measurement tests against
522 reference instruments was the first assessment of the sensors in real atmosphere with
523 natural meteorological parameters variability (temperature, humidity, wind speed and
524 direction, etc.) The laboratory tests were of primary importance to quantify the sensors;
525 responses to rapid atmospheric changes (i.e., humidity, pollutants levels). Mobile
526 measurements were necessary as the project goal is to use the sensors for personal
527 exposure, to be worn by volunteers.

528 •A large number of units of each sensor was tested to quantify the reproducibility and to
529 eliminate problems arising from a single deficient unit.

530 •A robust multi-metric static measurement assessment with the SET algorithm was
531 conducted in order to be as rigorous as possible in the assessment.

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

537

538

539 **ACKNOWLEDGEMENTS**

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

545

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., Moshhammer, H., Neuberger, M. and Hitzemberger, 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., Draher, 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

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

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

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

620 Janssen, N.A., Hoek, G., Simic-Lawson, M., Fischer, P., van Bree, L., ten Brink, H., Keuken, M.,
621 Atkinson, R.W., Anderson, H.R., Brunekreef, B. and Cassee, F.R. (2011). Black carbon as an
622 additional indicator of the adverse health effects of airborne particles compared with PM₁₀ and
623 PM_{2.5}. *Environ. Health Perspect.* 119(12): 1691–1699.

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

627 Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M.,
628 Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S. and Buckley, K. (2016).
629 Community air sensor network (Cairsense) project: evaluation of low-cost sensor performance
630 in a suburban environment in the south-eastern United States. *Atmos. Meas. Tech.* 9(11): 5281–
631 5292.

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

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

638 *Meteorological Organization*, 1215.

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

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

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

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

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

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

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

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

660 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 O₃/NO₂ 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*

684 *International Conference on Smart Objects and Technologies for Social Good, Goodtechs 55-*
685 *58.*

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

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

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

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

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

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

699

1 **Table titles**

2 **Table 1.** Tested sensor specifications.

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

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

11

12

13

14 **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

15

16

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

21 **Table 3.** SET results for every sensor, the units are "ppb" for gases, ng m^{-3} for BC and $\mu\text{g m}^{-3}$ for PM. Pollutant sensor: measured
 22 pollutant and sensor's name; M : mean concentration; Match: match score; RMSE: root mean squared error; r : Pearson correlation
 23 coefficient; t : Kendall correlation coefficient; S : Spearman correlation coefficient; Pres: presence parameter; LFE: low frequencies
 24 energy parameter; IPI: SET integrated performance index; SensTB: sensor time base; RefTB: reference time base; Start time:
 25 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

1

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

4 **Fig. 2.** BC and NO₂ time series in static measurements.

5 **Fig. 3.** Mobility tests for NO₂ sensors. The three sensors C are plotted in shades of green, the
6 three sensors B in shades of orange, the reference Airparif stations are plotted in black.

7 **Fig. 4.** Mobility tests for PM₁₀ sensors. The three sensors F are plotted in shades of blue, the three
8 sensors B in shades of orange, the reference Airparif stations are plotted in black.

