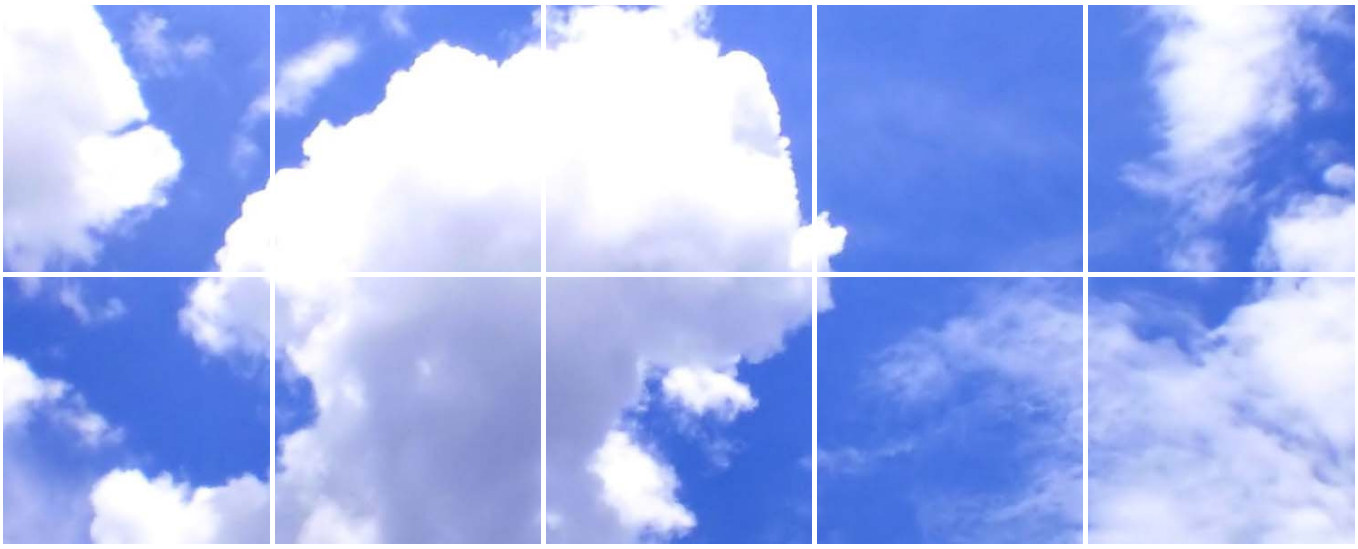
AEA



Field Investigation of Inter-Laboratory Variation in NO_2 Diffusion Tube Measurements 2009

Report to Defra and the Devolved Administrations

ED48673045
Issue 1
May 2010

Title	Field Investigation of Inter-Laboratory Variation in NO_2 Diffusion Tube Measurements 2009
Customer	Defra and the Devolved Administrations
Customer reference	Contract no. RMP 2877
Confidentiality, copyright and reproduction	This report is the Copyright of Defra and the Devolved Administrations and has been prepared by AEA Technology plc under contract to Defra dated November 2005. The contents of this report may not be reproduced in whole or in part, nor passed to any organisation or person without the specific prior written permission of Defra and the Devolved Administrations. AEA Technology plc accepts no liability whatsoever to any third party for any loss or damage arising from any interpretation or use of the information contained in this report, or reliance on any views expressed therein.
File reference	ED48673045
Reference number	AEAT/ENV/R/2936 – Issue 1

AEA group
551.11 Harwell Business Centre
Didcot
Oxfordshire
OX11 0QJ

t: 0870 190 6518
f: 0870 190 6377

AEA is a business name of AEA Technology plc

AEA is certificated to ISO9001 and ISO14001

Author	Name	Alison Loader
Project Team	Name	Alison Loader, Jaume Targa, Tony Clark, Paul Willis, Sarah Winne
Approved by	Name	Paul Willis
	Signature	
	Date	24/05/10

Executive summary

This report summarises a three-month field investigation of measurement of ambient nitrogen dioxide (NO_2) concentration using Palmes-type diffusion tubes, carried out on behalf of Defra and the Devolved Administrations. The objective was to investigate whether the harmonisation of diffusion tube preparation and analysis (introduced in January 2009) has reduced inter-laboratory variation in diffusion tube performance, compared to that observed in a similar study prior to harmonisation, in 2007. Tubes were exposed at the London Teddington air quality monitoring site at the National Physical Laboratory (NPL), co-located with a chemiluminescence analyser (which is defined as the reference method for NO_2.)

Seven of the eight laboratories that took part in the original 2007 study participated in this study (the eighth no longer analyses diffusion tubes). The study was carried out at the same site as the 2007 study, at the same time of year and using the same exposure duration (four weeks).

Two diffusion tube preparation methods were used:

- 50% TEA in acetone, grids dipped into solution and allowed to dry before tube assembly
- 20% TEA in water, solution pipetted onto grids resting in caps before tube assembly

Four participants prepared tubes using the acetone method, five used the water method and two used both.

The precision of the results, and their accuracy (with respect to the reference method) were compared with those obtained in the 2007 study. In the case of the two laboratories using both preparation methods, the precision and accuracy of the two methods were also compared with each other.

Precision of tubes prepared using the 50% TEA/acetone method did not appear to have improved in 2009 compared with the 2007 study (based on the results from the three laboratories that used this method in the 2007 and 2009 studies). Precision of tubes prepared using the 20% TEA/water method was better than in the 2007 study (based on five laboratories). However, it is not possible to directly attribute this to the harmonisation process, because data from a larger ongoing monthly field intercomparison indicates that mean precision has improved over the period 2006 – 2008. It is likely that the results of the present study simply reflect this general improvement.

Diffusion tubes prepared by both methods exhibited substantial over-read (positive bias) with respect to the automatic analyser in 2009. This is in contrast to the 2007 study in which there was a mixture of over- and under-estimation, and in which the mean bias was within +/- 5% for both methods. While this might at first appear to indicate worse performance, diffusion tubes are known to be affected by several sources of interference that tend to produce positive bias. So, the 2009 results are more consistent with the expected behaviour of diffusion tubes in the field. By contrast, substantial negative bias (as observed in many of the 2007 results) often results from inefficient extraction and is something that the harmonisation process aimed to eliminate.

There was some evidence of slightly improved inter-laboratory agreement in the case of the 50% TEA/acetone method. There was evidence of a greater improvement in inter-laboratory agreement in the case of the 20% TEA/water method: it may be that harmonisation has been more beneficial in the case of the 20% TEA/water method, as there was previously more potential for variation in this tube preparation technique.

Neither of the two preparation methods gave significantly better performance. The 50% TEA/acetone method appeared to provide slightly better precision, but the 20% TEA/water method gave consistently higher results and therefore may have the advantage of better collection of NO_2. However, the difference was only significant at the 95% confidence level in one case out of six.

The analytical uncertainties on individual measurements, as quoted by the participants, varied considerably and were in some cases comparable with the precision of replicate measurements. This should be investigated further.

Table of contents

1 Introduction

This report presents the results of a three-month field intercomparison study of inter-laboratory variation in the results of Palmes-type NO_2 diffusion tubes, analysed by seven different analytical laboratories. The study was carried out by AEA in collaboration with the National Physical Laboratory, in 2009.

1.1 Background

Palmes-type diffusion tubes are widely used by UK Local Authorities for indicative monitoring of nitrogen dioxide (NO_2). However, considerable variation in performance was observed between NO_2 diffusion tubes prepared and analysed by different laboratories, and it was suspected that variation in tube preparation and analysis methods was one of several contributing factors.

Therefore, Defra and the Devolved Administrations commissioned AEA and Air Quality Consultants to set up and manage a Working Group on harmonisation of NO_2 diffusion tube preparation and analysis methods. This work was undertaken during 2006 and 2007 as part of the Defra contract RMP 2877, for Support to Local Authorities for Air Quality Management.

As part of the work of this Working Group, in 2007 AEA carried out a three-month field intercomparison trial of tubes prepared using five different preparation methods[1]. The aim was to test the precision and accuracy of Palmes-type NO_2 diffusion tubes prepared using five different combinations of triethanolamine (TEA) solution and method of application. Eight laboratories participated in the study. The trial comprised three four-week exposure periods, during which batches of six replicate tubes of each type, from each laboratory, were exposed at an urban background air quality monitoring site at the National Physical Laboratory in Teddington, London. They were co-located with an automatic chemiluminescence analyser (defined within the EU as the reference method for NO_2).

The 2007 study concluded that there was no significant difference in the results of tubes prepared by the two methods that were in widespread use, i.e.

- 50% TEA in acetone, grids dipped into solution and allowed to dry before tube assembly
- 20% TEA in water, solution pipetted onto grids resting in caps before tube assembly

None of the other three methods tested appeared to offer any significant improvement over these. Therefore the Working Group recommended that laboratories should continue to use either one of the above methods, in the absence of any further information.

However, this study also clearly illustrated the problem of inter-laboratory variation: in many cases the results obtained by two laboratories using the same methods differed more than the results obtained by the same laboratory using the two different methods. This highlighted the importance of harmonising not only diffusion tube preparation method, but also analytical procedures.

The Working Group produced (in February 2008) a Practical Guidance document[2] aimed at both laboratories working with diffusion tubes, and end users - in particular, Local Authorities using diffusion tubes for Local Air Quality Management purposes. This Guidance set out a harmonised method for preparation and analysis of diffusion tubes, and it was intended that all suppliers and analysts of diffusion tubes used for Local Air Quality Management (LAQM) in the UK should implement the harmonised method by 1st January 2009.

1.2 Objectives

The present study aimed to repeat the relevant elements of the 2007 field intercomparison trial, with the objective of investigating whether the harmonisation of diffusion tube preparation and analysis has reduced inter-laboratory variation in diffusion tube performance, compared to that observed in 2007.

2 Experimental Details

2.1 Summary

The investigation involved exposure of diffusion tubes prepared by different laboratories alongside an automatic chemiluminescent analyser (which is defined by the EU as the reference method for NO_2.)

The trial comprised three exposure periods, each of nominally four weeks, during the period April to July 2009. These were as follows:

1 Period 1: 24th April – 22nd May 2009 (28 days)
2 Period 2: 22nd May – 19th June 2009 (28 days)
3 Period 3: 19th June – 16th July 2009 (27 days).

