PAPER

Check for updates

Standardisation of a European measurement method for organic carbon and elemental carbon in ambient air: results of the field trial campaign and the determination of a measurement uncertainty and working range†

Richard J. C. Brown, [ID] *[a] Sonya Beccaceci,[a] David M. Butterfield,[a] Paul G. Quincey,[a] Peter M. Harris,[a] Thomas Maggos,[b] Pavlos Panteliadis,[c] Astrid John,[d] Aleksandra Jedynska,[e] Thomas A. J. Kuhlbusch,[fg] Jean-Philippe Putaud[h] and Angeliki Karanasiou[i]

The European Committee for Standardisation (CEN) Technical Committee 264 'Air Quality' has recently produced a standard method for the measurements of organic carbon and elemental carbon in $PM_{2.5}$ within its working group 35 in response to the requirements of European Directive 2008/50/EC. It is expected that this method will be used in future by all Member States making measurements of the carbonaceous content of $PM_{2.5}$. This paper details the results of a laboratory and field measurement campaign and the statistical analysis performed to validate the standard method, assess its uncertainty and define its working range to provide clarity and confidence in the underpinning science for future users of the method. The statistical analysis showed that the expanded combined uncertainty for transmittance protocol measurements of OC, EC and TC is expected to be below 25%, at the 95% level of confidence, above filter loadings of 2 μg cm^{-2}. An estimation of the detection limit of the method for total carbon was 2 μg cm^{-2}. As a result of the laboratory and field measurement campaign the EUSAAR2 transmittance measurement protocol was chosen as the basis of the standard method EN 16909:2017.

Environmental significance

Particulate matter (PM) in ambient air remains a huge health concern. Compositional information on PM is essential to judge pollutant increments in more polluted areas, assess the possible contribution from long-range transport of air pollutants, support source apportionment analysis and for the understanding of specific pollutants. As an important indicator of general air quality and as an emerging health-relevant metric, the concentration of organic and elemental carbon in PM is an important but challenging measurement. This study describes the development of a standardised method for this measurement that will greatly improve the accuracy, stability and comparability of these measurements across Europe. The improved data produced using this method, which must be used by all EU Member States, will provide a better evidence base for more effective regulation to improve air quality.

1. Introduction

As a consequence of the adverse impact on human health and on the environment of poor air quality, the European Union has been active in attempting to reduce the exposure of the population to several pollutants across all Member States. Directive 2008/50/EC, Annex IV, requires the measurement of elemental and organic carbon (EC/OC) and anions/cations in the $PM_{2.5}$

[a]*National Physical Laboratory, Teddington, TW11 0LW, UK. E-mail: richard.brown@npl.co.uk; Tel: +44 (0)20 8943 6409*

[b]*National Centre for Scientific Research "Demokritos", Institute of Nuclear & Radiological Sciences & Technology, Energy & Safety, Environmental Research Laboratory, 15310 Ag. Paraskevi, Athens, Greece*

[c]*Public Health Service (GGD) Amsterdam, Department of Air Quality, Nieuwe Achtergracht 100, 1018 WT Amsterdam, Netherlands*

[d]*Institute for Energy and Environmental Technology e.V. (IUTA), Air Quality & Aerosols, Bliersheimer Str. 58-60, 47229 Duisburg, Germany*

[e]*TNO Utrecht, Princetonlaan 6, 3584 CB Utrecht, Netherlands*

[f]*Federal Institute for Occupational Safety and Health (BAuA), Friedrich-Henkel-Weg 1-25, 44149 Dortmund, Germany*

[g]*CENIDE, University Duisburg-Essen, Carl-Benz-Str. 199, 47057 Duisburg, Germany*

[h]*European Commission, Joint Research Centre, Directorate for Energy, Transport and Climate, Air and Climate Unit, Via E. Fermi 2749, I-21027 Ispra (VA), Italy*

[i]*Institute of Environmental Assessment and Water Research, Spanish Research Council, IDAEA-CSIC, 08034 Barcelona, Spain*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7em00261k

This journal is © The Royal Society of Chemistry 2017

Environ. Sci.: Processes Impacts, 2017, **19**, 1249–1259 | 1249

fraction of ambient particulate matter in background locations.[1] These background measurements are essential to judge the enhanced levels in more polluted areas (such as urban background, industry related locations, traffic related locations), assess the possible contribution from long-range transport of air pollutants, support source apportionment analysis and for the understanding of the generation and composition of specific pollutants such as particulate matter. It is also essential for the increased use of modelling in urban areas. As such, it is essential that the methods used across Europe by Air Quality Reference Laboratories are fit for purpose. They also have to be harmonized to ensure that measurements are stable over time, allowing the assessment of trends, comparable between locations, so that differences between countries can be assessed, and are made with as low an uncertainty as possible. To meet these requirements, the European Commission issued Mandate M/503 "Ambient air quality" for the development of "standards concerning automated measurements of particle matter in ambient air and the measurement of its chemical composition (organic and elemental carbon, inorganic components)." Any European Standard developed is likely to be adopted as the reference method in Europe requiring its use by all Member States. Although the method must be applicable for background sites to satisfy Directive 2008/50/EC, there is also a strong need that the same sampling and analysis method should also be applicable for suburban and urban background sites as well as, if possible, for roadside sites, to ensure maximum comparability and utility of the methods across Europe.

The task of Working Group 35 (WG35) of CEN/TC 264 'Air Quality' of the European Committee for Standardisation (CEN) was to prepare a new European Standard on the measurement of airborne elemental carbon (EC) and organic carbon (OC) in $PM_{2.5}$ in accordance to Directive 2008/50/EC and Mandate M/503. No European standard existed for these measurements and furthermore no traceable certified reference materials are available for EC and OC analysis. In addition, as shall be discussed, the scientific distinction between EC and OC is challenging and arguably not possible in absolute terms. Different thermal–optical protocols are available for the measurement of carbon in ambient air and for the split between OC and EC. The most used protocols in Europe are IMPROVE_A,[2] EUSAAR_2 (ref. 3) and NIOSH. The NIOSH-like protocols (NIOSH-5040, NIOSH-840, NIOSH-850, and NIOSH-870) are all modified versions of the Birch and Cary[4] and Birch et al.[5] protocols.

Recently published work comparing measurement protocols[6–9] concluded that whilst the Total Carbon (TC, as sum of OC and EC) concentrations measured with different protocols are comparable, the split between OC and EC can be significantly different. The studies showed that the split between OC and EC is most sensitive to the exact method used whereas TC is relatively insensitive to the details of the protocol. For example, lower temperature protocols such as EUSAAR_2 and IMPROVE_A lead to higher EC concentrations than the NIOSH-like protocols. However, the different composition of the aerosol, as well as the optical correction (i.e. transmittance correction and reflectance correction) can also influence the comparability of the different protocols.[10,11] Where measurands are operationally defined, such as the split between OC and EC, and significant differences between different methods exist, agreement of a standard method with low sensitivity to input parameters, and high repeatability and reproducibility, becomes even more important. Ultimately, this means choosing one method based on the results of the laboratory and field trials described in this paper.

