Results of the inter-laboratory comparison exercise for TC and EC measurements (ref.: OCEC-2016-1)

Summary

The European Centre for Aerosol Calibration (ECAC) with ACTRIS-2 completed (March 2016) an inter-laboratory comparison for the measurement of total carbon (TC), elemental carbon (EC) and organic carbon (OC) in particulate matter collected on filters. The aim of this comparison was to evaluate the performances of the measurement method (i.e. reproducibility and repeatability) and of individual laboratories (biases).

This exercise was based on ambient PM2.5 and PM10 aerosol samples collected on quartz fiber filters at regional background sites in Norway, Germany, and Spain, and at one urban background station in Greece. A solution of phthalic acid prepared at the JRC (the inter-laboratory comparison exercise coordinator) was also distributed.

Thirteen laboratories responsible for the aerosol chemical speciation at the EMEP or ACTRIS stations located in their countries (i.e. Cyprus, Czech Republic, Finland, France, Germany, Greece, Italy, Norway, Poland, Spain and Sweden) participated running their usual thermal-optical EUSAAR_2 protocol with their usual analytical instrument (Sunset lab. OCEC aerosol analyser).

Still in absence of a suitable certified reference material for atmospheric OC and EC the assigned values for TC loadings and EC/TC ratios in the test samples were calculated as the robust average values among all participants (outlier excluded). The assigned value for the concentration of phthalic acid was determined from primary gravimetric and volumetric measurements.

Measurement method performance: for TC determination, repeatability and reproducibility relative standard deviations ranged from 2% to 5% and from 6% to 10%, respectively. For the determination of the EC/TC ratio, repeatability and reproducibility relative standard deviations ranged from 4% to 7% and from 11% to 24%. Repeatability and reproducibility standard deviations show an inverse dependence on TC loadings and on EC/TC ratios.

Laboratory performance: for both TC loadings and EC/TC ratios, laboratories' performances were assessed in terms of z-scores, calculating the *standard deviation for proficiency assessment* (σ^*) *from the data obtained in the round of the proficiency testing scheme.*
For TC loadings, only one outlier and two stragglers were identified; and 80% of all entries is within 10% from the assigned TC concentration value and 95% within 15%.
Regarding EC/TC ratios, all ten stragglers and outliers were produced by three participants. Only 50% of all entries is within 10% of the assigned value and 85% is within the 25% of the assigned EC/TC ratio.
Although the contribution of localized sample heterogeneities and /or contaminations to biased data cannot be totally excluded, the random scheme adopted to distribute sub-samples was such that the recurrence of stragglers or outliers (more than two) for single laboratories most probably indicates an unsatisfactory laboratory performance as compared to the other participants.

Laboratories showing unsatisfactory precision (both in terms of repeatability and reproducibility) or significant biases for several test samples shall carefully examine their operating procedures and instrumental set-up and identify appropriate corrective actions with the help of ECAC staff if needed.

Introduction

Total carbon (TC), including Organic Carbon (OC) and Elemental Carbon (EC) is a relevant constituent of the fine fraction of particulate matter (PM), both from the perspective of health risks due to inhalation and indication of air pollution sources. For these reasons requirements for measuring EC and OC in $PM_{2.5}$ at rural background locations have been included in Air Quality Directive 2008/50/EC.

The directive states that measurements should be made in a manner consistent with those of the cooperative programme for monitoring and evaluation of the long range transmission of air pollutants in Europe (EMEP). Thermal-optical analysis has been recognized as the most suitable method for the determination of EC and OC collected on filters (see Technical Specification by the CEN/TC265 WG35) and the thermal protocol EUSAAR-2 with a transmittance optical correction for pyrolysis -already adopted in the EMEP manual for sampling and analysis- has been recently selected as the European standard thermal protocol (prEN16909).

The *European center for aerosol calibration* within the European project ACTRIS-2 has organized in 2016 an inter-laboratory comparison exercise (ILCE) (ref. OCEC-2016-1) among thirteen applicants. In particular, this ILCE involves the laboratories, all ACTRIS-2 partners and associated partners, in charge of OC and EC measurements at the EMEP stations in Cyprus, Czech Republic, Finland, France, Germany, Greece, Italy, Norway, Poland, Spain and Sweden.

1 Organization

1.1 Samples, sub-samples and sub-sample homogeneity

In lack of certified reference material for atmospheric OC and EC, this ILCE made use of ambient (outdoor) PM aerosol collected with high-volume samplers on quartz fiber filters at 4 sites across Europe (Table 1). Upon receipt at JRC, filters were stored in a refrigerator.

Table 1: filter test samples used for the inter-laboratory comparison

Sampling location				Sample collection		
Station	Country	Symbol	Site type	Period	Size fraction	Filter type
Melpitz	Germany	D	rural	Winter 2015	PM2.5	Munktell MK360
Demokritos	Greece	G	urban	Winter 2015	PM2.5	MCV-QF1
Montseny	Spain	E	rural	Winter 2015	PM2.5	Whatman QMH
Birkenes	Norway	N	rural	Apr. 2000	PM10	Whatman QMA

Rectangular filter punches of ca. 3.6 cm x 1.8 cm were randomly distributed to participants 1-13 to allow them to triplicate measurements. As the first set of test samples sent to Participant 4 got lost, they analysed for samples 4638, DEM2, T6694 and T6696 punches directly taken from the left over of the original big filter.

The homogeneity of these test samples was investigated by the JRC on distinct filter samples collected at each location. From each sample, ten subsamples of 1 cm² were taken along two perpendicular axes across the filter surface and analysed for their TC, OC and EC contents. The filter homogeneity was assessed as the standard deviation of the average of the 10 replicate analyses. This leads to an upper limit for the filter homogeneity since it includes the repeatability of the JRC laboratory (< 3 and 6% for TC and EC, respectively). The homogeneity is better than 6 and 11% for TC and EC/TC, respectively (Table 2). If sampling at each location occurred under repeatable conditions, it can be assumed that the test samples had similar homogeneities.

Table 2: homogeneity of the deposits on filters collected with the samplers used to produce the eight test filters. Analyses were performed with the protocol EUSAAR_2 and charring correction by transmittance monitoring.

Test sample and origin	Homogeneity for TC (%)	Homogeneity for EC/TC (%)
46_ _ MEL (D)	2.5	4.0
DEM _ DEM(G)	5.7	7.6
T66_ _ Mont (E)	3.9	10.2
A21_ BIR (N)	4.7	11.0

An aqueous solution of phthalic acid was also distributed to the participants to assess the uncertainty of the instrument calibration constant determination. The solution was prepared by dissolving a precisely known mass of pure phthalic acid (≥ 99.5%) in a precisely known volume of ultra-pure water (resistivity ≥18.2 mΩ cm).

1.2 Participants

Participants were selected among applicants to ECAC choosing (in the interest for the scientific community) in a first place laboratories which submit TC and EC data to the EBAS database and laboratories which could most benefit from the outcome of this exercise in term of their expertise development.
The list of the thirteen participants is reported in Table 3. For brevity, the number assigned to each participant will be used in the remainder of the document.

1.3 Sample shipment and reporting of results

Test samples were shipped to all participants (except "local" participant 10) on 8th Feb. 2016 via postal mail at ambient temperature without temperature record in closed petri dishes. Participants were asked to report TC and EC concentration, in µg C cm^{-2} units with three decimal digits, from three replicates of test ambient PM samples, by the end of March 2016. In addition, participants were asked to report the OC content of 10 µl of a phthalic acid solution precisely prepared and traceable to primary measurements.

