# AQMesh Fixed Sensor Network

# Data Quality Assurance and Control Procedures

# Table of Contents

## Document Development

This document was prepared by Jim Mills, ex Managing Director at Air Monitors, in consultation with staff from Breathe London (BL) partners (Environmental Defense Fund, the University of Cambridge, Cambridge Environmental Research Consultants, and National Physical Laboratory). A team from the National Physical Laboratory, led by Dr Nick Martin, helped ensure that this QA/QC plan followed generally accepted practices suitable for the estimated performance standards of the AQMesh sensors, if not those of reference networks. Dr Martin's team at NPL audited the application of the plan through random spot checks at the conclusion of the project (see **Appendix 2B**).

---

**Jim Mills**
ACOEM Air Monitors
2/3 Miller Court
Severn Drive, Tewkesbury,
Gloucestershire, GL20 8DN
jim.mills@acoem.com
Tel +44 (0) 1684 857530

## Introduction

The Breathe London fixed (or "stationary") network of air quality monitors consisted of over 100 small sensor systems ("pods") for the measurement of gaseous and particulate matter pollutants plus certain atmospheric parameters such as temperature, humidity and air pressure. For the project, pods were outfitted to measure nitrogen dioxide ($NO_2$), nitric oxide (NO), carbon dioxide ($CO_2$), various particulate matter size cuts ($PM_1$, $PM_{2.5}$, and $PM_{10}$); ten of the pods also measure ozone ($O_3$). The sensor pods were AQMesh and manufactured by Environmental Instruments Ltd. The sensors used in the pods were manufactured by Alphasense Ltd, SenseAir AB and Environmental Instruments Ltd. Sensors and pods underwent initial quality assurance at the respective manufacturing facility, by comparison with reference grade monitors. The devices were then subject to additional quality assurance and quality control (QA/QC) measures once deployed in the field and throughout their time in the network.

Data was published on the public Breathe London website in near-real time (within one hour of measurement) and, as such, all data disseminated by the project was considered "provisional" until all the procedures in this document were completed and applied to final, ratified datasets.[1]

This document outlines the QA/QC procedures for the Breathe London fixed sensor network data collection. The data were processed in stages, with a given stage adding one or more processing steps to the output of the previous stage. We first provide a summary of the stages of our QA/QC process then describe the underlying QA/QC steps in more detail. Finally, we describe the process for data ingestion from the ACOEM Air Monitors (AM) platform into the Breathe London website and data portal.

Such small, lower-cost sensor systems are relatively new and at the time of writing these procedures, there was no existing standards relating to their use. The European Committee for Standardization (CEN) Working Group 42 are developing such a standard at present which may be promulgated around 2021, and the current discussions within CEN are that the uncertainty requirement for small sensor systems carrying out indicative measurements will be in the region of 25% for each gaseous pollutant species and 50% for $PM_{10}$ and $PM_{2.5}$ mass fractions, at their respective limit values. This contrasts with the values for Reference or Equivalent monitors of 15% for gases and 25% for PM.

Conventional calibration procedures, using certified gases, are neither applicable nor cost effective for the Breathe London network, so the project team explored alternative methods to improve the accuracy of pod measurements. By using transfer standards ("gold" pods), as well as advanced data analysis techniques considering meteorology plus modelled outputs, we identified data that required further investigation and adjusted or redacted, as needed.

[1] Project data ($NO_2$ and $PM_{2.5}$) were available in near real-time at www.breathelondon.org

## Summary of QA/QC Stages

### Stage 0 (Factory Settings)

- Data at this stage is in the form that Air Monitors (AM) receives from the pod manufacturer, Environmental Instruments Ltd (EI Ltd), after application of factory QA/QC.
- Stage zero data represents raw measurements processed using results of factory testing and application of proprietary algorithms that convert sensor signals into pollution concentrations.
- Prior to shipping to the customer, each individual gas or PM sensor is characterised by EI Ltd. in terms of sensitivity and offset. This data is unique to each sensor and is used by the AQMesh processing algorithm to apply corrections for cross gas interferences and environmental conditions.
- Each data point generated by the pods is accompanied by a timestamp and a single fault code, determined by EI Ltd using a hierarchical system and applied by EI Ltd's server. Stage 0 flags are based on core sensor outputs, not computed concentrations. Later in the QA/QC process, additional flags may be added; all resulting flags would be presented as an array with each data point. Stage 0 data are also described as pre-scaled data (measurements before any scaling or offset is applied in later stages). Pre-scaled data is never modified in the AM platform except in extremely rare instances, so that data scaling can be revisited should the need arise. (see main text).

### Stage 1 (Empirically Verified)

- Stage 1 data is derived from Stage 0 data after the application of scaling adjustments, based on the results of co-location studies or the Network Calibration Method. The aim of both procedures is to evaluate and verify the field performance of sensors and bring their output as close as possible to that of a reference monitor or a transfer standard "gold" pod.
- Co-location studies are conducted with reference monitors or with a subset of pods qualified as transfer standards, also known as "gold" pods. Prior to initial placement at their intended location, about one-third of project pods were co-located with a reference monitor in Greater London for approximately 3 to 7 days.
- "Gold" pods are standard AQMesh pods which have been co-located at one or more reference monitoring locations, providing traceable evidence of the gold pod's performance. After characterisation, "gold" pods are moved adjacent to "candidate" pods across the network for approximately 7 days (actual duration will vary in order to achieve minimum data collection and quality objectives).

- After a reference or “gold” pod co-location, linear regression analysis is applied in order to determine the level of agreement between a candidate sensor’s output and the measurement from a reference instrument or “gold” pod. Scaling adjustments (slope and/or offset) are applied to the Stage 0 pre-scaled data if the regression fit’s slope and offset are statistically different (at a 95% confidence interval) than 1 and 0, respectively. If differences are smaller than those thresholds, then the Stage 0 pre-scaled data becomes Stage 1 data without further adjustment.
- Network Calibration Method: To maximise the number of sites for which we can publish preliminary data and to reduce operational costs, we have developed a novel Network Calibration Method primarily developed by the Cambridge University group, which aims to scale the entire network of sensors without the need for gold pods or reference co-locations for each pod. This approach involves selecting periods when non-local (aka regional) pollutant levels are likely to be relatively stable over the study area to determine relative pod sensitivities. The entire network is then scaled relative to one AQMesh pod whose measurements have already been scaled after co-location with a reference monitoring instrument.
- Stage 1 data is subdivided into four categories to distinguish the basis for the application of a scaling adjustment (reference co-location, “gold” pod study, network scaling) or the confirmation of pre-scaled outputs. The numbering of the Stage 1 sub-categories (1.1 to 1.4) is arbitrary and does not indicate hierarchical value or serial application.
    - Sub-category 1.1 represents confirmation of Stage 0 data – no adjustment is applied because co-location results for the corresponding sensor did not significantly differ from the traceable reference standard.
    - Subcategories 1.2 to 1.4 represent data scaled based on: co-location with a reference monitor (1.2), co-location with a gold pod (1.3), or network scaling (1.4).
- Since empirical scaling factors may be determined prior to subsequent data ratification of the reference network, we recognise there may be errors in the reference data that subsequently necessitate correction of sensor scaling factors. This is considered in Stage 4. Any other measurement artefacts during a field co-location would be reviewed once the ratified reference data were available.
- Scaling is normally applied within the Air Monitors server however it was decided that as more than one type of scaling process is to be used it was preferable that scaling is applied within the CERC platform immediately prior to data publication on the Breathe London platform.
- All Stage 1 scaled data points retain the same Stage 0 flag(s) assigned to the corresponding, pre-scaled measurement.