WG35, in collaboration with WG34 (determination of anions and cations in $PM_{2.5}$), conducted a laboratory and field trial campaign to validate the draft standard method and to set concentration ranges in which the method was validated and their corresponding uncertainties, although no target uncertainties are included in the Directive for the measurements of EC/OC and ions in $PM_{2.5}$. This paper summarises the results of the field trial using a novel statistical assessment of the results for field trials of this type. This output is used to calculate the uncertainty and working range of the standard method. It is hoped that this will provide clarity and confidence in the underpinning science for future users of the method in a similar way to publication of previous studies supporting the validation of standard methods produced within CEN/TC 264.[12,13]

2. Experimental

2.1 Laboratory and field trials

Prior to the field trial stage a laboratory comparison was performed to ensure that the agreed protocols to be tested in the field could be implemented satisfactorily at each analytical laboratory and to decide on an appropriate standardised calibration regime. The laboratory test was also a first step in performing an initial assessment of the likely comparability of the methods. It was important to ensure that the protocols produced sufficiently different results that it would be necessary to continue to use all three protocols during the field trial. However, the results of the laboratory tests were not used for the subsequent statistical analysis.

Field trials to validate proposed methods were performed by WG35 and took place around Europe during late 2013 and in 2014 at six locations with distinct and different pollution climates. The aim was to get as much information as possible on the performance of the draft standard method under different conditions. The sites were chosen to represent different ambient conditions, sources and concentration levels expected across Europe – see[14] for further details. The field trial was carried out in parallel with the field trial conducted by WG34. The site operators were either members of both groups or members of WG35 only. The details of the field trial campaigns can be found in Table 1.

Two Digitel DHA-80 high volume samplers (HVS) fitted with a $PM_{2.5}$ head and sampling at 30 m^3 h^{-1} were available for the duration of the trial. At the first site (Ispra, Italy) and at the last site (Cabauw, the Netherlands) the two samplers were run in parallel. As explained later, this was done in order to get an estimation of a between-sampler effect in the overall uncertainty. At the end of the sampling period at the first site, the two samplers were separated and sent to the other locations for

1250 | Environ. Sci.: Processes Impacts, 2017, **19**, 1249–1259

This journal is © The Royal Society of Chemistry 2017

Table 1 Details of the CEN/TC 264/WG35 field trial

	Italy	Spain	Germany	Germany	Netherlands	Netherlands
Country	Italy	Spain	Germany	Germany	Netherlands	Netherlands
Site operator	JRC-IES Institute for Environment and Sustainability, European Commission	IDAEA-CSIC, Institute of Environmental Assessment and Water Research	UBA, German Federal Environment Agency	IUTA, Institute of Energy and Environmental Technology	GGD, Public Health Service in Amsterdam, Department of Air Quality	TNO, Netherlands Organisation for Applied Scientific Research
Site location and details	Ispra, Rural background provisional monitoring station located on the JRC-IES site, situated in a semi-rural area (>20 km from large pollution sources)	Barcelona, Urban background site located 200 m from one of the main avenues of the city with approximately 120 000 vehicles per day	Waldhof, Rural background site located in the biggest north German forest and health environment with very little influence from local anthropogenic sources	Mülheim, Urban background site located in a residential area, but also within reach of a motorway (to the north), and industrial premises (to the east and south)	Amsterdam, Roadside site situated on one of the city's main street arteries and associated with approximately 15 000 vehicles per day	Cabauw, Rural background site situated in an agricultural area in the western part of The Netherlands, 44 km from the North Sea.
Site latitude; longitude	45° 49′ N; 8° 38′ E	41° 23′ N; 2° 07′ E	52° 48′ N; 10° 45′ E	51° 27′ N; 6° 52′ E	52° 22′ N; 4° 51′ E	51° 58′ N; 4° 55′ E
Sampling period	20 days, Oct 2013–Nov 2013	39 days, Feb 2014–Apr 2014	40 days, May 2014–Jul 2014	39 days, Feb 2014–Apr 2014	40 days, Jun 2014–Jul 2014	20 days, Sep 2014–Nov 2014

single sampler use and then co-located again at the last site of the field trial. Each sampler sampled for daily periods (from midnight to midnight) onto quartz filters (Pallflex® Tissuquartz™, 2500 QAT-UP, 150 mm). The sampler bodies were either placed outdoors or in air-conditioned cabins (maintained at 20–22 °C). At the beginning of the trial and after transport from one site to the next one, the samplers were leak-checked and the flowrate was checked and adjusted if necessary.

2.2 Sample analysis

After sampling, from each filter 4 punches were taken by the site operators using a 47 mm circular wad puncher and one punch sent to each of 4 different laboratories. The laboratories were GGD, TNO and IUTA (the same organisations as the site operators from Table 1) and in addition Demokritos: National Centre for Scientific Research in Greece. The punches were from radially symmetrical positions on the filter and therefore viewed to be equivalent – as demonstrated by previous studies.[15] This process was also performed for field blank filters and laboratory blank filters. Field blank filters are filters that are transported to the site as the other filters but are not sampled. The sub-samples obtained were analysed by each laboratory using three thermal–optical protocols with both transmittance and reflectance corrections: IMPROVE_A, EUSAAR_2 and NIOSH870. All thermal–optical analyses were carried out using the Sunset Thermal/Optical Carbon Analyzer (Sunset laboratory, USA), described in the literature by Birch and Cary (1996), using 1.5 cm^2 rectangular samples taken from the filter sub-samples. Table 2 describes the three protocols used for analysis. Standard sucrose solutions were used for instrument calibration and ongoing quality control. The sampling and the transport of the filters to and from the sites were carried out according to the European Standard EN 12341:2014 (ref. 16) for the measurements of PM_{10} and $PM_{2.5}$. Standard operating procedures were implemented by the working groups for the operation of the HVS and for the filter handling.

The analysis results from each laboratory were sent to the National Physical Laboratory (NPL, UK) for statistical analysis. These results were expressed as the mass of OC, EC and total carbon (TC – the sum of OC plus EC) per area of filter, $\mu g\ cm^{-2}$. NPL also calculated the ambient concentrations (in $\mu g\ m^{-3}$) measured during the field trial using the sample volumes provided by the site operators and the area of the sampled part of the filter. This helps to relate the performance characteristics of the method on a filter to ambient concentrations given a specific sampling volume.

