# Breathe London wearable sensor evaluation - PAM

Authors: Dr Benjamin Barratt and Shanon Lim

# Background – the importance of portable sensor testing

Personal air quality exposure assessment is a growing area of research whereby small portable sensors are provided to individuals to better estimate the air pollution people breathe. An important aspect of this type of research is the reliability, accuracy and precision of the sensors. Evaluation of portable sensors are typically completed with two tests, the first in a fixed location to test accuracy against an approved reference monitor and the second a mobile test to evaluate their ability to measure air pollution fluctuations in different environments.

Testing is carried out in urban, real world environments so key performance metrics can be determined to evaluate the sensor's fitness of purpose. Sensors are co-located with reference monitors at a dedicated air quality monitoring station for a sustained period of time to provide an assessment of the sensor's performance relative to a reference monitor and in response to changing field conditions. Besides evaluating sensor reliability, this extended evaluation period provides critical temporal data enabling the full characterisation of a sensor's performance in a specific type of dynamic outdoor environment where meteorology and concentrations of target and interfering species are subject to change.

The initial phase of a sensor's testing regime aims to evaluate the sensor's capacity for continuous, un-interrupted data capture, its inter-unit precision and the comparability of its data with reference monitor outputs. Raw data capture rates are used as an indicator of reliable sensor function and the robustness to withstand dynamic urban environments. If the above phase is satisfactorily passed, appropriate correction factors can be derived to calibrate the sensor against the specific reference monitor used during the test. After comparison to a fixed reference monitor, the sensors are also evaluated in a short mobile monitoring campaign to test how they respond in different pollution environments and commuting modes.

This document is one of a series detailing results of portable air quality sensor testing carried out as part of the selection process for the Breathe London Wearables study funded by the Greater London Authority.

## Introduction – sensor testing protocol

The Breathe London Wearables study is a public engagement campaign that aims to characterise London school children's exposure to air pollution and present this information in a way that the school community can understand, relate and act upon. In order to achieve the study's objectives, a suitable wearable air pollution sensor had to be identified, tested and selected. The sensor requirements were as follows:

1. Monitor $PM_{2.5}$ pollutant concentrations and GPS position at a time resolution of at least 1 minute. Monitored nitrogen dioxide ($NO_2$) concentrations were also desirable, but not essential.
2. Small and light enough to be carried by school children aged 5 – 11 years.
3. Battery life of at least 10 hours to cover a full school day.
4. Sufficiently low cost to allow at least 20 units to be deployed within a budget of £20,000.
5. Sufficiently robust and reliable to deliver valid results despite potentially rough treatment by children.
6. Demonstrable accuracy and precision sufficient to allow robust comparison between sensors and illustrate spatial variation in pollutant concentrations.

Six sensors appeared to meet these criteria and were selected for testing; (i) Plume Flow, (ii) Airbeam2, (iii) University of Cambridge PAM and (iv) Dyson wearable sensor. The suppliers of the two remaining sensor units were not able to supply test units in time for the trial, so these were dropped. A predefined testing protocol was followed for each sensor to ensure fair treatment and transferability of outcomes. The purpose of the protocol was to independently verify that the wearable sensor was able to demonstrate performance characteristics able to deliver the aims of the project. It also allowed us to identify sensor features and limitations, which would influence the design of the subsequent sensor deployments.

The testing protocol included two phases – a static test and a mobile test. The static test ran from 6 October to 29 October 2018. Three sensor units of each type were placed within a Stevenson's screen within one metre of the inlet of a $PM_{2.5}$ FDMS (Filter Dynamics Measurement System) reference monitor at the Marylebone Road kerbside research monitoring site ([www.londonair.org.uk/london/asp/publicdetails.asp?site=MY7](www.londonair.org.uk/london/asp/publicdetails.asp?site=MY7)). Sensor measurements were extracted from the units and a series of statistical tests performed on the data. The first 24 hours data were excluded to allow a settling in period.