## Stage 2 (Manual QA/QC)

***Stage 2.1***

- Stage 2.1 data has undergone manual QA/QC procedures conducted weekly by AM staff, including visual inspection of the sensor outputs with reference to nearby data from other pods and if available London Air Quality Network (LAQN) monitors. A basic credibility check is also conducted on any measurements which appear abnormal, e.g., extremely high concentrations which may indicate a sensor failure or malfunction but could also be a valid pollution event, or negative concentrations which are normally the result of a sensor fault or system failure.
- Once this manual process is completed each week, AM advises CERC of any devices that are suspect and those are temporarily removed from publication. Information relating to the reasons for redaction is entered into a spreadsheet, which is shared with CERC and the wider BL team. CERC will not publish any such suspect data, although further analysis in Stage 4 could subsequently restore such data as valid).
- If data passes this stage of quality assurance, then any previously unflagged data continues to the next stage: Stage 0 pre-scaled data and Stage 1 scaled data retain their “valid” flags.

***Stage 2.2***

- Stage 2.2 data is the result of a separate manual review by the QA/QC team at AM of the time series of historical data collected prior to a gold pod study in order to determine if the empirical adjustments from Stage 1 should be applied retrospectively to the historical data. It is assumed that such retrospective slope or offset adjustments are appropriate if there had not been any intervening sensor changes, or other material interventions that may have changed the output of the device.
- Stage 2.2 data also includes manually inspected historical data from pods with refitted power supplies or replacement sensors (see main text for a description of these issues) that have been deemed to be valid by the QA/QC team at AM in the absence of any discontinuities or other anomalies in the time series. The resulting date from which historical data can be used is passed to CERC and is incorporated into the project platform.
- Periodically, the AM QA/QC team manually reviews the credibility of observed concentration baselines in order to identify any outliers. Due to the labour-intensive nature of the above procedure, we have come to prefer a more systematic approach using information from the full network by use of the Network Calibration Method to periodically identify any outliers. This process is currently semi-automated and is being developed to provide higher levels of automation as the project continues in order to speed the process of redacting outlier data and to further reduce manpower and costs.

## Stage 3 (Automated QA/QC)

- In Stage 3, scaled data from Stage 2 are automatically reviewed against high and low limits. Additional flags are added at this point if the data exceeds any pre-set concentration limits
- Negative ozone measurements can occur at times when rapidly changing NOx pollutant levels can create an artefact related to the way ozone is measured. So, when NO and $NO_2$ levels are > 5 ppb (e.g. Ozone depletion episodes) ozone values are assigned a value of zero and are flagged as such.
- Data is then available to users of the Air Monitors application at www.airmonitors.net and via an API returning JSON data formats to the project partners.

## Stage 4 (Special Issues)

Identification of specific issues encountered in the project where we envision taking actions prior to finalising data. Some of these issues involve post-processing steps but are included in this document for context. Examples of special issues include accounting for changes in provisional LAQN data after ratification, correcting PM data for the effects of relative humidity or correction of $NO_2$ data for any residual ozone interference. Additional details are in the main text.

# Detailed Description of QA/QC Procedures

## Stage 0 - Factory Settings

*This is data in the form that Air Monitors (AM) receives from the pod manufacturer after application of factory QA/QC. Stage 0 data are also described as pre-scaled data (measurements before any scaling or offset is applied in later QA/QC stages).*

AQMesh “pods” consist of a number of solid-state, electrochemical and optical sensors each of which are quality assured by the sensor manufacturer (Alphasense Ltd, Senseair AB or Environmental Instruments Ltd) before being accepted by Environmental Instruments Ltd (EI Ltd), which then integrates the sensors into pods. EI Ltd receives these sensors with “calibration” data for each batch of sensors from the manufacturers and this data is entered into EI Ltd’s processing algorithms as the first stage in the characterisation of each individual sensor. The sensors are then assembled into each device (pod) and this “family” of sensors (each measuring a different pollutant) within each pod are then further tested as a system in a specially modified ambient air shelter outside of EI Ltd’s manufacturing facility in Stratford Upon Avon for a minimum period of 7 days (see Figure 1).


**Figure 1.** QA/QC Facility at EI Ltd, Stratford Upon Avon, UK

The raw data for the gas sensors is then subject to processing by the proprietary algorithms developed by EI Ltd in order to correct for any meteorological effects that may affect the sensors and any cross-gas interference effects between sensors. During this initial period of testing, the pods are allowed to stabilise for 2 days before the data is compared with reference monitors for a further 5-7 days.

The Optical Particle Counter (OPC) outputs are also processed by a similar algorithm which converts particle size distribution and particle number concentrations into mass concentration data. Differences in particle composition (shape, colour, refractive index and density), experienced at the factory in Stratford Upon Avon during the test period compared to those at the final monitoring location, can lead to errors in the mass determination. In this project, we carried out co-location studies with reference monitors or gold pods from Central London and applied revised slope and offset adjustments as required, in order to characterise the OPC to typical PM in Greater London.

The reference monitors in the EI Ltd test facility include a chemiluminescence NO/NOx monitor, a UV Photometer for Ozone, and a FIDAS 200 for PM mass fractions. $CO_2$ is calibrated using a transfer standard $CO_2$ monitor (SenseAir S30, traceable to a Picarro $CO_2$ monitor calibrated with certified $CO_2$ standard gas at Cambridge University).

The data output by these algorithms is adjusted according to the reference data at the EI Ltd test facility and is then designated as Stage 0 data and passed via an API (Application Program Interface) to a server managed by AM at www.airmonitors.net. Stage 0 data is designated as “pre-scaled” on the AM server.

EI Ltd does not provide separate calibration certificates for each pod but they do maintain a record of calibration for each pod and each individual sensor is expected to perform within their published “out of box” performance. Table 1 shows the specifications stated by the manufacturer.

