



Inter-comparability of analytical laboratories in quantifying polycyclic aromatic hydrocarbons collected from industrial emission sources

Adam Lilley¹ · Marc D. Coleman¹ · Sharon L. Goddard¹ · Ellie G. Mills¹ · Richard J. C. Brown¹ · Rod A. Robinson¹ · Martin J. Clack¹

Received: 3 August 2021 / Accepted: 6 February 2022 / Published online: 3 May 2022
© The Author(s) 2022

Abstract

We report an inter-laboratory comparison of analytical laboratories involved in the quantification of polycyclic aromatic hydrocarbons (PAHs) collected by sampling organisations from industrial stacks (e.g. waste incinerators). Four reference solutions were prepared containing nominally 10 ng/ml, 50 ng/ml, 200 ng/ml and 500 ng/ml of naphthalene, benzo[a]anthracene, chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]pyrene, indeno[1,2,3-cd]pyrene and dibenzo[a,h]anthracene prior to despatch to five analytical laboratories with quantification requested in accordance with ISO 11338-2. Across four of the laboratories (the 5th returned unusable data), significant deviations from the reference concentrations were found frequently in excess of the benchmarks of 37 %—from the validation data in ISO 11338-2—and 21 %—from the Environment Agency for England's Monitoring Certification Scheme. Also, much of the variance was systemic in nature indicating a possible issue with the quality of some of the stock solutions used by the laboratories for calibration. Whilst more proficiency testing would be welcomed to monitor and improve performance, this should be provided in addition to more support for analytical laboratories. A key mechanism of support is the standards themselves and there is a timely opportunity in that ISO/TC 146/SC 1 are due to revise ISO 11338. Possible improvements include full validation of high performance liquid chromatography and gas chromatography–mass spectrometry methods (to better understand what performance can reasonably be expected from laboratories), a requirement to correct results to individual laboratory PAH extraction efficiency, and a required uncertainty stipulated for the overall method (also aiding setting pass/fail criteria for proficiency testing).

Keywords ISO 11338 · Proficiency testing · Polycyclic aromatic hydrocarbons · Stack emissions · Industrial emissions directive

Introduction

Polycyclic aromatic hydrocarbons are organic hydrocarbon compounds that are composed of multiple aromatic (benzene) rings. These compounds are formed in combustion processes [1, 2], particularly from incomplete or uncontrolled combustion. All PAHs are classed as harmful to human health and the environment, and they degrade slowly in the environment and are considered as carcinogens or possible carcinogens [3]. There is evidence [4] to suggest biofuels and energy from waste (EFW) fuels, which are increasingly used in energy production, are major contributors to

PAH emissions and therefore (given toxicity and environmental persistence) accurate measurements of emissions from these sectors are becoming increasingly important. PAHs were included in the European Union's old Waste Incineration Directive [5], but then disappeared when this was superseded by the current Industrial Emissions Directive (IED [6]). There is mention in some Best Available Techniques Reference documents, but only in those concerning emissions into water, not air. Hence, legislative drivers are on national rather than European level, for example, the Environment Agency for England require PAH monitoring from most municipal waste incineration plants.

PAHs in stack emissions in Europe are monitored according to the international standards ISO 11338 (parts 1-2):2003 'Stationary source emissions—Determination of gas and particle-phase polycyclic aromatic hydrocarbons' [7, 8]. Sampling for PAHs in stack is carried out iso-kinetically

✉ Marc D. Coleman
marc.coleman@npl.co.uk

¹ National Physical Laboratory, Hampton Road,
Teddington TW11 0LW, UK

at multiple points across the diameter, because at typical stack temperatures ($<200\text{ }^{\circ}\text{C}$), the compounds can exist as solid particles or associated with other solid particulates and therefore might not be uniformly distributed in the gas stream. Three methods of sampling are described by ISO 11338: dilution, heated filter and condenser, cooled probe. All of these methods extract the sample, cool it rapidly and pass the sample through either a solid adsorber or a plane filter and solid adsorber. In the UK, sampling is usually carried out using the heated filter/condenser/adsorber method, although other sampling methods are used in other parts of Europe. The solid adsorber used can be either polyurethane foam (PUF) or more commonly proprietary adsorbents such as XAD®-2 or Porapak™ PS. With all these methods, the upstream parts of the apparatus are washed with solvents (acetone, hexane and then toluene) and then these washings, the solid adsorber and filter (if used) are all protected from light and transferred to the analytical laboratory in a cooled ($<7\text{ }^{\circ}\text{C}$) sealed container. The samples are then extracted using a suitable organic solvent (e.g. hexane) using a Soxhlet extractor (or other validated method of accelerated solvent extraction) and the solvent solution is then analysed by gas chromatography–mass spectrometry (GC–MS) or high performance liquid chromatography (HPLC).