1.4 Thermal-optical analysis

The thermal protocol EUSAAR-2 [Cavalli et al., 2010] with a transmittance optical correction for pyrolysis has been recently selected as the European standard thermal protocol for the measurements of TC, OC and EC in PM samples (prEN16909).

However, in Europe, the NIOSH protocol, and its variations [Peterson and Richard, 2002] are used by a few laboratories. Because of differences in temperatures and duration between EUSAAR-2 and NISOH, these two protocols are known to give significantly different results, with EC/TC ratios from NIOSH typically lower than those from EUSAAR_2. Participants were asked to analyze the test samples with the protocol normally used to produce their OC, EC and TC data. In this exercise all laboratories applied the EUSAAR_2 protocol (Table 5) with transmittance-based correction.

Table 3: List of participants in the inter-laboratory comparison 2015, and contact persons

Code	Participant	Acronym	Contact Person
1	Stockholm University - Department of Applied Environmaental Science	ACES	Hans Areskoug
2	Norsk institutt for luftforskning	NILU	Karl Espen Yttri
3	Institute of Environmental Engineering of the Polish Academy of Sciences	IEE	Barbara Mathews
4	University of Crete - Environmental Chemical Processes Laboratory	NOA	Nikos Mihalopoulos
5	Czech Hydrometeorological Institute	CHMI	Adela Smejkalova
6	Leibniz-Institut für Troposphärenforschung	TROPOS	Gerald Spindler
7	Instituto de Diagnóstico Ambiental y Estudios del Agua	IDAEA	Andres Alastuey
8	Laboratoire de Glaciologie et Geophysique de l'Environnement	LGGE	Jean-Luc Jaffrezo
9	Umweltbundesamt - Deutschland	UBA	Elke Bieber
10	Joint Research Centre - Institute for Environment and Sustainability	JRC	Fabrizia Cavalli
11	The Cyprus Institute	Cyl	Jean Sciare
12	National Center for Scientific Research "Demokritos"	NCSR	Lila Diapouli
13	University of Helsinki	UHEL	Mikko Aijala

Table 4: List of the analytical protocol and punch size used by each participant

Code	Participant	Instrument	Protocol	Punch size (cm^2)	Charring cor.
1	ACES	Sunset	EUSAAR_2	1.50	Transmittance
2	NILU	Sunset	EUSAAR_2	1.50	Transmittance
3	IEE	Sunset	EUSAAR_2	1.50	Transmittance
4	NOA	Sunset	EUSAAR_2	n.a.	Transmittance
5	CHMI	Sunset	EUSAAR_2	1.50	Transmittance
6	TROPOS	Sunset	EUSAAR_2	1.50	Transmittance
7	IDAEA	Sunset	EUSAAR_2	1.50	Transmittance
8	LGGE	Sunset	EUSAAR_2	1.50	Transmittance
9	UBA	Sunset	EUSAAR_2	1.50	Transmittance
10	JRC	Sunset	EUSAAR_2	1.00	Transmittance
11	Cyl	Sunset	EUSAAR_2	1.50	Transmittance
12	NCSR	Sunset on-line	EUSAAR_2	2.14	Transmittance
13	UHEL	Sunset on-line	EUSAAR_2	0.92	Transmittance

Table 5: Details of the two analytical protocols implemented by participants

	EUSAAR_2	
Carrier gas	Time (s)	Temp. (°C)
Helium	120	200
Helium	150	300
Helium	180	450
Helium	180	650
Helium		
Oxygen in Helium (2%)	120	500
Oxygen in Helium	120	550
Oxygen in Helium	70	700
Oxygen in Helium	80/110	850
Oxygen in Helium		
Oxygen in Helium		

2 Data evaluation

Ambient PM filter samples: In absence of certified reference material for atmospheric TC, OC and EC deposited on filters, the *measurement method performance* (par. 2.1) and *laboratory performances* (par. 2.2) were evaluated using atmospheric PM collected on filters as test samples.

In this report we focus on the *TC loadings* (in µg cm^{-2}) and *EC/TC ratios* reported by each participants for each test sample. TC represents the most robust (and protocol-independent) output of TOA analyses, while EC/TC ratios are free from biases in the total carbon determination, and reflect possible differences in the OC/EC split determination among participants.

On average, reported TC loadings ranged from 5 to 21 µg cm^{-2}, corresponding to atmospheric concentrations ranging from 1 to 5 µg m^{-3} collected for 24h at a face velocity of 54 cm s^{-1}. EC/TC ranged on average from 0.11 to 0.19. All results for TC (in µg cm^{-2}) and EC/TC ratios are presented in Annex 1 in Tables 1 and 2, respectively.

Note: While tabulating results from all participants, we observed abnormal results - both in TC and EC/TC ratio- for the majority of ambient PM filter samples from participants 12 and 13, both using a semi-continuous OCEC analyser. These values were considered irreconcilable with the other data and were not included in the following evaluations. Similar discrepancies were already observed in the previous inter laboratory comparison exercise and may arise from the attempt to operate in an off-line mode the field analyser. The two participants were informed and are further investigating this issue also with the manufacturer.

Aqueous solution of phthalic acid: This solution was used to assess the uncertainty of the instrument calibration constant determination. Results were analysed in terms of percentage differences from the assigned value.

Assigned values:

As *ambient PM collected* on filters was used as test samples, the true values for *TC and EC/TC loadings* were not known. The assigned value and its standard uncertainty for TC and EC/TC on each filter were calculated as the robust average among all participants (outlier excluded) (see Par 2.2).

For the *phthalic acid solution,* the assigned OC concentration value was calculated from the water volume used to make the solution, the mass of phthalic acid dissolved in this water volume, and the chemical formula of phthalic acid. The assigned value was 1.52 gC l^{-1} (traceable to primary measurements) with an expanded combined relative uncertainty (k = 2) of 1.0%.

2.1 TEST FILTER SAMPLES - Method performance

2.1.1 Data evaluation description

The assessment of the *method performance* aims at deriving, from the results of the present exercise, the precisions of the measurement method in terms of repeatability and reproducibility standard deviations. For this, the consistency of the dataset is evaluated, at first graphically, by means of Mandel's *h* and *k* statistics [ISO5725-2], for possible outliers (i.e. observations greater than the critical value at the 99% confidence level) or stragglers (i.e. observations greater than the critical value at the 95% confidence level but less or equal to the critical value at the 99% confidence level).

The Mandel's *k* parameter estimates the within-laboratory consistency (repeatability). The critical values for *Mandel's k* indicators (i.e. outlier and straggler) vary upon the number of replicate measurements. In this comparison exercise, all laboratories provided three replicates for every sample (except lab 2 for sample 4642 and T6696, lab 9 for sample T6694, and lab 11 for sample 4638). Thus *Mandel's k* was calculated for an average case of three replicates and compared to the critical values 2.01 (outlier) and 1.69 (straggler).

The Mandel's *h* parameter describes the between-laboratory consistency (reproducibility) and has been calculated for every laboratory and every sample. For an inter-laboratory comparison among eleven participants, the critical values for Mandel's *h* are 2.22 (outlier) and 1.82 (straggler).

To confirm the identified outliers and stragglers, statistical G_1-Grubbs' and Cochran's tests are applied for testing the between-laboratory and within-laboratory variances, respectively [ISO5725-2].

Based on the outcomes of above statistical treatments (G_1-Grubbs' and Cochran's tests), outliers are discarded for the calculation of the mean value, the repeatability and reproducibility standard deviations. Subsequently, the dependence of precision (i.e. repeatability and reproducibility) upon the mean values is investigated [ISO5725-2].

2.1.2 Results: Method performance for TC

Within-laboratory consistency. In Figure 1, the *Mandel's k* statistic values for TC are presented grouped by laboratory (Panel a) and, separately, by sample (Panel b).