**Table 1.** Performance specifications for each sensor as stated by the manufacturer

| Sensor | Limit of confidence#3 | Typical precision to ref#4 | Typical mean prescaled accuracy#5 |
|---|---|---|---|
| NO | < 5 ppb | >0.9 R2 | +/- 5 ppb |
| NO2 | < 10 ppb | >0.85 R2 | +/- 10 ppb |
| NOx | < 10 ppb | >0.9 R2 | +/- 10 ppb |
| O3 | < 5 ppb | >0.9 R2 | +/- 10 ppb |
| CO | < 50 ppb | >0.8 R2 | +/- 0.05 ppm |
| SO2 | < 10 ppb | >0.7 R2 | +/- 5 ppb |
| H2S | < 5 ppb | >0.7 R2 | +/- 5 ppb |
| CO2 | < 1 ppm | >0.9 R2 | +/- 30 ppm |
| **Sensor** | **Limit of detection** | **Typical precision to ref#4** | **Typical mean prescaled accuracy#5** |
| Pod temperature | 0.1°C | >0.9 R2 | +/- 2°C |
| Pressure | 1 mb | >0.9 R2 | +/- 5 mb |
| Humidity | 1% RH | >0.9 R2 | +/- 5% RH |
| Average noise#6 | 20 Hz to 20 kHz | >0.8 R2 | +/- 1 dB |
| Peak noise#6 | 20 Hz to 20 kHz | N/A | +/- 3 dB |
| Particle count | 0 particles | >0.9 R2 variable | N/A |
| PM1 (v2.5) | 0 µg/m3 | >0.85 R2 variable | +/- 15 µg/m3 variable |
| PM2.5 (v2.5) | 0 µg/m3 | >0.85 R2 variable | +/- 20 µg/m3 variable |
| PM10 (v2.5) | 0 µg/m3 | >0.75 R2 | +/- 30 µg/m3 variable |
| PM1 (v2.5h) | 0 µg/m3 | >0.9 R2 variable | +/- 5 µg/m3 variable |
| PM2.5 (v2.5h) | 0 µg/m3 | >0.9 R2 variable | +/- 5 µg/m3 variable |
| PM10 (v2.5h) | 0 µg/m3 | >0.85 R2 variable | +/- 5 µg/m3 variable |
| GPS | <0.5m | N/A | +/- 3m radius |

#3 Readings provided below this level, however due to interferences the level of uncertainty is greater than at higher levels of the target pollutant.
#4 Results based on field testing around the world versus certified reference or equivalence methods at hourly intervals, in extreme and varied conditions.
#5 Average variance to reference equivalence methods at hourly intervals from field testing around the world, in extreme and varied conditions.
#6 Peak noise is the highest recorded value over the gas reporting interval while average noise is calculated using all noise samples over the same period.

***System Flags Applied by the Manufacturer***

Each data point in Stage 0 is accompanied by a timestamp and single fault code which are determined by the manufacturer using a hierarchical system (Table 2). Later in the QA/QC process, during stages 2.1 and 3, additional flags may be added by the Air Monitors system. All flags are presented as an array with each data point via the AM API for ingestion by CERC.

**Table 2.** Summary of AQMesh Data Flags

| Status Code | Sensor State | Description | Effects |
|---|---|---|---|
| -1000 | Not Fitted | Sensor or component not fitted. | Coded flag in data feeds as there is no data to view. |
| -999 | Stabilising | Either when the POD has just been moved to a new location or manually instigated via server. | Values are redacted as they cannot be relied upon during this 2-day period and will remain non-viewable. |
| *-998* | *Rebasing* | *Typically, this is a 2-day period where local variables are calculated for use in the AQMesh algorithm are determined.* | *During the rebasing period the coded flag will remain, however upon completion of this process, valid data will be reinstated – Data will need to be re-retrieved for this period, so API pointers are reset.* |
| -997 | Optimising | When a pod is power-cycled for more than an hour i.e. Maintenance or power failure. | Values are redacted as they cannot be relied upon during this 1-hour period and will remain as non-viewable. |
| -996 | Failed | The system has detected that the sensor has failed. | Data classified as having a sensor fail is redacted & will remain as non-viewable |
| -995 | Cross Gas Error | If a sensor fails which is relied upon for the removal of interferences on another sensor, data from the reliant sensor becomes invalid. | Data will be redacted & remain non-viewable until the compensating sensor is replaced and produces good results. |
| -994 | No Data | Data points where the instrument has not recorded a reading. | Coded flag in data feeds as there is no data to view. |

| | | | |
|---|---|---|---|
| -993 | Destabilised | The system has detected that the sensor's output \ stability may be compromised due to odd fluctuations in temperature and pressure. | Readings for the prescribed period are redacted and are non-viewable until the conditions have normalised. |
| -992 | Extreme Environment | Following intensive testing of all electrochemical sensors we have determined the combination of extremes in climate in which the electrochemical sensors do not provide consistent outputs. As such precise and accurate measurement is not possible | Data classified as within the extreme ranges of environmental conditions and will be redacted and will remain as non-viewable. |
| -991 | Condensation | NDIR sensor has been affected by condensation on the detector | Data classified as being affected by condensation will be redacted and will remain as non-viewable. |
| Estimated Reading | Deliquescence | When not using the heated inlet option, outlying data points caused by hygroscopic particle size growth will be flagged following analysis of the particle count distribution | These readings are available but should be considered as potentially unreliable. |
| -893 | Mis-read | It is possible if the particle (or noise) sensor is unable to transfer valid data. | Occasional loss of data possible. Coded flag in data feed as there is no data to view. |
| -892 | Other fault zero | Due to the warm-up sequence and the timing of the event there is a chance that the particle counter is unable to provide a valid particle reading following a power-cycle and/or a change in pod settings. | Occasional loss of data possible. Coded flag in data feed as there is no data to view. |

***Field Stabilisation and Re-Basing***

A final step of factory-controlled QA/QC occurs once a pod is deployed in the field for the first time. When first powered up on location, pods can be programmed to undergo an initial period of stabilisation for 2 days in their new environment. This is recommended if a location's pollution levels or composition differ substantially from the factory at Stratford Upon Avon, which is a rural area with low pollution levels compared to much of London. During this period of stabilisation certain concentration data will be temporarily unavailable to view on the AM platform or via the API.

After the stabilisation period each device is "re-based". This is to allow an automated estimation of each sensor's true baseline at the new location. This is done by comparing 48 hours of variation in the concentrations and subsequently calculating the baseline using a proprietary algorithm. Once set, this baseline adjustment becomes part of a sensor's pre-scaled data and does not change thereafter, unless the user requests an additional period of re-basing (e.g. if a pod is moved to a location with a very different climate or pollution level). During this automated "re-basing" period certain sensor data will be unavailable until the process is completed; if the process completes successfully then the data is subsequently restored for the period since the re-basing was commenced.
In April 2019, Air Monitors identified that some sensor baselines were not in line with others in the network and raised this issue with the manufacturer. After investigation the manufacturer notified Air Monitors that the re-basing process had not completed successfully for some NO and $CO_2$ sensors during initial operation in London. Consequently, re-basing was repeated commencing on 18 April 2019 for all NO and $CO_2$ sensors across the network to correct this (and the results applied prospectively to new measurements and retrospectively to the entire time series). If a sensor must be replaced at any point due to failure, then the new sensor will be subject to re-basing automatically (see Figure 2 below).

All data collected during periods of stabilisation or re-basing are automatically flagged by EI Ltd via the API and in the www.airmonitors.net application. These flags are automatically removed after the stabilisation and re-basing are complete.  Any data remaining flagged as abnormal for stabilisation and re-basing will be held back from publication by CERC.