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

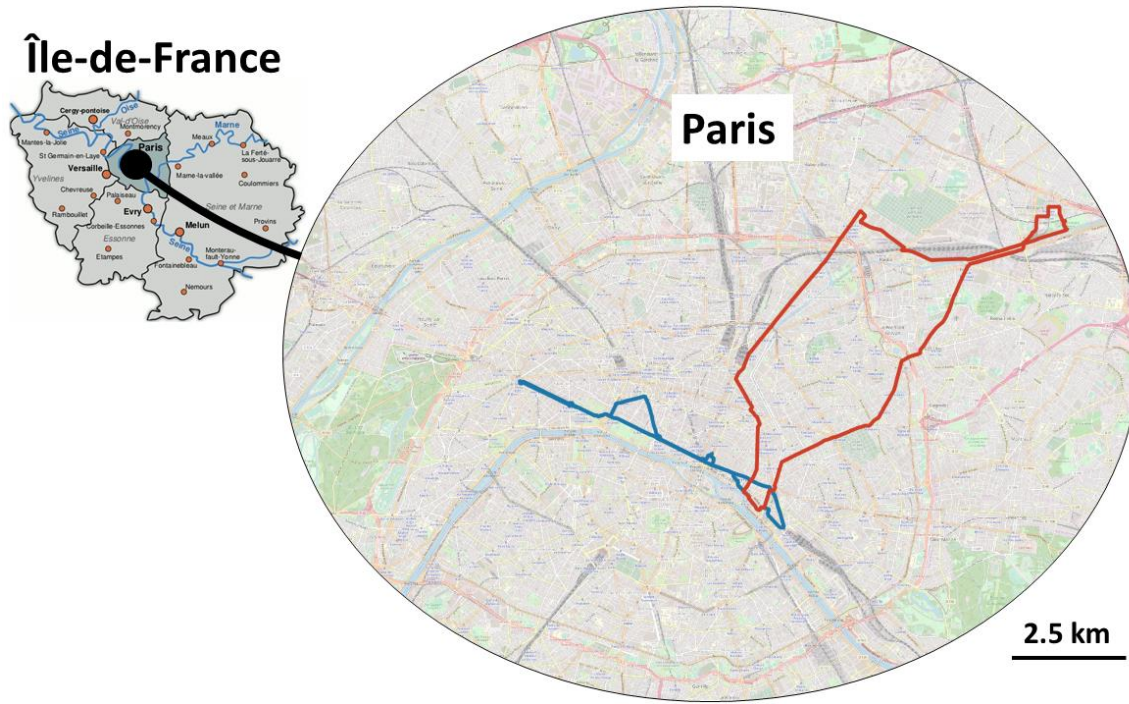
10 **Fig. 6.** Chamber controlled tests for BC sensor A, PM sensor F and NO₂ sensor C.

11 **Fig. 7.** Mobility test results.

12

13

14



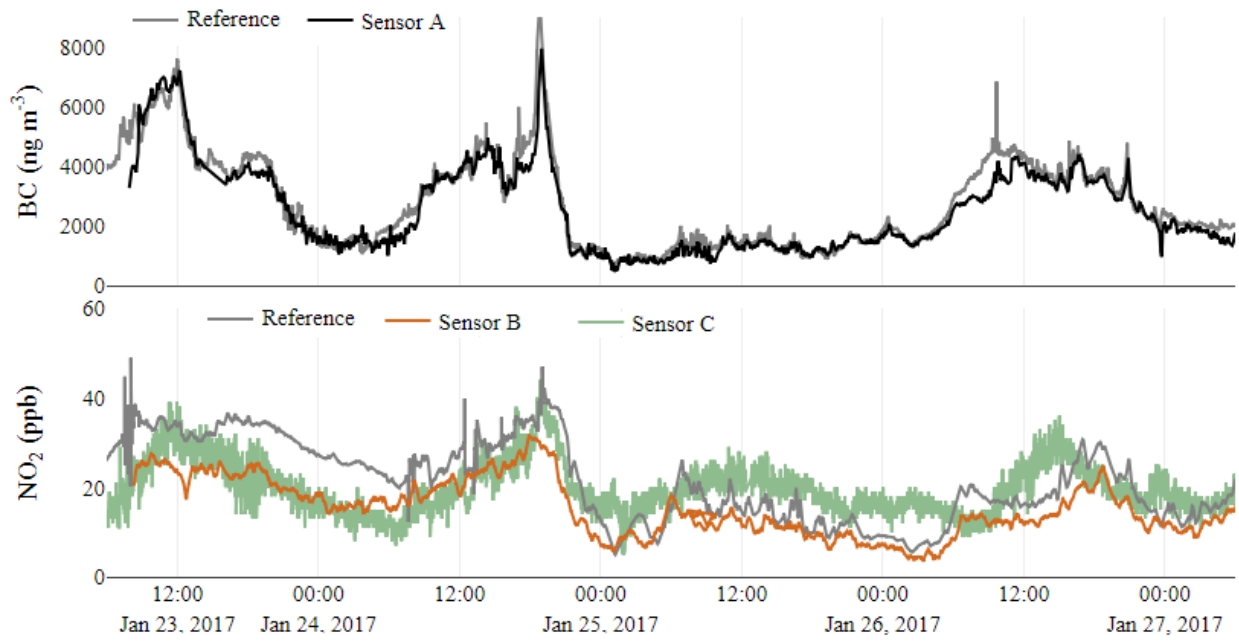
15

16

17

Fig. 1.