The trial was run at around the same time of year as the 2007 trial, and the exposure periods were the same. The same site was used: London Teddington, which is part of the UK's Automatic Urban and Rural Network (AURN). The National Physical Laboratory (NPL), who operate the site, carried out the tube changes, as they had in 2007.

Seven laboratories participated; all were commercial suppliers and analysts of Palmes-type NO_2 diffusion tubes, and all were participants in the original 2007 trial. (The eighth participant in the original trial has since ceased analysing diffusion tubes). The seven laboratories are identified here by the same identification numbers used in the 2007 study, Lab 2 to Lab 8 (Lab 1 having ceased operation).

Where the 2007 trial had tested five diffusion tube preparation methods, the 2009 study included only the two methods permitted by the harmonised method, i.e.

- 50% TEA in acetone, grids dipped into solution and allowed to dry before tube assembly
- 20% TEA in water, solution pipetted onto grids resting in caps before tube assembly

Not all participants were able to use both of these methods. In the 2009 study, two used only the acetone method, three used only the water method, and two used both.

- Lab2 used the acetone method only
- Lab 3 used the acetone method only
- Lab 4 used both methods
- Lab 5 used the water method only
- Lab 6 used the water method only
- Lab 7 used both methods
- Lab 8 used the water method only.

Each laboratory provided (for each method they were using), six replicate tubes for exposure, plus a travel blank, per period.

2.2 Diffusion Tube Preparation

Diffusion tubes were prepared according to the methods set out in the Practical Guidance. The triethanolamine (TEA) solution was to be prepared according to section 2.5 of the Practical Guidance, and applied to the grids as set out in section 2.6 of the same document.

2.3 Tube Labelling

A tube labelling convention was suggested to ensure that tube types were not mixed up. Tubes were to be assigned a unique identifier of the format:

XXXX-P1-Y-N

Where:
XXXX was the 3 - or 4 - digit laboratory identification code as used routinely in the Field Intercomparison. This was used to identify the laboratory.
P1 = period 1 (tubes for subsequent periods were labelled P2 and P3.)
Y = preparation method - eg "A" for acetone or “W” for water.
N = number of replicate tube, 1 to 6. The letters FB were used to denote the field blank (travel blank).
Some participants used other formats but these were clear and unambiguous, so were permitted.

2.4 Site

The site used was the London Teddington urban background monitoring site, shown in Figure 2.1. This site is part of the Defra Automatic Urban and Rural Network (AURN) and so is subject to rigorous standards of QA/QC. The Local Site Operator services for this site are provided by the National Physical Laboratory (NPL). The site is located on the roof of NPL’s London offices: this had the disadvantage that the site was a relatively exposed and windy location. However, for a study of this type it was essential to select a site which firstly had space to accommodate the large number of tubes, and secondly was inaccessible to the public, to prevent tube theft or tampering.

The four-weekly tube changes were carried out by NPL. NPL also recorded the exact exposure times of each set of tubes.

All tubes were exposed within 12m of the automatic analyser inlet. This is not ideal for a co-location study: ideally tubes should be within 5m of the analyser inlet. However, the site is a long way from any sources of NOx, so all tubes were likely to be exposed to similar NO_2 levels. Also, all tubes were likely to be exposed to similar conditions in terms of wind exposure, temperature and turbulence.

Figure 2.1 Tube exposure at London Teddington

2.5 Calculation of ambient concentration

After exposure, the diffusion tubes were returned as soon as possible to the laboratory that had prepared them, for analysis and calculation of measured ambient concentrations. The laboratories analysed the tubes and reported the results to AEA. Ambient concentration was calculated as follows:

$$C = \frac{1}{"s.rate"} \times \frac{m}{t}$$

where:
C = ambient concentration, ($\mu g\ m^{-3}$)
m = the mass of nitrite in the tube, as determined by analysis
t = exposure time,
"s. rate" = sampling rate: this is calculated from the tube dimensions and the diffusion coefficient of NO_2 in air, and is treated as a constant:

$$Sampling\ rate = \frac{D_{12}\, a}{l}$$

where:

a	=	the cross sectional area of the tube (dependent on tube manufacturer)
l	=	the length of the tube (dependent on tube manufacturer)
D_{12}	=	diffusion coefficient of gas 1 through gas 2 – in this case NO_2 through air.

As specified in the Practical Guidance document, all laboratories were using a value of D = 0.146 $cm^2\ s^{-1}$, which is based on a mean ambient UK temperature of 284K.

3 Results and Discussion

3.1 Automatic NO_2 Measurements

Results and data capture statistics, from the automatic NO_2 analyser at London Teddington, are shown in Table 3.1. All data have now been fully ratified.

Table 3-1 Automatic Analyser Results

Period	From	To	Mean NO_2 concentration, µg m^{-3}	Data Capture, %
Period 1	15:00 24th April 09	14:30 22nd May 09	13.7	99.4
Period 2	14:45 22nd May 09	15:00 19th June 09	15.9	99.6
Period 3	15:00 19th June 09	15:00 16th July 09	15.3	99.4

Although every effort was made to replicate the same conditions as in the 2007 study, it was obviously not possible to control ambient NO_2 concentration. Whereas in the 2007 study the 4-week mean concentrations for the three exposure periods were 38.6 µg m^{-3}, 18.7 µg m^{-3} and 22.8 µg m^{-3} respectively, the 4-week means for the three exposure periods in 2009 (shown in Table 3.1) were lower and varied less (all were within the range 10-20 µg m^{-3}). The 4-week mean NO_2 concentrations for the three periods in the 2009 study were 13.7 µg m^{-3}, 15.9 µg m^{-3} and 15.3 µg m^{-3} respectively.

Automatic analyser data capture was greater than 99% throughout the trial.

3.2 Diffusion Tube Measurements

Each of the participating laboratories supplied and analysed six replicate tubes (plus one travel blank) for each preparation method tested (one or both of the two methods).

Because the study aimed to investigate uncertainty, suspect results have not generally been discarded. The exceptions are where the tube was clearly damaged or contaminated: there were two such cases, one where a tube was broken in transit, another where the grid was contaminated by insect droppings.

In some cases (particularly in period 2 of 2009) the site operator put the designated travel blank out for exposure instead of one of the tubes numbered 1 to 6. As all these cases were clearly identified on the exposure sheets, it was possible to substitute the relevant tube results and this has been done.

The precision, or uncertainty, of each laboratory's tubes was assessed by comparing the coefficient of variation (CV) of each set of six replicate tube measurements. (The coefficient of variation, also known as the relative standard deviation, is the standard deviation expressed as a percentage of the mean). For consistency with the 2007 study, the comparison of CV was based on the mass of nitrite reported for each tube, rather than the calculated ambient NO_2 concentration.

On the basis of previous experience with diffusion tubes, the CV of six replicates would be expected to be within 10%. A CV greater than 20% usually indicates that there is a problem with some of the measurements, or that there is an outlying value.

Accuracy (sometimes termed bias) was assessed by comparison of the mean of the set of six replicate results with the "reference" result from the automatic analyser. This was assessed on the basis of the calculated mean NO_2 concentration. Statistics used in this report include:

- The standardised result, i.e. the ratio of the mean diffusion tube result to the reference measurement

- The percentage over-estimation or under-estimation, often termed “bias”, and calculated as (D-C)/C where C = concentration as measured by the chemiluminescent analyser and D is the concentration as measured by the diffusion tubes.
- The 95% confidence interval of the mean of the six replicate measurements. In the 2007 study, this was calculated by a Microsoft Excel spreadsheet: in the present study the 95% confidence interval of the mean is calculated using a slightly different formula, based on the t-distribution and suitable for small sample sizes such as n=6.

In accordance with the harmonised method, travel blank results have not been subtracted.

Table 3.2 shows the individual tube results, mean, standard deviation and CV for each laboratory in the 2007 study. (Only the 50% TEA/acetone dipping method and 20% TEA/water pipetting method are shown, although other methods were included in the 2007 study).

Table 3.3 shows the individual tube results, mean, standard deviation and CV for each laboratory in the 2009 study.

Table 3-2 Diffusion Tube Results 2007

Laboratory	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6	Mean	St Dev	CV%
Period 1									
Lab 1	35.0	31.9	28.2	28.7	27.2	24.4	29.2	3.74	12.8%
Lab 2	32.1	32.8	32.6	32.3	36.1	30.4	32.7	1.87	5.7%
Lab 3	39.9	38.2	37.5	38.8	39.5	38.2	38.7	0.88	2.3%
Lab 4 (acet.)	34.6	34.5	32.7	31.3	32.0	26.0	31.9	3.16	9.9%
Lab 4 (water)	31.8	29.0	32.2	28.1	30.2	29.7	30.2	1.59	5.3%
Lab 5	48.2	29.0	34.5	32.2	33.5	44.5	37.0	7.60	20.5%
Lab 6	20	28	38	30	31	38	30.8	6.77	21.9%
Lab 7	34.0	44.8	43.0	37.1	36.1	34.2	38.2	4.58	12.0%
Lab 8	33.0	31.3	36.6	33.2	34.2	33.9	33.7	1.74	5.2%
Period 2									
Lab 1	15.4	19.3	18.3	14.9	15.6	15.9	16.5	1.77	10.7%
Lab 2	Faulty tube	Faulty tube	Faulty tube	Faulty tube	Faulty tube	Faulty tube	-	-	-
Lab 3	21.0	21.3	21.3	21.8	22.5	19.7	21.3	0.94	4.4%
Lab 4 (acet.)	20.5	23.1	20.7	25.1	23.4	15.6	21.4	3.35	15.6%
Lab 4 (water)	20.9	23.4	24.1	24.9	21.4	23.3	23.0	1.54	6.7%
Lab 5	20.3	19.4	20.1	23.4	21.2	21.7	21.0	1.42	6.7%
Lab 6	21.0	20.0	20.0	19.0	21.0	16.0	19.5	1.87	9.6%
Lab 7	25.8	23.3	24.5	23.3	22.9	24.4	24.0	1.09	4.6%
Lab 8	Reject	reject	18.4	reject	18.5	16.9	17.9	0.90	5.0%
Period 3									
Lab 1	20.3	19.8	21.0	19.6	28.2	17.4	21.0	3.71	17.6%
Lab 2	21.0	23.8	21.5	23.3	20.4	19.6	21.6	1.65	7.6%
Lab 3	24.9	24.5	23.1	23.6	22.9	24.4	23.9	0.82	3.4%
Lab 4 (acet.)	24.0	26.4	26.5	28.0	28.8	27.2	26.8	1.65	6.2%
Lab 4 (water)	25.5	24.5	24.1	26.5	27.3	27.0	25.8	1.33	5.2%
Lab 5	29.7	29.2	35.2	32.7	29.4	30.1	31.1	2.42	7.8%
Lab 6	22.0	25.0	23.0	21.0	21.0	17.0	21.5	2.66	12.4%
Lab 7	24.7	25.1	27.8	25.5	26.9	24.5	25.8	1.31	5.1%
Lab 8	20.8	23.9	22.0	21.8	23.1	21.4	22.2	1.14	5.1%

Results were variously reported to 0, 1 or 2 decimal places: 1 decimal place is shown here.