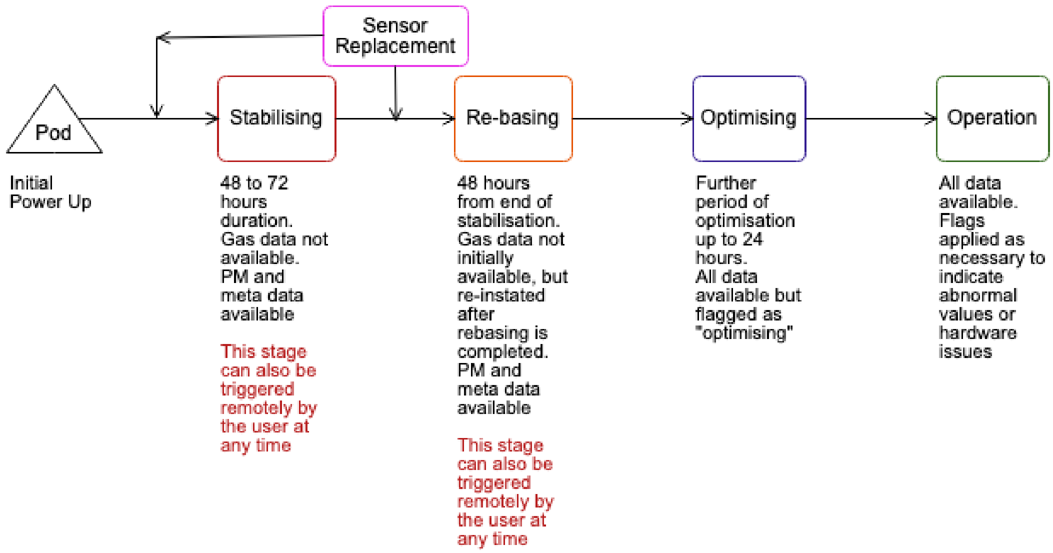


**Figure 2.** Initial start-up sequence

## Stage 1 - Empirically Verified

Stage 1 data is derived from Stage 0 data after the application of scaling adjustments based on results of co-location studies (with reference monitors or with a subset of pods qualified as transfer standards, also known as “gold” pods) or the Network Calibration Method. The aim of this process is to evaluate and verify the field performance of sensors and bring their output as close as possible to that of a reference monitor or a “gold” pod.

Stage 1 data is subdivided into four categories to distinguish the basis for the application of a scaling adjustment (reference co-location, “gold” pod study or the network-based calibration method) or the retention of Stage 0 (pre-scaled) outputs. The numbering of the Stage 1 sub-categories (1.1 to 1.4) is arbitrary and does not indicate hierarchical value or serial application.

> Sub-category 1.1 represents retention of Stage 0 data – no adjustment is applied because no scaling factor could be determined or co-location results for the corresponding sensor did not significantly differ from the traceable reference standard.
>
> Subcategories 1.2 to 1.4 represent data scaled based on co-locations with: a reference monitor (1.2), a gold pod (1.3), or the Network Calibration Method calibration (1.4).

***Co-location with Reference Monitors***

Prior to first deployment a three to seven-day field co-location within Greater London was conducted for 59 of the 100 Breathe London pods at one of four locations around the city. These locations were selected because they had reference equipment for one or more pollutants and there was enough space to co-locate the pods within 1m of the reference monitor inlets and at a similar height (Table 3 and Figure 3). The aim of this process is to evaluate and verify the field performance of sensors and bring their output as close as possible to that of a reference monitor. This process is intended to mimic the calibration process using traceable gases that is carried out periodically for reference instruments.

The duration of co-locations was intended to be long enough to capture a large range of concentrations, but in practice was limited by a desire to rapidly deploy the network of pods to host sites. All pods were not co-located prior to deployment due to time constraints and the availability of space at reference sites, but the data from the large subset of pods that were co-located provided valuable information about the typical performance of the pods used in the project. The co-locations we carried out at NPL in the early stages of the project were not used due to the low pollutant concentrations and suburban nature of the site and absence of reference NOx data at the location.

The Breathe London project's database logs the location and date/time of each co-location. After the conclusion of a co-location at a reference site, pods were transported to their final monitoring locations.

**Table 3.** Reference sites used for co-location Sites (each pod was co-located at only one site)

| Location | Classification | # Pods |
|---|---|---|
| Swiss Cottage | Kerbside | 10 |
| Holloway Road | Kerbside | 10 |
| Elephant & Castle | Urban Background | 13 |
| NPL Teddington | Suburban | 26 |

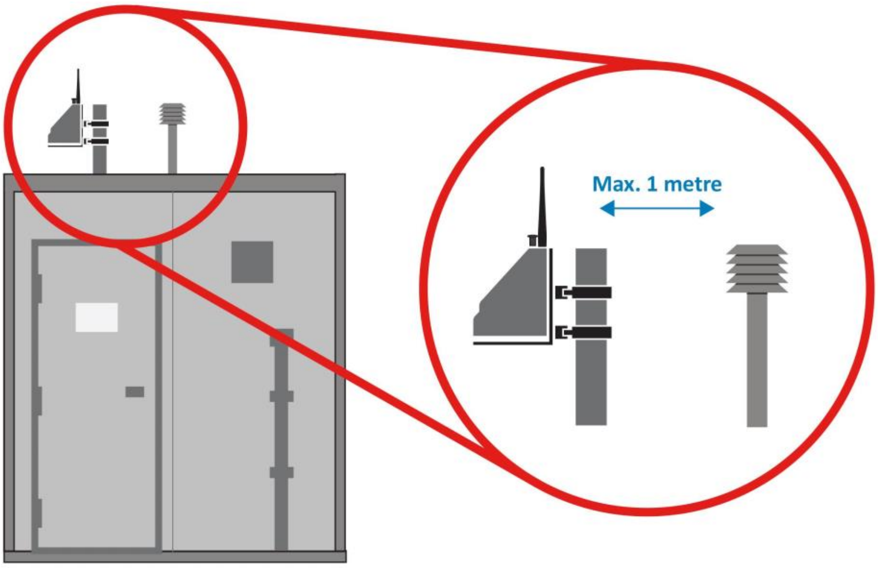


**Figure 3.** Co-Location of AQMesh with Reference Inlets

During a co-location, it is essential to have access to the reference data on the same average time basis as the data from the pods and to ensure that time stamps on all devices are fully synchronised to UTC. The time conventions used throughout the project were time ending. Data is then compared between the reference and candidate devices over a period of time (typically 3 to 7 days or more) and a linear regression analysis performed in order to determine the level of agreement ($R^2$) and any differences in the slope and intercept when compared to the pre-ratified reference data (see Figure 4).

Scaling adjustments (slope and/or offset) are applied to the Stage 0 data if the regression fit over the comparison period identifies a large enough difference. Considering the combined uncertainties in data from AQ Mesh sensors and reference instruments, scaling is applied only if the regression fit's slope and offset are statistically different (at a 95% confidence interval) than 1 and 0, respectively. If differences are smaller than those thresholds, then the pre-scaled data is considered verified in this Stage and used without further adjustment.

