Measurement Uncertainty for $PM_{2.5}$ in the Context of the UK National Network

James Allan, Roy Harrison and Richard Maggs on behalf of the Air Quality Expert Group

Measurement of $PM_{2.5}$

$PM_{2.5}$ is defined as the mass of particulate matter per unit volume of air passing a size-selective inlet with a 50% cut point efficiency at 2.5µm particle aerodynamic diameter. It contains a wide range of particle sizes and chemical components, and unlike measurements of trace gases does not have a single correct concentration value. This is because the sampling and subsequent conditioning of the sample affects the composition, and hence the mass. It is therefore dependent upon the sample collection, curation and weighing procedures, and is referred to as an operationally defined measurement. Regulatory agencies address this issue by defining highly constrained methods, designed to give repeatable measurements which are reproducible between different laboratories, and allowing alternative measurement techniques, if they can demonstrate "equivalence". In the case of $PM_{2.5}$, there are reference methods set by the European Union and the USEPA (the Federal Reference Method, FRM). Both rely upon collecting particles on a filter, conditioning at defined temperature and relative humidity, and weighing. As they differ in key aspects, the European and US methods do not give identical results. These demand sampling times typically of a day or more which are unsuitable for near real-time public information, and hence there are a number of continuous instruments giving much faster response and tuned to demonstrate equivalence to the reference method(s).

The manual reference methods are subject to artifacts. Sampling on filters can produce positive artifacts from the collection of trace gases such as sulphur dioxide and nitric acid on the filter, although the use of optimal materials and long sampling intervals minimises the problem. Negative artifacts are a bigger problem. These arise from the evaporation of semi-volatile components, notably ammonium nitrate, ammonium chloride and some organic compounds, during air sampling and sample conditioning. This is partly dependent upon the temperature and pressure drop across the filter during sampling and is not easily replicated in continuous instruments. Some components of $PM_{2.5}$ have bound water associated which is not included in the definition of $PM_{2.5}$, and the standard protocols remove it during a conditioning stage prior to weighing. The conditioning protocols differ between the European and FRM methods. The early continuous Tapered Element Microbalance (TEOM) instruments had a heated inlet to remove bound water, and typically measured only 80% of the reference method mass due to loss of other semi-volatile components. The Filter Dynamics Measurement System (FDMS) which removes water with a permeation dryer and conducts a separate estimation of non-water volatiles achieves equivalence to the reference method. Other instrumental methods, notably the Beta Attenuation Monitor (BAM) and Fidas optical monitor are also used in the AURN.

The European Reference Method (EN12341)

Air is drawn over a nominal sampling period of 24 hours through a size-selective inlet at 2.3 m^3/hour. It then passes through a pre-weighed filter of high collection efficiency (>99.5% at 0.3µm) which may be of PTFE, glass fibre or quartz, which must be maintained at within 5^0C of the ambient temperature if that exceeds 20^0C. Filters are then conditioned in the weighing room to constant weight for at least 48 hours at between 19 – 21^0C and 45 – 50% Relative Humidity prior to weighing under the same conditions of temperature and RH.

Instrumentation Used in the UK Networks

The European Reference method involves a manual gravimetric technique for monitoring particulate matter. However, this method does not allow for continuous monitoring of PM due to the time required to weigh sampled filters and for processing the data. Other techniques, which have demonstrated equivalence to these two European Reference methods in the field, are used instead. There are four methods used in the UK's AURN for monitoring ambient concentrations of PM that have demonstrated equivalence to the European Reference Method. Where the sites measure both PM_{10} and $PM_{2.5}$ the instruments used for both metrics are the same; however at some sites just one PM size fraction is monitored. There are four different instruments used for the detection of PM, and for each instrument type used in the AURN, there is only one manufacturer. Each instrument has demonstrated equivalence to the European Reference PM method; however, there are still differences in their specifications, for example temporal resolution and limits of detection.

The methods used on UK networks appear in Table 1, and the number of instruments according to site classification in Table 2. The TEOM method is gravimetric, depending upon the changing oscillation frequency of a quartz crystal microbalance with an air filter attached to its tip. Adjustments for semi-volatile constituents are made by the FDMS. Beta Ray Attenuation (BAM) uses the beta ray absorption of particles collected on a filter as a measure of particle mass. This has a dependence upon the atomic numbers of elements present, so there is some influence of sample composition, and corrections for semi-volatiles are made. The FIDAS optical instrument counts particles in different ranges of optical diameter and uses an internal algorithm to estimate particle mass. The Partisol sampler is a gravimetric method which draws air at 16.7 L/min ($1m^3$/hour) through a size-selective inlet and sequentially sampled filters. The collected particles are then manually conditioned and weighed as in the European Reference Method. The SEQ 47/50 operates according to the Reference Method in EN12341.