Table 3-3 Diffusion Tube Results 2009

Laboratory	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6	Mean	St Dev	CV%
Period 1									
Lab 2	15.0	17.1	18.2	16.5	11.9	14.6	**15.6**	15.0	14.3%
Lab 3	15.7	14.7	17.5	16.3	16.3	17.2	**16.3**	15.7	6.3%
Lab 4 (acet.)	14.7	14.7	15.6	14.5	13.4	14.7	**14.6**	14.7	4.8%
Lab 7 (acet.)	14.1	14.7	14.5	13.1	14.5	14.7	**14.3**	14.1	4.2%
Lab 4 (water)	16.8	16.4	15.8	16.7	16.5	16.8	**16.5**	16.8	2.3%
Lab 5	19.0	18.5	18.1	18.4	unexposed	17.9	**18.4**	19.0	2.4%
Lab 6	18.0	15.0	14.0	15.0	14.0	18.0	**15.7**	18.0	11.9%
Lab 7(water)	16.9	16.1	16.8	17.6	16.0	14.6	**16.3**	16.9	6.5%
Lab 8	17.0	15.4	15.5	16.1	15.1	16.5	**15.9**	17.0	4.6%
Period 2									
Lab 2	20.5	21.9	21.8	19.4	22.0	17.3	**20.5**	20.5	9.2%
Lab 3	22.0	21.0	23.0	23.0	22.0	22.0	**22.2**	22.0	3.4%
Lab 4 (acet.)	19.1	19.6	20.0	17.4	20.1	20.4	**19.4**	19.1	5.6%
Lab 7 (acet.)	18.5	20.2	20.2	19.1	18.7	19.3	**19.3**	18.5	3.9%
Lab 4 (water)	20.8	19.7	19.2	reject - 15.7	22.1	20.0	**20.4**	20.8	5.6%
Lab 5	25.3	22.2	23.2	24.0	23.4	23.2	**23.5**	25.3	4.4%
Lab 6	24.0	23.0	22.0	25.0	24.0	25.0	**23.8**	24.0	4.9%
Lab 7(water)	20.7	21.5	23.4	22.6	22.5	23.2	**22.3**	20.7	4.7%
Lab 8	20.2	21.6	22.0	21.0	18.6	20.6	**20.7**	20.2	5.8%
Period 3									
Lab 2	21.8	23.2	21.3	20.7	24.7	20.4	**22.0**	21.8	7.5%
Lab 3	19.0	19.0	18.0	18.0	17.0	17.0	**18.0**	19.0	5.0%
Lab 4 (acet.)	17.4	17.0	18.8	18.0	18.3	18.8	**18.1**	17.4	4.1%
Lab 7 (acet.)	17.2	18.3	17.6	19.1	18.4	17.2	**18.0**	17.2	4.0%
Lab 4 (water)	16.5	20.8	17.3	19.1	19.0	18.7	**18.6**	16.5	8.1%
Lab 5	21.3	21.2	22.1	21.7	21.5	22.2	**21.7**	21.3	1.9%
Lab 6	18.0	18.0	23.0	23.0	20.0	18.0	**20.0**	18.0	12.2%
Lab 7(water)	19.2	18.3	18.6	19.3	21.5	20.1	**19.5**	19.2	5.9%
Lab 8	22.3	21.7	21.5	22.7	21.9	17.3	**21.2**	22.3	9.3%

Results were variously reported to 0, 1 or 2 decimal places: 1 decimal place is shown here.

3.2.1 Analytical Uncertainty

Following the 2009 trial, the laboratories were asked to estimate the uncertainty on a diffusion tube analysis (i.e. analytical uncertainty only, as would be expected for an artificially spiked tube, not including any errors relating to exposure-related factors). The measurement uncertainty quoted by the laboratories varied considerably, from ±2.5% to ±10.9%;

Lab 2: ±7%
Lab 3: ±4%
Lab 4: ±5%
Lab 5: ±8%
Lab 6: no value given (median value of ±5% therefore used as an estimate in this study.)
Lab 7: ±10.9%
Lab 8: ±2.5%

Laboratories were not asked for this information in the 2007 study. The measurement uncertainty at that time was not necessarily the same as in 2009, as the harmonisation process has involved changes to most laboratories' procedures. But in the absence of any better information, and in the interests of taking a consistent approach, the same analytical uncertainty values have been used for the 2007 data.

Figures 3.1, 3.2 and 3.3 show each laboratory's results in periods 1, 2 and 3 of the 2007 study respectively. The (2009) analytical uncertainty on each individual measurement is shown by a black error bar. These figures also show the mean of all six replicates, with the 95% confidence interval of the mean represented as a red error bar.

Figures 3.4, 3.5 and 3.6 show the same information for periods 1, 2 and 3 of the 2009 study.

It should be noted that for some participants (for example, Lab 7) the quoted analytical uncertainty on individual measurements is comparable to, and in some cases higher than, the 95% confidence interval on the mean. Therefore the 95% confidence interval is of limited usefulness in identifying significant differences between mean results, as is highlighted in subsequent sections of this report.