Two outliers (lab/sample: 1/DEM2; 4/4638) and eight stragglers (lab/sample: 3/4642; 4/T6694 and T6696; 7/DEM3 and A212; 8/4638; 10/T6694; 11/4642) were identified (Fig. 1a). Cochran's test confirmed one outlier (lab/sample: 1/DEM2).

It should be noted that participant 4 analysed for 4638, DEM2, T6694 and T6696 punches taken from different part of the original big filters. The poor repeatability obtained for samples 4638, T6694 and T6696 could therefore be more affected by deposit heterogeneities across the big filters.

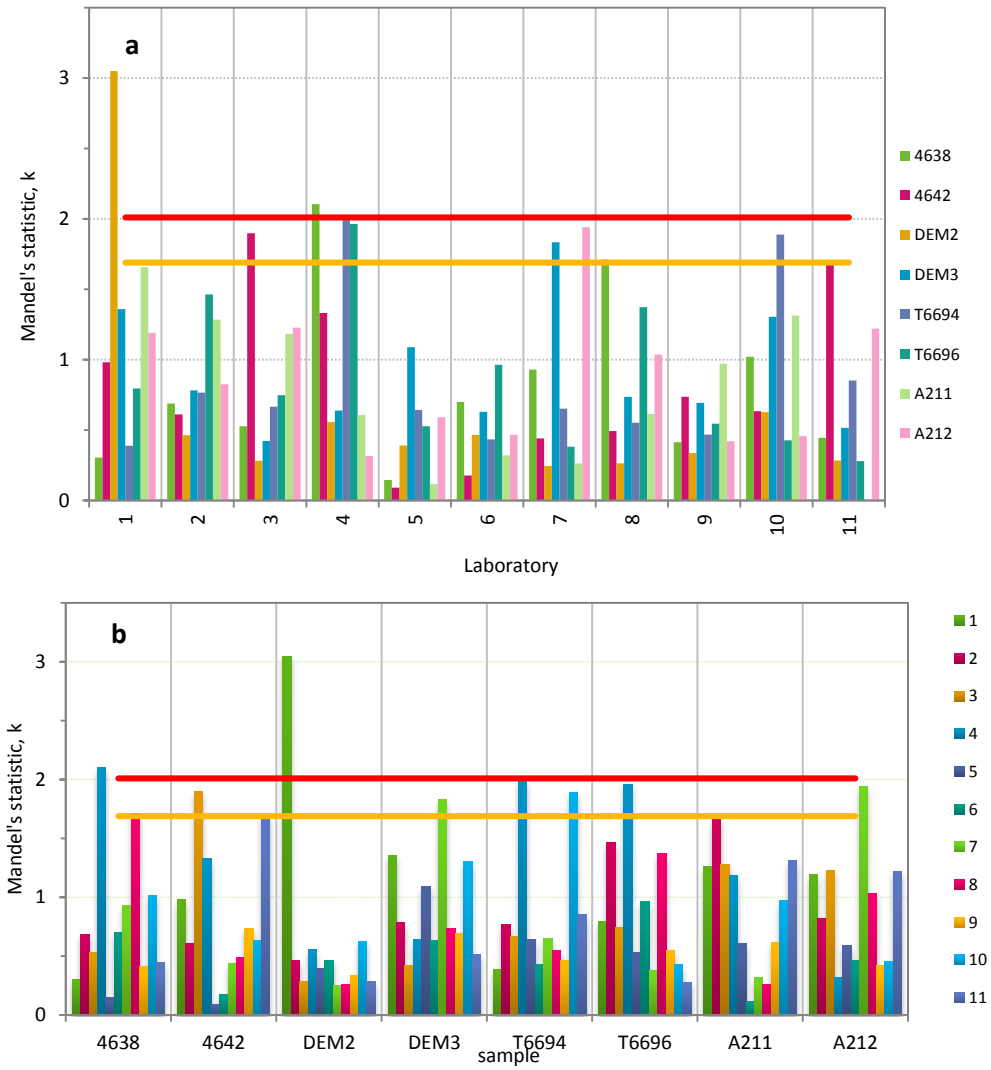

Figure 1. *Mandel's k* statistic values for within laboratory consistency on TC data, grouped by laboratory (panel a) and by sample (panel b). For eleven laboratories and three replicates, k values should be < 2.01 at the 1% significance level (red line) and < 1.69 at the 5% significance level (orange line).

Between-laboratory consistency. In Figure 2, the *Mandel's h* statistic values are presented grouped for each laboratory (Panel a) and, separately, for each sample (Panel b).
In the TC dataset, one outlier (lab/sample: 4/DEM2) and two stragglers (1/DEM3; 3/T6694) were identified (Fig. 2a). The Grubbs' test does not confirm any outlier.

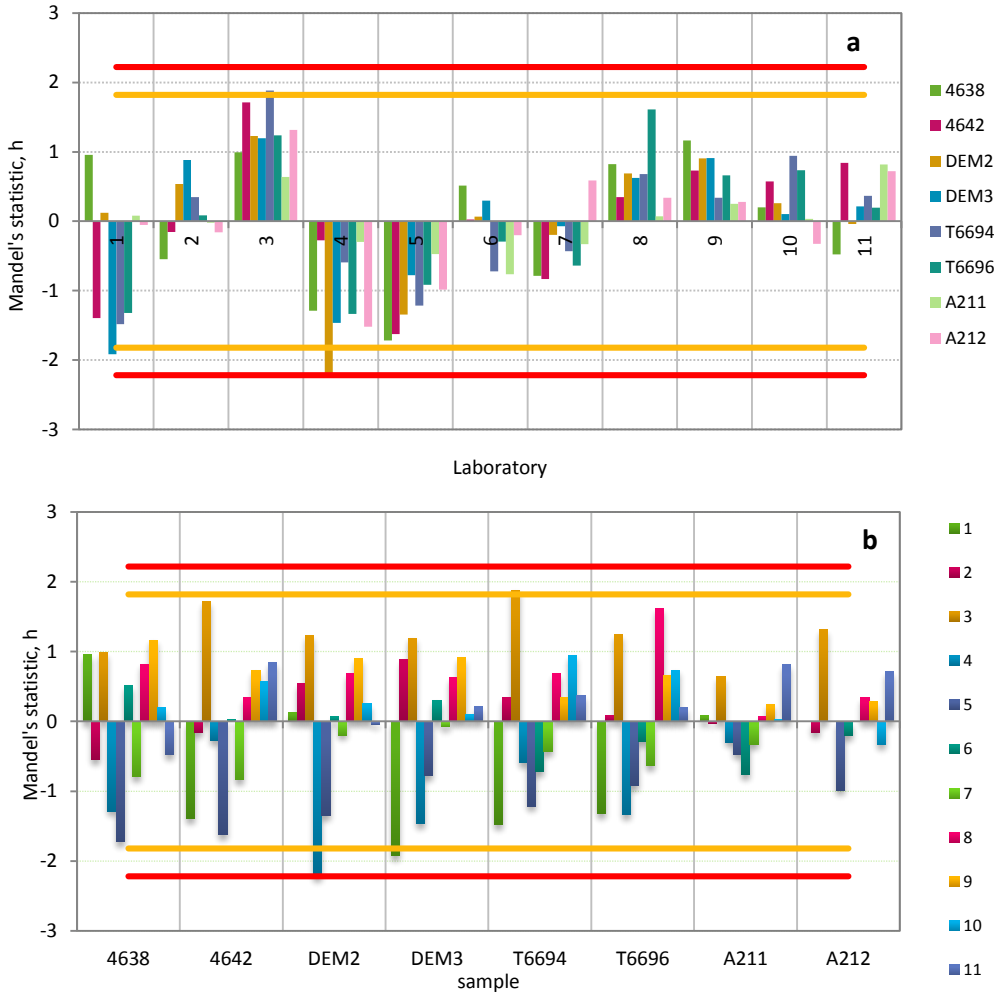

Figure 2. *Mandel's h* statistic values for between-laboratory consistency on TC data, grouped by laboratory (panel a) and by sample (panel b). For eleven laboratories, *h* values should be < 2.22 at the 1% significance level (red line) and < 1.82 at 5% significance level (orange line).