2.3 Uncertainty estimation from the field trial process

The results from each sampled filter analysed by the 4 laboratories were statistically analysed to calculate the uncertainty in a single measured result. The uncertainty analysis was performed on the whole data set and by concentration range. The overall uncertainty includes inter-laboratory variability, between laboratory variability and between sampler variability. Laboratory blanks, which are filters taken from the filter box

Table 2 Temperature (T), time (s) and gas usage details for each sequential stage of the different thermal–optical protocols used during this field trial

Analysis stage, with gas used	EUSAAR_2 T/°C, t/s	IMPROVE_A T/°C, t/s	NIOSH870 T/°C, t/s
(1) He	200, 120	140, 150–580[a]	310, 80
(2) He	300, 150	280, 150–580[a]	475, 80
(3) He	450, 180	480, 150–580[a]	615, 80
(4) He	650, 180	580, 150–580[a]	870, 110
(5) He	No heating, 30	—	550, 45
(6) 2% O$_2$ in He	500, 120	580, 150–580[a]	550, 45
(7) 2% O$_2$ in He	550, 120	740, 150–580[a]	625, 45
(8) 2% O$_2$ in He	700, 70	840, 150–580[a]	700, 45
(9) 2% O$_2$ in He	850, 80	—	775, 45
(10) 2% O$_2$ in He	—	—	850, 45
(11) 2% O$_2$ in He	—	—	870, 110

[a] The residence time at each step depends upon when a well-defined carbon peak has evolved. '—' indicates a stage that is skipped in the given protocol.

without being transported to site, were analysed to determine detection limits. Field blanks, taken to site but not exposed, were also analysed as a quality control measure for the field samples to ensure levels of contamination were kept under control, but were not used as part of the statistical analysis. For the purpose of the data analysis described below, the data were divided into three data sets as follows:

(1) Data set 1 relates to those sites for which a single sampler was used. The sampler used at a particular site may be the same as that used at a different site or it may be different from the samplers used at all other sites. The data set is used to investigate between- and within-laboratory effects and to quantify those effects.

(2) Data set 2 relates to those sites for which two samplers were used. As for data set 1, a sampler used at a particular site may be the same as that used at a different site, or it may be different from the samplers used at all other sites. The data set is used to investigate a between-sampler effect and to quantify that effect.

(3) Data set 1 & 2 were combined into data set 3 and split into four concentration ranges to determine uncertainty over the range of concentrations.

The objective of the data analysis is to evaluate the standard uncertainty to be associated with an individual measured value of concentration, which combines the within-laboratory, between-laboratory and between-sampler effects. For data set 3 only the within-laboratory and between-laboratory effects were calculated. The methods employed for these calculations are elaborated upon below.

2.3.1 Calculating between- and within-laboratory variability. Let there be S sites identified by the index i, $i \in \{1, ..., S\}$ and L laboratories identified by the index j, $j \in \{1, ..., L\}$. Suppose measurements are made by all L laboratories at site i on each of the days identified by the index k, $k \in K_i$. The measured values of concentration are then denoted by:

$$x_{ijk}, k \in K_i, j \in \{1, ..., L\}, i \in \{1, ..., S\} \tag{1}$$

The data set is balanced in the sense that all laboratories make measurements at all sites. However, the days K_i on which measurements are made can be different from one site to another. The data processing comprises two stages. In a first stage, the L measured values corresponding to each site and each day are processed to remove outlying values, and normalised to remove effects associated with the factors of site, sampler and time. Outlier rejection was performed by removing values that were in excess of 3.5 robust standard deviations away from the median of the data set, as described in ref. 17. Missing values were also removed. Approximately 90% of the maximum possible data set remained after this process. In a second stage, an ANalysis Of VAriance (ANOVA) is applied to the resulting data set corresponding to all sites and all days to decide whether a between-laboratory effect exists. If the effect exists, a calculation of the between- and within-laboratory standard deviations is undertaken. If the effect does not exist, a calculation of the within-laboratory standard deviation only is made. In either case, the within-laboratory standard deviation describes the repeatability standard deviation for the laboratories, which is assumed to be the same for all laboratories. Data normalisation involves the following steps:

(1) Evaluate the average

$$\bar{x}_{ik} = \frac{1}{L_{ik}} \sum_{j \in J_{ik}} x_{ijk} \tag{2}$$

where L_{ik} is the number of indices in J_{ik} (where J_{ik} contains the indices of the laboratories for which measured values are retained after the stage to remove outlying values);

(2) Shift the measured values by the average $\bar{x}_{ik}$, and scale the shifted values by $\bar{x}_{ik}$ to obtain normalised values

$$v_{ijk} = \frac{x_{ijk} - \bar{x}_{ik}}{\bar{x}_{ik}}, \quad j \in J_{ik} \tag{3}$$

The average value calculated in step 1 is taken as a reference or consensus value for the concentration of OC, EC or TC at the specific site i on the specific day k. The aim of shifting the measured values is to remove, at least approximately, any dependence of measured concentration on the time of measurement, the site at which the measurement is made, and the sampler that is used to make the measurement. The aim of scaling the shifted values is to remove, at least approximately, any dependence of the variability of measured concentration on the value of concentration. The application of scaling is based on the assumption that the repeatability standard deviation is proportional to the measured concentration, *i.e.*, the relative repeatability standard deviation is approximately constant. Whilst it is acknowledged that this model is only valid for a limited measurement range, away from the detection limit, the assumption remains fit-for-purpose for the subsequent statistical analysis performed in this study. The normalised data is reported as a fraction (or percentage) of the reference value. Following data normalisation, the data is denoted by

$$v_{ijk}, k \in K_i, j \in J_{ik}, i \in \{1, \ldots, S\} \tag{4}$$

Since there is no mathematical interest in the factors of site and time, an equivalent representation of the data is

$$y_{lr}, r \in \{1, \ldots, R_l\}, l \in \{1, \ldots, L\} \tag{5}$$

which groups the measured values by laboratory. Here, R_l is the number of retained measured values for laboratory l over all sites and days, and each y_{lr} equates to one of the normalised values v_{ijk}. An ANOVA is used to test the null hypothesis that the averages for the laboratories are equal, $i.e.$, there is no laboratory effect. The ANOVA calculation must account for the fact that the data set may be unbalanced, because the numbers of measured values can be different from one laboratory to another. The function "anovan", which is provided in Matlab's Statistics Toolbox, can be used to perform a (multiway) ANOVA and allows for an unbalanced data set. The function returns a p value that is compared with a critical value p_c, usually 0.01 or 0.05, chosen before the analysis. If $p \geq p_c$, the null hypothesis of no laboratory effect is accepted at a $100p_c\%$ level of confidence. Otherwise, the null hypothesis is rejected at that level of confidence. The within-laboratory variance (squared standard deviation) is calculated as the pooled variance s^2 of the laboratory values, and is given by:

$$s^2 = \frac{1}{\left(\sum_{l=1}^{L} R_l\right) - L} \sum_{l=1}^{L} (R_l - 1) s_l^2 \tag{6}$$

where

$$\bar{y}_l = \frac{1}{R_l} \sum_{r=1}^{R_l} y_{lr}, \ s_l^2 = \frac{1}{R_l - 1} \sum_{r=1}^{R_l} (y_{lr} - \bar{y}_l)^2 \tag{7}$$