Figure 3-1 Diffusion tube measurements in period 1 of 2007 study

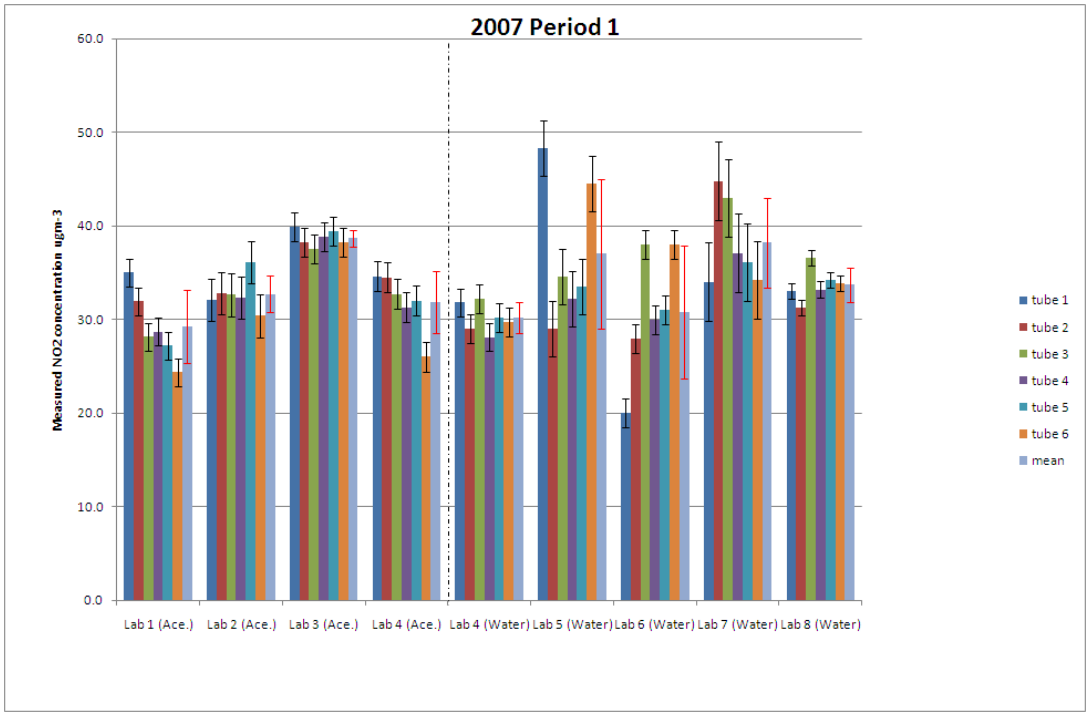


Figure 3-2 Diffusion tube measurements in period 2 of 2007 study

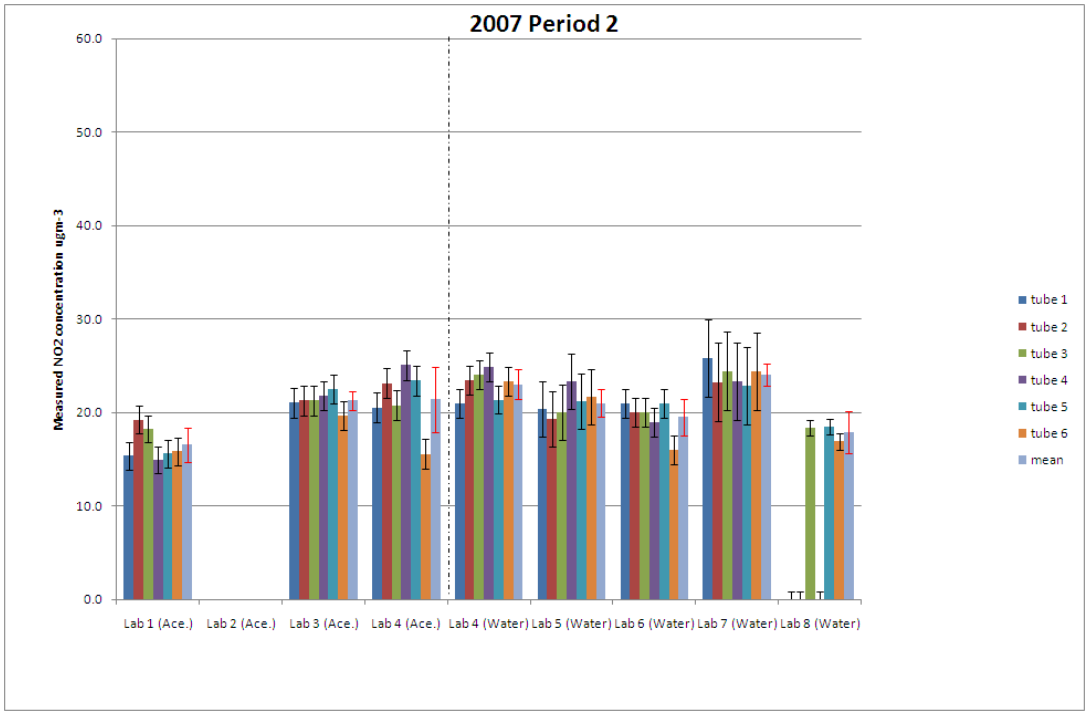


Figure 3-3 Diffusion tube measurements in period 3 of 2007 study

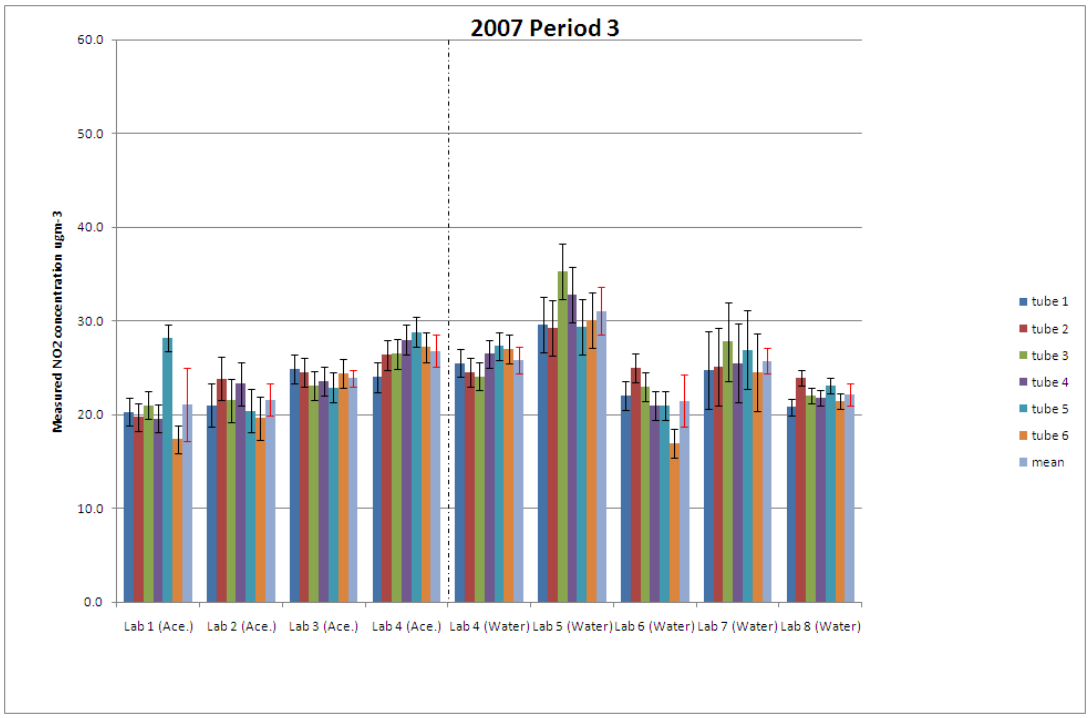


Figure 3-4 Diffusion tube measurements in period 1 of 2009 study

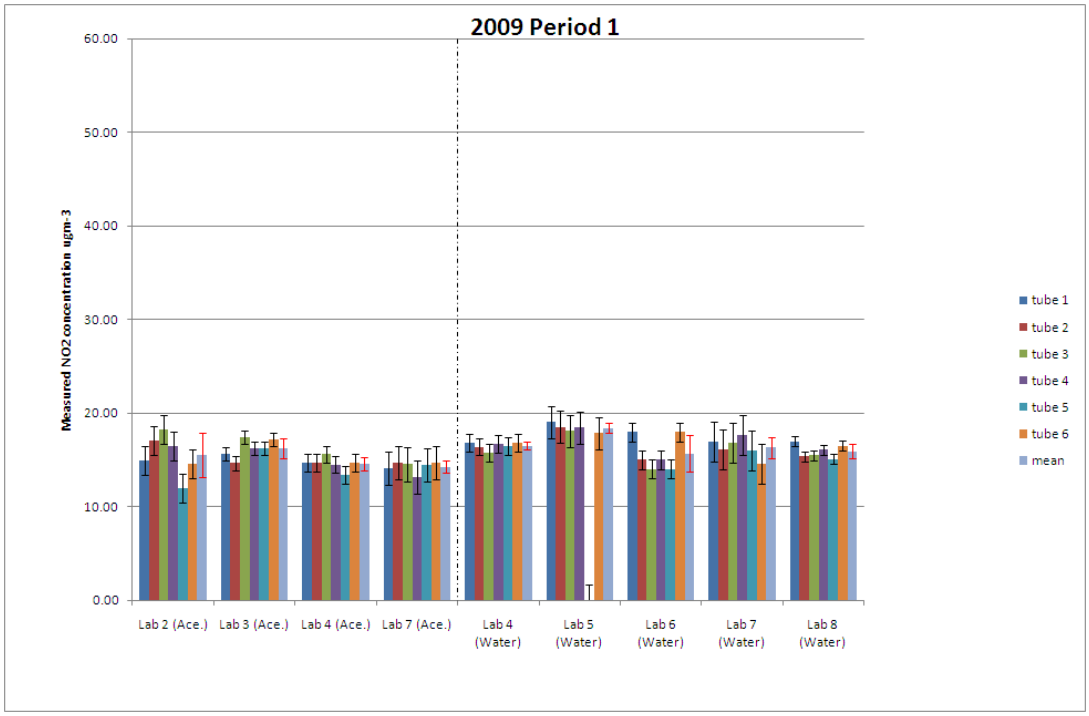


Figure 3-5 Diffusion tube measurements in period 2 of 2009 study

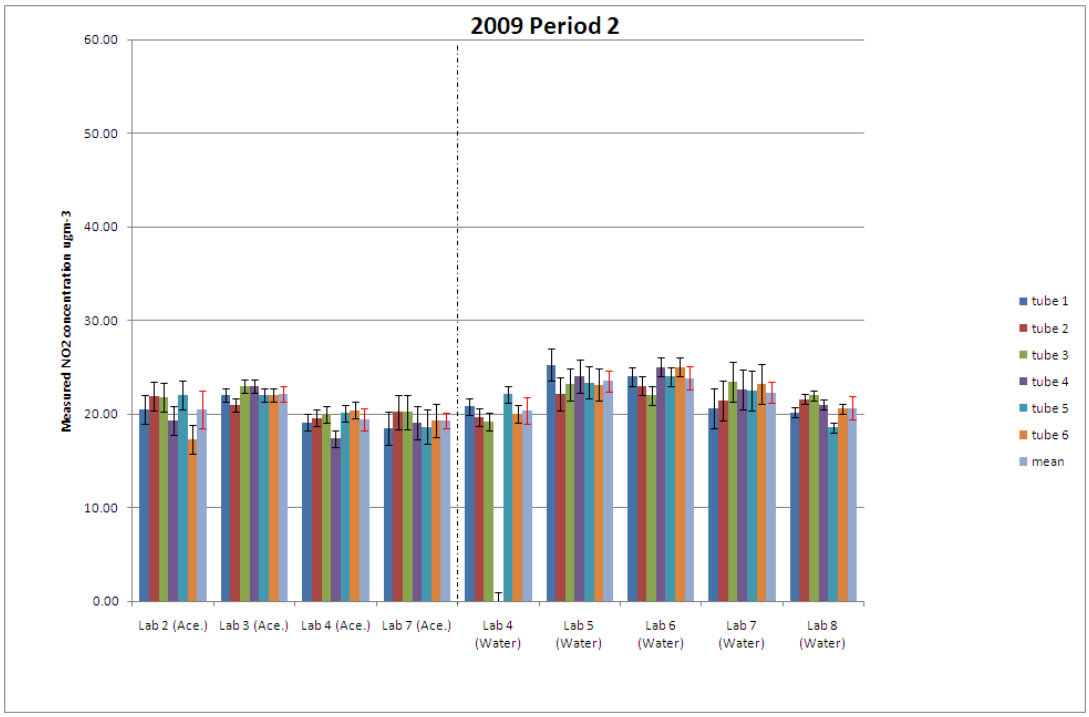
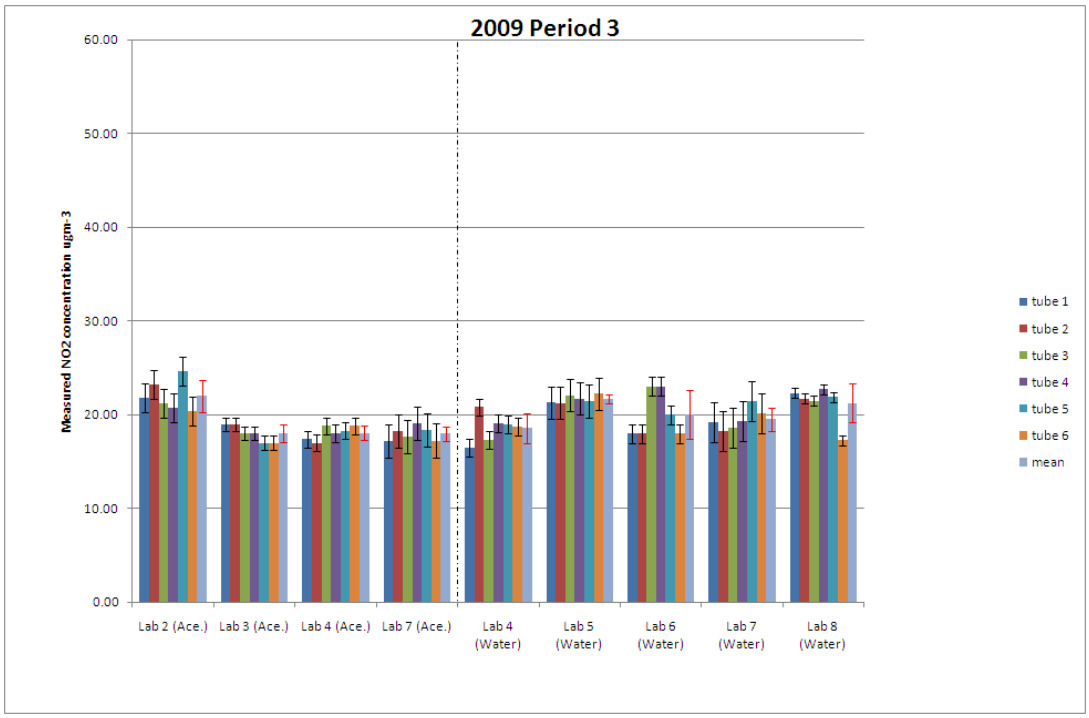