From the retained values and for each sample separately, the mean value, the repeatability (s_r) and reproducibility (s_R) standard deviations were calculated. The general means and values of s_r and s_R for the eight test filter samples are listed in Table 6. Both repeatability and reproducibility relative standard deviations show an inverse dependence on TC. Combining the repeatability and reproducibility relative standard deviation for the EUSAAR-2 protocol obtained during the 2014, 2015 and the present ILCE, we observe that the method precision (both sR and sr) for TC measurement becomes exponentially poorer toward lower TC contents i.e. 10 µgC / cm² (Fig. 3).

Table 6: General mean**,** repeatability (s_r) and reproducibility (s_R) standard and relative standard deviations for TC.

test sample	general mean	sr		sR	
	µgC / cm²	µgC / cm²	%	µgC / cm²	%
4638	21.5	0.6	2.8	1.4	6.4
4642	7.6	0.4	5.2	0.7	9.3
DEM2	13.5	0.3	1.9	0.9	7.0
DEM3	9.8	0.2	2.0	0.9	9.5
T6694	11.1	0.3	2.5	0.9	7.9
T6696	5.8	0.3	4.6	0.5	8.9
A211	4.9	0.3	5.3	0.5	10.4
A212	6.1	0.2	3.8	0.5	7.7

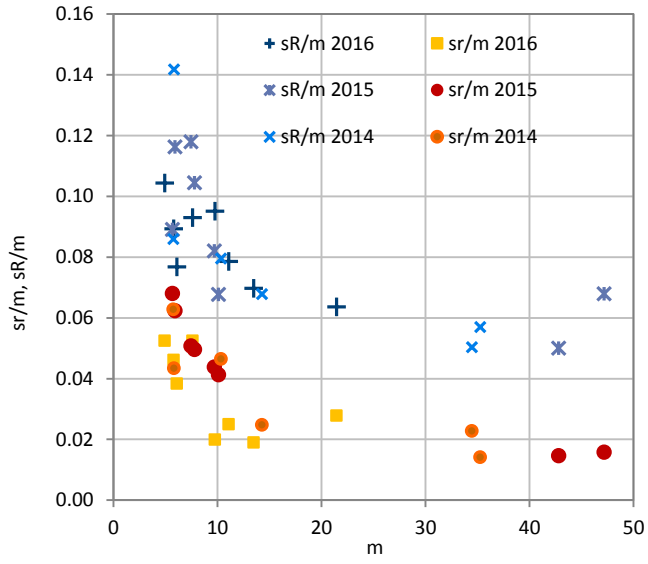

Figure 3. Repeatability and reproducibility relative standard deviation for the EUSAAR-2 protocol for TC measurement obtained during the 2014, 2015 and the present inter-laboratory comparison.

2.1.3 Results: Method performance for EC/TC

Within-laboratory consistency. In Figure 4 the *Mandel's k* statistic values are presented grouped for each laboratory (Panel a) and, separately, for each sample (Panel b). In the EC/TC dataset, four outliers (lab/sample: 1/4642, T6694 and T6696; 8/4638) and two stragglers (lab/sample: 1/DEM3; 4/4642) were identified (Fig. 4a). Cochran's test confirmed the entries 1/4642 and T6696 and 8/4638 to be outliers and 1/T6694 as stragglers.

Localized sample heterogeneities or contaminations cannot rigorously be excluded, but the occurrence of several stragglers and or outliers from a single laboratory (case of lab 1) most probably suggests unsatisfactory laboratory repeatability for the determination of the EC/TC ratio as compared to the other laboratories.

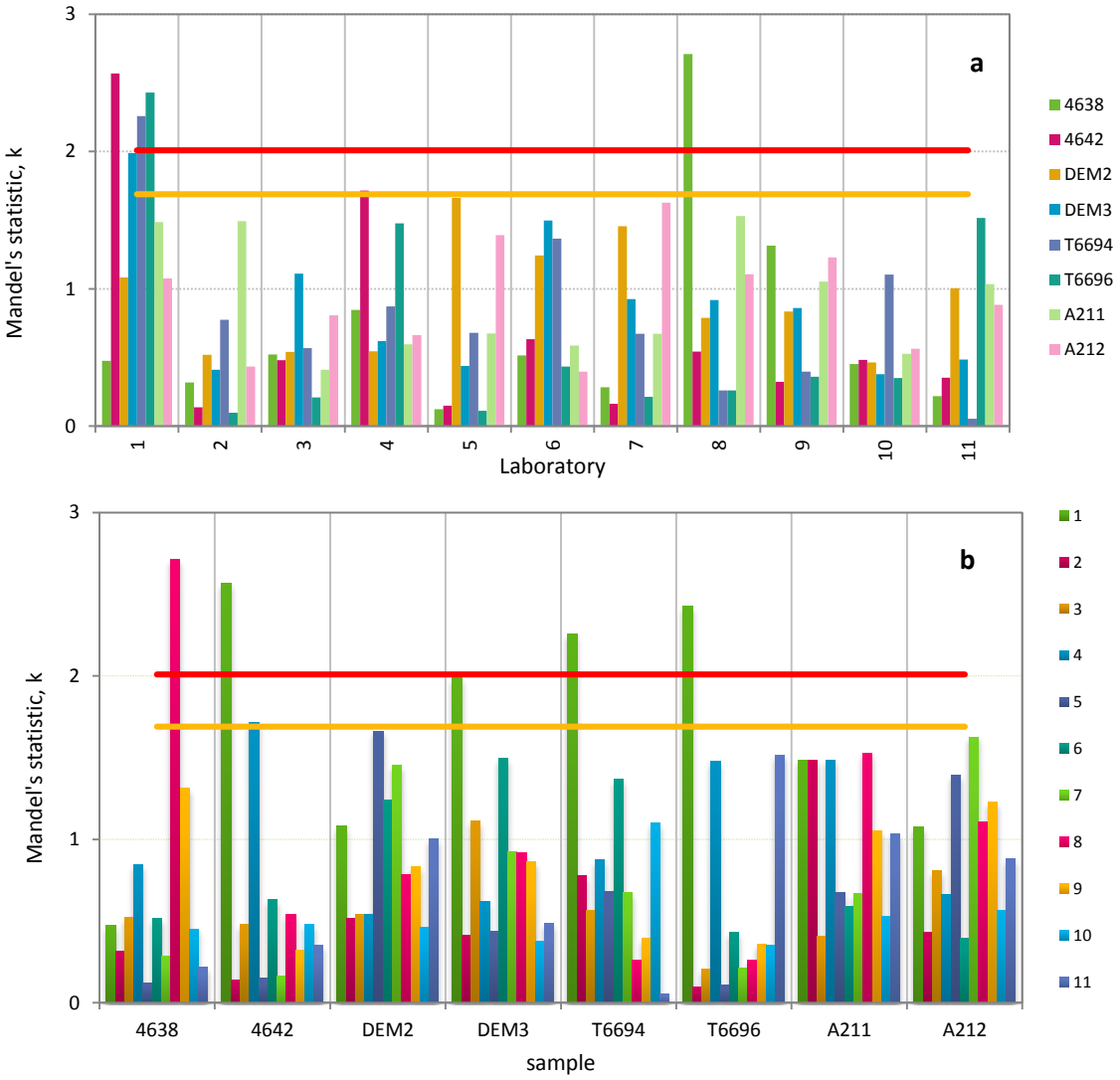

Figure 4. *Mandel's k* statistic values for within laboratory consistency on EC/TC ratio obtained from the entire database, grouped by laboratory (panel a) and by sample (panel b). For eleven laboratories k and three replicates values should be < 2.01 at 1% significance level (red line) and < 1.69 at 5% significance level (orange line).