If the ANOVA indicates the existence of a between-laboratory effect, then that effect is quantified by the standard deviation s_{lab} of the averages calculated for the laboratories, $i.e.$,

$$s_{\text{lab}}^2 = \frac{1}{L-1} \sum_{l=1}^{L} (\bar{y}_l - \bar{\bar{y}})^2 \tag{8}$$

where

$$\bar{\bar{y}} = \frac{1}{L} \sum_{l=1}^{L} \bar{y}_l \tag{9}$$

2.3.2 Calculating between-sampler variability. Let there be S sites identified by the index i, $i \in \{1, \ldots, S\}$, and L laboratories identified by the index j, $j \in \{1, \ldots, L\}$. Suppose measurements are made by all L laboratories at site i on each of the days identified by the index k, $k \in K_i$, using samplers identified by index f, $f \in F_i$. The measured values of concentration are then denoted by:

$$x_{ijkf}, f \in F_i, k \in K_i, j \in \{1, \ldots, L\}, i \in \{1, \ldots, S\} \tag{10}$$

The data set is balanced in the sense that all laboratories make measurements at all sites. However, the days K_i on which measurements are made, as well as the samplers F_i used, can be different from one site to another. In a similar way to that for data set 1, the data processing comprises two stages. In a first stage, considering the data for each sampler separately, the measured values corresponding to each site and each day are processed to remove outlying values. Then, considering the data for the samplers together, the values corresponding to each site and each day are normalised to remove effects associated with the factors of site and time, but preserving any sampler effect. In a second stage, an ANOVA is applied to the resulting data set corresponding to all sites, all days and all laboratories to decide whether a between-sampler effect exists. If the effect exists, a calculation of the between-sampler standard deviations is undertaken. Following data normalisation, the data is denoted by:

$$v_{ijkf}, f \in F_i, k \in K_i, j \in J_{ikf}, i \in \{1, \ldots, S\} \tag{11}$$

where J_{ikf} contains the indices of the laboratories for which measured values are retained for site i, day k and sampler f. Since there is no mathematical interest in the factors of site, time and laboratory, an equivalent representation of the data is as:

$$y_{fr}, r \in \{1, \ldots, R_f\}, f \in F_1 \cup \ldots \cup F_S \tag{12}$$

which groups the measured values by sampler. Here, R_f is the number of retained measured values for sampler f over all sites, days and laboratories, and each y_{fr} equates to one of the normalised values v_{ijkf}. As for data set 1, the data set is generally unbalanced, because the numbers of measured values can be different from one sampler to another. If an ANOVA indicates the existence of a between-sampler effect, then that effect is quantified by the standard deviation s_{sam} of the averages calculated for the samplers, $i.e.$,

$$s_{\text{sam}}^2 = \frac{1}{F-1} \sum_{f=1}^{F} (\bar{y}_f - \bar{\bar{y}})^2 \tag{13}$$

where

$$\bar{y}_f = \frac{1}{R_f} \sum_{r=1}^{R_f} y_{fr}, \ \bar{\bar{y}} = \frac{1}{F} \sum_{f=1}^{F} \bar{y}_f \tag{14}$$

and F is the total number of samplers. In the case that the null hypothesis of no sampler effect is accepted, the between-sampler standard deviation is taken to be zero.

2.3.3 Combined standard uncertainty. The relative standard uncertainty u_{rel} associated with an individual measured value of concentration is given by:

$$u_{\text{rel}}^2 = s^2 + s_{\text{lab}}^2 + s_{\text{sam}}^2 \tag{15}$$

which combines the standard deviations related to respectively within-laboratory, between-laboratory and between-sampler effects. The between sampler variability results showed that s_{sam}^2 was insignificant in most cases and was not included in the uncertainty calculation. The data analysis described, including the use of ANOVA and the calculations of the

standard deviations to quantify the various effects, depends on assumptions about the homogeneity of these effects for different concentrations, sites, laboratories, samplers and time. Graphical displays of the data, in which the data is plotted against these factors, can be useful to identify obvious departures from these assumptions. In cases that the assumptions do not hold, the results of the data analysis may not be reliable expressions of the various effects considered. The expanded combined uncertainty u_{rel} at a 95% level of confidence is then calculated as ku_{rel} where k is taken to be 2.

3. Results and discussion

3.1 Laboratory trials

Prior to the field trial stage a laboratory comparison was performed to ensure that the agreed protocols could be implemented satisfactorily at each analytical laboratory and to perform an initial assessment of the likely comparability of the methods. This was an important step to ensure that the protocols produced sufficiently different results that it would be necessary to continue to use all three protocols during the field trial. Each of the four analytical laboratories measured three real filter samples taken at the Amsterdam field trial site. Within each laboratory the results have been normalised to the average concentration on each filter, and then averaged across the three filters and the four laboratories. The results are shown in Fig. 1. This shows that the three protocols produce consistent results for measurement of TC, whilst there are significant differences in the results produced for the measurement of OC and EC. The NIOSH870 protocol seems most different from the others but there are also significant differences between the EUSAAR-2 and IMPROVE_A protocols for EC. As discussed later the lower concentrations of EC compared to OC routinely

encountered in ambient air explain the larger discrepancies between protocols for EC than for OC. The differences observed in these results demonstrated the need to use all three analysis protocols during the field trial. The laboratory trials were not used to generate any of the statistics discussed in Section 3 nor to form any opinions on the relative merits of each of the protocols.

3.2 Field trials

The average filter loading measured at the six locations are shown in Fig. 2(a)–(f).

3.3 Data set 1 – between-laboratory and within-laboratory variability

The analysis of variance for between-laboratory effect was carried out on the data set where only one sampler was present. For each thermal–optical protocol, the carbon concentrations in μg cm^{-2} were calculated using both transmittance (T) and reflectance (R) corrections, leading to six thermal–optical data sets to assess. Table 3 shows the p values from the analysis of variance for between-laboratory effects. If $p \geq p_c$, the null hypothesis of no laboratory effect is accepted at a $100p_c\%$ level of confidence. Otherwise, the null hypothesis is rejected at that level of confidence. It can be seen that only IMPROVE_A reflectance (EC), NIOSH870 reflectance (EC) and NIOSH870 transmittance (OC) show no significant between-laboratory effect at the 5% level of confidence. In all cases the between laboratory variability was calculated and included in the overall uncertainty calculations.