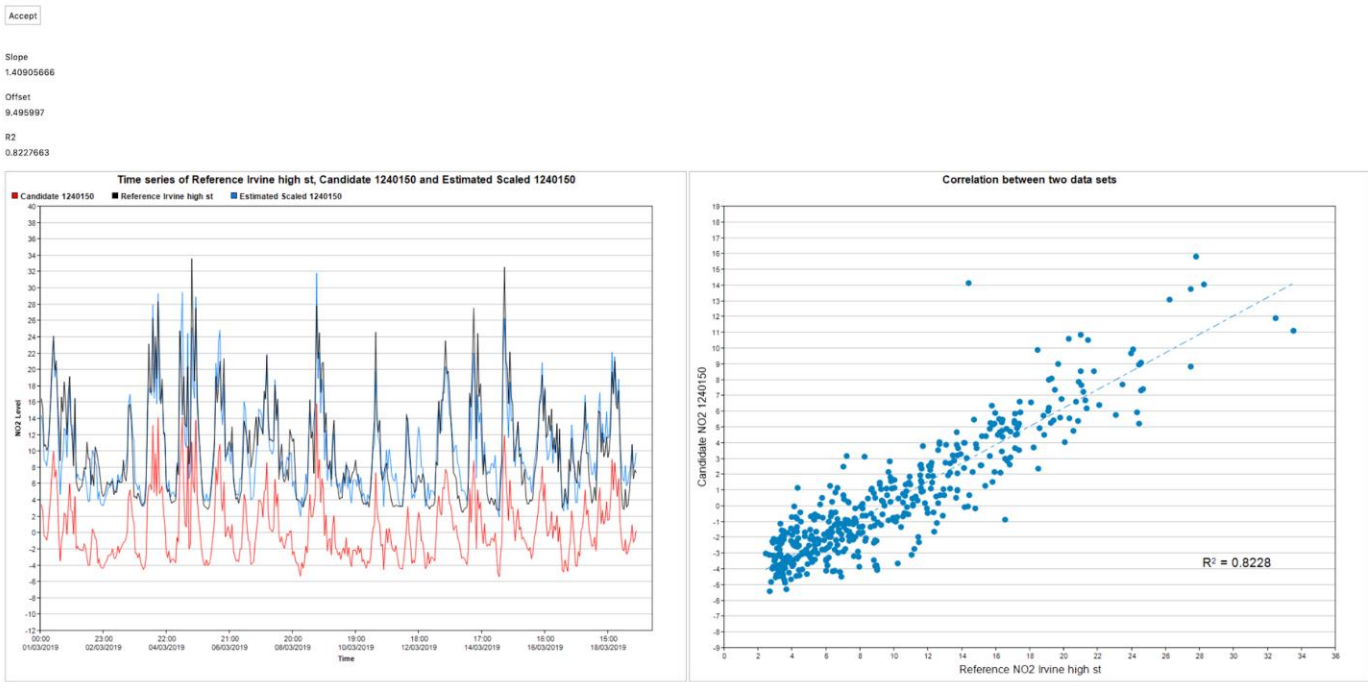


**Figure 4.** Typical results of a Co-Location Study with offset and slope adjustments indicated

Scaling adjustments based on co-locations are only applied to pods that are subsequently deployed at sites with similar characteristics to the reference site (see Table 3). This was not the case for some of the pods co-located at NPL, which were sited at urban locations.

We subcategorise Stage 1 data into four categories (1.1—1.4) to distinguish the basis for the application of a scaling adjustment or, alternatively, the confirmation of Stage 0 data. The resulting numbering of these four categories is arbitrary and does not imply any hierarchical value or serial application of scaling factors. Stages 1.1 and 1.2 are defined here and the remaining two are defined later in this section:

> ***Stage 1.1***
> represents confirmation of Stage 0 (pre-scaled) data – no adjustment is applied because co-location results for the corresponding sensor did not significantly differ from the traceable reference standard.

> ***Stage 1.2***
> Stage 0 data scaled based on the results of a co-location with a reference monitor.

***Co-location using a Transfer Standard (Gold Pods)***

Although one can return any network pod at any time to a reference location and re-compare the data in order to ensure ongoing accuracy of that device, this disrupts data capture and can also be time and labour intensive. A more efficient alternative method is to use a transfer standard ("gold" pod) approach, as described here.

Individual AQMesh pods can be designated as "gold pods". While in fact identical to the other pods, they are designated "gold" by virtue of an extended co-location with a trusted reference monitoring location. With sufficiently long co-location adjacent to the reference location ( >2 weeks to many months in this project), the "gold" pod becomes well characterised with respect to the reference and as such can be used to "transfer" this calibration information to the candidate pods around the network. To confirm the performance of gold pods after cycling through several candidate monitors within the network, they are periodically returned to the reference sites (ideally every 3 months) for a duration of at least 7 days to ensure that their performance has not changed in any significant manner during their use as a transfer standard. If any sensors in the "gold" pod fail at any time, the sensors are replaced, and the pod returned to the reference site for a new characterisation and adjustment where necessary, before being re-designated as a "gold" pod. The "gold" pod method is less labour intensive and less disruptive than a full reference co-location study and has proved to be an effective way of quality assuring networks of pods on an on-going basis.

After characterisation, the "gold" pod is placed adjacent to a candidate pod and the data compared for a period of 7 to 14 days, or longer if required, and a linear regression analysis performed as is the case with a reference co-location study, but in this case using the "gold" pod as the reference. The duration of the deployment is determined by the satisfaction of several criteria starting with a minimum number of valid data points (130 hourly values).

Once the minimum data threshold is reached, co-locations are evaluated with two statistical metrics to determine if the agreement between the candidate pod and reference site is sufficient to obtain an acceptable calibration. First, we require a reasonably good correlation between the gold pod and candidate site ($R^2 > 0.7$). Second, we require that the Root Mean Square Error (RMSE) is sufficiently low in comparison to the average concentrations during the co-location (normalised RMSE < 0.5). After these criteria are met, the same procedure used to calculate offset and slope for field co-locations is applied to the candidate bringing it in line with the "gold" pod and thus traceable to the reference site (poor correlations would trigger a technical evaluation of a pod's performance).

When Covid-19 confinement measures started in March 2020 it caused drastic reductions in NOx emissions. We observed uncharacteristically low NO and $NO_2$ concentrations throughout Greater London. Without a sufficient range of concentrations to compare, the co-location $R^2$ and normalised RMSE criteria were often not met during the typical 1-2 week period, and thus co-locations were extended until pollution levels rose enough to achieve acceptable calibrations that met the criteria.

**Table 4.** Gold Pod Locations. CD1 = Camden - Swiss Cottage, IS2 = Islington - Holloway Road, SK6 = Southwark - Elephant and Castle.

| Gold Pod # | Reference Sites | Classification |
|---|---|---|
| 2450099 | IS2, CD1 | Roadside, Kerbside |
| 2450079 | SK6, CD1, IS2 | Urban Bkg, Kerbside, Roadside |
| 2450091 | CD1, IS2 | Kerbside, Roadside |
| 2027150 | IS2, CD1 | Roadside, Kerbside |
| 2028150 | IS2, CD1 | Roadside, Kerbside |
| 2030150 | IS2, CD1 | Roadside, Kerbside |
| 2032150 | SK6, CD1, IS2 | Urban Bkg, Kerbside, Roadside |
| 2036150 | SK6, CD1, IS2 | Urban Bkg, Kerbside, Roadside |
| 2046150 | SK6, CD1, IS2 | Urban Bkg, Kerbside, Roadside |

Starting in April 2019, we rotated nine "gold" pods (Table 4) around the network (six of the pods were loaned to the project for this purpose by the University of Cambridge). Gold pod studies started with locations of high interest related to the ULEZ and to locations where we have evidence that the data from a pod is significantly different from nearby reference station measurements or predicted concentrations from the CERC air quality model. The goal is to eventually verify every pod in the network using this method and to supplement and verify the Network Calibration Method as described in Stage 1.4 below.