18



19

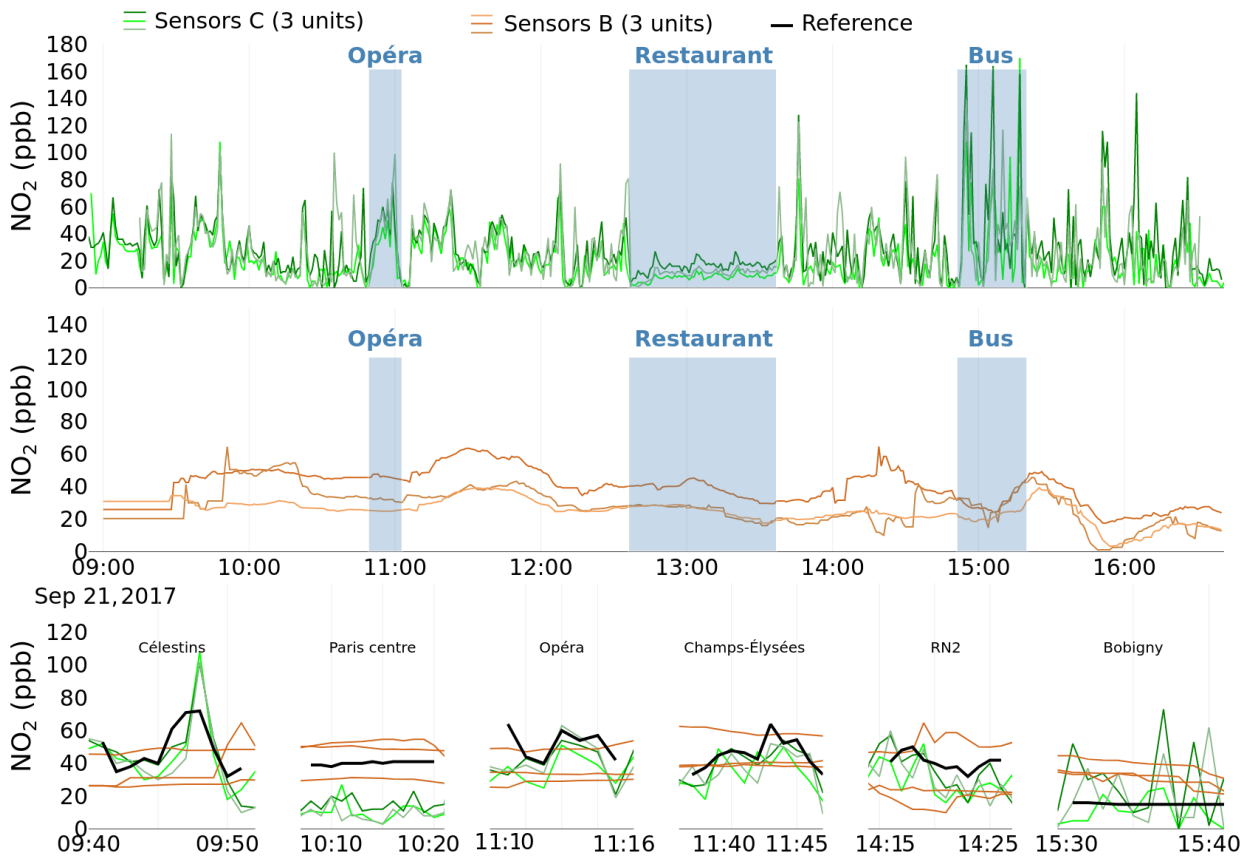
20

21

22

Fig. 2.