The mobile test was carried out on 29 October 2018. This comprised a one-hour test journey on a prescribed route across London from Marylebone Road to Waterloo, incorporating contrasting environments (parkland and busy congested traffic routes). The first half of the journey was carried out by foot, the second half in a diesel taxi. The sensors were assessed based on the inter-unit comparability and how the sensors responded in different pollution environments compared to expected spatial patterns.

To provide an overall assessment, each sensor was given a rating for aesthetics, bulk, setup, reliability, usability, precision, accuracy, GPS and cost. Double weighting was applied to precision and accuracy categories reflecting their importance. A separate report was produced for each unit type detailing performance against each test and their overall assessment rating.

This report gives the results of the evaluation of the PAM $PM_{2.5}$ sensor units. As the device also measured $NO_2$, a short note on the reliability, accuracy and precision of the $NO_2$ sensor is included at the end of the report.

# Results

## Capture rates (reliability)

This table describes the percentage of valid one-minute readings logged by the sensors. Data loss may be caused by breakdown of sensor, logging or communication system. The target is 100%.

**Table 1: Valid data capture rates (% based on 1-minute readings). Capture rates less than 90% are highlighted in red.**

| Week Commencing | PAM052 / % | PAM053 / % | PAM063 / % |
|---|---|---|---|
| **08-Oct-18** | 85 | 76 | 87 |
| **15-Oct-18** | 76 | 41 | 41 |
| **22-Oct-18** | 14 | 85 | - |
| **Full period** | 58 | 67 | 42 |

All PAM sensors had capture rates of less than 90%. It appeared that the sensors worked reasonably well sporadically on individual weeks with capture rates of all devices being higher than 85% on given weeks, however they did not have reliable capture rate over the entire three-week period. Given the pattern of missing data, it is likely that the gaps were caused by communications problems, rather than component failure, therefore, data may have been retrievable using other methods.

## Inter-unit correlations (precision)

Table 2 indicates the degree of correlation between the three sensors tested, describing the level of inter-unit precision. Precision is important to assess the likelihood that additional untested units perform in the same way as tested units and transferability of derived correction/scaling factors. Inter-comparability between devices is particularly important when comparing exposures between different individuals in studies. Results are presented as Reduced Major Axis correlation (RMA) coefficient ($R^2$). The target is 1.00.

**Table 2: Correlation coefficient ($R^2$) between units. Coefficients of less than 0.75 are highlighted in red.**

| $R^2$ (RMA) | PAM052 / % | PAM053 / % | PAM063 / % |
|---|---|---|---|
| **PAM052** | - | 0.98 | 0.98 |
| **PAM053** | 0.98 | - | 0.99 |
| **PAM063** | 0.98 | 0.99 | - |

All PAM sensors demonstrated an extremely high degree of precision (approximately 98% of the variation in one unit was explained by any other unit).

## Correlation coefficient against reference monitor (accuracy)

This table describes the degree of agreement between each sensor unit and the reference $PM_{2.5}$ monitor. The target is 1.00, which would indicate that 100% of the variation in $PM_{2.5}$ was described by the sensor unit.

**Table 3: Correlation coefficient ($R^2$) in comparison with reference monitors. $R^2$ values less than 0.75 are highlighted in red.**

| Week Commencing | PAM052 / % | PAM053 / % | PAM063/ % |
|---|---|---|---|
| **08-Oct-18** | 0.77 | 0.86 | 0.85 |
| **15-Oct-18** | 0.28 | 0.72 | 0.41 |
| **22-Oct-18** | 0.83 | 0.77 | n.a. |
| **Full period** | 0.51 | 0.73 | 0.73 |

Accuracy (in terms of correlation against the reference monitor) varied from week to week, with week 1 and week 3 (8 and 22 October 2018) showing high accuracy (0.77 to 0.86). Week 2 accuracy was comparatively poor for all sensors with accuracy between 0.28 to 0.72. This change is most likely due to variation in meteorological conditions. Week to week accuracy must be considered in the context of week to week correction, indicated in the next section.

