King's College London

University of London

Breathe London wearable sensor evaluation – AirBeam2

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 20 October to 28 October 2018. Typically, this fixed test should be conducted for a longer period to observe any meteorological effects on the sensor between different monitoring weeks, however due to customs delays and limited time available, only 8 days of fixed testing was able to be completed. 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). 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 AirBeam2 $PM_{2.5}$ sensor units produced by HabitatMap.

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	ABEAM001 / %	ABEAM002 / %	ABEAM003 / %
Full period	5	85	85

From the table above the ABEAM001 sensor failed to capture over 5% of data, the sensor was clearly defective and therefore it was not analysed in further assessment. The other sensors worked reasonably well with a capture rate of 85%, there were however short sporadic gaps in the data where the sensors failed to record.

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)	ABEAM001/ %	ABEAM002 / %	ABEAM003 / %
ABEAM001		n.a.	n.a.
ABEAM002	n.a.		0.97
ABEAM003	n.a.	0.97	

The remaining two AirBeam2 sensors demonstrated an extremely high degree of precision, where 97% 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	ABEAM001/ %	ABEAM002 / %	ABEAM003 / %
Full period	n.a.	0.91	0.93

Accuracy (in terms of correlation against the reference monitor) from the two sensors was high varying at 0.91 and 0.93.

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	ABEAM001	ABEAM002	ABEAM003
Full period	n.a.	0.79	0.75

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. The two sensors under-read in comparison with the reference monitor and had scaling factors of 0.75 and 0.79.

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	ABEAM001	ABEAM002	ABEAM003
Full period	n.a.	-1.5	-1.8

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 slight baseline under-estimation for $PM_{2.5}$ of just under 2 $\mu g\ m^{-3}$.

Hourly mean time series

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

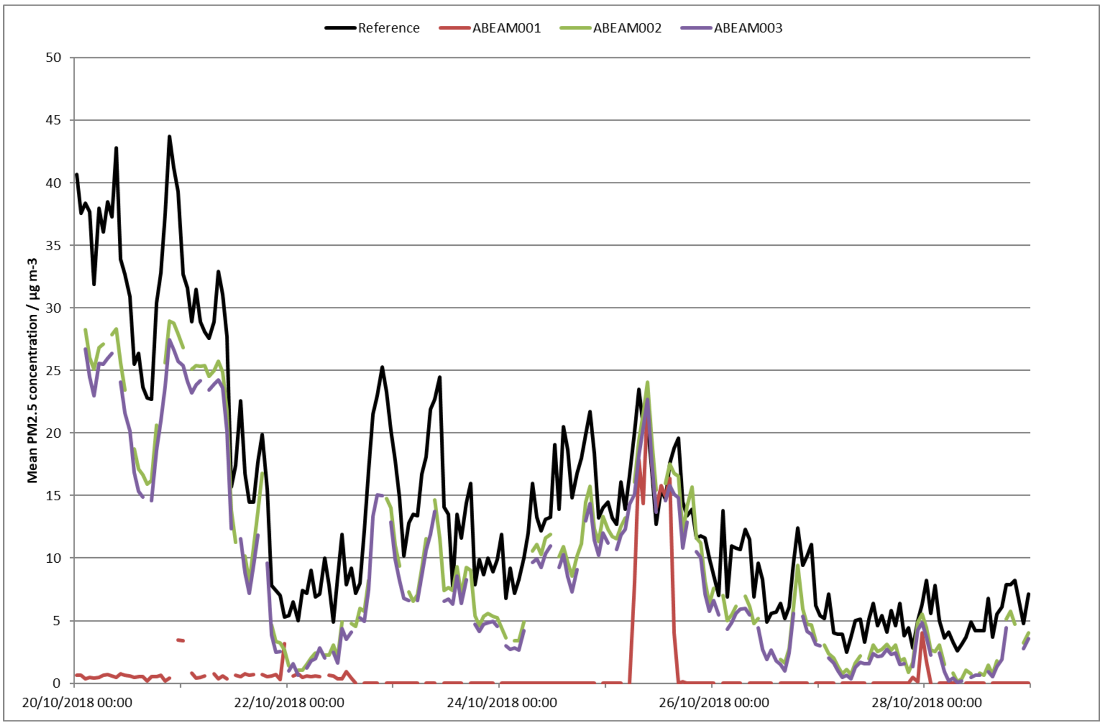

Figure 1: Time series chart of hourly mean sensor and reference $PM_{2.5}$ concentrations over the test period prior to application of scaling factors.