Whilst there is a relative dearth of literature with respect to the measurement of PAHs from stacks and flues, there are several reports discussing the influence upon emissions of different process types and abatement technologies. Modern coal-fired power stations (large combustion plants) employ a series of abatement technologies to remove pollutants including PAHs: flue gas desulphurisation (FGD), electrostatic precipitators (ESP), selective catalytic reduction (SCR) [9]. Studies (carrying out sampling and analysis in accordance with the California Air Resources Board Method 429) of these technologies have shown that whilst removing other PAHs, SCR can actually generate 3- and 4-ring PAHs and ESP 5-ring PAHs [4]. These are important observations as showing not only the change in total emissions but any influences on the amount fraction distributions are important: since it would clearly be undesirable to decrease a total emission in absolute concentration terms yet actually increase the toxicity.

With respect to municipal waste incinerators (MWIs), the technologies of activated carbon injection + bag filter and catalytic filter [10–13] have been compared. The latter showed greater removal efficiency of PAHs from the stack than the former and furthermore resulted in significantly less secondary pollution. This being due to the catalytic filter destroying gas-phase PAH whereas the activated carbon injection + bag filter converted gas-phase PAH to solid-phase, i.e. in essence the pollution fundamentally remained [4]. In contrast to the technologies mentioned above for coal-fired combustion, no significant changes in amount fraction

distributions were found for either of the MWI technologies. In terms of comparing large combustion plant to MWI—post the above described abatements—on a per tonne of fuel burnt basis, the emissions for MWIs are approximately five-fold greater: although it should be noted that MWI fuel can vary significantly, so such a ratio is far from fixed.

Both for legislative compliance and for performance testing of abatement technologies confidence in PAH measurements is key. To this end, we report results of a study where standard solutions of PAHs in hexane were despatched to five arbitrarily chosen European analytical laboratories routinely involved in measuring PAHs in samples collected from stacks. Providing samples in hexane solute enabled the quantitative step of ISO 11338-2 to be isolated eliminating any influence due to sampling and analytical extraction. The deviations between the laboratories are discussed to begin to deconvolve the contributions of measurement variability to the overall standard method, thereby allowing better understanding of the overall monitoring capability of ISO 13388.

Experimental

Four sample solutions were prepared comprising of eight PAHs (Table 1) of nominal concentrations 10 ng/ml, 50 ng/ml, 200 ng/ml and 500 ng/ml in hexane. Samples were prepared volumetrically in accordance with the instructions for calibration solutions in EN 15549 [14] and internal procedures for the dilution of certified reference materials under National Physical Laboratory's (NPL's) ISO/IEC 17025 [15] accreditation. Furthermore, staff used to carry out the sample preparations held the necessary competencies, awarded via successful assessment in NPL internal audits. Samples were prepared volumetrically from dilution of a certified standard stock solution (Chemical Products for Analysis, France). Calibrated gas-tight syringes were used to transfer aliquots of the stock solution and sub-dilutions into amber volumetric flasks (Grade A). The flasks were then filled to the mark with hexane (HPLC grade, Fisher Scientific, UK).

Table 1 PAH compounds (and their associated abbreviations) used in the inter-laboratory comparison exercise

PAH	Abbreviation
Naphthalene	Nap
Benzo[a]anthracene	B[a]A
Chrysene	Chr
Benzo[b]fluoranthene	B[b]F
Benzo[k]fluoranthene	B[k]F
Benzo[a]pyrene	B[a]P
Indeno[1,2,3-cd]pyrene	Ind
Dibenzo[a,h]anthracene	D[a,h]A

Each inter-comparison sample solution was prepared in a single amber volumetric flask (20 mL). Sub-samples for each laboratory were dispensed from the original flask by Pasteur pipette into amber GC vials. The sealed vials were then stored in a fridge before distribution to the analytical laboratories by refrigerated courier. The contracted laboratories were requested to refrigerate all samples prior to quantification, carry out quantification in accordance with ISO 11338-2, and report results in units of ng/mL. Whilst all the analytical laboratories had procedures for the implementation of ISO 11338-2, not all held ISO/IEC 17025 accreditation for said implementation. However, as well recognised and experienced laboratories, they nonetheless represented suitable participants from which to gain an understanding of the performance possible in following the method laid out in ISO 11338-2. In addition, one laboratory was sent two sets of samples (Laboratory A) and was requested to analyse this second set 3 months after the first. The aim of this was to gain some measure of laboratory internal reproducibility, since at the very least, the instrument would be calibrated with a new (i.e. independent) set of calibration standards, amongst other possible temporal variables. Laboratories were aware that received samples were part of a comparison exercise, but not of the concentrations in the samples or of whom the other participants were.