Figure 3-6 Diffusion tube measurements in period 3 of 2009 study



3.3 Precision

3.3.1 50% TEA in Acetone Method

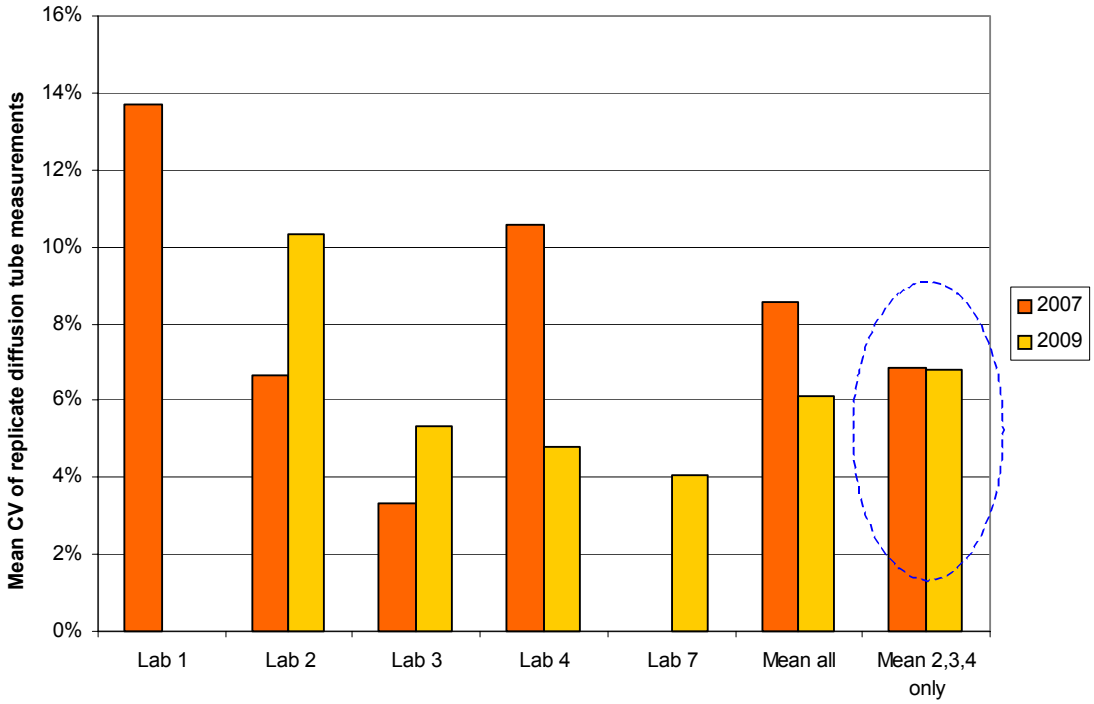
Table 3.4 summarises the precision for the acetone method, in terms of the coefficient of variation of the six replicate measurements for each laboratory from each period. Results from both the 2007 and 2009 studies are included. The mean precision, averaged over all labs, appears at first sight to have improved from 8.56% in 2007 to 6.13% in 2009: however, this is not a valid comparison as neither Lab 1 nor Lab 8 took part in both studies. When comparing the mean CV for Lab 2, 3, and 4 only (the three laboratories that participated in both years), there is little change: from 6.85% to 6.82%. When comparing the CV for the individual laboratories, Lab 2 and (in particular) Lab 3 had worse precision in 2009 than in 2007: only Lab 4's precision was better in 2009. These results are illustrated in Figure 3.7.

There is therefore no evidence that the changes introduced by the harmonisation process have improved precision for laboratories using the method of dipping the grids in a 50% solution of TEA in acetone.

Table 3-4 Comparison of Precision (CV) 2007 and 2009, 50% TEA/acetone (dipping) method

Lab	2007 P1	2007 P2	2007 P3	Mean 2007	2009 P1	2009 P2	2009 P3	Mean 2009
Lab 1	12.78%	10.67%	17.63%	13.69%				
Lab 2	5.72%	No data	7.61%	6.67%	14.25%	9.21%	7.50%	10.32%
Lab 3	2.22%	4.36%	3.36%	3.31%	6.33%	3.59%	6.13%	5.35%
Lab 4	10.04%	15.65%	6.03%	10.57%	4.69%	5.48%	4.23%	4.80%
Lab 7					4.20%	3.90%	4.05%	4.05%
Mean all				**8.56%**				**6.13%**
Mean 2,3, 4 only				**6.85%**				**6.82%**

Figure 3-7 Comparison of Precision (CV) 2007 and 2009, acetone method



3.3.2 20% in TEA/ water method

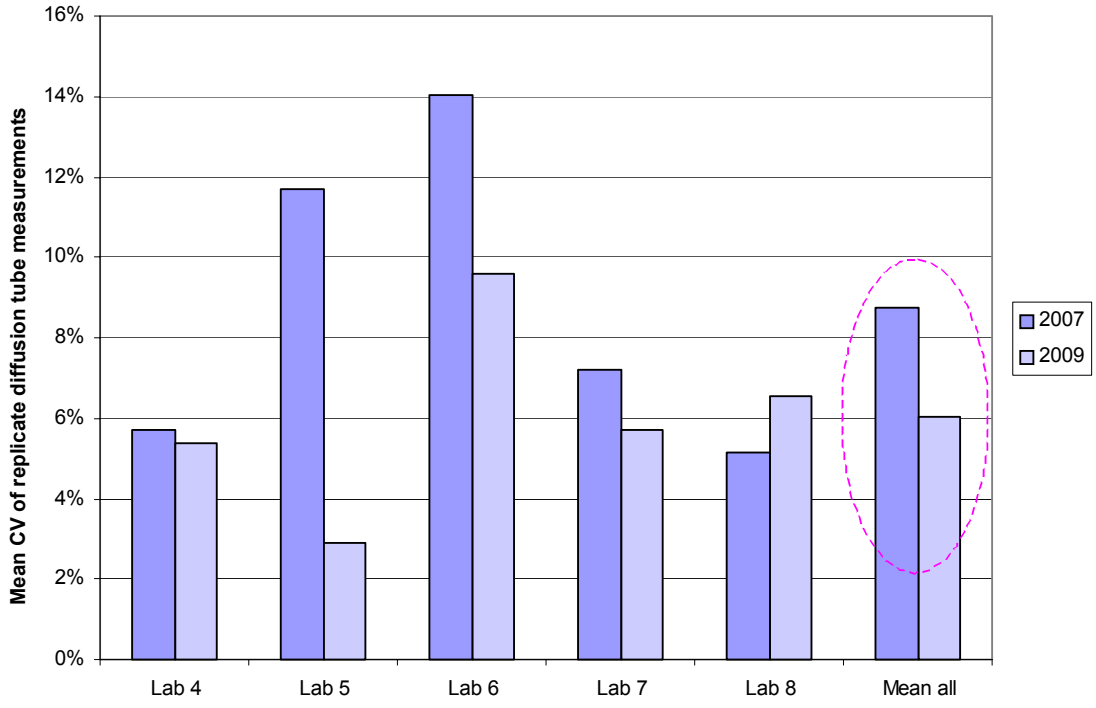
Table 3.5 summarises the precision for the water method (20% TEA in water, solution pipetted onto grids). For this method, five laboratories took part in the 2007 and 2009 studies. Mean precision in the 2009 trial is better than that in the 2007 trial, and four of the five laboratories showed improved precision in 2009 compared with their results in 2007. These results are illustrated in Figure 3.8.

It therefore appears from this short-term study that precision of diffusion tubes prepared by the 20% TEA/water pipetting method has improved between 2007 and 2009. However, it is important to note that this is not necessarily attributable to the harmonisation process: information from the ongoing monthly field intercomparison in central London[3] indicates that diffusion tube precision has generally improved on average over the three years 2006 - 2008.

Table 3-5 Comparison of Precision (CV) 2007 and 2009, 20% TEA/water (pipetting) method

Lab	2007 P1	2007 P2	2007 P3	Mean 2007	2009 P1	2009 P2	2009 P3	Mean 2009
Lab 4	5.31%	6.71%	5.07%	5.69%	2.41%	5.33%	8.34%	5.36%
Lab 5	20.54%	6.74%	7.80%	11.69%	2.39%	4.42%	1.89%	2.90%
Lab 6	20.93%	8.96%	12.23%	14.04%	10.60%	4.96%	13.21%	9.59%
Lab 7	11.99%	4.55%	5.09%	7.21%	6.51%	4.71%	5.88%	5.70%
Lab 8	5.10%	5.17%	5.14%	5.14%	4.57%	5.88%	9.22%	6.56%
Mean all				**8.75%**				**6.02%**

Figure 3-8 Comparison of Precision (CV) 2007 and 2009, water method



3.4 Accuracy

Accuracy is expressed here in terms of the percentage over-read or under-read with respect to the reference method. As explained above, this is often referred to as “bias”, and calculated as

Bias = (D-C)/C (expressed as a percentage)

- where C = concentration as measured by the chemiluminescent analyser and D is the concentration as measured by the diffusion tubes.

3.4.1 50% TEA in Acetone Method

Table 3.6 summarises the accuracy of the diffusion tube measurements, for the 50% TEA/acetone (dipped grids) method. In both years, mean over- or under-read relative to the reference analyser (“bias”) varied considerably from month to month. Laboratories exhibited a mixture of under- and over-read in 2007. Negative values are highlighted in blue in this table. In 2009, no tubes under-read. The mean bias for the three laboratories that participated in both years (i.e. Lab 2, Lab 3 and Lab 4) was 0.3% in 2007, and 23.2% in 2009. So while the mean of the 2007 results was close to zero, the mean of the 2009 results showed a substantial positive bias. This might initially be viewed as an indication of worse performance; however, this is not the case for diffusion tubes. There are known sources of interference affecting diffusion tubes (such as wind and UV effects); these result in positive bias, so the observed positive bias is consistent with their expected behaviour. By contrast, negative bias often results from inefficient extraction (something that the Working Group sought to eliminate), so the fact that there was only positive bias in the 2009 results should not be viewed as indicative of a problem.