23



24

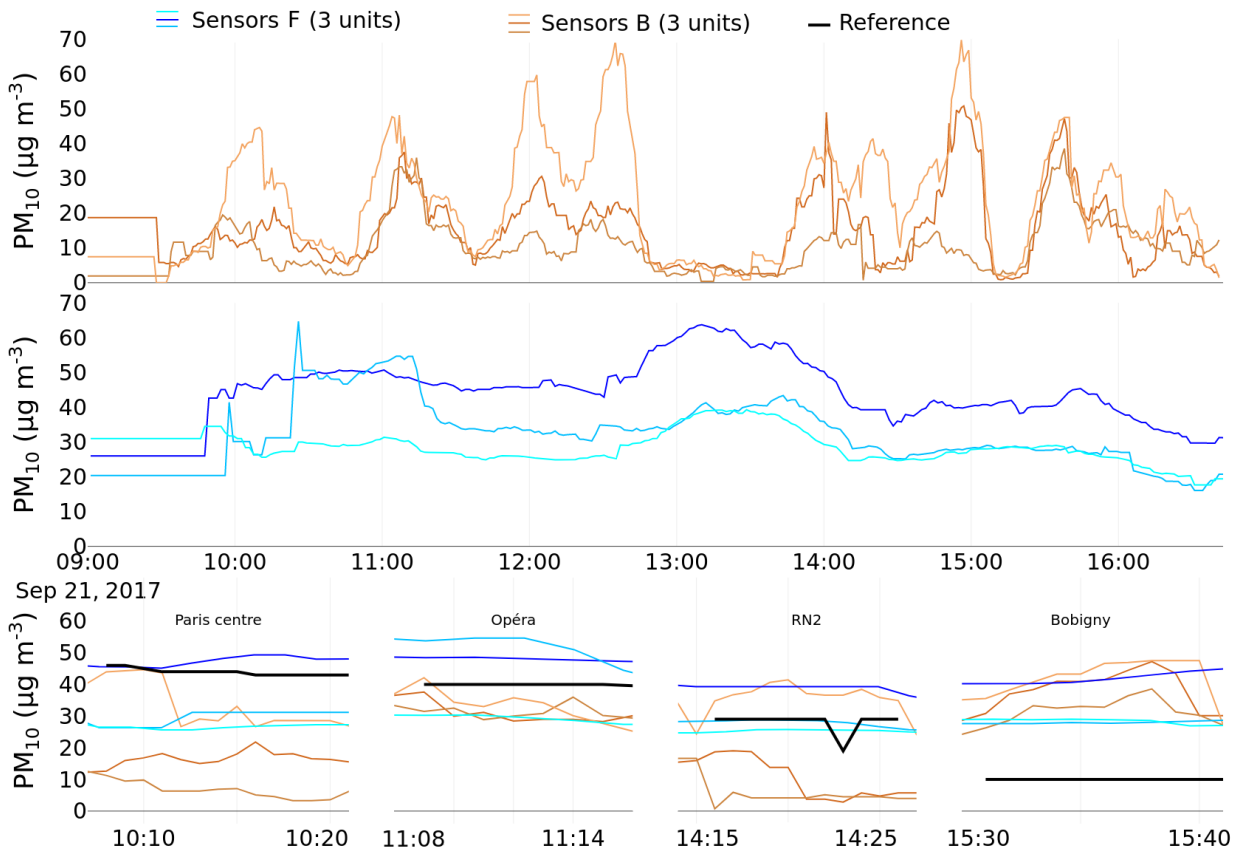
25

26

27

Fig. 3.