Table 1: Instruments used in the UK Automatic Urban and Rural Network (AURN).

Instrument	Technique	Time resolution for PM measurements	Number of sites in AURN for PM_{10}	Number of sites in AURN for $PM_{2.5}$
Tapered element oscillating microbalance (TEOM) with Filter Dynamics System (FDMS)	TEOM1400 FDMS-8500B-CB	1 hr	6	5
Beta Ray Attenuation	BAM 1020	1 hr	42	31
Optical particle size spectrometer with conversion to mass concentration	FIDAS 200	15 mins	47	47
Active Sampling	Partisol 2025	1 day	1	1
Gravimetric Sampler	SEQ 47/50	1 day	1	0

Table 2: Instrument numbers used for $PM_{2.5}$ measurement in the AURN according to site classification.

For $PM_{2.5}$ Monitoring

Location type	TEOM1400 FDMS – 8500B-CB	BAM 1020	FIDAS 200	Partisol 2025
Background	2	14	38	0
Traffic	3	15	7	1
Industrial	0	2	2	0

The Guidance for Demonstration of Equivalence

The Directive is based on the adoption of common pollutant measurement methods by all Member States to provide for comparability in pollution climates across Europe in respect of reporting compliance. A Member State can deviate from adopting the “reference” methods through an exercise of undertaking comparisons between its preferred method (a “candidate” method) with that of the “reference” method. The Guidance for Demonstration of Equivalence 2010[1] provides for a standard protocol by which this can be achieved.

For **particulate samplers** the following provides a summary of the key steps:

Laboratory test

- Applicable to automated filter weighing changers leading to deviations in filter storage conditions specified in the standard
- Filter weighing provisions which use different conditions to those specified in the standard

Field test

- Operation of candidate and reference instruments side by side
- Between sampler variability of two instruments of the same type / model and comparability between candidate and reference methods
- Initial studies should be based on a minimum of 4 comparisons at 2 sites which vary in climatic conditions (i.e. summer and winter) and also vary in particulate composition (i.e. traffic / background/ industrial, et)
- Operation of instruments should reflect the adoption of procedures aligned to compliance network management operations

A copy of the original 2006 Equivalence Programme is available on UK-AIR at the following url:

https://uk-air.defra.gov.uk/assets/documents/reports/cat05/0606130952_UKPMEquivalence.pdf

Updated analysis of the original data was required in 2010 when the original GDE guidance was updated. The report is available at the following url:

https://uk-air.defra.gov.uk/assets/documents/reports/cat14/1101140842_Assessment_of_UK_AURN_PM_Equipment_against_2010_GDE.pdf

Expanded uncertainty is calculated at the Limit Value concentration and expressed as W_{CM}%, the maximum allowable value of which is 25%.

A definition of expanded uncertainty is provided as follows (cited from EN 16450: 2017:

[1] https://ec.europa.eu/environment/air/quality/legislation/pdf/equivalence.pdf

3.12
expanded uncertainty
quantity defining an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand

Note 1 to entry: The fraction may be viewed as the coverage probability or level of confidence of the interval.

Note 2 to entry: To associate a specific level of confidence with the interval defined by the expanded uncertainty requires explicit or implicit assumptions regarding the probability distribution characterized by the measurement result and its combined standard uncertainty. The level of confidence that may be attributed to this interval can be known only to the extent to which such assumptions may be justified.

[SOURCE: ISO/IEC Guide 98-3:2008 [7]]

In practice: a combined uncertainty (expressed in $\mu g/m^3$) is calculated as the square root of the combined square of the bias component and the square of the random component. The combined uncertainty is then divided by the limit value and multiplied by 100 to express as a percentage. This is multiplied by a coverage factor of 2 in order to have 95% confidence.