Results and discussion

The laboratories chosen to analyse the samples all perform regular PAH testing on stack, water or soil samples so this analysis would be part of their routine work. Laboratories A, B, and D used GC–MS and laboratory C used HPLC. Both analytical methods are permitted and considered as equivalent by the standard. The results along with uncertainties estimated by each laboratory are shown in Table 2. Note that, the data returned from the 5th laboratory could not be used in this study as for every PAH in every solution they reported that the detected concentration was below the limit of detection.

As can be seen, the results show considerable deviation between laboratories and also within laboratory. Generally, laboratories C and D demonstrate closer agreement with the reference values than laboratories A and B. Also, laboratories C and D show both positive and negative deviations from the reference values, whereas for some species, laboratories A and B show systematic negative deviations, these being particularly large in some instances for the latter. The duplicate samples sent to laboratory A show that reported concentrations are consistently and markedly higher on the second analysis than the first.

To give some context to the inter-laboratory comparison (ILC), it is important to consider both typical abundance and

toxicity of the PAHs emitted from stacks. From an examination of historical stack emissions data collected by the NPL Emissions Team, the most abundant PAH found is naphthalene (Table 3). The toxic equivalency factors (TEFs) for PAHs in air (Nisbett & La Goy 1992 [16], there are other factors for PAHs in water and soil) give an indication of the potential to cause harm relative to benzo(a)pyrene. From the TEFs, it is seen that there is a relatively broad range of toxicity, for example, a 1 mg dose of benzo(a)pyrene is potentially as harmful as a 1000 mg dose of naphthalene. Hence, to take account of this, TEF equivalent concentrations are calculated from the product of the measured concentration in the stack gas and the applicable TEF. As naphthalene is the most abundant measurand (from both an absolute and also sometimes from an TEF equivalent concentration perspective), it might be expected that the laboratories would optimise the set-up of their instrumentation for naphthalene. Although individual laboratory's approach to instrument set-up is unknown, the results do appear consistent with this premise with all the laboratories producing results for naphthalene closer to the associated reference values than for other species. Alternatively, or in addition, the better performance for naphthalene could be due to it eluting from the column first giving generally a clean, sharp peak and being less at risk of instrumental drift than species with longer retention times. Also, it is not untypical for calibration solutions to contain impurities, particularly of the heavier PAHs, hence giving a noisier baseline at longer retention times potentially impacting the quantification of the larger PAHs.

In terms of what performance might be expected, ISO 11338-2 does not stipulate an uncertainty requirement (in contrast to other standards in the emissions area), so as PAHs are not included in the Industrial Emissions Directive there is no general requirement for uncertainty. There is, however, in Annex E of ISO 11338-2, a summary of the performance characteristics of the HPLC method, which gives a standard deviation of reproducibility of 6.9 %–37 % (no such data are provided for GC–MS). Hence, in the absence of a specific requirement, 37 % is an appropriate benchmark that can be considered in order to give some context to the comparability of the data. On a more localised basis, there is then the Environment Agency for England who stipulate national requirements of ≤ 15 % precision and ≤ 15 % bias (both standard uncertainty requirements ($k = 1$, 68 % confidence)) under their MCERTS (Monitoring Certification Scheme) standard 'Performance standard for laboratories carrying out testing of samples from stack emissions monitoring' [17]. Whilst such requirements only apply in England (although in practice most of the UK complies), such values nonetheless provide an indication of the level of performance considered by at least one local competent authority to be necessary for quantifying captured PAH emissions. Combining these values as a simple root sum of squares gives 21 %. From Fig. 1,

Table 2 Reference concentrations with concentrations reported by anonymised laboratories with associated uncertainties expressed at a 95 % level of confidence ($k=2$)