28



29

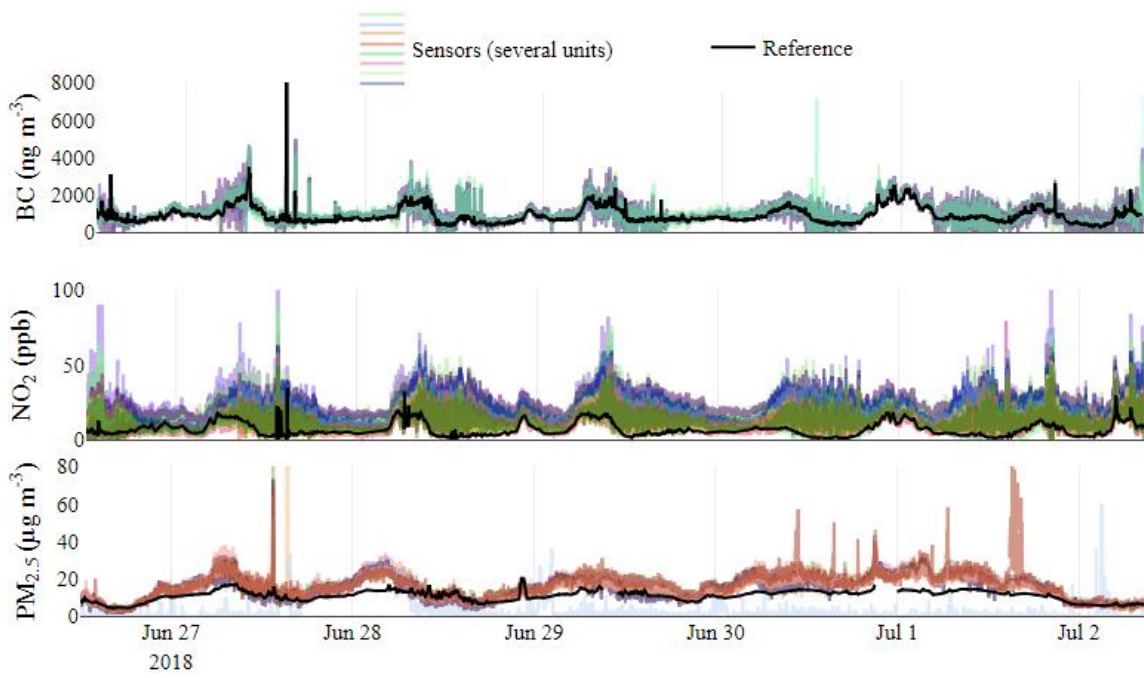
30

31

32

Fig. 4.