## Scaling factor relative to reference monitor (scale correction)

This table shows the multiplication factor required to scale the sensor to the reference monitor. This is calculated using linear regression (y = mx + c, where m is the scaling factor, x is the reference monitor, c is the offset and y is the portable sensor). The target for m is 1.0.

**Table 4: Scaling factor relative to reference monitors (based on hourly readings).**
**Sensor reporting capture rates less than 50% or with a correlation coefficient less than 0.5 are marked 'n.a.'**

| Week Commencing | PAM052 | PAM053 | PAM063 |
|---|---|---|---|
| **08-Oct-18** | 0.50 | 0.91 | 0.94 |
| **15-Oct-18** | n.a. | n.a. | n.a. |
| **22-Oct-18** | n.a. | 0.52 | n.a. |
| **Full period** | 0.44 | 0.62 | 0.96 |

Correction factors (offset and scaling), are a normal part of an instrument scaling procedure, but to be effective, they must be stable over time and across a range of ambient conditions. The greater the accuracy (correlation against reference monitor) the more reliable the scaling correction factor will be.

All sensors under-read in comparison with the reference monitor by a factor of between 0.4 and 1.0. The inconsistency in this scaling factor most likely reflects week to week variation in meteorological conditions and particulate composition. This variation in conditions will affect particle density and size distribution, which are essential components in the conversion from particle number count (measured by the sensor hardware) to particle mass (reported as the desirable metric, $PM_{2.5}$). There were also some weeks which failed to meet the

criteria of greater than 50% capture rate and correlation coefficient greater than 0.5, reflecting poor accuracy and reliability.

## Offset from reference monitor (offset correction)

This table shows the mean offset difference between the sensor and the reference monitor calculated using linear regression (y = mx + c, where c is the offset). The target for c is 0.

**Table 5: Offset from reference monitors (based on hourly mean readings).**

| Week Commencing | PAM052 | PAM053 | PAM063 |
|---|---|---|---|
| **08-Oct-18** | -3.1 | -8.6 | -8.4 |
| **15-Oct-18** | n.a. | n.a. | n.a. |
| **22-Oct-18** | n.a. | 0.6 | n.a. |
| **Full period** | -1.3 | -2.3 | -7.8 |

Offset correction is the second component of the instrument correction procedure. These values indicate that the sensor units have a consistent positive or negative offset from zero. The greater the accuracy (correlation against reference monitor) the more reliable the offset correction factor will be. The sensor units had a baseline under-estimation of $PM_{2.5}$ of between 1 and 9 µg m$^{-3}$.