	Reference		Lab A	Lab A (+ 3 months)	Lab B	Lab C		Lab D	
	ng/ml	U/ng/ml	ng/ml	ng/ml	ng/ml	ng/ml	U/(ng/ml)	ng/ml	U/(ng/ml)
Nap	10.4	0.5	10	12.1	10	10.05	3.02	8.9	2.8
	51.1	1.4	48	51	50	46.85	14.06	53.5	3
	205.2	5.33	193	205	210	199	29.85	218.5	5
	514.5	13.3	518	583	520	532.5	79.88	552	18
Chr	10.3	0.45	8.6	11.1	8.6	10.6	3.18	10.3	1.7
	50.9	1	39	47	39	51.3	15.39	53.6	2
	204.5	3.6	173	205	173	207.5	31.13	214.5	6.3
	512.9	9	470	482	470	549	82.35	546	17
B[a]A	10.3	0.46	7.8	11.4	9	10.05	3.02	10.6	2
	50.8	1.1	38	46	39	46.85	14.06	54.2	2.4
	204	4.04	173	200	160	199	29.85	215.9	8.3
	511.5	10.1	468	472	440	532.5	79.88	545	23
B[b]F	10.3	0.52	8.6	11.5	4	10.25	3.08	10.6	1.9
	50.8	1.62	41	45	5	50.4	15.12	53.7	2
	204	6.25	177	184	27	202	30.30	214.1	6.2
	511.6	15.6	467	479	75	537.5	80.63	544	16
B[k]F	10.4	0.48	11	13.7	6	9.805	2.94	10.6	2.4
	51.1	1.28	48	52	10	47.3	14.19	55.4	3.1
	205.3	4.79	190	199	42	190.5	28.58	221	14
	514.9	11.9	490	501	110	505.5	75.83	572	30
B[a]P	10.3	0.57	8	11.7	1	10.5	3.15	10.34	0.8
	50.8	1.97	38	46	7	50.5	15.15	53.6	1.1
	204.2	7.69	161	185	31	203	30.45	212.7	6.4
	511.9	19.2	438	460	90	539	80.85	543	20
Ind	10.3	0.46	6.3	9.3	2	9.41	2.82	10.3	1.7
	50.7	1.09	31	41	4	45.55	13.67	52.5	1.9
	203.6	3.97	148	190	25	182.5	27.38	206.7	7.3
	501.5	9.9	416	433	79	491.5	73.73	520	22
D[a,h]A	10.3	0.49	5.3	8.5	8	9.285	2.79	10.5	2.6
	50.8	1.36	28	39	48	44.5	13.35	51.4	2.7
	203.9	5.16	142	197	170	181	27.15	205	6.9
	511.4	12.9	403	426	480	479	71.85	520	20

Uncertainties reported as 22% of value for Laboratory A (values not shown), 20% of value for Laboratory B (values not shown)

it is seen that for naphthalene, all the observed deviations are within the benchmarks of 37 % and 21 %. Also, there is no clear evidence of systematic bias in the results from any of the different laboratories—for naphthalene, the repeatability of each laboratory is larger than any biases between them (although both components are still relatively small). As naphthalene is the most commonly found PAH and is usually found in the highest concentrations it could be said to be most significant determinant. However, as mentioned above, there is a large variation in the toxicity of individual PAH's. Hence, the TEF concentrations shown in Table 3 are possibly the more important metric. Taking the mean PAH emissions across eight waste incinerators in the UK

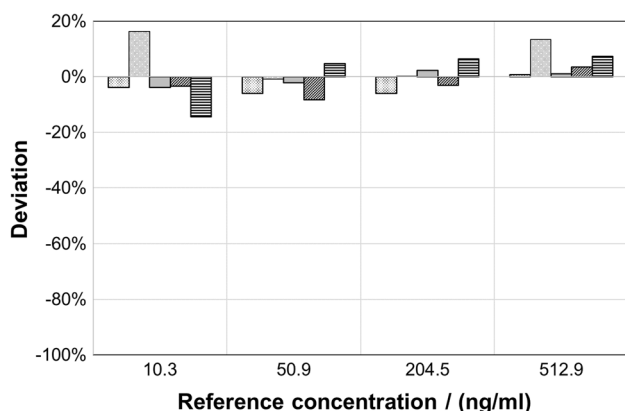
and applying associated TEFs gives an industry representative abundance, which shows that TEF emissions of benzo[a]pyrene, benzo[b]fluoranthene and benzo[k]fluoranthene (albeit with the latter based only on detection limit data) are threefold more abundant than naphthalene. Therefore, from a toxicity perspective, it is arguably of greater importance to examine the performance for these other compounds.

The results for benzo[a]pyrene are much more varied as can be seen in Fig. 2. Whilst two of the laboratories did produce results well within the ISO 11338-2 and MCERTS benchmarks, there is however evidence of a systematic bias present between laboratories and across the concentration range. Here, it is clear that the biases are much larger than

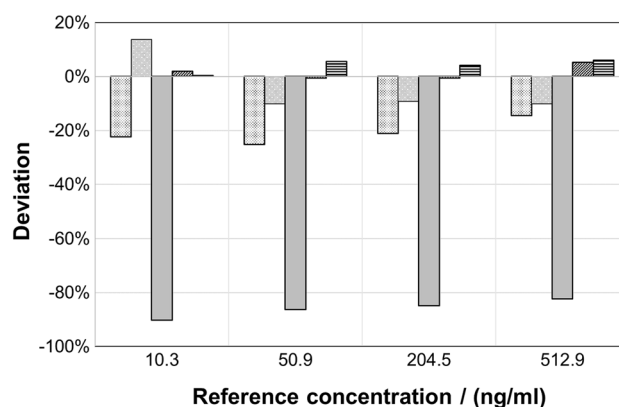
Table 3 Mean emitted concentrations and associated toxic equivalency factor concentrations across eight UK incineration plants acquired in the course of routine monitoring work by NPL's ISO/IEC 17025 accredited Emissions Team

