Quantifying the uncertainties in thermal-optical analysis of carbonaceous aircraft engine emissions: An interlaboratory study

Timothy A. Sipkens[1], Joel C. Corbin[1], Brett Smith[1], Stéphanie Gagné[1], Prem Lobo[1,#], Benjamin T. Brem[2,3], Mark Johnson[4], Gregory J. Smallwood[1]

[1]Metrology Research Centre, National Research Council Canada, Canada
[2]Empa, Swiss Federal Laboratories for Materials Science and Technology, Switzerland
[3]Laboratory of Atmospheric Chemistry, Paul Scherrer Institut, Switzerland
[4]Rolls-Royce plc, Derby, United Kingdom
[#]Current address: Office of Environment and Energy, Federal Aviation Administration, Washington, D.C. 20591, USA

Correspondence to: Timothy A. Sipkens (timothy.sipkens@nrc-cnrc.gc.ca)

Abstract. Carbonaceous particles, such as soot, make up a notable fraction of atmospheric particulate matter and contribute substantially to anthropogenic climate forcing, air pollution, and human health. Thermal-optical analysis (TOA) is one of the most widespread methods used to speciate carbonaceous particles and divides total carbon (TC) into the operationally defined quantities of organic carbon (OC; carbon evolved during slow heating in an inert atmosphere) and elemental carbon (EC). While multiple studies have identified fundamental scientific reasons for uncertainty in distinguishing OC and EC, far fewer studies have reported on between-laboratory reproducibility. Moreover, existing reproducibility studies have focused on complex atmospheric samples. The real-time instruments used for regulatory measurements of the mass concentration of aircraft engine non-volatile particulate matter (nvPM) emissions are required to be calibrated to the mass of EC, as determined by TOA of the filter-sampled emissions of a diffusion flame combustion aerosol source (DFCAS). However, significant differences have been observed in the calibration factor for the same instrument based on EC content determined by different calibration laboratories. Here, we report on the reproducibility of TC, EC, and OC quantified using the same TOA protocol, instrument model (Sunset Laboratory Model 5L), and software settings (auto split-point: Calc405) across five different laboratories and instrument operators. Six unique data sets were obtained, with one laboratory operating two instruments. All samples were collected downstream of an aircraft engine after treatment with a catalytic stripper to remove volatile organics. Between-laboratory contributions made up a majority of the within-filter uncertainties for EC and TC, even for these relatively well-controlled samples. Overall, expanded (k = 2) uncertainties due to measurement reproducibility correspond to 17 %, 8.0 %, and 12 % of the nominal values for EC, OC, and TC, respectively, and 6.8 % in the EC/TC ratio. These values are lower than previous studies, including atmospheric samples without volatile organic removal, and therefore likely represent lower limits for the uncertainties of the TOA method.

1 Introduction

Carbonaceous particles contribute to both natural and anthropogenic climate forcing, air pollution, and human health impacts. The aviation industry remains a notable source of these particles, and air transportation continues to expand. Unlike CO_2, particulate matter (PM) emissions from the aviation industry contain larger uncertainties, as is their effect on contrails and cloud formation (Righi et al., 2021). For aircraft engine emissions, thermal-optical analysis (TOA) is currently the reference standard for calibrating the instruments used to measure the mass concentration of non-volatile particulate matter (nvPM) emitted by aircraft engines (SAE, 2018; Lobo et al., 2015b; Lobo et al., 2020). However, questions remain open regarding the uncertainties and associated metrology (referring to the establishment of uncertainties by way of interlaboratory comparisons and the traceability) of these measurements.

In TOA, the total carbon (TC) mass collected on a quartz filter is measured in two distinct phases. First, the total carbon mass evolved from a sample during controlled heating in an inert environment is considered organic carbon (OC), while the remainder, heated in an oxidizing environment, is considered EC, after correction for pyrolysis (Birch and Cary, 1996). If the mass fraction of carbon in OC (40–80%; (Turpin and Lim, 2001; Bae et al., 2006)) or in EC (90–98%; (Figueiredo et al., 1999; Singh and Vander Wal, 2020; Corbin et al., 2020)) is known, these quantities can then be used to estimate the total mass of carbonaceous particles on the filter.