## Hourly mean time series

A time series chart comparing each sensor against the reference monitor over the three-week testing period is presented prior to and following application of the full period correction factors.
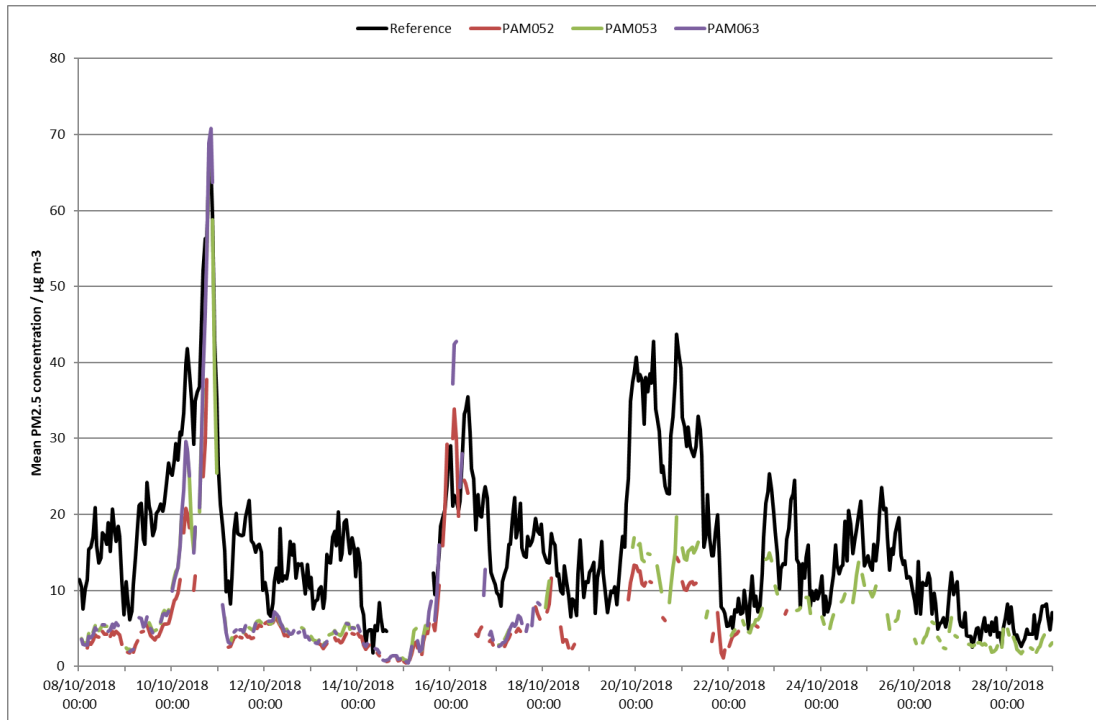


**Figure 1: Time series chart of hourly mean sensor and reference PM$_{2.5}$ concentrations over the three week test period prior to application of scaling factors.**
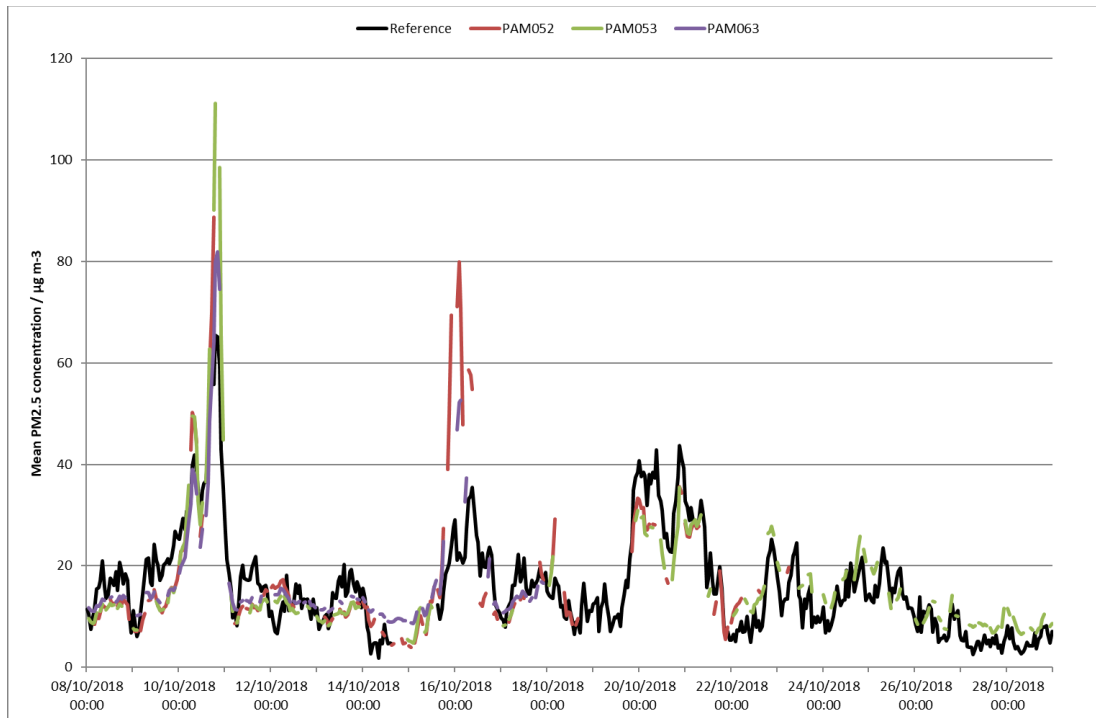