Between-laboratory variability and within-laboratory variability were calculated for the six thermal–optical protocols and the results are shown in Table 4 and 5.

The combined variability – calculated from the between-laboratory variability and within-laboratory variability added in quadrature – was also calculated at each field trial site and the results are shown in Table S1 of the ESI.†

3.4 Data set 2 – between-sampler variability

The between-sampler variability was calculated using the data set from the sites where the two samplers were running in parallel. No significant between-sampler variability was found for TC. The OC IMPROVE_A protocol, with the transmittance optical correction, gave a possibly significant between-sampler variability (5.5% level) resulting in an inter-sampler variability of 1.94% (2σ, at a 95% level of confidence). For EC the three protocols showed a significant to very significant between-sampler variability: EUSAAR_2 (T), 3.2% level, IMPROVE_A (T), 0.3% level & NIOSH870 (T) 0.5% level, resulting in a between-sampler variability of 5.8, 6.0 and 8.8% (2σ, at a 95% level of confidence) respectively. The origin of the significant between-sampler variability is unclear, but it may have only been detected due to the lower repeatability of these 3 protocols, 18%, 22% & 32% (2σ, at a 95% level of confidence). The other three protocols, EUSAAR_2 (R), IMPROVE_A (R) & NIOSH870 (R) had within-laboratory repeatability in excess of 100% (2σ, at a 95% level of confidence) making the determination of
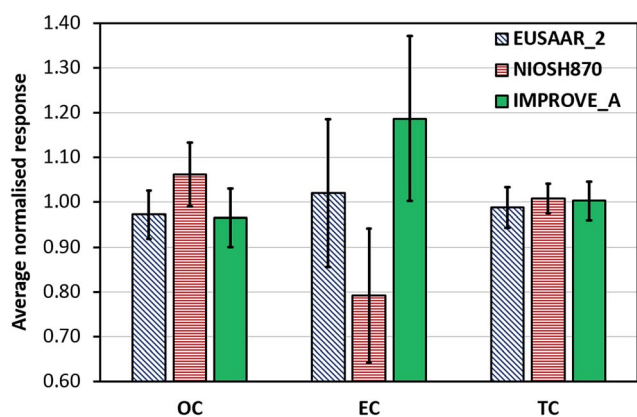

Fig. 1 The average normalised response across three filter samples measured at the four analytical laboratories for the three analysis protocols used: EUSAAR_2 (blue, diagonal stripes), NIOSH870 (red, horizontal stripes) and IMPROVE_A (green, solid bar). The error bars represent the standard deviation of the responses expanded using the relevant t-factor to give a level of confidence of approximately 95%. (The approximate average TC, OC and EC loadings, respectively, for the three filters studied were, in μg cm^{-2}: [24, 22, 2], [24, 15, 9], and [14, 12, 2].)
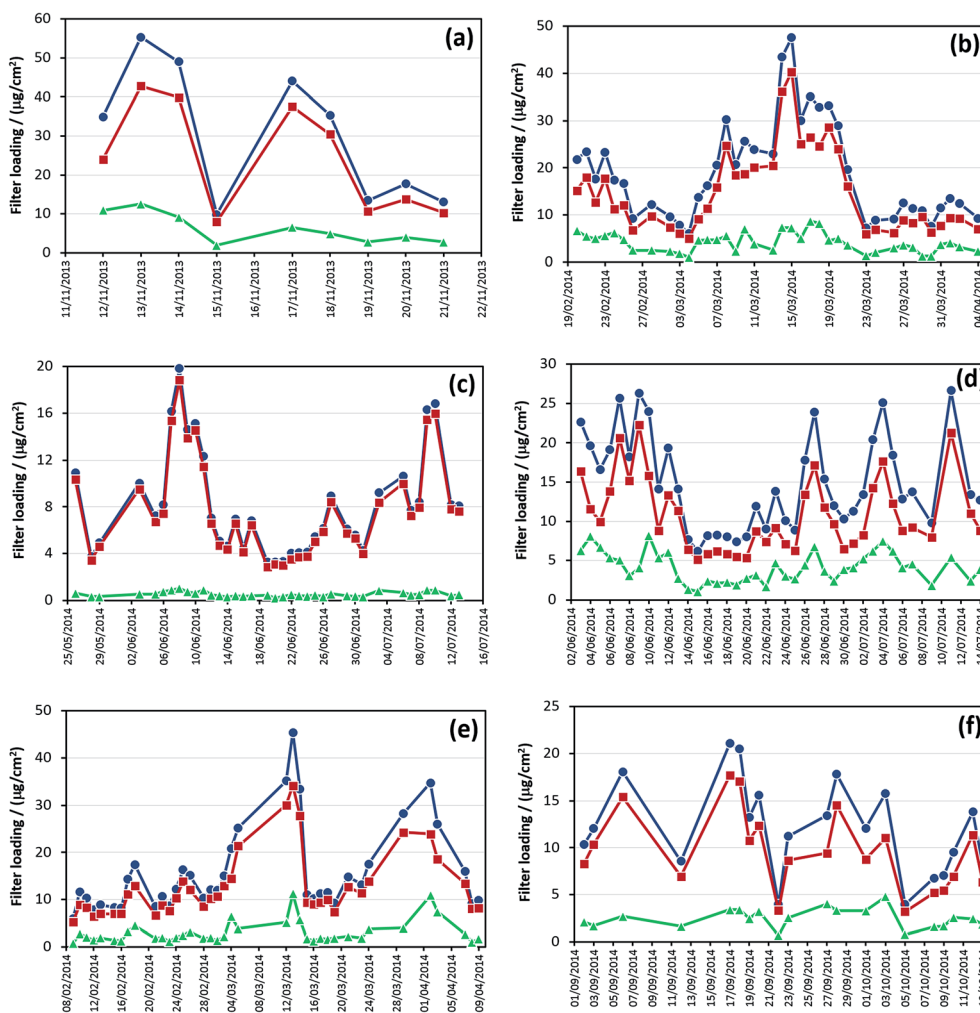
Fig. 2 The filter loadings of TC (blue circles), EC (green triangles) and OC (red squares) measured during the field trials at (a) Ispra, (b) Barcelona, (c) Waldhof, (d) Amsterdam, (e) Mülheim, (f) Cabauw, averaged across all protocols and all analysis laboratories. Outliers and missing values have been removed. At Ispra and Cabauw only days where valid samples were produced by both samplers operating in parallel are shown.

between-sampler repeatability impossible. Where significant between-laboratory variability was detected it was insignificant compared to the within-laboratory repeatability and between-sampler variability was not included in the overall uncertainty calculations.