Table 3-6 Comparison of Accuracy (as % "bias") 2007 and 2009, 50% TEA/acetone (dipping) method

Lab	2007 P1	2007 P2	2007 P3	Mean 2007	2009 P1	2009 P2	2009 P3	Mean 2009
Lab 1	-24.3%	-11.5%	-7.7%	-14.5%				
Lab 2	-15.2%	No data	-5.2%	-10.2%	13.5%	28.8%	43.7%	28.6%
Lab 3	0.2%	13.7%	4.8%	6.2%	18.6%	39.4%	17.6%	25.2%
Lab 4	-17.5%	14.5%	17.6%	4.9%	6.6%	22.2%	18.0%	15.6%
Lab 7					4.2%	21.5%	17.4%	14.4%
Mean all				**-3.4%**				**21.0%**
Mean 2,3, 4 only				**0.30%**				**23.2%**
Max – Min bias, all	24.5%	26.0%	25.3%	**25.3%**	**14.5%**	**17.9%**	**26.3%**	**19.5%**
Max – min bias, 2,3,4 only	17.7%	0.8%	22.8%	**13.7%**	**12.1%**	**17.2%**	**26.0%**	**18.4%**

Table 3.6 also shows the range of bias (maximum – minimum) for each exposure period. (The range of bias is compared for each exposure period separately, rather than across all three periods in each year, because bias can vary considerably from month to month due to environmental or meteorogical factors.) This gives an indication of how closely the various participants' results agreed over the same period.

For all laboratories, the average range of bias in the 2009 trial (19.5%) was smaller than that in the 2007 trial (25.3%). However, neither Lab 1 nor Lab 7 participated in both studies. If only Lab 2, Lab 3 and Lab 4 (i.e. the laboratories that participated in both studies) were considered, the mean range of bias in 2009 was actually greater than in 2007 (18.4% compared with 13.7%). On this basis it does not appear that harmonisation has improved inter-laboratory agreement for the 50% TEA/acetone method.

Figure 3.9 compares the bias for acetone tubes in both years. In the 2009 trial, there appeared to be close agreement between Lab 4 and Lab 7: with the exception of period 3, there was also reasonable agreement between Lab 2 and Lab 3. It may be useful to further investigate whether there are any remaining differences between procedures used by these pairs of laboratories.

Figure 3-9 Comparison of percentage bias of 50% TEA/acetone diffusion tubes in 2007 and 2009 3-month field studies (each bar is mean of 6 replicates)

3.4.2 20% in TEA/ water method

Table 3.7 summarises the accuracy of the diffusion tube measurements, for the 20% TEA/water (pipetted) tubes. In both years, mean over-/under-read ("bias") varied considerably from month to month. Negative values are highlighted in blue in this table. As in the case of the acetone method, laboratories exhibited a mixture of under- and over-read in 2007, with a mean of 4.3%. In 2009, all tubes over-read compared to the reference method. The mean bias, based on all five laboratories, was 30.7%. So, as in the case of the acetone method tubes above, the magnitude of the bias was greater in 2009, but more consistent with what is expected of diffusion tubes exposed in the field.

Table 3.7 also shows, for each exposure period, the range of bias (minimum to maximum). The spread of bias in each exposure period was typically lower in 2009, possibly indicating better inter-laboratory agreement for the 20% TEA in water method. The range of bias was lower in 2009 (20.7%) than in 2007 (31.9%). This would appear to indicate an improvement in inter-laboratory agreement for this method.

Figure 3.10 compares the bias observed for this method in both years.

With the unharmonised 20% TEA/water method, there was potential for variation in terms of pipetting procedures, volume of solution applied, application technique etc., whereas for the acetone method the grids were simply submerged in the solution. It may be that harmonisation has been more beneficial in the case of the water method, as there was previously more variation.

Table 3-7 Comparison of Accuracy (as % "bias") 2007 and 2009, 20% TEA/water (pipetting) method

Lab	2007 P1	2007 P2	2007 P3	Mean 2007	2009 P1	2009 P2	2009 P3	Mean 2009
Lab 4	-21.8%	24.3%	13.2%	5.2%	20.4%	28.1%	21.4%	23.3%
Lab 5	-4.1%	13.5%	36.2%	15.2%	34.3%	48.0%	41.6%	41.3%
Lab 6	-20.1%	5.4%	-5.7%	-6.8%	14.4%	49.9%	30.7%	31.7%
Lab 7	-1.0%	29.9%	13.0%	13.9%	19.2%	40.2%	27.4%	28.9%
Lab 8	-12.7%	-3.1%	-2.8%	-6.2%	16.3%	30.0%	38.8%	28.4%
Mean all				**4.3%**				**30.7%**
Max – Min bias	20.8%	32.9%	41.9%	**31.9%**	19.9%	21.8%	20.3%	**20.7%**

Figure 3-10 Comparison of percentage bias of 20% TEA/water diffusion tubes in 2007 and 2009 3-month field studies (each bar is mean of 6 replicates)

3.5 Inter-laboratory agreement

Section 3.4 highlighted that there appeared to be a reduction in the range of bias, relative to the automatic analyser result, for the 20% TEA/water method (though not for the 50% TEA/acetone method). Inter-laboratory agreement was investigated further by attempting to identify cases where, for any two laboratories using the same method, the mean of the 6 replicates was significantly different.

Several approaches were used, and these are discussed below.

3.5.1 Comparison of 95% Confidence Interval

One approach to identifying cases where two laboratories' results are significantly different is to compare the ranges covered by the 95% confidence interval of the mean. If these overlap, this indicates no significant difference (at the 95% confidence level) between the mean results obtained by the two laboratories concerned.

However, in this case, this method has two limitations:

(i) It is based on the assumption that either the population the laboratory results are drawn from is normally distributed, or that the sample size is sufficiently large. This assumption may not hold, especially with a sample size of only six.

(ii) It does not take account of the analytical uncertainty on each of the six replicate measurements, which varied considerably between the various laboratories in the 2009 study, and was in some cases as high as 10.9%. This information is not available for the 2007 study.

Therefore, this approach is used here ***for indicative purposes only***; the number of cases where two laboratories using the same preparation method got mean results that differed at the 95% confidence level was calculated for 2007 and for 2009 and compared. This is used as an indication of whether inter-laboratory agreement has changed in the intervening period. The 95% confidence intervals were calculated using equation 1: this is the formula applicable where the sample size is small[4]:

$$\textbf{95\% confidence interval} = t_{(n-1,\ 0.05)} \times s/\sqrt{n} \qquad \textbf{Eqn. 1}$$

where t is the critical value of the t statistic for a probability of 0.05 and n-1 degrees of freedom. In most cases in this study, n = 6 and $t_{(5,\ 0.05)}$ = 2.571. Acetone tubes were not compared with water tubes or vice-versa.

Figures 3.1, 3.2 and 3.3 (for 2007) and Figures 3.4, 3.5 and 3.6 (for 2009) include the mean result, with the 95% confidence interval shown by the red error bars.

Acetone Method

Table 3.8 and Table 3.9 show, for the 2007 and 2009 trials respectively, cases where the 95% confidence interval of the mean did and did not overlap. Only Lab 2, Lab 3 and Lab 4 are included here, as Lab 1 and Lab 7 did not participate in both years.

Table 3-8 2007 trial, acetone method: overlap of 95% confidence intervals of mean diffusion tube results

	Period 1 2007			Period 2 2007			Period 3 2007		
	Lab2	**Lab3**	**Lab4**	**Lab2**	**Lab3**	**Lab4**	**Lab2**	**Lab3**	**Lab4**
Lab 2		**X**	✓		**-**	-		✓	**X**
Lab 3			**X**			✓			**X ***
Lab 4									

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

Table 3-9 2009 trial, acetone method: overlap of 95% confidence intervals of mean diffusion tube results

	Period 1 2009			Period 2 2009			Period 3 2009		
	Lab2	**Lab3**	**Lab4**	**Lab2**	**Lab3**	**Lab4**	**Lab2**	**Lab3**	**Lab4**
Lab 2		✓	✓		✓	✓		**X**	**X**
Lab 3			✓			**X**			✓
Lab 4									

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

In 2007, for this particular method, there were four instances where two laboratories' mean results differed significantly on this basis out of seven possible pairs of laboratories (i.e. 57% of pairings). In 2009, there were three such instances out of 9 possible pairs (i.e. 33%). This indicates better inter-laboratory agreement in 2009 than in 2007, for the 50% TEA/acetone method.

In both trials, there were some instances where although the 95% confidence intervals did not overlap there was an overlap between the two sets of six replicate results. This would be expected: the 95% confidence intervals relate to the *mean* result. As a more stringent test, these cases were excluded, and the number of instances where any two laboratories' results differed significantly again compared. In this case, there were three (out of seven possible pairs, i.e. 43%) in 2007 and three (out of nine possible pairs, i.e. 33%) in 2009.

Water Method

Table 3.10 and Table 3.11 show, for the 2007 and 2009 trials respectively, cases where the 95% confidence interval of the mean did and did not overlap.

Table 3-10 2007 trial, water method: overlap of 95% confidence intervals of mean diffusion tube results

	Period 1 2007					Period 2 2007					Period 3 2007				
Lab	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**
Lab 4		✓	✓	**X**	**X***		✓	✓	✓	**X**		**X**	**X***	✓	**X**
Lab 5			✓	✓	✓			✓	**X ***	✓			**X**	**X**	**X**
Lab 6				✓	✓				**X**	✓				**X**	✓
Lab 7					✓					**X**					**X**
Lab 8															

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

Table 3-11 2009 trial, water method: overlap of 95% confidence intervals of mean diffusion tube results

	Period 1 2009					Period 2 2009					Period 3 2009				
Lab	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**
Lab 4		**X**	✓	✓	✓		**X**	**X***	✓	✓		**X**	✓	✓	✓
Lab 5			**X***	**X**	**X**			✓	✓	**X**			✓	**X***	✓
Lab 6				✓	✓				✓	**X***				✓	✓
Lab 7					✓					✓					✓
Lab 8															

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

In 2007, for the water method, there were 14 instances where any two laboratories' mean results differed significantly on this basis out of 30 possible pairs of laboratories (i.e. 47% of cases). In 2009, there were 10 such instances out of 30 possible pairs (i.e. 33% of cases). This indicates better inter-laboratory agreement in 2009 than in 2007, for the 20% TEA/water method.