**Figure 2: Time series chart of hourly mean sensor and reference PM$_{2.5}$ concentrations over the three week test period following application of scaling factors.**

Values were scaled using the correction factors for the full period (Scaled = (Raw - c) / m).  It can be seen from Figure 2 that these factors produce a reasonable representation of $PM_{2.5}$ at low concentrations as recorded by the reference monitor. However, at high levels the correction factors overestimated $PM_{2.5}$ levels, suggesting that these correction factors cannot be used in all pollution environments.

## Mobile monitoring evaluation

Mobile monitoring was conducted for a period of just over an hour. A map of the route and concentrations is shown below (Figure 3). To provide different modes of travel and environments the first half of the journey was completed by walking through a park and congested street canyon and the second half completed by taxi. All three PAM devices were taken on the mobile monitoring campaign, however only PAM053 appeared to record concentrations.
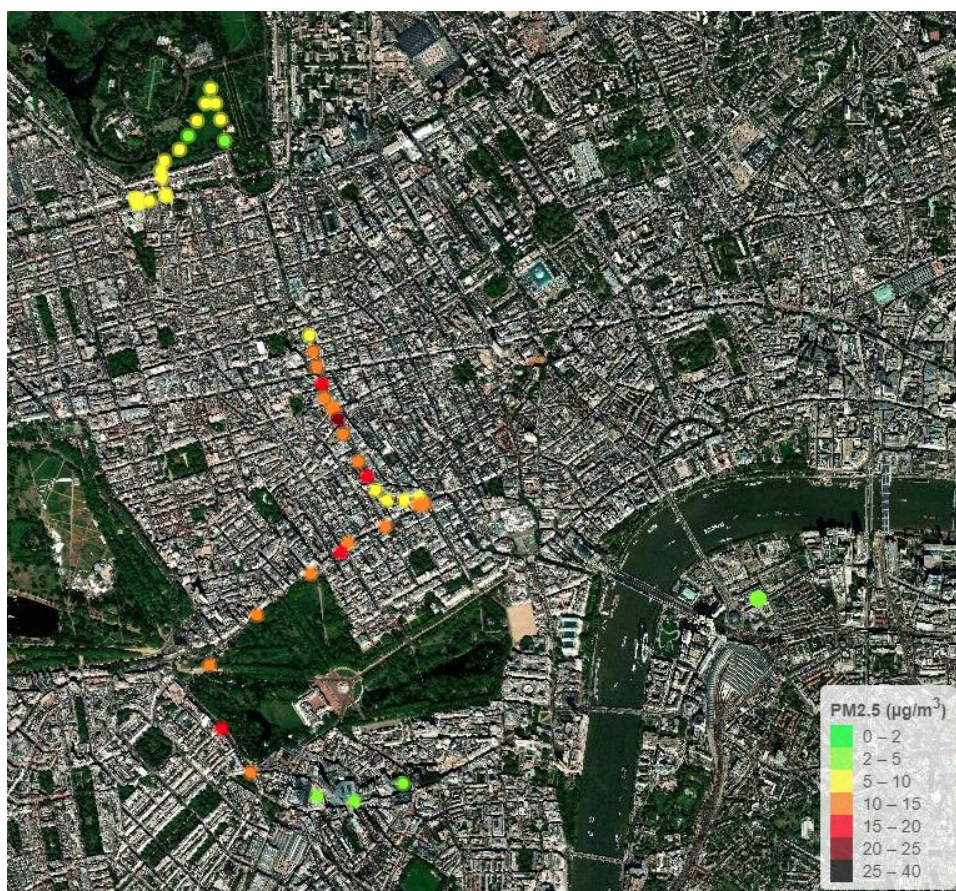