If a candidate pod is found to have a significantly different response than the "gold" pod, then a slope and offset adjustment will be applied as done in a reference co-location (Stage 1.2) to the data from that point in time onward.

Here we introduce the additional subcategories of Stage 1 data which distinguish pods whose data was verified using a gold pod:

***Stage 1.3***

Stage 0 data which have been scaled after a gold pod co-location.

***Stage 1.4***

Stage 0 data which have been scaled based on the Network Calibration Method is described in a separate attachment and summarised briefly here. This novel methodology is being submitted for publication in a peer-reviewed journal. To maximise the number of sites for which we can publish preliminary data (e.g., when we don't have valid co-location data) and to reduce operational costs, we have applied a Network Calibration Method developed primarily by the Cambridge

University group, which aims to scale the entire network without the need for gold pods or co-location.

A statistical criterion was applied to network calibration results, where scaling factors were only applied with sufficiently high covariance (> 0.5).

The separation of local sources immediately adjacent to a pod site from the non-local background (aka regional) pollutant levels, which are often consistent over substantial distances (10 – 50+ kilometres), allows scaling of pods across the network. This approach involves selecting periods when non-local pollutant levels are likely to be relatively stable over the study area to determine relative pod sensitivities. The entire network is then scaled relative to an AQMesh pod which has been or is currently co-located with a reference monitoring instrument. This “scale separation” methodology has been previously demonstrated by Heimann et al. (2015) and Popoola et al. (2018). Results using this method continue to be evaluated throughout the project by comparisons with calibration factors (slopes and offsets) derived from direct co-locations and the results to date have been generally in agreement. For NO data, where atmospheric chemistry is more complex, preliminary scaling uses a hybrid of the network-based calibration and direct co-location analysis by applying a network-based offset to prescaled pod values before calculating the gain using a fixed-zero intercept linear regression. Where direct co-locations are not available for NO, we use the same hybrid method with a generic gain correction of 0.81 (the average gain of all co-located NO sensors against reference instruments).

***Linkages to Other Stages***

In the Breathe London project, retrospective adjustments based on either reference or gold pod co-locations were considered in Stage 2. Specifically, in Stage 2.2, each pod’s historical time series is evaluated to determine whether scaling adjustments from the gold pod study should be applied to the historical data (generally done in the absence of a change of sensors or any other changes that may have affected data output in that time period).

Records of any changes of slope adjustments or offsets are automatically managed within the AM server database, passed over to the CERC database with every data point, and are additionally logged in a project spreadsheet to allow cross referencing.

Since empirical scaling factors may be determined prior to subsequent data ratification of the reference network, we recognise there may be errors in the reference data that subsequently necessitate correction of sensor scaling factors. This is considered in Stage 4.

Any other measurement artefacts during a field co-location would be reviewed once the ratified reference data are available.

All Stage 1 scaled data points retain the same Stage 0 flag(s) assigned to the corresponding, pre-scaled measurement (Stage 0 flags are based on core sensor outputs, not computed concentrations).

Once scaling has been determined either by co-location or by network-based calibration, the scaled data values are used by CERC as the basis for publication onto the Breathe London public website and data portal (see Appendix). Notwithstanding, pre-scaled data is also stored allowing subsequent correction of scaling should that be necessary due to revision of the reference data after ratification processes are completed or for other reasons.

## Stage 2 - Manual QA/QC

Stage 2.1 data has undergone manual QA/QC procedures conducted weekly by trained AM staff, including visual inspection of the sensor outputs with reference to nearby data from other pods and if available LAQN monitors. A basic credibility check is also conducted on any measurements which appear abnormal, e.g., extremely high concentrations which may indicate a sensor failure or malfunction but could also be a valid pollution event, or negative concentrations which are normally the result of a fault.

Stage 2.2 data is the result of a separate manual review by the QA/QC team at AM of the time series of historical data collected prior to a gold pod study, in order to determine if the empirical adjustments from Stage 1 should be applied retrospectively to the historical data.

### ***Stage 2.1***

Once deployed in the field, AQ Mesh pods will normally remain stable and provide good data quality for 6 - 12 months with minimal intervention. This statement is based on experience where pods have been co-located and then co-located again after a period, where no significant changes in slope or offsets have occurred. However, because it is possible that sensor characteristics may change over extended time periods, it is important to check data outputs regularly to ensure that sensor performance remains consistent over time. Additionally, it is also useful to review data to identify malfunctions that may be missed by the manufacturer's automated error flags.

Stage 2.1 data is defined as data that has undergone a manual review process conducted weekly by AM staff that seeks to identify the occurrence of power interruptions, missing data, negative readings, elevated or depressed baselines, or other interferences, flatlining, or any data which appears abnormal in the judgment of AM's data team (this final category is subjective and is always referred to colleagues for confirmation before action is taken or data redacted or flagged). If data passes this stage of quality assurance, then any previously unflagged data continues to the next stage: Stage 0 pre-scaled data and Stage 1 scaled data retain their "valid" flags.

Once this manual process is completed each week, AM advises CERC of any devices that are suspect and those are temporarily removed from publication. Information relating to the reasons for redaction is entered into a spreadsheet, which is shared with CERC and the wider BL team. CERC will not publish any such suspect data, although further analysis in Stage 4 could subsequently restore such data as valid).

***Stage 2.2***

Stage 2.2 data is the result of a separate manual review by the QA/QC team at AM of the time series of historical data collected prior to a gold pod study, in order to determine if the empirical adjustments from Stage 1 should be applied retrospectively to the historical data. It is assumed that such retrospective slope or offset adjustments are appropriate provided there have not been any intervening sensor changes, or other material interventions that may have changed the output of the device.

If data passes Stage 2.2 of quality assurance, then any previously unflagged data continues to the next stage: Stage 0 pre-scaled data and Stage 2.2 scaled (historical) data retain their "valid" flags.

***Early Stage Power Supply Issues***

It was discovered during Manual QA/QC review in the early stages of deployment that the use of certain mains electricity power transformers can affect the quality of data produced at certain locations (where AC/DC transformers convert 230V AC to 12V DC which then power the pods). Specifically, spikes (pulsing) in the PM data were found to occur as the pod's cellular modem was powered up each hour, in readiness to transmit data to the server. Other effects of Electro Magnetic Current (EMC) included induced noise in PM and/or gas data. There was also evidence that these induced currents could be picked up by some of the electrochemical sensors resulting in a baseline shift (sensitivity (slope) did not deem to be affected). Importantly, no measurable impact was observed on the $CO_2$ optical sensor or any of the meteorological sensors such as temperature, pressure or humidity. These effects were not consistent at each of the affected sites and at some sites there was no observed effect at all.