Measurement Uncertainty and Ongoing Demonstration of Equivalence of Existing AURN $PM_{2.5}$ Measurement Methods: Candidate Methods v Reference Method

The Directive provides for focus on two primary data quality objectives to be achieved by instruments deployed for $PM_{2.5}$ measurement in the UK. These are:

1) Data capture: For compliance reporting, the Directive requires that data capture be at least 90 %. The European Commission have subsequently released Guidance that allows for 5 % maintenance time, and stipulates that data capture should be at least 85 %.
2) Expanded uncertainty (W_{CM}%) of 25% to be achieved within each compliance year where W_{CM}% is calculated at the Limit Value.

In order to provide a means of evidencing the achievement of the expanded uncertainty for each instrument type a programme of “ongoing equivalence” is undertaken with comparisons between AURN deployed instruments and that of the EU reference method. This has used two sites located at London Teddington and Manchester Piccadilly. The number of sites at which ongoing equivalence should be undertaken is determined by EN16450:2017 (a relevant extract of which is provided below):

8.6.2 Number, duration and frequency of comparisons

There is a requirement for ensuring the on-going quality of the particulate measurement results obtained using the Automated Measurement System (AMS). This is particularly important because the type approval test procedure depends only on field tests between the reference method and AMS, and there is limited QA/QC that can be carried out on a routine basis (flow calibration, calibration of temperature and pressure sensors).

In addition, the type approval tests and suitability evaluation have been performed under a limited range of particulate compositions, which may not continue to be representative for the actual conditions. Therefore, it is necessary that periodic parallel measurements are performed with the reference method to confirm that the equivalence claims are still valid. For these measurements only one implementation of the reference method and one AMS are needed. The fraction of sites to be tested under this regime (with a minimum) will depend on the relative expanded uncertainty found by evaluating the combined data of the type-approval and suitability tests (see 8.2.1). The minimum requirements are given in the following Table.

Table 1: Requirements for on-going comparisons with the reference method.

W_{AMS}, in %	≤ 10	10 to 15	15 to 20	20 to 25
% of sites in network [a]	10	10	15	20
Number of on-going equivalence sites [a]	2	3	4	5

[a] The smaller of the two resulting numbers may be applied. The minimum number of on-going equivalence test sites is 2 for each AMS.

For example, when the relative expanded uncertainty for the AMS measurement results from the type-approval and suitability test results (see 8.2.1) is between 10 % and 15 %, comparisons shall be performed at a minimum of 2 or 3 sites – depending on the size of the monitoring network – during a full year. During this year a minimum of 80 valid data pairs shall be obtained. This may be achieved, e.g., by having the reference method sample every 4 days. One of the sites may be a location at which tests have been performed as a part of the initial suitability tests. Other sites shall be different from the initial test locations and may be changed each year to increase the coverage of the monitoring network. The sites shall be representative of the various conditions that are typical for the network.

8.6.3 Evaluation of test results

The results of these tests shall be evaluated yearly using data collected over the previous 3-year period, when available, using the approach described from 7.5.8 onwards. When the resulting uncertainty falls into a different category, the extent of tests for the next year shall be changed accordingly.

When the uncertainty is > 25 %, corrective actions shall be taken. These may include a recalibration of the method.

It may be favourable at a certain stage to use the data obtained to voluntarily recalibrate the method in order to reduce uncertainty, and, consequently, the extent of verification testing. The data used shall then fulfil the requirements given in 7.5. The investigations and actions shall be fully documented.

Summary of Data Capture and Expanded Uncertainties from the ongoing equivalence programme (2017 – 2019) for PM_{10} and $PM_{2.5}$ instruments:

Notes:

- *T = Teddington*
- *M = Manchester*
- *Figures in **bold** do not meet the data quality criteria (data capture and/or expanded uncertainty)*

Measurements are referenced against $PM_{2.5}$ collected on filters and then weighed under lab conditions that comply with the CEN standard for gravimetric weighing