It is well known that the widely variable properties of carbonaceous materials leads to significant uncertainties in the separation of TC into OC and EC using TOA (Watson et al., 2005; Lack et al., 2014). In particular, inorganic carbonates may generate spurious signals; soot may partly vaporize at the OC stage; materials such as tarballs or highly-oxidized organics may generate EC signals; and inorganic compounds may catalyze the formation of EC or confound the optical quantification of pyrolysis (Corbin et al., 2020). It is also well known that different temperature ramp protocols lead to differences in the OC/EC ratio reported by TOA (e.g. Bautista et al., 2015; Schauer et al., 2003; Cavalli et al., 2010; Brown et al., 2017; Cheng et al., 2010; Giannoni et al., 2016; Wu et al., 2016; Cheng et al., 2012).

Less well studied are the uncertainties in TOA across multiple laboratories. Interlaboratory studies allow for an assessment of measurement reproducibility (changing laboratories, instruments, and operators), rather than simply repeatability (e.g., replicate measurements performed by the same operator). Here, the few reproducibility studies that exist have often focused on atmospheric aerosols or surrogates thereof. Schmid et al. (2001) analyzed urban air pollution samples from Berlin, Germany, using 9 different protocols obtained in 17 different laboratories. They reported relative standard deviations for between-laboratory uncertainty on their TC measurements of 6.7–11 %, with between-laboratory contributions making up 87–96 % of the overall variance. Schauer et al. (2003) evaluated EC and OC reproducibility for filter samples of Asian and North American air pollution, as well as secondary organic aerosol, reporting between-laboratory standard deviations of 12–22 % for EC and 3.6–12 % for OC. They additionally evaluated the reproducibility of the EC/OC division (split point) for various other samples, focusing on this ratio after identifying it as a major source of uncertainty. Ten Brink et al. (2004) sampled rural air pollution in Germany and analyzed the filters in four different laboratories, reporting less than 10% variability in TC and EC. In a pan-

European study, Panteliadis et al. (2015) gathered results from 17 different laboratories to determine a reproducibility standard error of 20–26 % for EC and 12–15 % for TC. Finally, Brown et al. (2017) reported a combined standard error of < 13 % for a reproducibility study between four laboratories. The known technical shortcomings of TOA instruments cannot explain the magnitude of these uncertainties (Boparai et al., 2008).

We note that neither Schmid et al. (2001), Ten Brink et al. (2004), nor Panteliadis et al. (2015) presented a detailed statistical analysis of OC concentrations, and reported up to a factor of two difference between OC measured by different protocols. This is related to the fact that the accurate quantification of OC in atmospheric samples is extremely difficult, due to the potential vaporization and/or adsorption of volatile organic compounds during and after sampling, especially for low filter loadings, and even when attempting to measure these artifacts (discussed below). This difficulty is one of the reasons that emissions testing protocols typically specify the removal of volatile OC by devices such as catalytic strippers, which remove all volatiles (typically at 350 °C) prior to filter collection. Consequently, any carbon measured as OC on downstream filters must either represent pyrolysis products or contamination. Importantly, Corbin et al. (2020) showed that once gas-phase contamination is accounted for, the remaining OC is also measured by in-situ (filter-free) techniques, and is therefore not a sampling or TOA artifact.

Overall, despite a very significant body of work on the fundamentals and statistical uncertainties behind TOA measurements, there have been few studies where the sample was (*i*) non-volatile, (*ii*) taken from the same or identical filter, and (*iii*) of known composition. Here, we present an intercomparison study where the same filters were punched six times for analysis by five different laboratories, after sampling aircraft engine exhaust treated at 350 °C with a catalytic stripper. Identical instrument models and protocols were used by all laboratories. Our study provides a general estimate of the between-laboratory uncertainty of TOA analyses from similar emissions tests, and acts as a lower limit for the TOA reproducibility in atmospheric studies where additional uncertainties are introduced (for example, by way of having multiple sources or differing thermal protocols).