These effects were attributed to EMC (or RF) being picked up by the cellular modem antenna and being transferred to the low current circuitry of the Optical Particle Counter (OPC). In order to resolve the problem, new high-quality digital switching Power Supply Units (PSU's) with a high level of EMC/RF filtering were procured. The new PSUs were specified with more than 3x the capacity required and could be used to power 2 pods simultaneously during co-location exercises. While slightly more than half of the project pods were affected (including an intended gold pod at NPL), we decided to err on the side of safety, by fitting the new PSUs to every pod in the network not powered with a solar panel. This process took place during February to April 2019.

***Hardware Maintenance***

During operation it is possible that individual sensors may fail, inlets may become blocked by foreign matter, bugs etc and water ingress can occur during extreme weather. The system is designed to minimise these events, but it is always possible that things can go wrong. Pods can also be damaged accidentally or vandalised causing data to become unreliable or the device to go offline.

We therefore visit pods as regularly as possible, within budgetary constraints, to carry out visual inspection and to clean and check inlets and their associated weather protection. In BL, these hardware checks are performed whenever a site is visited for any reason; thus, every site was visited at least once during the first 4-6 months after the original installation. Some more than this due to power or other issues. Within the first year we expect that all pods would receive at least two such visits. Where deficiencies are found parts are cleaned or replaced as required and reports placed on-line in order to inform network management procedures and data quality control.

***Sensor Replacements***

When a sensor fails it is usually evident by a change in baseline or flatlining of data. This is detected by algorithms at the manufacturer's server (EI Ltd) after a 48-hour period. They are also often spotted by the Stage 2 manual QA/QC checks before the automated alerts are issued. As of March 2019, less than 5% of sensors have failed. Sensors tend to fail more in winter wet conditions than in summer dry conditions (80% of all sensor failures typically occur between October and January).

When sensors fail, replacement sensors are fitted as soon as possible, and new sensor serial numbers entered into the www.airmonitors.net application. This in turn informs the manufacturer's server that the sensor has been physically changed and the server will then automatically apply the new sensor characteristics within its processing algorithm. There will be a period of stabilisation and re-basing due to the sensor replacement and this may also affect other sensors in the same device due to the cross-gas corrections which are applied. Sensor replacements are also logged with a date and time on the Breathe London database for reference in future analyses.

If individual sensors were to fail during deployment, it will be necessary to repeat the stabilisation or re-basing of those sensors and due to the possible cross gas interference effects (e.g., any $NO_2$ sensor failure in a pod may affect $O_3$ results and vice versa). We will also repeat the co-location exercise where possible in order to obtain the best quality data possible or at least use a gold pod to verify that the sensor change had no detrimental effect on the data quality.

## Stage 3 - Automated QA/QC

Prior to Stage 3, every data point is flagged in Stage 0 according to one or more of the 19 conditions listed in Table 2 above, as determined by the pod diagnostics or the processing stage applied by the server. If the data is deemed to be valid (i.e. absence of flags 1-19), then scaling factors (offset and slope), if needed, are applied in Stages 1 or 2 (where offset and slope values are determined by co-location studies network-based calibration, or manual QA/QC review for historical data).

In Stage 3, the Air Monitors server automatically applies additional flags where appropriate to data from Stage 2. First, data is reviewed against high and low limits. Additional flags 20-26 (Table 5) are added if measurements exceed any pre-set concentration limits.

| **Description** |
|---|
| Above Hi Limit |
| Above Hi Hi Limit |
| Below Lo Limit |
| Below Lo Lo Limit |
| $PM_1 > PM_{2.5}$ |
| $PM_1 > PM_{10}$ |
| PM2.5 > PM10 |

**Table 5.** Air Monitors Data Flags

The final three flags applied above are to warn of any data points where the sub fractions of PM exceed the upper fraction. (e.g. PM2.5>PM10) this should not occur in a perfect scenario but can occur due to imperfections in the conversion from size distributed particle count data to mass concentration. We currently do not redact data based on these x>y flags but use them to spark further investigation if they occur frequently.

The Hi-Lo values currently used for each channel relative to concentration limit flags are shown in Table 6 below.

| AQMesh Pod Parameter | Hi-Hi | Hi | Lo | Lo-Lo |
|---|---|---|---|---|
| NO ppb | 1000 | 800 | -5 | -10 |
| $NO_2$ ppb | 500 | 300 | -5 | -10 |
| $CO_2$ ppm | 2000 | 1200 | 300 | 200 |
| $O_3$ ppb | 150 | 120 | -5 | -10 |
| PARTICLE_COUNT p/cm$^3$ | 300 | 200 | 0 | -5 |
| $PM_{10}$ µg/m$^3$ | 1000 | 500 | 0 | -10 |
| $PM_{2.5}$ µg/m$^3$ | 500 | 300 | 0 | -10 |
| TEMP C | 50 | 40 | -10 | -20 |
| Relative Humidity % | 101 | 100 | 20 | 0 |
| AIRPRES mb | 1200 | 1055 | 925 | 800 |
| VOLTAGE V | 5 | 4 | 2.8 | 2 |

**Table 6.** Hi/Lo limit values

Flags in Table 5 are combined with those from prior stages (the EI Ltd Flags from Stage 0). Flags are presented as a machine-readable array at this stage and may include a factory (EI Ltd) flag and any additional flags generated by AM in Stage 3.

Data is then available to users of the Air Monitors application at www.airmonitors.net or via an API returning JSON data formats to the project partners. Stage 0 and Stage 1 data remain available via the Air Monitors API complete with all flags and metadata, such as Temperature, Pressure and Humidity. The Air Monitors API also allows for further averaging of the 1-minute data from the pods as required and data can be specified in relative units (ppb or ppm) or in absolute units (µg/m$^3$ or mg/m$^3$).

Stage 3 data is passed to CERC for publication on the project website and for use in conjunction with the CERC model. After receiving the data and the data flags as an array via the API, CERC applies a 'valid from' date for each pod as advised by AM through the Master Spreadsheet and data before this point is redacted. Any data flagged as in table 6 above may also be redacted at this stage, although. As described in the Appendix, CERC then passes the data on for publication on the www.BreatheLondon.org website as one hour averaged data in near real time (1 hour lag).

## Stage 4 - Special Issues

In this section we identify specific issues encountered in the project where we envisioned taking actions prior to finalising data in Year 2 of the project. Some of these issues involve post-processing steps but are included in this document for context.

### *Power Supply Issues with EMF and RF interference*

It was observed by manual QA/QC during the first few months of operation that at certain locations the PM sensor output was subject to regular pulsing. This was subsequently identified as EMF interference on the system during data transmission. Steps were taken to correct this by replacing the power supplies with ones that did not exhibit this problem. The investigation and remedy took several months to complete and as such not all historical data may be recoverable. There may also be some effect on gas channel data due to EMF prior to them being exchanged, however we are still assessing the extent and if some of that data can also be recovered. We will evaluate ways to recover affected data, such as the possibility of filtering out EMF-induced spikes in PM readings that occur hourly when the data transfer uplink occurs.