Instrument	2017		2018		2019		Initial Testing
	Data Capture / %	Expanded Uncertainty / %	Data Capture / %	Expanded Uncertainty / %	Data Capture / %	Expanded Uncertainty / %	Expanded Uncertainties / %
PM_{10} FDMS 8500	94.5 (T)	13.8 (T)	96.6 (T) 98.1 (M)	14.7 (T) 22.1 (M)	91.9 (T) 92.8 (M)	11.6 (T) **25.6 (M)**	9.1 (5.6 to 31.6)
$PM_{2.5}$ FDMS 8500	**80.2 (T)**	10.2 (T)	96.2 (T) 94.7 (M)	**29.1 (T)** **50.3 (M)**	93.2 (T) 92.1 (M)	**30.5 (T)** **51.2 (M)**	15.6 (7.7 to 31.4)
PM_{10} Partisol 2025	98 (T)	14.3 (T)	Not Operating		Not Operating		8.0 (6.0 to 22.3)
$PM_{2.5}$ Partisol 2025	99 (T)	15.4 (T)	Not Operating		Not Operating		18.8 (11.3 to 30.0)
PM_{10} Unheated BAM 1020	86.6 (T)	**31.6 (T)**	**84.4 (T)** 87.3 (M)	22.4 (T) 7.5 (M)	**78.7 (T)** 96.1 (M)	**27.6 (T)** 12.8 (M)	10.4 (8.9 to 39.7)
$PM_{2.5}$ Smart Heated BAM 1020	93.6 (T)	14.6 (T)	85.5 (T) 90.4 (M)	**26.0 (T)** **59.5 (M)**	**81.3 (T)** 96.5 (M)	17.8 (T) **47.1 (M)**	12.6 (7.4 to 23.9)
PM_{10} Smart Heated BAM 1020	Not Operating		88.7 (T) **33.5 (M)**	13.7 (T) 6.1 (M)	**69.1 (T)** 95.2 (M)	23.4 (T) 6.4 (M)	13.8 (9.8 to 22.4)
Fidas 200 (Measuring PM_{10})	Not Operating		99.6 (T) 99.9 (M)	9.1 (T) 7.4 (M)	97.5 (T) 100.0 (M)	11.5 (T) 7.0 (M)	7.5 (5.7 to 19.1)
Fidas 200 (Measuring $PM_{2.5}$)	Not Operating		99.6 (T) 99.9 (M)	10.3 (T) **34.9 (M)**	97.5 (T) 100.0 (M)	9.1 (T) **28.0 (M)**	9.3 (8.5 to 22.4)

The following headline observations are made:

- In 2017, all but the PM_{10} Unheated BAM 1020 met the 25% Expanded Uncertainty requirement, and all but the $PM_{2.5}$ FDMS 8500 met the 85 % Data Capture requirement.
- In 2018 there were six continuous instruments in both Teddington and Manchester, making a total of twelve continuous instruments. Of these, all but the PM_{10} Unheated BAM 1020 at Teddington, and the PM_{10} Smart Heated BAM 1020 at Manchester met the 85 % Data Capture requirement. All six PM_{10} instruments met the 25 % Expanded Uncertainty requirement. However, of the six $PM_{2.5}$ instruments, only the Fidas 200 at Teddington met the 25 % Expanded Uncertainty requirement. In each case, this was due to a high slope with the Equivalent Method reading greater than the Reference Method.
- In 2019 there were again six continuous instruments in both Teddington and Manchester. Of these, all but the three BAM 1020s at Teddington met the 85 % Data Capture requirement and the low data capture for these was due to persistent leaks found during the leak tests. The list of instruments failing the 25 % Expanded Uncertainty requirement was similar to that during 2018, but there were differences. For PM_{10}, all but the PM_{10} FDMS 8500 at Manchester and PM_{10} Unheated BAM 1020 at Teddington passed the 25 % criterion. However, of the six

$PM_{2.5}$ instruments, only the Fidas 200 and $PM_{2.5}$ Smart Heated BAM 1020 at Teddington met the 25 % Expanded Uncertainty requirement. In each case, this was due to a high slope with the Equivalent Method reading greater than the Reference Method.

In both 2018 and 2019, differences between sites are stark and show that failure to achieve the necessary criteria is more evident at Manchester. The precise reasons for this are unclear but observations from the site operative indicated that it may be an influence of market stall cooking for the street food market held in Manchester Piccadilly Gardens. Closer investigation into the effect of the market stalls indicates that one stall was affecting concentrations from July to September 2018 whereupon it was moved. However, such differences were still observed in 2019 so the precise reason is still unclear as to why differences in achieving the expanded uncertainty are so stark between sites, although it is likely related to differences in the $PM_{2.5}$ composition.

Measurement Uncertainty and Ongoing Demonstration of Equivalence of Existing AURN $PM_{2.5}$ Measurement Methods: Intercomparison of Instrumental Methods

The data may also be used to compare instrument types when run together.