Between-laboratory consistency. Figure 5 shows the *Mandel's h* statistic values for EC/TC ratio calculated on the entire database for each laboratory (Panel a) and, separately, for each sample (Panel b).

Four outliers (lab/sample: 1/DEM2; 6/DEM3; 11/4638 and A211) and three straggler (1/4642, T6694; 11/A212) were identified. Grubbs' test identifies only 6/DEM3 as outliers.

Participants 1 and 11 reported each three anomalous values. Unlike participant 11, participant 1 also produced four anomalous values in terms of repeatability. Although localized sample heterogeneities or contaminations cannot be rigorously excluded, the recurrence of stragglers and outliers from a single laboratory most probably suggests unsatisfactory reproducibility for EC/TC ratio determination as compared to the other participants.

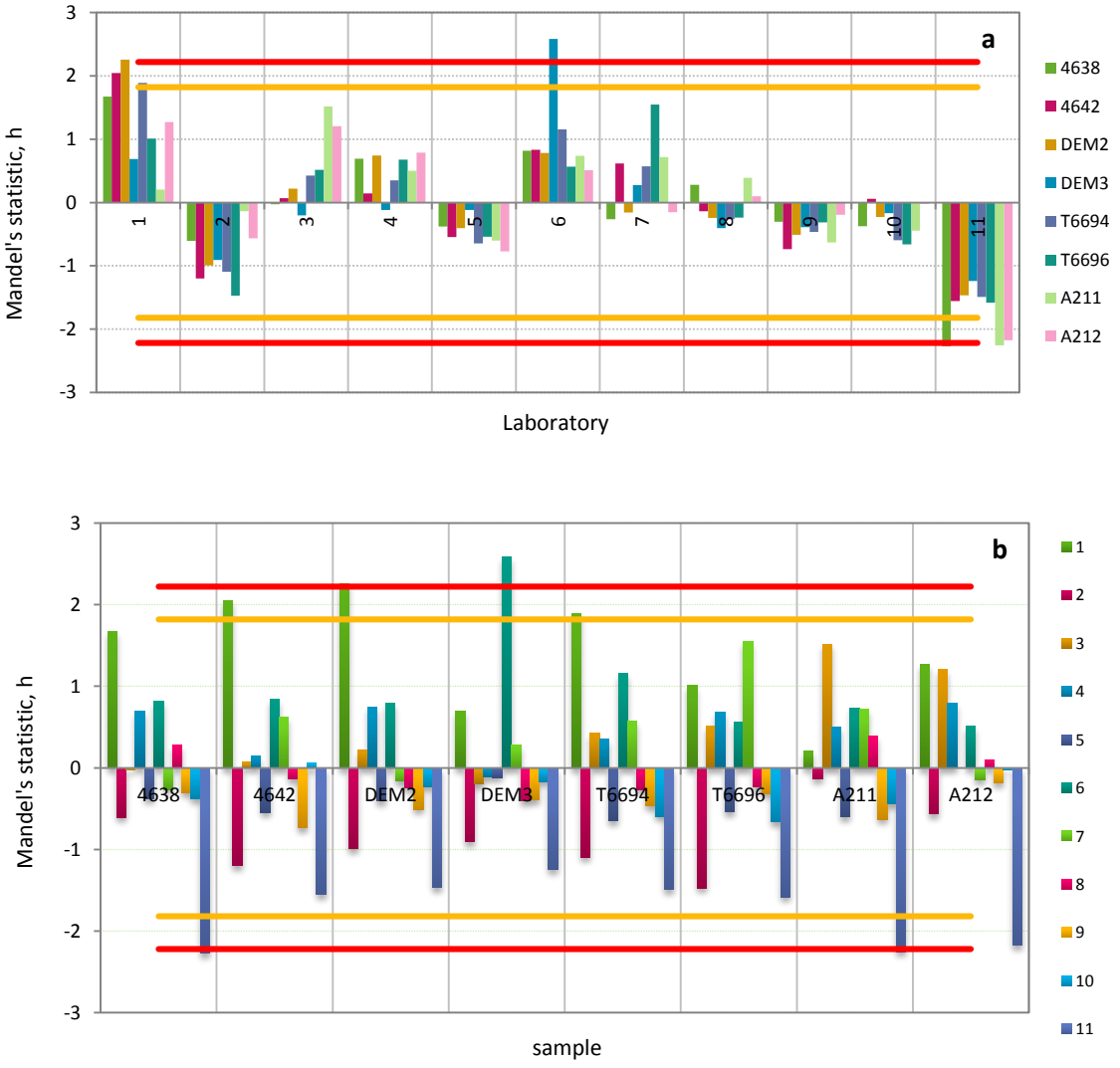

Figure 5. *Mandel's h* statistic values for between laboratory consistency on EC/TC ratio obtained from the entire database, grouped by laboratory (panel a) and by sample (panel b). For eleven laboratories h values should be < 2.22 at 1% significance level (red line) and < 1.82 at 5% significance level (orange line).

The entries identified as outliers by the statistical tests were discarded from the dataset, and the mean value, the repeatability (s_r) and the reproducibility (s_R) standard deviations for EC/TC were calculated for each sample from the retained values (Table 7). Perhaps due to the small variability in EC/TC ratios among the eight test samples, no clear dependence of the EC/TC ratio precision upon the EC/TC ratio values can be identified. Combining the repeatability and reproducibility relative standard deviation for the EUSAAR-2 protocol obtained during the 2014, 2015 and the ILCE, we observe that the method precision (both sR and sr) for EC/TC ratio measurement becomes exponentially poorer toward lower EC/TC ratio. i.e. 0.05 (Fig. 6).

Table 7: General mean, repeatability (s_r) and reproducibility (s_R) standard and relative standard deviations for EC/TC.

test sample	general mean	sr		sR	
	ratio	ratio	%	ratio	%
4638	0.19	0.01	4.2	0.02	11.9
4642	0.16	0.01	4.7	0.02	11.3
DEM2	0.15	0.01	5.2	0.02	15.8
DEM3	0.10	0.00	4.9	0.02	16.6
T6694	0.15	0.01	3.8	0.03	17.2
T6696	0.12	0.01	6.1	0.02	19.2
A211	0.13	0.01	7.0	0.03	24.4
A212	0.12	0.01	6.0	0.02	19.3

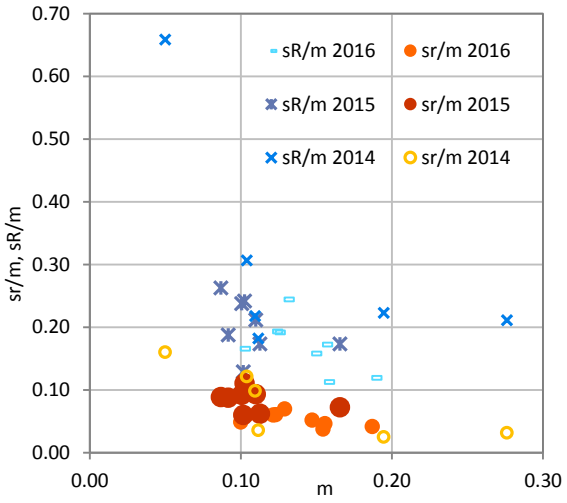

Figure 6. Repeatability and reproducibility relative standard deviation for the EUSAAR-2 protocol for EC/TC measurement obtained during the 2014, 2015 and the present inter-laboratory comparison.