Species	TEF	Mean 8 tests		
		N/D	Concentration in stack gas / ng/m ³	TEF conc. / ng/m ³
Napthalene	0.001		5762	5.8
Acenaphthylene	0.001		63	0.1
Acenaphthene	0.001		233	0.2
Fluorene	0.001		441	0.4
Phenanthrene	0.001		754	0.8
Anthracene	0.01		129	1.3
Fluoroanthene			148	
Pyrene	0.001		310	0.3
Benzo(a)anthracene			23	
Chrysene			17	
Benzo(b)fluoranthene	0.8		19	15.0
Benzo(k)fluoranthene	0.8	*	14	11.2
Benzo(a)pyrene	1		17	17.1
Indeno(1,2,3-cd)pyrene		*	14	
Dibenzo(a,h)anthracene		*	13	
Benzo(g,h,i)perylene		*	13	

Non detected compounds (N/D) are reported at the method detection limit. Toxic Equivalency Factors (TEFs) referenced against benzo[a]pyrene taken from [16]. Rows in bold indicate compounds included in the inter-laboratory comparison reported herein

**Fig. 1** Relative deviations between reported and reference values for naphthalene measurement by anonymised laboratories A (□), A analysing a second set of samples 3 months after the first (▨), B (▩), C (■) and D (▤)

the within-laboratory repeatability. The results are similar for benzo[a]anthracene, benzo[b]fluoranthene, benzo[k]fluoranthene and indeno[1,2,3-cd]pyrene (Table 4). However, dibenzo[a,h]anthracene gives a different spread of results as shown in Fig. 3, but still demonstrating systematic bias present between laboratories and across the concentration range. In this case, it is laboratory A that produces results outside the benchmark of 21 % for the two higher concentrations and beyond the benchmark of 37 % for the two lower concentrations. Whilst all laboratories except

**Fig. 2** Relative deviation between reported and reference values for benzo[a]pyrene measurement by anonymised laboratories A (□), A analysing a second set of samples 3 months after the first (▨), B (▩), C (■) and D (▤)

laboratory D evidence a negative systematic bias. Results for Chrysene (Table 4) were similar to naphthalene, although with larger deviations from the reference values. A calibration issue of some sort is one way that such systemic biases could be observed given that the biases correlate both at different concentration levels and also across different PAH species. Although, in order to determine this with confidence the laboratories would need to carry out internal investigations. In any case, regardless of cause, it is clear that the

Table 4 Relative deviations from reference summarised by laboratory, by species for each test concentration level and by species overall

	Reference / ng/ml	Lab A deviation / %	Lab B deviation / %	Lab C deviation / %	Lab D deviation / %	Mean species deviation by test concentration / %	Mean species deviation / %
Nap	10.4	−4	−4	−3	−14	6	5
	51.1	−6	−2	−8	5	5	
	205.2	−6	2	−3	6	4	
	514.5	1	1	3	7	3	
Chr	10.3	−17	−3	3	0	6	7
	50.9	−23	8	1	5	9	
	204.5	−15	3	1	5	6	
	512.9	−8	3	7	6	6	
B[a]A	10.3	−24	−24	−13	−2	16	15
	50.8	−25	−25	−23	−8	20	
	204	−15	−15	−22	−2	14	
	511.5	−9	−9	−14	4	9	
B[b]F	10.3	−17	−61	0	3	20	26
	50.8	−19	−90	−1	6	29	
	204	−13	−87	−1	5	26	
	511.6	−9	−85	5	6	26	
B[k]F	10.4	6	−42	−6	2	14	22
	51.1	−6	−80	−7	8	26	
	205.3	−7	−80	−7	8	25	
	514.9	−5	−79	−2	11	24	
B[a]P	10.3	−22	−90	2	0	29	28
	50.8	−25	−86	−1	6	29	
	204.2	−21	−85	−1	4	28	
	511.9	−14	−82	5	6	27	
Ind	10.3	−39	−81	−9	0	32	32
	50.7	−39	−92	−10	4	36	
	203.6	−27	−88	−10	2	32	
	501.5	−17	−84	−2	4	27	
D[a,h]A	10.3	−49	−22	−10	2	21	15
	50.8	−45	−6	−12	1	16	
	203.9	−30	−17	−11	1	15	
	511.4	−21	−6	−6	2	9	
Mean laboratory deviation / %		18	45	7	5	19	

For all mean calculations absolute values of individual deviations are used

biases are of a level in many of the measurements in excess of the requirements of at least one local competent authority.