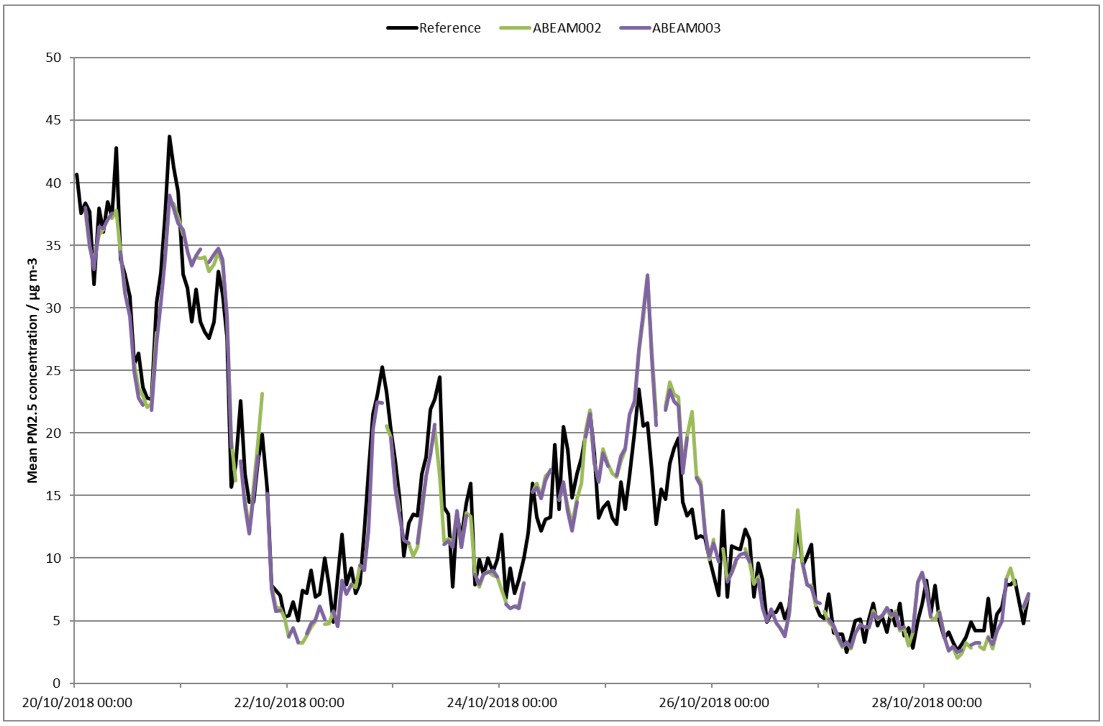

Figure 2: Time series chart of hourly mean sensor and reference $PM_{2.5}$ concentrations over the test period following application of scaling factors. ABEAM001 is excluded due to this sensor being defective.

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 $PM_{2.5}$ as recorded by the reference monitor for the majority of the monitoring period. However, there was a slight overestimation of $PM_{2.5}$ levels on the 25 October 2019, possibly indicating different meteorological conditions. There were also small sporadic gaps in measurements which highlights the reliability of the instrument.

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. In the absence of a mobile reference monitor only a sensibility check and inter-unit comparisons (precision) could be made (Table 6).

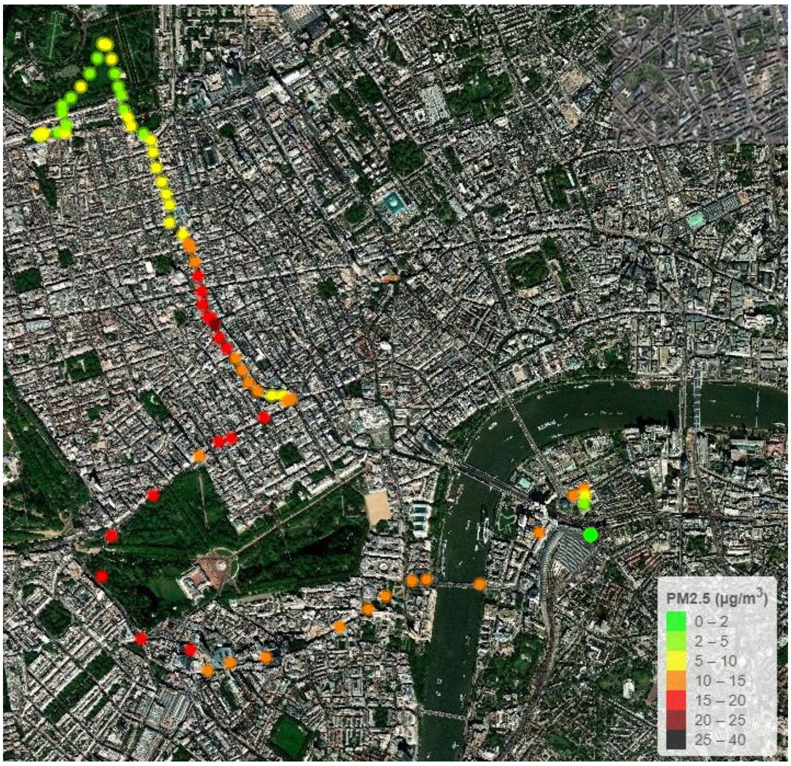

Figure 3: Map of mobile monitoring test for ABEAM003, from Marylebone Road (North) to Stamford Street (South), London. Each dot represents one-minute of concentration.

The spatial pattern recorded by the units was 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.

Table 6: $PM_{2.5}$ correlation coefficient (R^2) between units for mobile monitoring campaign. Coefficients of less than 0.75 are highlighted in red.