Table 3 Value of p obtained from the analysis of variance for between-laboratory effect for all the thermal–optical protocols (T indicates transmittance correction, and R reflectance correction). Values considered not significant at the 5% level of confidence are in italics and underlined

Protocol	EC	OC	TC
EUSAAR_2 (T)	0.0003	0.0006	0.0003
IMPROVE_A (T)	0.0002	<0.0001	<0.0001
NIOSH870 (T)	<0.0001	_0.2399_	<0.0001
EUSAAR_2 (R)	0.0003	0.0012	0.0003
IMPROVE_A (R)	_0.0519_	<0.0001	<0.0001
NIOSH870 (R)	_0.3544_	<0.0001	<0.0001

3.5 Data set 3 – uncertainty over the measured concentration range

Results for each analysis protocol from the complete data set were ordered by concentration and then split into four concentration bins with an equal number of measurements in each concentration bin. The combined expanded uncertainty (2σ, 95% level of confidence) due to between-laboratory variability and within-laboratory variability were calculated for each concentration bin, the results of which are shown in Fig. 3–5.

The uncertainty displayed by these plots is generally quite consistent in absolute terms across the Q4 to Q2 range, with some protocols showing a clear decrease as filter loadings decrease. For all protocols a much larger uncertainty is seen for Q1, at low filter loadings. These trends are consistent with the expected characteristic empirical concentration-dependent uncertainty function of analytical measurement of the form: $\sigma = \sqrt{\alpha^2 + (\beta c)^2}$ where σ represents uncertainty, c is concentration and α and β are constants.[18] The uncertainty for the Q1 concentrations becomes limited by the detection limit. The

This journal is © The Royal Society of Chemistry 2017

Environ. Sci.: Processes Impacts, 2017, **19**, 1249–1259 | **1255**

Table 4 Between-laboratory variability for the six thermal–optical protocols (*T* indicates for transmittance correction, and *R* reflectance correction)

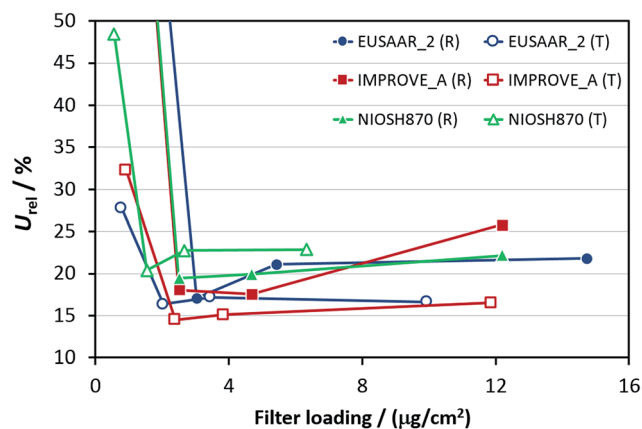| Component	Analysis protocol	Between-laboratory variability, %
EC	EUSAAR_2 (*T*)	2.2
	IMPROVE_A (*T*)	2.9
	NIOSH870 (*T*)	9.0
	EUSAAR_2 (*R*)	13
	IMPROVE_A (*R*)	67
	NIOSH870 (*R*)	>100
OC	NIOSH870 (*T*)	0.6
	EUSAAR_2 (*T*)	1.5
	EUSAAR_2 (*R*)	1.5
	IMPROVE_A (*T*)	2.1
	IMPROVE_A (*R*)	2.3
	NIOSH870 (*R*)	3.3
TC	EUSAAR_2 (*T*)	1.3
	EUSAAR_2 (*R*)	1.3
	NIOSH870 (*R*)	1.6
	NIOSH870 (*T*)	1.6
	IMPROVE_A (*T*)	2.4
	IMPROVE_A (*R*)	2.4

Fig. 3 Combined uncertainty at the 95% level of confidence for the measurement of elemental carbon (EC) across the concentration range observed in the field trial using the measurement protocol indicated. The *x*-axis plots the median of each concentration range. The results for all three measurement protocols when using the reflectance correction are off scale at >100%.

uncertainties observed across the different protocols are rather consistent. However, a couple of significant patterns are clear. For OC and EC measurements the use of a transmittance correction results in much lower uncertainties than a reflectance correction. In particular for EC the reflectance correction is unable to detect small EC concentrations, especially below 1 μg cm^{-2}. The reflectance correction results in uncertainties of over 100% at low concentrations – these are unacceptably high for routine measurement. For EC measurement the NIOSH870 transmittance correction protocol results in a much larger uncertainty than the transmittance correction protocols for

EUSAAR_2 and IMPROVE_A. In all cases the combined uncertainty is less than 25% for filter loadings above 2 μg cm^{-2}. For TC measurement the uncertainty differences between the protocols are not significant.

3.6 Method detection limit

Laboratory blank filters were repeatedly analysed by each laboratory and the results pooled to determine the method detection limit, *D*, in μg cm^{-2} for each protocol according to the following equation:

$$D = \sigma \times t_{99,n-1} \qquad (18)$$

where: σ = the standard deviation of the repeated measurements, $t_{99,n-1}$ = the Student *t* factor for a level of confidence of

Table 5 Within-laboratory variability for the six thermal–optical protocols (*T* indicates for transmittance correction, and *R* reflectance correction)

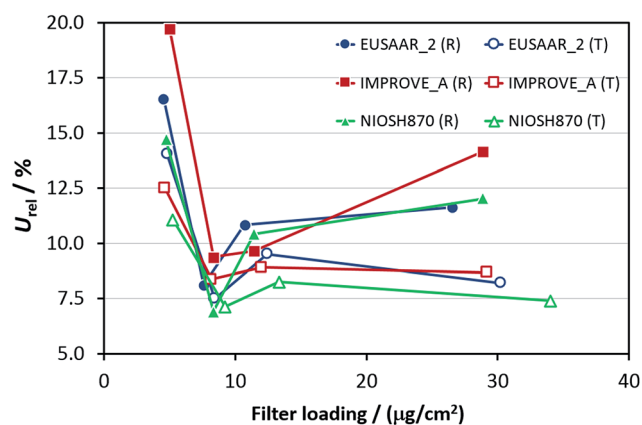| Component	Analysis protocol	Within laboratory variability, %
EC	EUSAAR_2 (*T*)	9.4
	IMPROVE_A (*T*)	12
	NIOSH870 (*T*)	16
	EUSAAR_2 (*R*)	56
	IMPROVE_A (*R*)	>100
	NIOSH870 (*R*)	>100
OC	NIOSH870 (*T*)	4.3
	NIOSH870 (*R*)	5.1
	IMPROVE_A (*T*)	6.4
	EUSAAR_2 (*T*)	6.7
	IMPROVE_A (*R*)	7.0
	EUSAAR_2 (*R*)	7.1
TC	NIOSH870 (*R*)	3.5
	NIOSH870 (*T*)	3.5
	IMPROVE_A (*T*)	5.6
	IMPROVE_A (*R*)	5.6
	EUSAAR_2 (*T*)	5.9
	EUSAAR_2 (*R*)	5.9

Fig. 4 Combined uncertainty at the 95% level of confidence for the measurement of organic carbon (OC) across the concentration range observed in the field trial using the measurement protocol indicated. The *x*-axis plots the median of each concentration range.