33



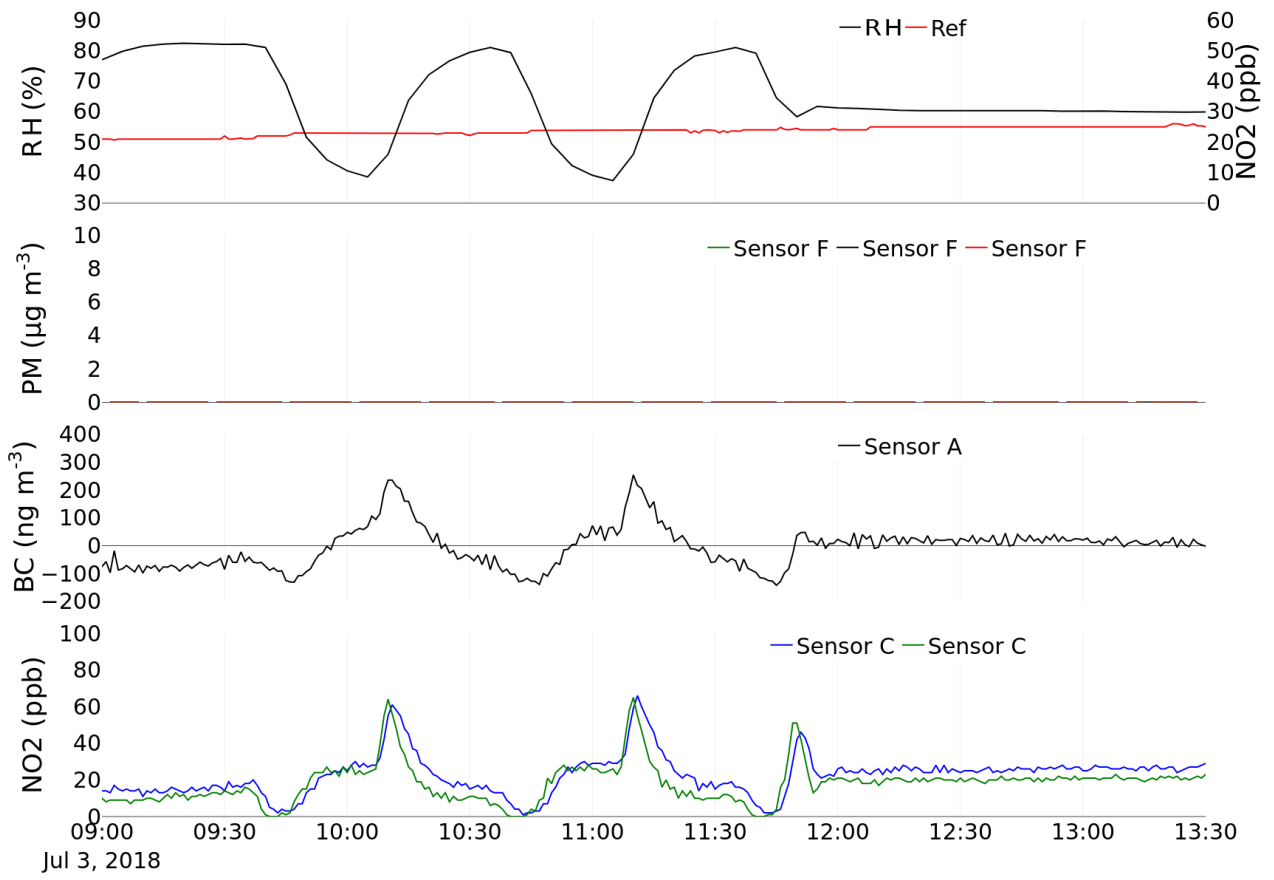
34

35

36

Fig.5.

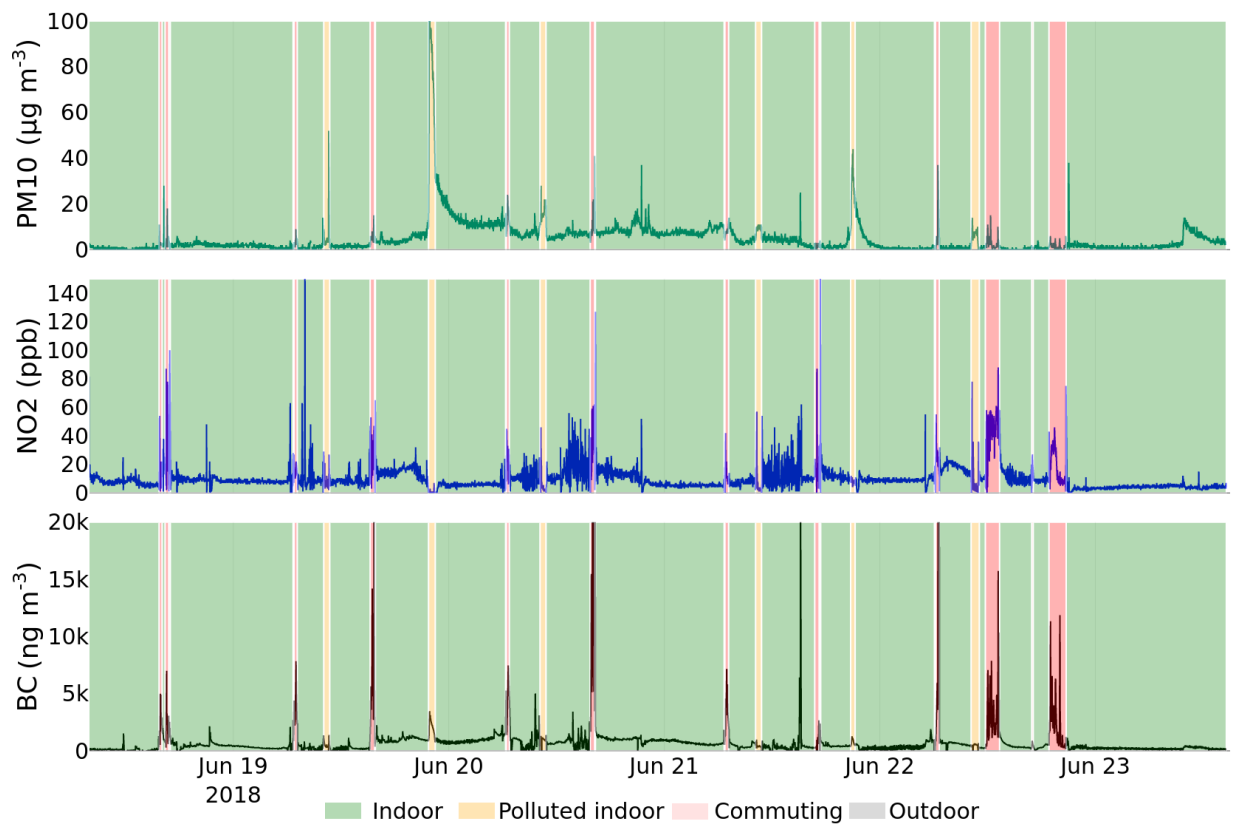
37
38



39
40
41
42

Fig. 6.

43
44
45



46
47
48

Fig. 7.