2.2 FILTER TEST SAMPLES - Laboratory performance

2.2.1 Data evaluation description

The assessment of the *laboratory performance* aims at describing the laboratory bias compared to the assigned value associated with its standard deviation. Each participant's performance is determined in terms of *z-scores,* a measure of the deviation from the assigned value. To calculate *z-scores*, an assigned value and its standard deviation have to be determined for each test sample.

- *Determining the assigned value*: Among the available methods for determining the assigned value, the approach of the *consensus value from participants to a round of a proficiency testing scheme* was chosen, in absence of a reference or certified reference material. With this approach, the assigned value X for each test sample used in the ILCE is the robust average calculated, with a recursive algorithm, from the results reported by all participant (outlier excluded) (See ISO 13528:2005(E), Annex C).

- *Determining the standard deviation for proficiency assessment*: Among the available methods for determining the standard deviation for proficiency assessment (σ*), the approach of calculating σ* *from data obtained in a round of a proficiency testing scheme* was chosen. With this approach, σ* is the robust standard deviation calculated, with a recursive algorithm, from the results reported by all participant (outlier excluded) testing (See ISO 13528:2005(E), Annex C.

These approaches might become statically ineffective [ISO 13528:2015 (E)], for example, if the number of participant is lower than twenty. To verify their reliability the robust mean and its standard deviation were also calculated applying the Q/Hampel method (Ref.?). The obtained values do not significantly differ from those obtained by the *consensus value from participant results*, in Table 8, which are then used for the following elaboration.

For each laboratory and test sample, the *z-score* was calculated as:

$$z = (x_i - X)/ \sigma^*$$

where x_i is the result from the participant i; X is the assigned value for the sample; and σ^* is the standard deviation for proficiency assessment.

When a participant reports an entry that produces a bias greater than +3 z or less than -3 z (i.e. deviating from the assigned value for more than 3 standard deviations), this entry is considered to give an "action signal". Likewise, a laboratory bias above +2 z or below -2 z (i.e. deviating from the assigned value for more than 2 but less than 3 standard deviations) is considered to give a "warning signal". A laboratory bias between -2 z and +2 z indicates a satisfactory laboratory performance with respect to the standard deviation for proficiency assessment.

2.2.2 Results: Laboratory performance for TC

The assigned values X and the related standard deviations for proficiency assessment σ* calculated from the entire database (outliers excluded) for each sample, are reported in Table

8. Following ISO13528, σ* were calculated *from data obtained in a round of a proficiency testing scheme*.

Table 8: Assigned values and standard deviations for proficiency assessment σ**from data obtained in a round of a proficiency testing scheme* for TC.

		4638	4642	DEM2	DEM3	T6694	T6696	A211	A212
assigned value	µg/cm2	21.7	7.6	13.6	9.8	11.1	5.8	4.9	6.1
standard	µg/cm2	1.4	0.7	0.7	0.9	0.9	0.5	0.5	0.4
deviation	%	6.3	9.0	4.9	9.1	7.9	9.1	10.3	6.1
2σ*	%	13	18	10	18	16	18	21	12
3σ*	%	19	27	15	27	24	27	31	18

Figure 7 shows z-scores calculated from σ*. One outlier (lab/sample; 4/DEM2) and two stragglers (lab/sample: 1/DEM3; 4/A212) can be identified.
For each sample, seven to nine out of eleven participants (showed deviations from the assigned values within +/- 1 σ* as listed in Table 8 (i.e. within 1 z-score).
80% of all entries is within 10% from the assigned value and 95% within 15%.
Participants 3, 4, 5, 8 and 9 show the systematic (i.e. for all test samples) tendency of underestimating or overestimating the assigned TC concentrations. A more accurate determination of the calibration constant (e.g. implementing where possible CO_2 calibration) would correct this tendency.

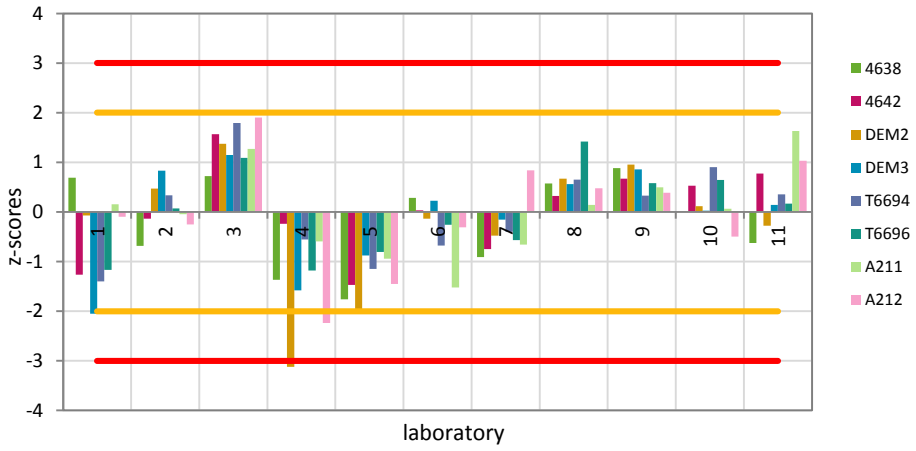

Figure 7. z-scores for TC calculated using σ* *from data obtained in a round of a proficiency testing scheme*.

2.2.3 Results: Laboratory performance for EC/TC

The assigned values, X, and the related standard deviations for proficiency assessment, σ*, are reported in Table 9. Following ISO13528, σ* were calculated *from data obtained in a round of a proficiency testing scheme* and corresponding z-scores are shown in Figure 8.

Table 9: Assigned values and standard deviations for proficiency assessment σ**from data obtained in a round of a proficiency testing scheme* for EC/TC.

		4638	4642	DEM2	DEM3	T6694	T6696	A211	A212
assigned value	ratio	0.18	0.16	0.14	0.10	0.15	0.12	0.13	0.12
standard	ratio	0.01	0.02	0.02	0.01	0.03	0.03	0.02	0.02
deviation	%	6	12	12	11	17	22	19	17
2σ*	%	12	23	24	22	34	43	38	33
3σ*	%	18	35	36	33	51	65	57	50

Four outliers (lab/sample: 1/4638 and DEM2; 6/DEM3; 11/4638) and six stragglers (1/4642 and DEM3; 6/4638; 11/DEM3, A211, A212) can be identified. For each sample, six to nine out of eleven laboratories showed deviations from the assigned values within +/- 1 σ* as listed in Table 9 (i.e. within 1 z-score).

Only 50% of all entries is within 10% of the assigned value and 85% is within the 25% of the assigned value.

A contribution of filter heterogeneities to poor laboratory performances cannot be completely excluded. However, all outliers and stragglers were produced by participants 1, 6 and 11. The recurrence (more than two) of stragglers or outliers for single laboratories as observed in this exercise most probably suggest biases in EC/TC determination compared to the other laboratories. Participants (1 and 11) showing large biases (|z-scores|> 2) shall carefully examine their procedures and identify appropriate corrective actions that are likely to prevent the recurrence of such results in the future.

The majority of participants, i.e. lab 1, 2, 3, 4, 6, 9 and 11 show the systematic (i.e. for all test samples) tendency of underestimating or overestimating the assigned EC/TC ratio. A more solid and stable in time instrument set-up in terms of i) laser stability; ii) FID response in He and He/O2 phases; iii) temperature calibration and iv) transit time would correct this behavior and reduce the observed variability in EC/TC ratio determination.