In order to obtain some indication of reproducibility, laboratory A analysed two sets of samples 3 months apart. It is seen (Table 2) that there is significant deviation between the results, with the second set of results showing consistently higher concentrations. Any sample degradation over time would have produced lower concentrations. Had a seal on a vial been broken some of the hexane could have evaporated over time making the solution more concentrated and producing higher values. However, this would need to have occurred on not just one but all four vials despatched to the laboratory and this seems unlikely. It is also pertinent to note

that as part of NPL's ISO/IEC 17025 accreditation for the implementation of EN 15549 (PAHs in particulate matter in ambient air) that it has been successfully demonstrated that PAH extracts are stable for well in excess of 3 months. Hence, there is confidence that the observed deviation is indeed a measure of laboratory reproducibility. However, as data are only available for one laboratory from this study, it isn't possible to say if this magnitude of within-laboratory reproducibility is typical, but what is clear is that temporal performance of individual laboratories is an area where further study is needed.

The variability of the results observed raises questions about the level of data comparability that ISO 13388 can

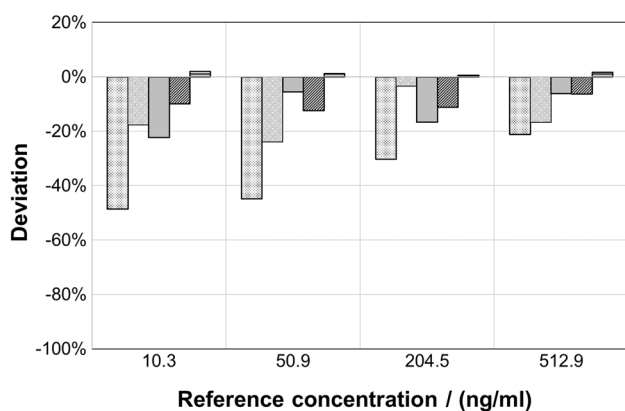


Fig. 3 Relative deviation between reported and reference values for dibenzo[a,h]anthracene measurement by anonymised laboratories A (□), A analysing a second set of samples 3 months after the first (▤), B (▤), C (▤) and D (▤)

provide and also concerns that the success or failure of demonstrating compliance with the emission limit value could in some cases come down to which analytical laboratory has been selected. It is noted that this study has involved four analytical laboratories (as the data returned from the 5th was unusable), which is not a large number. However, this is more representative than might initially appear given there are far fewer analytical laboratories providing PAH analysis services than for other species emitted from industrial processes, such as SO_2 . For example, if samples had been sent to one UK laboratory, this would, in effect, have encompassed the entire UK capability as there is currently only one UK analytical laboratory ISO/IEC 17025 accredited for the implementation of the analytical element of ISO 11338-2. Also, all the laboratories included in this study provide analytical services to multiple sampling teams so in effect what is seen here is the analytical variance that many sampling teams, and therefore emission measurements, are subject to. Therefore, due to the number of sampling teams linked to these laboratories, a reasonable portion of the emissions community is sampled, and this is a reasonable basis on which to assess if ISO 11338-2 is providing sufficient control of the analytical method.

This study (by design) isolates the quantitative step in ISO 11338-2 and tests the comparability of PAH in hexane solution. Hence, the variance of the entire measurement method will be significantly greater than that discussed so far due to the other key elements of the method: pumping gas sample from the stack, solvent washing apparatus upstream of the filter for PAH, analytical laboratory extraction of trapped PAH from the filter and solid adsorbent. The first of these two elements are variables attributable to sampling teams rather than analytical laboratories and hence beyond the scope of this paper, which is focussed upon the analytical elements of PAH monitoring. However, the latter is clearly a

variable associated with the analytical laboratory and so it is important to consider what impact this may have.