1256 | *Environ. Sci.: Processes Impacts*, 2017, **19**, 1249–1259

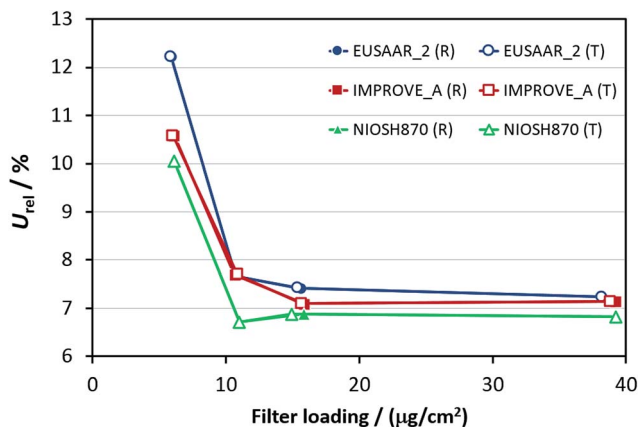This journal is © The Royal Society of Chemistry 2017

Fig. 5 Combined uncertainty at the 95% level of confidence for the measurement of total carbon (TC) across the concentration range observed in the field trial using the measurement protocol indicated. The *x*-axis plots the median of each concentration range. The (R) and (T) data for the three different protocols is almost completely overlapping.

99% and $n-1$ degrees of freedom, $n =$ the number of repeated measurements.

The absolute blank values varied according to filter type, but were around 2 μg cm^{-2} for TC, almost entirely contributed by the OC content. Because of the low sensitivity of the methods for very low EC concentrations – particularly for the reflectance mode correction but also to some extent for transmittance mode correction – detection limits for EC were also calculated using sampled filters with low concentrations discounting those with a zero content. A cut-off of 0.4 μg cm^{-2} was chosen for the filters used as this represented twice the detection limit achieve for the EC transmittance mode using blank filters. Eqn (18) was used for these calculations. Table 6 shows the detection limits calculated using these method.

Table 6 Methods detection limits calculated for the analysis protocols described. EC (blank) and EC (sample) refer to detection limits calculated using the different methods described in the text. The OC and TC detection limits were calculated using the blank method. Values of n for the OC, EC (blank) and TC determinations were 14 for the EUSAAR_2 and IMPROVE_A protocols and 12 for the NIOSH870 protocol. For the EC (sample) determinations values of n varied for each protocol, from top to bottom of the table, as: 21, 59, 39, 29, 42, 82. Differences in detection limits for OC and EC arise because of differences in the optical correction and data processing between the different protocols. Differences in the detection limits for TC are not significant but would arise because of the slightly different repeatability characteristics of the three protocols

| Analysis protocol | Detection limit/(μg cm^{-2}) | | | |
	OC	EC (blank)	EC (sample)	TC
EUSAAR_2 (R)	1.9	0.017	0.25	1.9
EUSAAR_2 (T)	1.8	0.18	0.10	1.9
IMPROVE_A (R)	2.2	0.001	0.11	2.2
IMPROVE_A (T)	2.1	0.24	0.10	2.2
NIOSH870 (R)	2.8	0.002	0.23	2.8
NIOSH870 (T)	2.3	0.66	0.17	2.8

During the drafting of the standard method it was proposed that the maximum OC content of laboratory blank filters must be 2 μg cm^{-2}. Given the detection limits achieved this seems like just about the lowest requirement that could be specified. The EUSAAR_2 protocol just achieves this, whereas the IMPROVE_A protocol just fails to meet this requirement. The NIOSH870 protocol shows much higher detection limits and would not be able to confidently assess maximum OC content at a loading of 2 μg cm^{-2}. The detection limits determined using reflectance correction protocol were marginally higher than those found using the transmittance correction protocol – excluding the known measurement issues with the reflectance correction values for EC (blank).

It is important to note that this definition of the detection limit is only dependent on the repeatability of the analytical response of a blank sample and the sensitivity of the analytical method (such that this analytical response is then converted into a filter loading). The detection limit is only indirectly dependent on the absolute values obtained in that there is a relationship between the analytical intensity and repeatability. However the absolute size of the blank is an important quality control criterion that is applied in the final method to demonstrate that appropriate filters are being used for sampling. Care must be used when quoting results around or below detection limits[19] but it is common for these to be calculated in such a way as part of European standard methods. The detection limit played no part in the choice of protocol for the final standard. The data in Table 6 was used, in general terms, to provide a lower working range for the standard method (based on the measurements reported) and also provide a quality control criterion for the filters used in the method.

4. Conclusions

This paper has presented a summary of the result of the CEN/TC264/WG35 field trials and summarised the techniques developed to derive an uncertainty, working range and detection limit from these data. These results have fed directly into the published European Standard EN 16909:2017.[20] It is likely that this standard will be adopted as the reference method in Europe requiring its use by all Member States. This standard method addresses the analysis of PM$_{2.5}$ and smaller particles. Although the method must be applicable for background sites to satisfy Directive 2008/50/EC, there is also a strong need that the same sampling and analysis method should also be applicable for suburban and urban background sites as well as, if possible, for roadside sites, to ensure maximum comparability and utility of the methods across Europe. The measurement of OC and EC in larger particles and higher filter loadings that will be encountered at these sites requires further development because of the various measurement artefacts that are likely to be encountered. The measurement of OC/EC in larger particles is a topic that WG35 is likely to address in future.

The analysis shows that the reflectance protocols for all three methods are not able to detect very low levels of elemental carbon and are unable to deliver the detection limits required and therefore it is clear that the use of transmittance protocols is preferred, particularly at low concentrations where the

reflectance protocol is unable to detect small EC concentrations, especially below 1 μg cm^{-2}. The results from the transmittance protocols show no protocol clearly performing better in term of repeatability and reproducibility than the others across all of the site types and concentrations encountered during the field validation tests – although the uncertainty of the NIOSH870 protocol for EC seemed to be significantly larger than the other protocols at low concentrations. The laboratory tests showed that there are significant differences between the protocols in terms of the split between OC and EC – although as discussed these measurands are currently operationally defined by the methods themselves and the different characteristics of the correction algorithms to deal with charring of the OC components during heating. The differences in the split for OC and EC are larger between NIOSH870 and the other two protocols than they are between EUSAAR_2 and IMPROVE_A.