R^2 (RMA)	ABEAM001 / %	ABEAM002 / %	ABEAM003 / %
ABEAM001	-	n.a.	n.a.
ABEAM002	n.a.	-	0.96
ABEAM003	n.a.	0.96	-

The two PM sensors demonstrated a high degree of precision in the mobile monitoring test, with 96% of the variation in one unit was explained by the other unit. This precision is further observed in the timeseries of the mobile monitoring test presented in Figure 4. However, despite the good precision it was notable that ABEAM003 always read slightly higher compared to ABEAM002. 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 15:50.

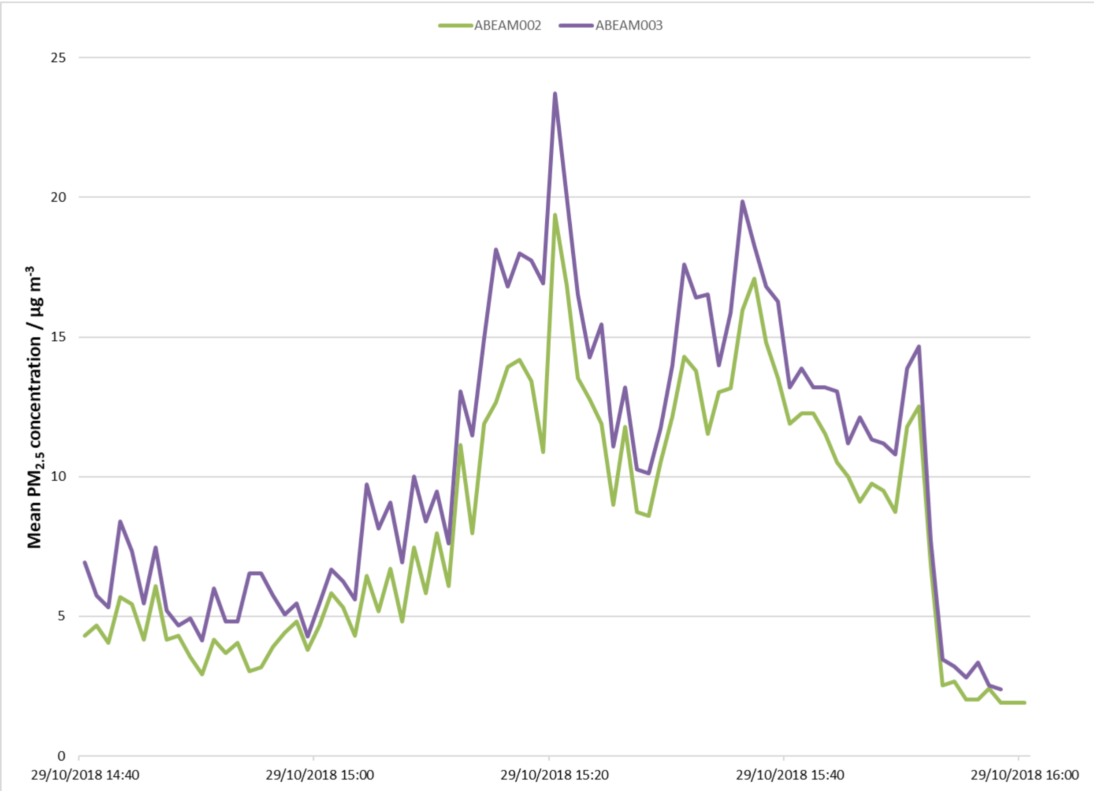

Figure 4: Time series chart of one-minute sensor $PM_{2.5}$ concentrations over the one-hour mobile monitoring campaign.

Overall sensor evaluation for $PM_{2.5}$

Table 7: 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
AirBeam2	**3**	**2**	**2**	**3**	**3**	**6**	**4**	**4**	**3**	**30**

The AirBeam2 $PM_{2.5}$ sensor is a good low-cost monitor. It rated highly in precision, usability, cost, GPS, reliability and aesthetics. The device requires a smartphone to use in mobile mode and the app used to run the sensor does crash sporadically. This does reduce its practicality for use in mobile monitoring campaigns. However, when used in a fixed location it can be run without a smartphone and was found to be reliable. The monitoring data collected is typically publicly available on its online repository, this could raise potential privacy issues when conducting studies. The unit is commercially available and can easily be used by members of the public who are interested in air pollution monitoring.

Conclusion

Three portable AirBeam2 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 eight days and a one-hour mobile monitoring test assessing the portable sensors performance in different pollution environments.

One sensor was found to be defective consistently reporting zero values, therefore only two sensor units could be tested. The two sensor units 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 short sporadic gaps in data recorded over the 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, with appropriate correction. One of the issues with the device is that a smartphone is required to record the data and there were slight issues with the reliability of the app. In its current format the sensor would be more applicable for use in community engagement to raise awareness on high air pollution environments or for use by individuals interested in conducting their own monitoring.

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.

Note from manufacturer: Since this evaluation was completed, HabitatMap, the non-profit that manufactures the AirBeam2 and operates the AirCasting platform, has updated the firmware running on the AirBeam2 to improve data capture rates and updated the AirCasting app to improve stability and reduce the frequency of crashes.