Summary of Expanded Uncertainties from the ongoing equivalence programme (2017 – 2019) for $PM_{2.5}$ instrument comparability between instruments:

Note

- *T = Teddington*
- *M = Manchester*

n = number of observations

Candidate instrument on y-axis	**FDMS 8500**	-		
	FIDAS 200	(T) 2018: 22.9 (n=353) (M) 2018: 16.2 (n=229) (T) 2019: 22.9 (n=341) (M) 2019: 23.9 (n=329) (T) 2020: 17.1 (n=320) (M) 2020: tbc	-	
	Heated BAM	(T) 2014: 25.0 (n=133) (T) 2015: 15.5 (n=291) (T) 2016: 11.6 (n=316) (T) 2017: 11.0 (n=281) (T) 2018: 15.9 (n=286) (M) 2018: 22.9 (n=244) (T) 2019: 12.9 (n=294) (M) 2019: 16.7 (n=328) (T) 2020: 10.4 (n=314) (M) 2020: tbc	2014: n/a 2015: n/a 2016: n/a 2017: n/a (T) 2018: 15.7 (n=287) (M) 2018: 12.8 (n=212) (T) 2019: 15.6 (n=298) (M) 2019: 21.1 (n=351) (T) 2020: 14.5 (n=317) (M) 2020: 14.0 (n=317)	-
		FDMS 8500	**FIDAS 200**	**Heated BAM**
		Candidate instrument on x-axis		

The following headline observations are made:

- FDMS $PM_{2.5}$ comparisons with FIDAS 2000 show achievement of 25% W_{CM} , but towards the higher level of the accepted criteria
- FDMS $PM_{2.5}$ comparison with Heated BAM have shown improvement in initial comparisons over that observed in the initial roll out of instruments (2014) due to operational knowledge and understanding. Subsequent year to year comparisons between instruments are fairly consistent in performance achieving W_{CM} of between 11.0 – 15.9%
- FIDAS 2000 comparison with Heated BAM are limited currently due to later adoption of the FIDAS but show consistent achievement of W_{CM} at ~ 15.6%.

- No Partisol comparisons are provided due to withdrawal of instrumentation from the network after 2018
- Additional comparisons between instruments at Manchester are currently ongoing with more variability shown, possibly due to the influences of the food market. 2020 represents an improvement in comparisons for previous years, possibly reflecting the impact of lockdown and the fact that no market was generally present from March 2020 onwards.

It should be stressed that the equivalence comparisons presented here are strictly relevant to current $PM_{2.5}$ and the slopes and R^2 statistics of the comparisons with the reference method are strongly influenced by the data points at higher concentrations. Noting that there are significant deviations from the 1:1 line in the <10 µg m^{-3} regime in some datasets, the implication is that under a lower-concentration scenario in the future, the equivalence of these technologies may be less good. Furthermore, while the case of the Manchester Piccadilly site being influenced by food cooking does not undermine the equivalence as currently defined, it does illustrate the sensitivity of the comparison to different particle compositions. This may prove important because under a future scenario where $PM_{2.5}$ from other sources (e.g. transport, domestic burning and regional pollution) is successfully reduced, cooking may come to represent a higher fraction of urban $PM_{2.5}$, particularly in areas with high numbers of restaurants and street vendors. Taking the Manchester Piccadilly comparison as an extreme case, it could be expected that the different measurement techniques could significantly diverge under this scenario. But taken more generally, if any other hypothetical emerging sources prove to become significant in the future, these may also result in instrumental biases if not properly characterised. Taken together, these highlight the need for ongoing ambient benchmarking exercises, with particular attention being paid to the role of $PM_{2.5}$ composition in any instrumental divergence.

Conclusions

1. The expanded uncertainty data, which represent both random and systematic divergences, show substantial differences (typically up to around 25%) both between instrumental monitors and the reference method, and between different instrumental monitor types.
2. The comparisons of monitor types when viewed graphically (not shown) show apparent systematic differences between monitor types. Tests of variations between monitors of the same type have not been conducted, so this factor cannot be accounted for.
3. Any change to the instrument type at a single site will introduce a discontinuity into the dataset, making the evaluation of changes in air quality more challenging.
4. Changes in instrument types at one or more sites will affect the ability to evaluate changes in mean concentrations across a number of sites, as might be used in a Population Exposure Reduction Target.
5. A rigorous statistical analysis is needed to establish the uncertainties in data reported for an individual site, and across an aggregate of sites, and to define the smallest change in concentration which could be determined with confidence.
6. Compositional changes in $PM_{2.5}$, especially in terms of the semi-volatile components, are liable to affect the mass measurements in ways which it is not currently possible to quantify.