Given no significant differences in performance between the three transmittance protocols – especially when the differing performance at different test locations was considered (see the ESI†) – the majority decision within the WG was to recommend the EUSAAR_2 transmittance protocol. This was not only because this would ensure consistency and continuity of data with the pre-existing EUSAAR, EMEP and ACTRIS European monitoring activities, but also because many air quality reference laboratories across Europe already used the protocol.

The expanded uncertainty of a single measurement result (at the 95% confidence interval) for the transmittance protocols is less than 25% above filter loadings of 2 μg cm^{-2}. For concentrations below this the relative uncertainty is greater than 25% at which point the content and variability of the blank filter and therefore the detection limit becomes a limiting factor. Further detailed guidance on calculation of uncertainties was provided in the published standard. From the detection limit data obtained during the study recommendations were made for qualifying filter materials suitable for use with the published standard. A working range for the standard was produced based on the filter loadings encountered and successfully measured during the laboratory and field campaigns.

It is hoped that the detail provided in this paper will give clarity and confidence in the underpinning science for future users of the standard method and will also provide a detailed framework for assessing the uncertainty and method performance characteristics for future field trials of this type. This study has also highlighted the uncertainty with which data can be obtained using this method and therefore by corollary imposes a limitation of the level of certainty with which conclusions about trends in filter loadings and comparability of filter loadings between locations can be made.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 European Commission, Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, *Official Journal of the European Union*, 2008, **L152**, 1–44.

2 J. C. Chow, J. G. Watson, A. Chen, M. C. Chang, N. F. Robinson, D. Trimble and S. Kohl, The IMPROVE_A temperature protocol for thermal/optical carbon analysis: maintaining consistency with a long-term database, *J. Air Waste Manage. Assoc.*, 2007, **57**, 1014–1023.

3 F. Cavalli, M. Viana, K. E. Yttri, J. Genberg and J.-P. Putaud, Toward a standardised thermaloptical protocol for measuring atmospheric organic and elemental carbon: the EUSAAR protocol, *Atmos. Meas. Tech.*, 2010, **3**, 79–89.

4 M. E. Birch and R. A. Cary, Elemental Carbon-Based Method for Monitoring Occupational Exposures to Particulate Diesel Exhaust, *Aerosol Sci. Technol.*, 1996, **25**, 221–241.

5 M. E. Birch, D. Dahmann and H. H. Fricke, Comparison of two carbon analysis methods for monitoring diesel particulate levels in mines, *J. Environ. Monit.*, 1999, **1**, 541–544.

6 P. Panteliadis, T. Hafkenscheid, B. Cary, E. Diapouli, A. Fischer, O. Favez, P. Quincey, M. Viana, R. Hitzenberger, R. Vecchi and D. Saraga, ECOC comparison exercise with identical thermal protocols after temperature offset correction: instrument diagnostics by in-depth evaluation of operational parameters, *Atmos. Meas. Tech.*, 2015, **8**, 779–792.

7 A. T. Bautista, P. C. B Pabroa, F. L. Santos, L. L. Quirit, J. L. B. Asis, M. A. K. Dy and J. P. G. Martinez, Intercomparison between NIOSH, IMPROVE_A, and EUSAAR_2 protocols: Finding an optimal thermal–optical protocol for Philippines OC/EC samples, *Atmos. Pollut. Res.*, 2015, **6**, 334–342.

8 Y. M. Han, L. W. Chen, R. J. Huang, J. C. Chow, J. G. Watson, H. Y. Ni, S. X. Liu, K. K. Fung, Z. X. Shen, C. Wei and Q. Y. Wang, Carbonaceous aerosols in megacity Xi'an, China: Implications of thermal/optical protocols comparison, *Atmos. Environ.*, 2016, **132**, 58–68.

9 C. Wu, X. H. Huang, W. M. Ng, S. M. Griffith and J. Z. Yu, Inter-comparison of NIOSH and IMPROVE protocols for OC and EC determination: implications for inter-protocol data conversion, *Atmos. Meas. Tech.*, 2016, **9**, 4547–4560.

10 M. Giannoni, G. Calzolai, M. Chiari, A. Cincinelli, F. Lucarelli, T. Martellini and S. Nava, A comparison between thermal-optical transmittance elemental carbon measured by different protocols in PM$_{2.5}$ samples, *Sci. Total Environ.*, 2016, **571**, 195–205.

11 Y. Cheng, K. B. He, F. K. Duan, Z. Y. Du, M. Zheng and Y. L. Ma, Ambient organic carbon to elemental carbon ratios: Influence of the thermal–optical temperature protocol and implications, *Sci. Total Environ.*, 2014, **468**, 1103–1111.

12 R. J. C. Brown, N. Pirrone, C. Van Hoek, F. Sprovieri, R. Fernandez and K. Toté, *J. Environ. Monit.*, 2010, **12**, 689–695.

13 R. J. C. Brown, N. Pirrone, C. Van Hoek, M. Horvat, J. Kotnik, I. Wangberg, W. T. Corns, E. Bieber and F. Sprovieri, *Accredit. Qual. Assur.*, 2010, **15**, 359–366.

14 VDI/DIN-Commission on Air Pollution Prevention (KRdL) – Standards Committee, Results of European laboratory and field tests, https://www.vdi.de/technik/fachthemen/reinhaltung-der-luft/europaeische-und-internationale-normung/results-of-european-lab-and-field-validation-tests/, accessed August 2017.

15 R. J. C. Brown, K. E. Jarvis, B. A. Disch, S. L. Goddard and A. S. Brown, Spatial inhomogeneity of metals in particulate matter on ambient air filters determined by LA-ICP-MS and comparison with acid digestion ICP-MS, *J. Environ. Monit.*, 2009, **11**, 2022–2029.

16 European Committee for Standardisation, EN 12341:2014, *Ambient Air Quality: Standard gravimetric measurement method for the determination of the PM10 or PM2.5 mass concentration of suspended particulate matter*, CEN, Brussels, 2014.

17 S. Beccaceci, R. J. C. Brown, D. M. Butterfield, P. M. Harris, R. P. Otjes, C. van Hoek, U. Makkonen, M. Catrambone, R. Fernández Patier, M. M. G. Houtzager and J.-P. Putaud, Standardisation of a European measurement method for the determination of anions and cations in PM2.5: results of field trial campaign and determination of measurement uncertainty, *Environ. Sci.: Processes Impacts*, 2016, **18**, 1561–1571.

18 M. Thompson, Uncertainty functions, a compact way of summarising or specifying the behaviour of analytical systems, *TrAC, Trends Anal. Chem.*, 2011, **30**, 1168–1175.

19 R. J. C. Brown, The use and abuse of limits of detection in environmental analytical chemistry, *Sci. World J.*, 2008, **8**, 796–801.

20 European Committee for Standardisation, EN 16909:2017, *Ambient air – Measurement of elemental carbon (EC) and organic carbon (OC) collected on filters*, CEN, Brussels, 2017.