2 Methods

2.1 Experimental protocol

Sampling was performed in accordance with SAE ARP6320A (SAE, 2018), with the experimental setup shown schematically in Figure 1. Emissions were collected from the exhaust of a helicopter turboshaft engine using a single point sample probe, in a subsequent study to MANTRA (reported by (Olfert et al., 2017)), on the same model of engine and in the same facility. The sample stream was mixed with heated dilution air before passing through a catalytic stripper (Catalytic Instruments CS15). The sample flow was split to pass through a pair of Dekati diluters (DI-1000, operated with HEPA filtered compressed air) and a pair of cyclones, each with a 1.0 μm cutoff at 50 LPM, before being directed through a sampling manifold. Samples were distributed from the manifold to a suite of instruments, including other instruments for online mass quantification (e.g.,

as in Corbin et al. (2020)) and TOA. Particles for TOA were collected on quartz filters in stainless steel filter holders. The quartz filters were then sealed in Analyslide Petri dishes (28145-473) and kept at room temperature until analysis.

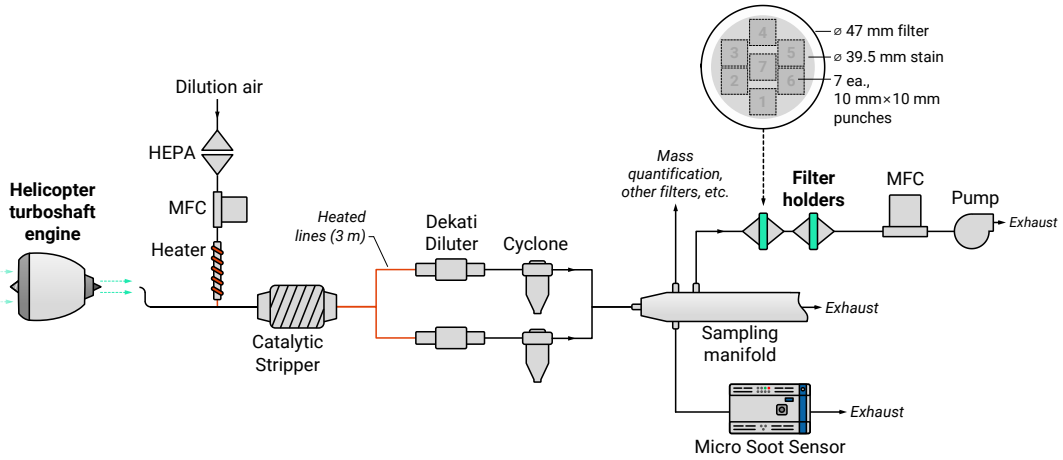

Figure 1. Schematic of the experimental setup where emissions from a helicopter turboshaft engine transit to filter holders containing the quartz filters used for thermal-optical analysis (TOA). Cyclones had a 1.0 µm cutoff at 50 LPM, with the actual sample flow rate for each cyclone being 56 LPM. MFC stands for mass flow controller, while HEPA refers to a HEPA filter. The diluter/cyclone system is consistent with (SAE, 2018). Inset at the top, right depicts the punch positions on the filter. Note that the angular position of the punches on the filter was not constrained.

Samples were composed of 20 filters, with five filters each, at mass concentrations of approximately 50, 100, 250, and 500 µg/m^3 (based on measurements made by a AVL Micro Soot Sensor on a separate parallel line connected to the sampling manifold). To compensate for the different mass concentrations used for loading the filters, the sampling time durations were adjusted such that the mass loadings of nvPM on each filter were similar for all 20 filters. All nvPM samples were collected at high power conditions for the Gnome engine. All the samples loaded at mass concentration of 50, 100 and 250 µg/m^3 of nvPM were obtained with the engine operating at a steady 22,000 rpm. To produce the higher nvPM concentration required for the samples loaded at mass concentration of 500 µg/ m^3, these samples were obtained with the engine operating at a steady 23,000 rpm. Saffaripour et al. (2020) demonstrates that there is no significant change in the morphology of the particles from the same engine model for such modest changes in the rotation speed. Five laboratories compliant with ISO 17025 (demonstrating competence) for TOA were selected for this intercomparison. Each of the laboratories was instructed to take one (or two, in the case of one laboratory) punches from each of the twenty filters. Seven punches were possible on each filter with an allowance of one spare punch per filter in addition to one (or two) per laboratory, arranged in a ring of six with one central punch as shown in the inset to Figure 1. Punch positions on each sample were implicitly randomized by not otherwise providing further instruction to the laboratories. While this introduces a slight risk in the case of uneven filter loading, symmetry in the sampler and random filter orientations would minimize such risks in all but the center punch. The loading of