### *Accounting for changes in provisional LAQN data after ratification*

Prior to finalising the Breathe London dataset we will recalculate any scaling factors based on LAQN reference site data if the latter data changed after ratification.

### *Fog effects on PM data.*

During fog the Optical Particle Counters used in the pods are unable to differentiate small water droplets from particles. Currently we are using the manufacturers “deliquescence” flag to identify when high humidity conditions are affecting particle size distributions. This is working in some situations but is not yet consistent across all pods in all conditions. We are working with the manufacturer and project partners to improve this. As fog events are not common and only affect the data for relatively short periods, this should not have a major impact on overall data capture rates. The use of visibility and humidity data from the Heathrow airport appeared suitable to identify these periods.

***PM-specific Humidity Adjustment***

A separate issue from the above effect of fog and relates to the size growth of certain particles during periods of high humidity. Correction strategies for achieving this continue to be developed by Prof. R. Jones' group at Cambridge but were not finalised in time for preparation of the final dataset.

***Anomalous behaviours in $NO_2$ concentrations***

*Apparent non-linear behaviours at low concentrations.* Results of co-locations with reference monitors for $NO_2$ exhibit some departure from a linear regression fit at low concentrations. This, we believe after evaluation and consultation with the manufacturer, is due to an ozone artefact, which may become more severe over time. The final $NO_2$ dataset includes a first-order correction for ozone cross-interference, which is applied to scaled (Stage 1) data.

*Departure from established scaling at high T.* The project team noticed a high bias in $NO_2$ measurements during periods characterised by the very high temperatures observed in London during the summer of 2019. This may be the result of degraded sensor performance and/or an interference from ozone. No explicit correction was identified for this issue, although the correction for ozone cross-interference may partially address the effect.

***Using network information to evaluate baseline anomalies***

We are examining the use of the Network Calibration Method to identify sensors whose baseline may be systematically biased or whose performance may indicate a malfunction.

It is also possible to use this method to correct for any residual ozone artefacts affecting $NO_2$ concentrations as the ozone filter material on the $NO_2$ sensors degrade over time. This potential error in the $NO_2$ data occurs when the $NO_2$ concentrations are relatively low and the ozone concentrations high. The Cambridge team and colleagues at NPL continue to develop this methodology which reduces this potential error considerably.

# Appendix

## Data Management: Data Flags and Validation

The flow of data from the pods, via the EI server and the AM Server (www.airmonitors.net), to the project partners CERC is achieved by two API's in series as shown below. Documentation is available online at:
https://api.airmonitors.net/3.5/documentation?key=C55638AM

Each data point received by Air Monitors is accompanied by a single fault code which is determined by the manufacturer using a hierarchical system. (See Table 2) Additional flags are applied by AM in Stages 2 or 3. The above flags are passed by the Air Monitors system via API to CERC as a JSON string.

For the platform, CERC continually maintains and updates a Google cloud database that contains all the 1-minute prescaled data, flags and scaling factors. Nothing is deleted. Before publishing the data, CERC redact data according to the flag rules in the table and calculate the hourly average (based on a data validity requirement of 85%) for publication on the platform and in data analysis. We use the prescaled data so that if scaling factors need to be changed retrospectively, we can do that. Similarly, flags can be changed retrospectively, after receiving manual instruction from AM to re-ingest previously ingested data.

## Air Monitors to CERC Data Interface (API)

Most of the above flags would result in the data being withheld from live publication. However, there are circumstances where publication may be allowed such as situations which are a result of data causing a flag which is within the uncertainty of measurement. For example, where data is "Optimising" or perhaps above a Hi Limit where the data is subsequently found to be valid. The final decision on whether to publish or not will be taken by the group and a set of rules will then be applied by CERC.

| Status Code | Sensor State | Description | Effects |
|---|---|---|---|
| -1000 | Not Fitted | Sensor or component not fitted. | Coded flag in data feeds as there is no data to view. |
| -999 | Stabilising | Either when the POD has just been moved to a new location or manually instigated via server. | Values are redacted as they cannot be relied upon during this 2-day period and will remain non-viewable. |
| *-998* | *Rebasing* | *Typically, this is a 2-day period where local variables are calculated for use in the AQMesh algorithm are determined.* | *During the rebasing period the coded flag will remain, however upon completion of this process, valid data will be reinstated – Data will need to be re-retrieved for this period, so API pointers are reset.* |
| -997 | Optimising | When a pod is power-cycled for more than an hour i.e. Maintenance or power failure. | Values are redacted as they cannot be relied upon during this 1-hour period and will remain as non-viewable. |
| -996 | Failed | The system has detected that the sensor has failed. | Data classified as having a sensor fail is redacted & will remain as non-viewable |
| -995 | Cross Gas Error | If a sensor fails which is relied upon for the removal of interferences on another sensor, data from the reliant sensor becomes invalid. | Data will be redacted & remain non-viewable until the compensating sensor is replaced and produces good results. |
| -994 | No Data | Data points where the instrument has not recorded a reading. | Coded flag in data feeds as there is no data to view. |
| -993 | Destabilised | The system has detected that the sensor's output \ stability may be compromised due to odd fluctuations in temperature and pressure. | Readings for the prescribed period are redacted and are non-viewable until the conditions have normalised. |

| | | | |
|---|---|---|---|
| -992 | Extreme Environment | Following intensive testing of all electrochemical sensors we have determined the combination of extremes in climate in which the electrochemical sensors do not provide consistent outputs. As such precise and accurate measurement is not possible | Data classified as within the extreme ranges of environmental conditions will be redacted and will be non-viewable. |
| -991 | Condensation | NDIR sensor has been affected by condensation on the detector | Data classified as being affected by condensation will be redacted and will remain as non-viewable. |
| Estimated Reading | Deliquescence | *When not using the heated inlet option, outlying data points caused by hygroscopic particle size growth will be flagged following analysis of the particle count distribution* | *These readings are available but should be considered as potentially unreliable.* |
| -893 | Mis-read | It is possible if the particle (or noise) sensor is unable to transfer valid data. | Occasional loss of data possible. Coded flag in data feed as there is no data to view. |
| -892 | Other fault zero | Due to the warm-up sequence and the timing of the event there is a chance that the particle counter is unable to provide a valid particle reading following a power-cycle and/or a change in pod settings. | Occasional loss of data possible. Coded flag in data feed as there is no data to view. |
| AM Array | Above Hi Limit | Concentrations exceed likely upper limit | Data is subjected to further scrutiny and may be redacted |
| AM Array | Above Hi- Hi Limit | Concentrations exceed upper sensor range | Data is redacted |

| | | | |
|---|---|---|---|
| AM Array | Below Low Limit | Concentrations exceed likely lower limit | Data is subjected to further scrutiny and may be redacted |
| AM Array | Below Low-Low Limit | Concentrations exceed lower sensor range | Data is redacted |

**Table 7.** Data Status Flag Array (and treatment for publication)