**Figure 3: Map of mobile monitoring test for PAM053, from Marylebone Road (North) to Stamford Street (South), London. Each dot represents one-minute of concentration.**

The spatial pattern recorded by the unit was somewhat consistent with expectations, with peak concentrations recorded in the street canyon and in congested traffic. Lower concentrations were recorded in the park and free flowing traffic. The lowest concentrations were recorded indoors within an office at the end of the mobile test. There were however some issues with reliability of the GPS as locations were not recorded for the entire monitoring campaign indicated by the gaps of points in Figure 3.

The timeseries of the mobile monitoring test presented in Figure 4. Increases in concentrations could be observed when walking in a busy street canyon around 15:10 to 15:30 and while traveling by taxi at 15:30 to 15:40. There is uncertainty as to why exposures dropped before the end of the taxi trip at around 15:40.



**Figure 4: Time series chart of one-minute sensor PM$_{2.5}$ concentrations over the one-hour mobile monitoring campaign.**

## Overall sensor evaluation for PM$_{2.5}$

**Table 6: Scoring results out of 55 (0 – 5, 0 is low, accuracy and precision given double weighting):**

| Sensor | Aesthetics | Bulk | Setup | Reliability | Usability | Precision | Accuracy | GPS | Cost | Total Score |
|---|---|---|---|---|---|---|---|---|---|---|
| **PAM** | 1 | 1 | 3 | 2 | 4 | 4 | 6 | 2 | 1 | 24 |

Results of the PAM PM$_{2.5}$ sensor came back as a good sensor, due to its high scores in precision, accuracy, setup and usability. The device is comparatively large to other sensors tested and would be inconvenient for a child to carry, therefore it scored poorly in bulk and aesthetics fields. The GPS data during the mobile test recorded sporadically and therefore also had a low score in this category. The cost of the commercial version of the sensor unit was higher than specified in the budget (£1,000 per unit), so units would have to be loaned to be included in the study, therefore it also scored low in the cost category.

## NO₂ sensor evaluation

**Table 7: Summary of NO₂ sensor evaluation against reference monitor at Marylebone Road for three-week period from 8 to 29 October 2019.**

|  | PAM052 | PAM053 | PAM063 |
|---|---|---|---|
| **Capture rate** | 60% | 85% | 42% |
| **Correlation ($R^2$)** | 0.71 | 0.73 | 0.68 |
| **Scaling factor** | 0.12 | 0.09 | 0.10 |
| **Offset** | 10.3 | -1.9 | -7.3 |

As the NO₂ sensor evaluation was not the primary aim of this report, only summary results are presented. Results for the three-week monitoring period are shown in Table 8. Correlation with the chemiluminescent reference monitor was reasonable for the three sensors over the monitoring period (0.68 to 0.73), however PAM052 and PAM063 had poor capture rates of less than 60%. While the correlation was reasonable, the scaling factor was substantially below 1, indicating the sensors recorded very low NO₂ values compared to reference.