the filters was visually homogeneous, which further supports this decision. Further, even if there was a bias, for instance due to handling of the filter, it is important to capture this as part of the interlaboratory variability, as this would be representative of real-world measurements. Quartz filters adsorb gas phase organic artifacts, and following the procedure outlined in Corbin et al. (2020), the data from TOA of the quartz filter in the second filter holder shown in Figure 1 is used to correct the OC and TC measurements from the front filter for the gas phase organics that were adsorbed on the front filter.

The determination of TC, OC, and EC was quantified using the same TOA protocol, instrument model, and software settings for all participating laboratories. In all cases, analysis took place on Sunset Laboratory Model 5L analyzers (analogous ILCs have yet to be performed on the other commercially-available instrument, the Sunset Laboratory Model-4 Semi-Continuous OC-EC Field Analyzer). The protocol for aircraft engine emissions, a refinement based upon NIOSH 5040 (SAE, 2018; Lobo et al., 2015a) with the final oxidizing temperature step at a higher temperature of 930 C and a longer duration to ensure complete oxidation of the particles was used to perform the analysis, with the EC/OC split determined automatically by the instrument software (Sunset Laboratory, Calc405). The protocol and sample data are shown in Figure 2.

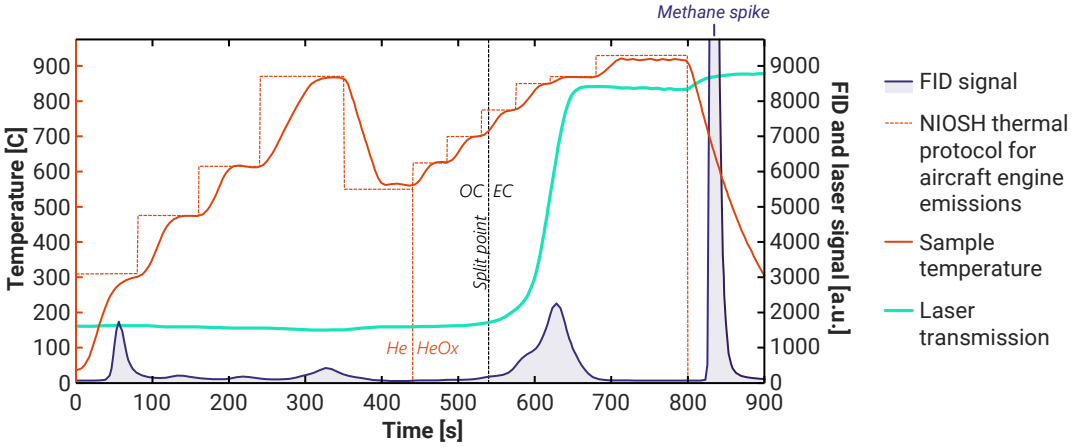

Figure 2. Representative example of a TOA thermogram for nvPM emissions collected from the engine used in this study. Shown are the thermal protocol for aircraft engine emissions (SAE, 2018; Lobo et al., 2015a), the sample temperature, the FID signal, and the laser transmission measurement. HeOx corresponds to 2% oxygen in He. FID is a flame ionization detector. The methane spike corresponds to the introduction of methane that is used for calibration after analysis.