As outlined in the Introduction, PAH is extracted by analytical laboratories from the filter and solid adsorbent using a suitable solvent (e.g. hexane) and a Soxhlet extractor (or other validated method) for GC-MS or HPLC quantification. The efficiency of this extraction is tested by spiking samples with small amounts of determinands carrying a deuterium marker. ISO 11338-2 requires the recovery to be between 50 % and 150 %, which is a broad tolerance and reflects that such extractions are challenging. Table 5 shows the extraction efficiencies for two laboratories analysing two different samples and it is seen that the efficiencies vary significantly, and indeed in the case of the latter with a mean extraction efficiency of 36%, it fails to comply with ISO 11338-2. In the analogous situation of measuring PAHs in ambient air EN 15549, ‘Air quality —Standard method for the measurement of the concentration of benzo[a]pyrene in ambient air’ [14] requires results to be corrected for extraction efficiency (for GC-MS-based analysis), which should result in a decrease in variance between laboratories. However, it is significant that ISO 11338-2 neither requires laboratories to correct results in this way nor take account of such efficiencies in the uncertainty budget. Consequently, this potentially leaves results generated under ISO 11338-2 subject to the full extent of this variance. Hence, whilst the data comprising Tables 4 and 5 are not directly linked, it is clear that the effect of extraction efficiency can only serve to increase variance and reduce the frequency with which laboratories meet the ISO 11338-2 and MCERTS benchmarks of 37 % and 21 %. If ISO 11338-2 were to require correction of results for extraction efficiency, this would serve to decrease variance between laboratories and have the added benefit of harmonising the approaches taken in the two standards: important since both data sources are used together in modelling.

It is noteworthy that at its recent plenary meeting in September 2020, ISO/TC 146/SC 1 passed a resolution [18] for the revision of ISO 11338-1 and ISO 11338-2. Hence, this revision would seem a timely opportunity to revise the normative text of ISO 11338-2 to require correction of data to extraction efficiency and consider further alterations to the method with the aim of facilitating an improvement in analytical laboratory comparability.

Conclusions

Through an inter-laboratory comparison, it has been shown that there is significant variation in PAH quantification by analytical laboratories following the method laid out in ISO 11338-2. Frequently, this variation was in excess of the benchmarks of 37 %—standard deviation of reproducibility from the validation data of HPLC

Table 5 Relative recoveries of PAHs reported by two ISO/IEC 17025 accredited analytical laboratories analysing two different samples

Species	Sample 1				Sample 2			
	N/D	Recovery / %	Mass in sample / ng	Mass in sample if corrected for recovery / ng	N/D	Recovery / %	Mass in sample / ng	Mass in sample if corrected for recovery / ng
Napthalene		117	1493	1276		23	6500	28,261
Acenaphthylene		102	35	34		29	100	345
Acenaphthene			13			58	400	690
Fluorene			24			39	300	769
Phenanthrene		85	452	532		44	830	1886
Anthracene	*		5		*	35	30	30
Fluoroanthene		65	40	62		38	200	526
Pyrene			83			37	320	865
Benzo(a)anthracene	*	74	5	5	*	38	40	40
Chrysene	*	83	5	5	*	36	30	30
Benzo(b)fluoranthene	*	109	5	5	*	35	30	30
Benzo(k)fluoranthene	*	100	5	5	*	34	30	30
Benzo(a)pyrene	*	65	5	5	*	29	30	30
Indeno(1,2,3-cd)pyrene	*	58	5	5	*	32	30	30
Dibenzo(a,h)anthracene	*	69	5	5	*	27	30	30
Benzo(g,h,i)perylene	*	54	5	5	*	35	30	30

Example of impact on reported concentrations if results corrected to relative recovery. * non-detected (N/D) species reported by laboratories at respective detective limits. Rows in bold indicate compounds included in the inter-laboratory comparison reported herein

under ISO 11338-2 - and 21 %—from the Environment Agency for England's MCERTS scheme. Moreover of the PAHs studied whilst the best performance was found for naphthalene, once toxic equivalency concentrations were calculated (product of the measured concentration and the associated toxic equivalency factor), it was found that there was significantly poorer performance for several PAHs with a threefold higher toxic equivalency concentration than naphthalene, i.e. the performance was poorer for the PAHs with the greater environmental impact. Given that many of the deviations observed correlated across laboratories, species, and concentration test levels, it was clear that the biases were systemic in nature. Without laboratory internal investigation, it was not possible to pinpoint the issues, but certainly calibration and/or quality of stock solution (i.e. is the supplier ISO/IEC 17025 accredited? Can SI traceability be demonstrated?) seemed likely sources. The current level of variance is of concern as process site operator compliance with site permits could come down to which analytical laboratory the sampling organisation despatch the samples to for analysis. Clearly this is an unsatisfactory situation to site operators, local competent authorities and the health of society, which is of course what the system is put in place to protect. Moreover, when the remaining key elements of the measurement method are considered—namely, extraction efficiency and sample collection from

the stack—these can only serve to increase the variance seen here raising further concern. Certainly, more proficiency testing of, analysis (similar to what has been described here), PAH extraction + analysis and, stack sampling + PAH extraction + analysis, would be welcome to monitor and encourage the best performance possible under ISO 11338. However, analytical laboratories also need supporting and with ISO/TC 146/SC 1 passing a resolution in September 2020 to revise ISO 11338-1 and ISO 11338-2 this presents a timely opportunity to update these standards. Elements, for example, that could be improved include full validation of GC–MS and HPLC methods (to better understand what performance can reasonably be expected from analytical laboratories), a requirement to correct results to extraction efficiency as is done in the ambient air sector, and a required uncertainty stipulated for the implementation of the method (which would also aid in setting pass/fail criteria for proficiency testing).