As in the case of the acetone method, there were instances in both trials where there was an overlap between the two sets of six replicate results despite there being no overlap between the 95% confidence intervals of the two means. If these instances are not counted, the number of instances where any two laboratories' results differed significantly was 11 (out of 30 possible pairs, i.e. 37%) in

2007 and six (out of 30 possible pairs, i.e. 20%) in 2009. This still indicates that inter-laboratory agreement had improved between 2007 and 2009, for this method.

3.5.2 Welch-Satterthwaite t-test

Significant differences were further investigated using a two-tail, small-sample (Welch-Satterthwaite) t-test to compare the laboratories' results. The same limitations apply as were highlighted in section 3.5.1 above: the sample size is small, the samples may not be normally distributed, and the uncertainty on individual measurements is not taken into account. **Therefore, this test is also used only for indicative purposes.**

The test statistic t was calculated as shown in Equation 2; this calculation assumes that the data are not paired in any way, does not assume that the values are normally distributed, or that the two sets of six values have equal standard deviations.

$$t = (m_2 - m_1) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)} \qquad \text{Eqn. 2}$$

- Where m_1 and m_2 are the means of the two sets of six replicate measurements being compared (based on acetone and water respectively), s_1 and s_2 are the standard deviations of the two sets of values, and n_1 and n_2 are the number of replicates in each case (usually six).

The number of degrees of freedom for each test is given by Equation 3:

$$((s_1^2/n_1) + (s_2^2/n_2))^2 / ((s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)) \qquad \text{Eqn. 3}$$

This gives non-integer values which are rounded to the nearest integer when looking up t in statistical tables.

Acetone tubes were not compared with water tubes or vice-versa.

This t-test identified a considerable number of cases where two laboratories' diffusion tubes (prepared using the same method) gave significantly different mean results – more than the confidence interval approach used above.

Acetone Method

For the acetone method, in 2007, there were five instances (out of seven possible pairs of laboratories excluding Lab 1 and Lab 7) where two laboratories' mean results differed significantly on the basis of the t-test (i.e. 71%). This is shown in Table 3.12. In 2009, there were four such instances out of nine possible pairs (i.e. 44%), as shown in Table 3.13. This indicates better inter-laboratory agreement in 2009 than in 2007, for the 50% TEA/acetone method.

However, in some of these cases although the t-test indicated a significant difference, there was an overlap between the two sets of six replicate results. These cases are indicated by an asterisk. As a more stringent test, these cases were excluded. This left just two instances in 2007 (out of seven pairs, i.e. 29%), and three instances in 2009 (out of nine pairs, i.e. 33%) where the mean results obtained by two laboratories using the acetone method differed significantly on the basis of the t-test. On this basis, the test does not show an improvement in inter-laboratory variation, for the acetone method.

Table 3-12 2007 trial, acetone method: Result of Welch-Satterthwaite t-test comparing mean diffusion tube results

	Period 1 2007			Period 2 2007			Period 3 2007		
	Lab2	Lab3	Lab4	Lab2	Lab3	Lab4	Lab2	Lab3	Lab4
Lab 2		X	✓		-	-		X*	X
Lab 3			✓			X*			X*
Lab 4									

✓ = t-test indicates no significant difference at 95% confidence level.
X = t-test indicates means significantly differ at 95% confidence level. Asterisk indicates that actual datasets overlapped.

Table 3-13 2009 trial, acetone method Result of Welch-Satterthwaite t-test comparing mean diffusion tube results

	Period 1 2009			Period 2 2009			Period 3 2009		
	Lab2	Lab3	Lab4	Lab2	Lab3	Lab4	Lab2	Lab3	Lab4
Lab 2		✓	✓		✓	✓		X	X
Lab 3			X*			X			✓
Lab 4									

✓ = t-test indicates no significant difference at 95% confidence level.
X = t-test indicates means significantly differ at 95% confidence level. Asterisk indicates that actual datasets overlapped.

Water Method

For the water method, in 2007, there were 17 instances (out of 30 possible pairs of laboratories, i.e. 57%) where two laboratories' mean results differed significantly on the basis of the t-test. This is shown in Table 3.14. In 2009, there were 14 such instances out of 30 possible pairs (i.e. 47%), as shown in Table 3.15. This indicates better inter-laboratory agreement in 2009 than in 2007, for the 20% TEA/water method.

As in the case of the acetone method, in some of these cases, although the t-test indicated a significant difference, there was an overlap between the two sets of six replicate results. Excluding such cases left 11 instances in 2007 (out of 30 pairs, i.e. 37%), and three instances in 2009 (out of 30 pairs, i.e. 20%) where the mean results obtained by two laboratories using the water method differed significantly on the basis of the t-test. This again indicates better inter-laboratory agreement in 2009 than in 2007.

Table 3-14 2007 trial, water method: Result of Welch-Satterthwaite t-test comparing mean diffusion tube results

	Period 1 2007					Period 2 2007					Period 3 2007				
Lab	4	5	6	7	8	4	5	6	7	8	4	5	6	7	8
Lab 4		✓	✓	X*	X		X*	X*	✓	X		X	X*	✓	X
Lab 5			✓	✓	✓			✓	X*	X			X	X	X
Lab 6				✓	✓				X	✓				X	✓
Lab 7					✓					X					X*
Lab 8															

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

Table 3-15 2009 trial, water method. Result of Welch-Satterthwaite t-test comparing mean diffusion tube results

	Period 1 2009					Period 2 2009					Period 3 2009				
Lab	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**	**4**	**5**	**6**	**7**	**8**
Lab 4		**X**	✓	✓	✓		**X**	**X***	X*	✓		**X**	✓	✓	X*
Lab 5			**X***	**X**	**X**			✓	✓	**X**			✓	**X***	✓
Lab 6				✓	✓				X*	**X***				✓	✓
Lab 7					✓					X*					✓
Lab 8															

✓ = 95% confidence intervals overlap, i.e. means the same.
X = no overlap of 95% confidence intervals, i.e. means differ. Asterisk indicates that actual datasets did overlap.

The above results indicate that inter-laboratory agreement for the acetone method improved slightly between 2007 and 2009. For the water method, the results indicate that inter-laboratory agreement has improved more markedly.

3.5.3 Consideration of Analytical Uncertainty

Neither of the above approaches take into account the analytical uncertainty on individual diffusion tube measurements, as discussed in section 3.2.1. It would not be valid to conclude that two sets of results were significantly different, if the difference was less than the analytical uncertainty.

The analytical uncertainty reported in 2009 varied considerably, ranging from ± 2.5% to ± 10.9%. It was not clear in every case how this parameter had been calculated. The laboratories were not asked for this information in the 2007 study, and the uncertainty may not have been the same in 2007 as it was in 2009.

Therefore it is difficult to deal with the analytical uncertainty in a robust way. The most stringent approach is to say that, when comparing any two laboratories' results, the difference between the two mean results should only be considered significant if there is no overlap between the lowest result *(minus analytical uncertainty)*, obtained by the laboratory with the higher mean, and the highest result *(plus the analytical uncertainty)*, obtained by the laboratory with the lower mean. That is, when comparing any two laboratories' results in Figures 3.1 - 3.6, not only the datasets themselves, but also the black error bars (representing the analytical uncertainty) should not overlap.

This makes the conditions for "significance" much stricter: the number of cases in which two laboratories' results differ significantly on this basis was reduced to just one: in period 3 of the 2007 study, Lab 5 got significantly higher results than Lab 8, using the water method.

The results above *indicate* that inter-laboratory agreement has improved between 2007 and 2009, particularly for the water method. However, the large uncertainty on individual measurements, and the fact that measurement uncertainty in the 2007 trial was not investigated, makes it difficult to draw reliable conclusions as to whether the difference is statistically significant.

3.6 Comparison of Water and Acetone Methods

Two of the seven participating laboratories used both the preparation methods included in the 2009 trial, i.e.

- 50% TEA in acetone, grids dipped into solution and allowed to dry before tube assembly
- 20% TEA in water, solution pipetted onto grids resting in caps before tube assembly

These were the laboratories designated Lab 4 and Lab 7.

Table 3.16 compares the precision (expressed as the coefficient of variation) obtained by each of these two laboratories, for the two methods.

Table 3-16 Comparison of Precision between Preparation Techniques, 2009 only

		CV P1	CV P2	CV P3	mean CV
Lab 4	acetone	4.8%	5.6%	4.1%	4.8%
Lab 7	acetone	4.2%	3.9%	4.0%	4.1%
Lab 4	water	2.3%	5.5%	8.1%	5.3%
Lab 7	water	6.5%	4.7%	5.9%	5.7%

On average, CV was marginally lower (i.e. precision better) for the acetone method than for the water method. However, this was not consistently the case – for example in period 1 Lab 4 obtained better precision using the water method.

Table 3.17 compares the results obtained by each of these two laboratories, for the two methods.

Table 3-17 Comparison of Results Obtained Using the Two Preparation Techniques, 2009 only - measured concentration (µg m^{-3})

Lab	Type	Period	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6	Mean	St Dev	95% Confidence Interval
Lab 4	Acetone	1	14.7	14.7	15.6	14.5	13.4	14.7	**14.6**	0.70	0.739
Lab 4	Water	1	16.8	16.4	15.8	16.7	16.5	16.8	**16.5**	0.38	0.398
Lab 7	Acetone	1	14.10	14.67	14.53	13.13	14.47	14.73	**14.27**	0.60	0.630
Lab 7	Water	1	16.90	16.08	16.80	17.64	15.96	14.55	**16.32**	1.06	1.115
Lab 4	Acetone	2	19.1	19.6	20.0	17.4	20.1	20.4	**19.43**	1.09	1.147
Lab 4	Water	2	20.8	19.7	19.2	Reject*	22.1	20.0	**20.36**	1.13	1.302
Lab 7	Acetone	2	18.48	20.22	20.22	19.08	18.65	19.30	**19.32**	0.75	0.791
Lab 7	Water	2	20.65	21.47	23.41	22.58	22.49	23.17	**22.30**	1.05	1.102
Lab 4	Acetone	3	17.4	17.0	18.8	18.0	18.3	18.8	**18.05**	0.74	0.773
Lab 4	Water	3	16.5	20.8	17.3	19.1	19.0	18.7	**18.57**	1.51	1.581
Lab 7	Acetone	3	17.22	18.26	17.64	19.07	18.37	17.24	**17.97**	0.73	0.763
Lab 7	Water	3	19.15	18.30	18.60	19.34	21.45	20.13	**19.50**	1.15	1.203

** rejected tube – insect droppings on grid.*

Table 3.18 expresses the observed differences in terms of accuracy (bias) relative to the reference automatic analyser.