Mobile monitoring was also conducted at the same time as the PM sensor. A map of the route and NO₂ concentrations is shown below (Figure 6). The spatial pattern of NO₂ should be similar to the PM sensors, as both pollutants have similar sources in the London urban environment. However, PAM053 did not appear to show similar spatial patterns to PM and had periods of missing data.
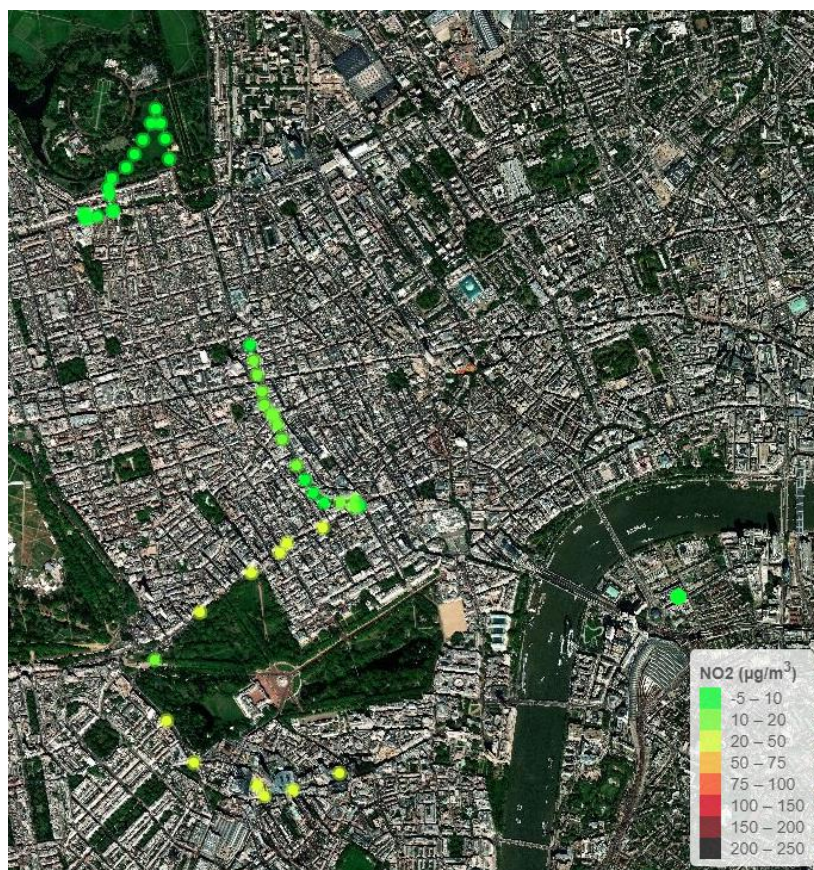


**Figure 5: Map of NO₂ mobile monitoring test for PAM053, from Marylebone Road (North) to Stamford Street (South), London. Each dot represents one-minute of concentration.**

# Conclusion

Three portable PAM sensors were put through rigorous testing to assess their suitability for personal exposure monitoring campaigns. The evaluation consisted of two tests, comparison to a fixed reference monitor for a period of three weeks and a one-hour mobile monitoring test assessing the portable sensors performance in different pollution environments.

The sensor unit showed very good precision between sensors and reasonable accuracy when measuring PM compared to the fixed reference monitor. There were slight issues with reliability as there were sporadic gaps in data recorded over the three-week fixed testing period. The mobile monitoring test revealed a good response to PM levels in different pollution environments. From these tests the PM sensors were precise and could be used in personal exposure monitoring when comparing between individuals, however corrected data appeared to overestimate PM at high concentrations. The bulk of the device made it unsuitable to carry for children, but it could be used in studies for adults where the sensor does not need to be carried for long periods.

Unusually for wearable units, the $NO_2$ sensors had good accuracy in comparison to the reference monitor and demonstrated good inter-comparability.

It should be noted that this is a research grade unit that is not marketed as usable 'out of the box' but requires additional software to process data outputs. Therefore, different users will obtain different performance characteristics dependent on the nature of their processing software.

It is important to note that these tests were only performed in a London pollution environment and it is likely in different regions and cities with different pollution sources the correction factors and accuracy would be different. For use in other environments similar tests would need to be run against reference monitors located in similar environments. To develop these devices further, additional post processing of raw data could be used to improve the accuracy and precision of the sensors.

*Note: Results applicable to the version of the sensor tested at that time (October 2018). Any changes in the software algorithm used to convert sensor signals into pollutant concentrations would require retesting.*