Acknowledgements We gratefully acknowledge funding from the EMPIR programme co-financed by the Participating States and from the European Union's Horizon 2020 research and innovation programme, and from the UK's Department for Business, Energy and Industrial Strategy National Measurement System under the Energy and Environment Programme.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Gachanja AN, Maritim PK (2019) Polycyclic aromatic hydrocarbons determination. In: Worsfold P, Poole C, Townshend A, Miró M (eds) Encyclopedia of analytical science, 3rd edn. Academic Press, Cambridge, Massachusetts, pp 328–340. <https://doi.org/10.1016/B978-0-12-409547-2.14107-1>
- Moldoveanu SC (2019) Pyrolysis of Hydrocarbons. In: Moldoveanu SC (ed) Pyrolysis of Organic Molecules, 2nd edn. Elsevier, Amsterdam, Netherlands, pp 35–161. <https://doi.org/10.1016/B978-0-444-64000-0.00002-0>
- IARC working group on the evaluation of carcinogenic risks to humans, Somenon-heterocyclic polycyclic aromatic hydrocarbons and some related exposures IARC Monographs on the Evaluation of Carcinogenic Risks to Humans/World Health Organization (2010) 92 International Agency for Research on Cancer
- Hsu WT, Liu MC, Hung PC, Chang SH, Chang MB (2016) PAH emissions from coal combustion and waste incineration. *J Hazard Mater* 318:32–40
- Directive 2000/76/EC of the European Parliament and of the Council of 4 December 2000 on the Incineration of Waste. *OJ C* L332: 91–111
- Directive 2010/75/EU of the European Parliament and of the Council of 24 November 2010 on industrial emissions (integrated pollution prevention and control). *OJ EU* L334: 17–119
- ISO 11338-1:2003 Stationary source emissions — Determination of gas and particle-phase polycyclic aromatic hydrocarbons — Part 1: Sampling. International Organization for Standardization, Geneva
- ISO 11338-2:2003 Stationary source emissions — Determination of gas and particle-phase polycyclic aromatic hydrocarbons — Part 2: Sample preparation, clean-up and determination. International Organization for Standardization, Geneva
- Wang R, Zhang J, Liu J, Guijian L (2013) Levels and patterns of polycyclic aromatic hydrocarbons in coal-fired power plant bottom ash and fly ash from Huainan, China. *Arch Environ Contam Toxicol* 65:193–202
- Weber R, Plinke M, Xu Z, Wilken M (2001) Destruction efficiency of catalytic filters for polychlorinated dibenzo-*p*-dioxin and dibenzofurans in laboratory test and field operation – insight into destruction and adsorption behaviour of semivolatile compounds. *Appl Catal B Environ* 31:195–207
- Šyc M, Pekárek V, Fišerová E, Punčochář M, Karban J, Prokeš O (2006) Catalytic filter application in the Termizo municipal waste incineration plant in Liberec. *Organ Compd* 68:1232–1235
- Tirler W, Mair K, Donegà M, Voto G, De Carli A (2010) PCDD/F removal with catalytic filters in a municipal solid waste incineration. *Organ Compd* 72:1561–1564
- Fritsky KJ, Kumm JH, Wilken M (2001) Combined PCDD/F destruction and particulate control in a baghouse: experience with a catalytic filter system at a medical waste incineration plant. *J Air Waste Manage Assoc* 51:1642–1649
- EN 15549:2008 Air quality—Standard method for the measurement of the concentration of benzo(a)pyrene in ambient air. Comité Européen de Normalisation, Brussels
- ISO/IEC 17025:2017 General requirements for the competence of testing and calibration laboratories. International Organization for Standardization, Geneva
- Nisbet CT, LaGoy PK (1992) Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs). *Regul Toxicol Pharmacol* 16:290–300. [https://doi.org/10.1016/0273-2300\(92\)90009-X](https://doi.org/10.1016/0273-2300(92)90009-X)
- Performance standard for laboratories carrying out testing of samples from stack emissions monitoring. Environment Agency, version 3, December 2018. www.gov.co.uk
- Resolution 3, ISO/TC 146/SC 1/ N 1379, Resolutions of the 40th meeting of ISO/TC 146/SC 1 ‘Stationary source emissions’, September 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.