Table 3-18 Comparison of Accuracy between Preparation Techniques, 2009 only

		bias P1	bias P2	bias P3	mean bias
Lab 4	acetone	6.6%	21.5%	-0.3%	9.3%
Lab 7	acetone	4.2%	20.8%	-0.7%	8.1%
Lab 4	water	20.4%	27.3%	2.6%	16.8%
Lab 7	water	19.2%	39.3%	7.7%	22.1%

Both laboratories got consistently higher results from tubes prepared using the 20% TEA/water method compared with those prepared using the 50% TEA/acetone method. The 95% confidence interval was used to investigate whether the difference was significant for either laboratory.

Figure 3.11 compares the mean results obtained by Lab 4, using tubes prepared by the acetone method and the water method, in each period of the 2009 study. The error bars represent the 95% confidence interval (calculated by the same method as above). Figure 3.12 shows the same comparison for Lab 7.

Figure 3-11 Comparison of Results, Acetone and Water Methods, Lab 4

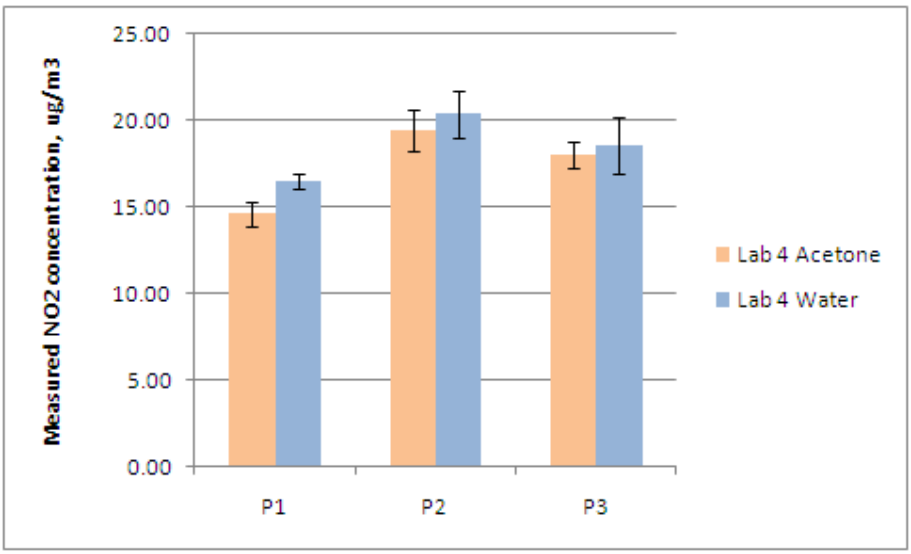
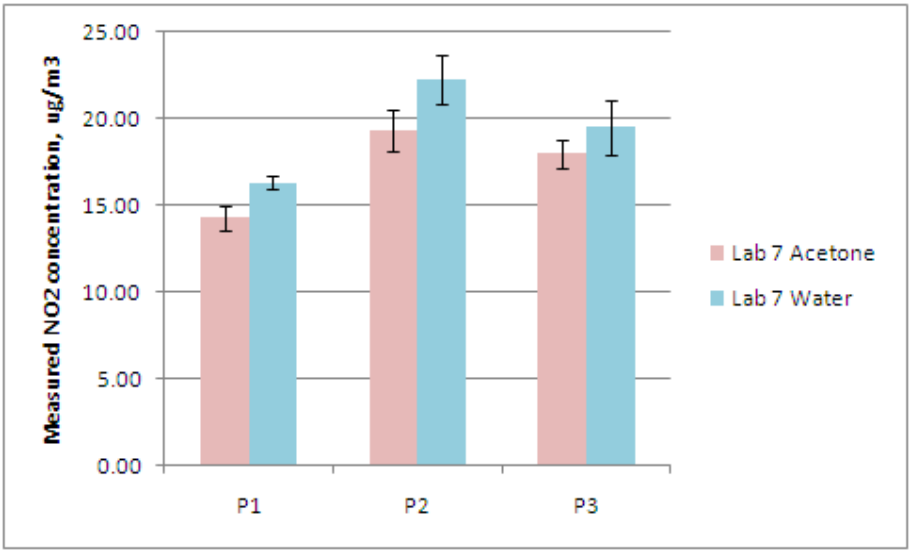


Figure 3-12 Comparison of Results, Acetone and Water Methods, Lab 7



There were three instances where a laboratory appeared to obtain significantly different results (on this basis) using 50% TEA/acetone tubes versus 20% TEA/water tubes. These were: period 1 for both Lab 4 and Lab 7, and period 2 for Lab 7 only.

However, it is also important to consider the analytical uncertainty. Lab 4 quoted an analytical uncertainty of 5%, and Lab 7 quoted an analytical uncertainty of 10.9%. Even if the 95% confidence intervals of the mean results did not overlap, it would not be valid to conclude that there was genuinely a significant difference between the two sets of results, if the difference between the means was less than the analytical uncertainty quoted by the laboratory.

For Lab 4, in period 1, the difference between the mean obtained by the water tubes and the mean obtained by the acetone tubes was greater than the analytical uncertainty.

Lab 7 quoted a higher analytical uncertainty, and in both periods 1 and 2, the difference between the mean obtained by the water tubes and the mean obtained by the acetone tubes was less than the analytical uncertainty.

Therefore, although the 50% TEA/water method appeared to give consistently higher results than the 50% TEA/acetone method, when analytical uncertainty was taken into account the difference was only statistically significant in one case out of six.

4 Conclusions and Recommendations

The conclusions of this small-scale and short-term study are as follows:

1. The precision of diffusion tubes prepared using the 50% TEA in acetone (dipped grids) method does not appear to have improved between 2007 and 2009.

2. The precision of diffusion tubes prepared using the 20% TEA in water (pipetting) method *does* appear to have improved between 2007 and 2009. However, this is not necessarily attributable to the harmonisation process: information from the ongoing monthly field intercomparison in central London indicates that diffusion tube precision has generally improved on average over the three years 2006 - 2008.

3. The typical bias (relative to the automatic analyser) exhibited by tubes prepared using the 50% TEA/acetone method was higher (more positive) in the 2009 study than in 2007. Whereas in 2007 many tubes exhibited negative bias, in 2009 no tubes of this type underestimated. While there was a substantial overall positive bias (mean 21%), this is consistent with the expected behaviour of diffusion tubes in the field. By contrast, negative bias (as observed in many of the 2007 results) often results from inefficient extraction (something that the Working Group sought to eliminate).

4. The typical bias (relative to the automatic analyser) exhibited by tubes prepared using the 20% TEA/water method was also higher (more positive) in 2009. The mean bias was 31% and no tubes under-estimated. This contrasts with the 2007 study, in which negative bias was widespread. Again, this is more consistent with what is expected of diffusion tubes exposed in the field.

5. The range of bias exhibited in any one period by tubes prepared using the 20% TEA/water method was smaller in 2009 than in 2007 (mean 21% in 2009 compared with 32% in 2007). This indicates that inter-laboratory variation has improved, for tubes prepared by the water method. The same was not observed for tubes prepared by the 50% TEA/acetone method, for which the range of bias was similar in both years.

6. It is possible that harmonisation has been more beneficial in the case of the 20% TEA/water method, as there was previously more potential for variation in terms of pipetting procedures, volume of solution applied, application technique etc., whereas for the 50% TEA/acetone method the grids were coated in TEA by simply submerging them in the solution and then allowing them to dry.

7. Two of the participating laboratories prepared and analysed tubes by both the 50% TEA/acetone method and the 20% TEA/water method. Both these laboratories obtained slightly better precision using the acetone method.

8. Both the above two laboratories obtained consistently higher results with the 20% TEA/water method than with the 50% TEA/acetone method, indicating that the former may provide better absorption of NO_2. However, the differences were only statistically significant at the 95% confidence level in one instance out of six comparisons.

9. There is still not a strong case for recommending either one of the two preparation methods tested here, in preference to the other. While the 50% TEA/acetone method may provide slightly better precision, the 20% TEA/water method may have the advantage of better collection of NO_2 (although not significantly).

10. The analytical uncertainties reported by the participating laboratories varied considerably and were in many cases as large as the CV of the six replicate measurements. This is something that should be investigated further. Parallel trials involving field exposure together with laboratory-based tank exposure and/or artificially doped tubes may be useful.

5 Acknowledgements

AEA would like to thank the participating laboratories, who supplied their tubes free of charge, and the National Physical Laboratory (NPL) who allowed use of their Teddington monitoring site. N. A. Martin, E. J. Leming and S. D. Jarvis of NPL carried out the tube changes, and provided comments on the report. NPL's contribution to the project was funded by the Chemical and Biological Metrology Programme of the Department for Business, Innovation and Skills (BIS).

6 References

1. A Loader "Field Trial of Preparation Techniques for NO_2 Diffusion Tubes" AEA Report AEAT/ENV/R/2503 - Issue 1, prepared for Defra and the Devolved Administrations, November 2007.

2. Defra Working Group on Harmonisation of Diffusion Tubes *"Diffusion Tubes for Ambient NO2 Monitoring: Practical Guidance for Laboratories and Users".* Report to Defra and the Devolved Administrations, AEA report number AEA/ENV/R/2504, Feb 2008.

3. A Loader, J Targa "Three Year Multi-Laboratory Urban Field Intercomparison of NO_2 Diffusion Tubes" – poster presented at AAMG conference on "Measuring Air Pollutants by Diffusive Sampleing and Other Low Cost Monitoring Techniques", Krakow, Poland.15^{th} – 17^{th} September.

4. GM Clarke, D Cooke "A Basic Course in Statistics" - Edward Arnold, London, 1981. p245.

AEA group
551.11 Harwell
Didcot
Oxfordshire
OX11 0QJ

Tel: 0870 190 6518
Fax: 0870 190 6377