Of the six sets of measurements considered, two belonged to a single laboratory and analyst, and are denoted in subsequent figures and discussion as Laboratory 1A and 1B. The remaining laboratories contributed a single set of data and are numbered in ascending order in terms of the average EC measurement across all the filters. (Two of the laboratories were commercial service providers and did not contribute scientifically to the work.)

2.2 Statistical treatment

Results are analyzed using a hierarchical random effects model. In this framework, measurements, Y_{ijk}, are modeled as a combination of effects:

$$Y_{ijk} = F_j + L_{ij} + E_{ijk} \quad (1)$$

where *i*, *j*, and *k* denote the *i*th laboratory, *j*th filter, and *k*th repeat and the remaining quantities are random variables accounting for various effects or biases. The quantity F_j is a filter-specific effect, accounting for any inconsistency in the loading of the filters. The quantity L_{ij} is the effect or bias for each laboratory and represents a systematic shift in the measurements made by that laboratory for the *j*th filter. The remaining term, E_{ijk}, represents the additional random error in the individual measurements reported by each laboratory, i.e., the mismatch between the expected laboratory bias and the actual measurement. As is a typical convention, uppercase letters are used here to denote a random variable, while lowercase letters are used to denote the realization of the variable. Thus, l_{ij} is the realization of L_{ij} and corresponds to the bias specifically for the *i*th laboratory on the *j*th filter. This quantity will be a positive value if a laboratory has a bias above the filter-specific effect and vice versa. If there is no such bias in any of the laboratories, all of the l_{ij} will be zero. The statistical model is shown schematically in Figure 3 for a single filter.

A given laboratory reported their measurements, denoted as y_{ijk} (i.e., a realization of Y_{ijk}), and an associated laboratory-reported standard deviation, denoted as $\tilde{s}_{ij}$. The uncertainty values reported by the different laboratories are automatically generated by analysis software for the Sunset Laboratory Model 5L, which is a combination of the limit of detection of the instrument (0.2 μg/cm^2) and a percentage based upon prior statistical analysis of duplicate filter punches (±5%). For these reasons, the laboratory-reported uncertainties do not truly represent repeatability, incorporating a wider range of effects. Further, given limitations in the number of punches available for each filter in this study and that the test is destructive, only a single measurement is available for each laboratory-filter combination here. As such, the model is hereafter stated without the subscript *k* dependence:

$$Y_{ij} = F_j + L_{ij} + E_{ij} \ . \quad (2)$$

The use of single measurements also complicates a simple interpretation using ISO 5725-2 (ISO, 2019), as was applied by Panteliadis et al. (2015).