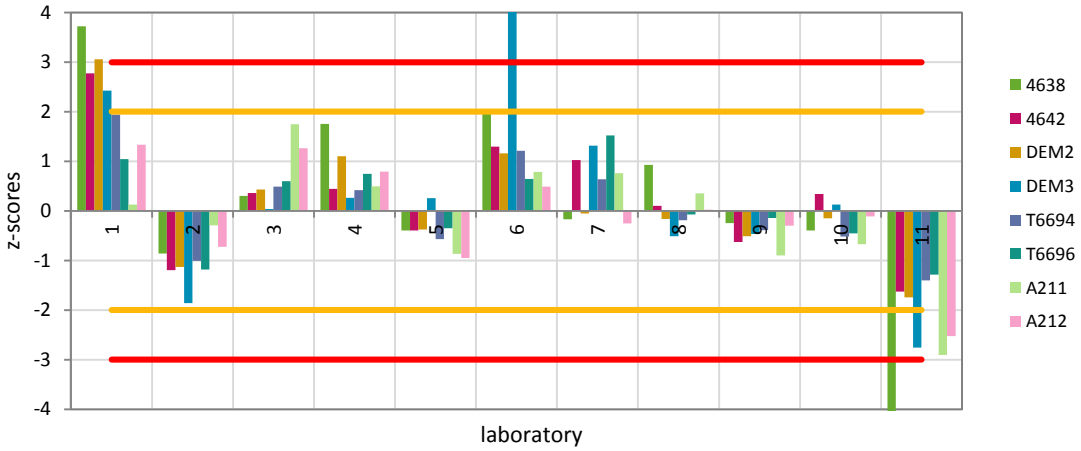

Figure 8. z-scores for EC/TC ratio calculated using σ* *from data obtained in a round of a proficiency testing scheme*.

2.3 PHTHALIC ACID SOLUTION – Percentage differences

Participants were asked to report the OC content of 10 µl phthalic acid solution. This included the analysis of samples prepared by spiking a pre-cleaned filter punch with 10 µl solution. This is the procedure normally used by laboratories to determine and verify the FID calibration constant.

Figure 9 shows the percentage differences from the assigned value (1.52 ± 0.02 gC l^{-1}, calculated from primary mass and water volume measurements) for each participant. The observed percentage differences range -11% to +13%, while six out of eleven laboratories reported OC deviating from the assigned value by less than ±5%. Since each phthalic acid solution flask was not checked individually, deviations from the assigned value of the standard solutions cannot be completely excluded.

This exercise did not aim at identifying systematic tendency of a laboratory to underestimate or overestimate the C content of analysed samples but rather to highlight the potential uncertainty (and variability) that can affect TC determination, when the spiking procedure is applied to determine the FID calibration constant.

It is recommended to implement the calibration with CO_2 injections where possible, or to carefully revise the accuracy of all steps involved in the external solution spiking procedure (calibration of the pipette volume, complete deposition of the volume onto a punch filter, drying etc.).

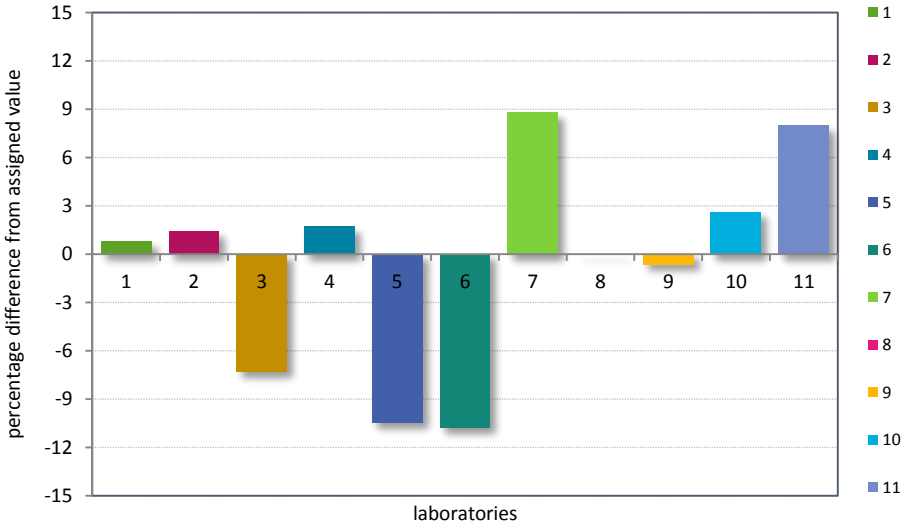

Figure 9. Phthalic acid solution –percentage differences from the assigned value, i.e. the C concentration of the test solution calculated from the mass of phthalic acid and the volume of ultra-pure water used to make the solution.

Conclusions

This inter-laboratory comparison involved thirteen participants applying all thermal-optical analysis with the EUSAAR_2 protocol. Abnormal results from two participants, both using a semi-continuous OCEC analyses, could not be included in the statistical evaluations of this ILCE.
Assigned values for TC loadings and EC/TC ratios in the filter test samples were calculated as averages among all participants, after excluding outliers.
The measurement method ***repeatability and reproducibility for TC*** ranged from 2% to 5% and from 6% to 10% (as one relative standard deviation), respectively. For the ***EC/TC ratio, repeatability and reproducibility*** ranged from 4% to 7 and from 11% to 24% (as one relative standard deviation), respectively. Combining the repeatability and reproducibility relative standard deviation for the EUSAAR-2 protocol obtained during the 2014, 2015 and the present ILCE, we observed that the method precision (both sR and sr) becomes exponentially poorer toward lower TC contents i.e. 10 µgC / cm² and EC/TC ratio. i.e. 0.05.
Stragglers and outliers in terms repeatability of the determination of EC/TC ratio were produced mainly by a single participant. Although the contribution of localized sample heterogeneities and /or contaminations to biased data cannot be totally excluded, the random scheme adopted to distribute sub-samples was such that the recurrence of stragglers or outliers for single laboratories most probably indicates an unsatisfactory laboratory performance as compared to the other participants.
Laboratory performances were assessed for both TC loading and EC/TC ratio determinations based on z-scores, applying as *standard deviation for proficiency assessment* (σ*) the one calculated *from data obtained in a round of a proficiency testing scheme.*
For TC loadings, only one outlier and two stragglers were identified; and 80% of all entries is within 10% from the assigned TC concentration value and 95% within 15%.
Participants 3, 4, 5, 8 and 9 show the systematic (i.e. for all test samples) tendency of underestimating or overestimating the assigned TC concentrations. A more accurate determination of the calibration constant (e.g. implementing where possible CO_2 calibration) would probably correct this tendency.
Regarding EC/TC ratios, all ten outliers and stragglers were produced by three participants (1, 6 and 11). Only 50% of all entries is within 10% of the assigned value and 85% is within the 25% of the assigned EC/TC ratio.
The majority of participants, i.e. lab 1, 2, 3, 4, 6, 9 and 11 show the systematic (i.e. for all test samples) tendency of underestimating or overestimating the assigned EC/TC ratio. A more solid and stable in time instrument set-up in terms of i) laser stability; ii) FID response in He and He/O2 phases; iii) temperature calibration and iv) transit time would correct this behavior and reduce the observed variability in EC/TC ratio determination.
The participants showing important systematic or random biases shall carefully examine their procedures and instrument set-up, and identify corrective actions (seeking for advice from ECAC if needed) which could prevent the recurrence of such results in the future.

References

EU Directive 2008/50/EC on ambient air and cleaner air for Europe.

FprCEN/TR 16243. Ambient air quality - Guide for the measurement of elemental carbon (EC) and organic carbon (OC) deposited on filters. CEN, Brussels, 2011.

Error! Reference source not found., **Error! Reference source not found.**. CEN, Brussels, 2016.

Peterson, M.R. and Richards, M.H. (2002), Thermal-optical-transmittance analysis for organic, elemental, carbonate, total carbon, and OCX2 in PM2.5 by the EPA/NIOSH method in: Proceedings, Symposium on Air Quality Measurement Methods and Technology-2002, edited by Winegar, E. D. and Tropp, R. J.Air & Waste Management Association, Pittsburgh, PA.

Cavalli F., Putaud J., Viana M., Yttri K., Genberg J., Toward a Standardised Thermal-Optical Protocol for Measuring Atmospheric Organic and Elemental Carbon: The EUSAAR Protocol. Atmospheric Measurement Techniques 3 (1); p. 79-89, 2010.

ISO 13528. Statistical methods for use in proficiency testing by inter-laboratory comparisons. ISO, Geneva, 2005.

ISO 13528. Statistical methods for use in proficiency testing by inter-laboratory comparisons. ISO, Geneva, 2015.

ISO 5725-2. Accuracy (trueness and precision) of measurement methods and results -- Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. ISO, Geneva, 1994.

Annex 1. Numerical results reported by participants

Table 1: Total carbon loadings (µg/cm^2)

Laboratory	4638	4642	DEM2	DEM3	T6694	T6696	A211	A212
	22.477	7.140	14.240	7.767	9.845	5.027	5.355	5.763
1	22.681	6.688	15.101	7.954	9.748	5.092	4.700	6.210
	22.834	6.371	11.471	8.290	9.959	5.435	5.005	6.276
	21.200	7.330	13.800	10.700	11.300	6.120	5.410	5.910
2	20.400	7.670	14.300	10.400	11.200	5.550	4.720	6.250
	20.700		13.800	10.600	11.600		4.620	5.920
	22.636	9.468	14.766	10.757	12.654	6.590	5.626	6.872
3	23.048	8.001	14.426	10.909	12.453	6.182	5.234	6.520
	22.440	8.505	14.517	10.890	12.815	6.336	5.896	7.091
	18.583	8.022	11.329	8.280	10.643	5.190	4.284	5.374
4	21.054	7.228	11.387	8.463	10.207	4.631	4.820	5.254
	19.855	7.034	11.958	8.518	10.913	5.712	4.813	5.238
	19.344	6.562	12.545	8.976	9.904	5.369	4.524	5.567
5	19.336	6.631	12.063	8.882	10.058	5.228	4.579	5.447
	19.191	6.579	12.374	9.288	10.253	5.519	4.283	5.724
	21.926	7.540	13.236	9.888	10.408	5.571	4.189	6.065
6	21.813	7.679	13.799	10.107	10.416	5.450	4.179	5.878
	22.576	7.617	13.641	10.094	10.616	5.958	4.132	6.070
	20.305	7.268	13.497	9.624	10.907	5.603	4.654	5.927
7	20.011	7.054	13.195	9.372	10.702	5.393	4.509	6.558
	21.068	6.925	13.295	10.078	10.554	5.505	4.653	6.811
	21.750	8.011	13.943	10.309	11.660	6.452	5.002	6.371
8	22.099	7.798	14.086	10.483	11.784	6.955	5.086	6.026
	23.639	7.624	14.272	10.198	11.486	6.214	4.949	6.495
	22.988	7.832	14.526	10.447	11.271	6.086	5.239	6.220
9	23.130	7.938	14.129	10.618	11.451	6.259	5.326	6.197
	22.656	8.379	14.212	10.714		5.959	5.017	6.378
	22.197	7.669	13.983	10.025	11.909	6.043	5.261	5.838
10	21.041	8.143	13.277	9.562	12.352	6.268	4.862	6.050
	21.894	8.042	13.920	9.976	11.329	6.094	4.795	5.918
		7.643	13.260	10.067	11.413	5.825	5.437	6.381
11	20.668	8.877	13.569	9.872	11.602	5.973	5.754	6.299
	21.039	7.835	13.566	9.927	11.141	5.861	6.118	6.831
12	23.616	11.738	19.154	16.104	18.198	13.176	12.222	5.850
	23.450	12.130	18.575	17.854	18.362	11.454	13.544	5.940
13	67.989	67.011	36.011	60.946	66.500	42.598	53.174	35.891
	42.663	46.739	48.022	64.000	43.935	19.717	41.457	47.141

Table 2: Elemental carbon / total carbon (ratios)

Laboratory	4638	4642	DEM2	DEM3	T6694	T6696	A211	A212
	4.900	1.552	2.803	0.905	2.061	0.768	0.642	0.917
1	5.145	1.503	2.873	1.041	1.836	0.876	0.669	0.929
	5.238	1.109	2.372	1.133	2.110	0.668	0.719	0.901
	3.610	0.990	1.670	0.880	1.400	0.560	0.600	0.650
2	3.560	1.020	1.790	0.850	1.470	0.500	0.650	0.690
	3.690		1.780	0.830	1.430		0.580	0.620
	4.102	1.492	2.282	1.156	2.052	0.907	0.991	1.028
3	4.287	1.285	2.127	1.046	2.056	0.864	0.937	1.011
	4.357	1.425	2.245	1.107	2.161	0.858	1.013	1.016
	3.960	1.144	1.826	0.876	1.800	0.786	0.593	0.778
4	4.041	1.240	1.922	0.900	1.635	0.689	0.706	0.730
	4.055	1.241	1.937	0.856	1.755	0.709	0.715	0.708
	3.503	0.964	1.549	0.947	1.386	0.604	0.494	0.558
5	3.464	0.995	1.744	0.901	1.408	0.596	0.481	0.627
	3.417	0.976	1.813	0.978	1.366	0.617	0.501	0.548
	4.407	1.302	2.040	1.198	1.823	0.800	0.660	0.798
6	4.434	1.364	2.325	1.273	1.956	0.751	0.616	0.803
	4.813	1.417	2.349	1.345	2.011	0.804	0.620	0.799
	3.631	1.250	1.964	1.142	1.810	0.904	0.673	0.780
7	3.718	1.235	1.740	1.033	1.856	0.886	0.706	0.738
	3.817	1.211	2.048	1.194	1.774	0.881	0.707	0.748
	3.667	1.223	1.923	1.041	1.701	0.764	0.626	0.754
8	3.992	1.214	1.955	0.989	1.752	0.823	0.764	0.722
	5.501	1.249	2.124	0.935	1.703	0.764	0.729	0.864
	3.909	1.146	1.869	1.004	1.587	0.734	0.587	0.709
9	4.004	1.159	1.992	0.973	1.649	0.713	0.528	0.682
	4.533	1.175	1.960	1.078		0.714	0.590	0.811
	3.838	1.285	1.971	1.01	1.622	0.661	0.628	0.679
10	3.838	1.288	1.941	1.001	1.663	0.713	0.566	0.742
	4.009	1.281	1.940	1.024	1.657	0.650	0.53	0.73
		0.938	1.634	0.686	1.326	0.428	0.305	0.410
11	2.804	1.115	1.507	0.704	1.341	0.623	0.301	0.459
	2.936	1.020	1.478	0.726	1.292	0.509	0.428	0.525
12	2.919	1.235	2.010	1.000	1.733	1.056	1.602	0.698
	3.028	1.226	1.547	1.023	1.283	0.712	1.869	0.792
13	1.978	2.033	3.772	2.489	3.196	1.663	1.554	1.772
	0.196	0.120	1.880	1.663	2.674	0.283	0.120	0.250