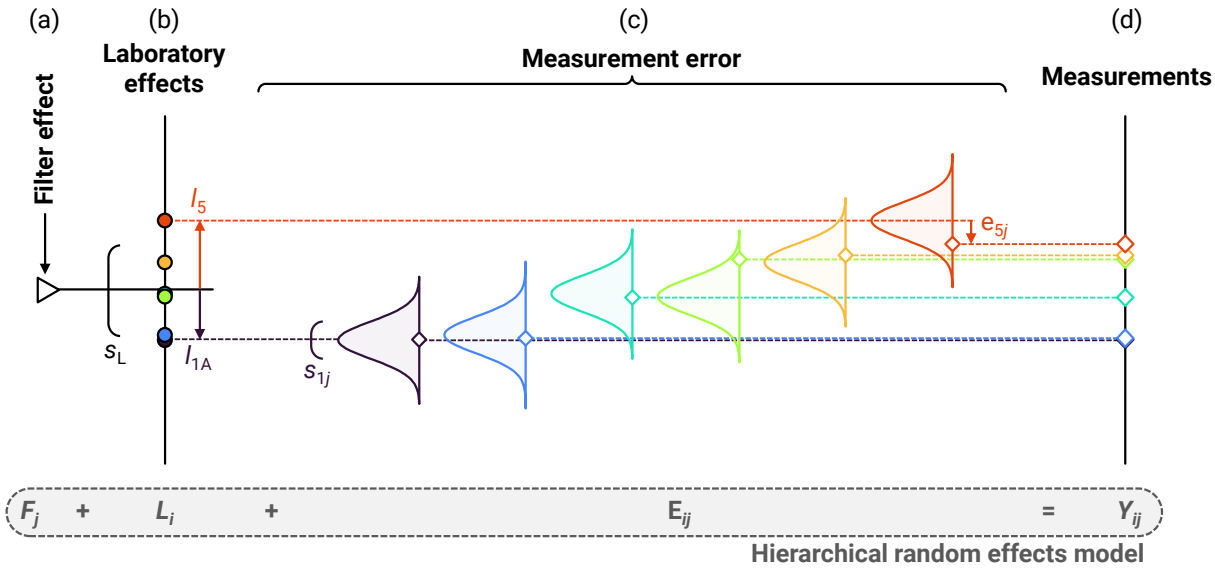

Figure 3. Schematic demonstrating the hierarchical random effects model used in the present analysis. Data corresponds to a single filter computed using the MCMC procedure described in the text.

Rather, a Bayesian method is employed. Realizations of the random effects models are obtained using a Markov Chain Monte Carlo (MCMC) approach, similar to the method presented by Melanson et al. (2018) and in the direction of the method used by Conrad and Johnson (2019). MCMC seeks to find the range of inputs, in this case the magnitude of various effects and uncertainties, that would cause some distribution of the observed measurements. The approach in this work attempts to overcome the limitations in the preceding paragraph by noting that, while there was variability across the filters, the laboratory effects are assumed to be consistent across the different filters (i.e., the laboratory effects are not specific to a given filter), such that the model is further simplified to,

$$Y_{ij} = F_j + L_i + E_{ij} \ , \tag{3}$$

removing the *j*th filter dependence from *L*. Combining this with an assumption of normally-distributed measurement errors, a likelihood can then be stated as

$$Y_{ij} \sim \mathcal{N}\left(f_j + l_i, s_{ij}^2 \right), \tag{4}$$

where s_{ij} are unknown within-laboratory uncertainties. This likelihood relates the various effects to the observed measurements. To restrict the solution space and improve convergence of the sampling algorithm, priors are also applied (encoding approximate information known before the statistical analysis) to these quantities, which are summarized in Table 1. Exponential priors are used for the variances, and Gaussian priors for the effects. These correspond to maximum entropy priors for variables given that variances have a point estimate and effects generally have point estimates and a spread, a priori. The set of s_{ij}, f_j, l_i and $\tilde{s}_L$ are sampled as part of the MCMC procedure, where the latter quantity, $\tilde{s}_L$, used in the prior placed on L_i, and is treated as a nuisance parameter (that is, it is allowed to change and contribute to variability in the sampling but not

explicitly included in the reported output). All of the data is considered in a single MCMC run, rather than separating the data into levels as in ISO 5725-2. To minimize the impact of a large burn-in period for the MCMC, the set of f_j were initiated about the average of y_{ij} for a single filter, while the l_i were initiated about the average y_{ij} over all of the filters after subtracting the average f_j. A total of 25,000 samples were generated, after thinning the MCMC data by a factor of 20 (to avoid short range correlation in the samples) spread across four independent chains. MCMC samples were realized using the Just Another Gibbs Sampler (JAGS) code (Hornik et al., 2003). Visual inspection of the samples indicated that the chains had converged. Further increasing the number of samples did not have an impact on the statistical outcomes. A brief comparison of the MCMC method to an application of the ISO 5725-2 method is presented in the Supplemental Information, with overall reproducibility holding similar values in most instances to those derived with the current method but with a different breakdown of the uncertainties. Note that the repeatability variance, s_r^2, is estimated from this procedure using the average of the within-laboratory variances, s_{ij}^2, roughly consistent with the ISO method, except that it is estimated in the MCMC procedure instead of computed directly from repeat measurements. Table 2 summarizes the different variances used in this work.

Table 1. Table of quantities related to the statistical treatment, including those on which likelihood and priors are directly stated. Note that measurements, y_{ij}, are the input to the MCMC procedure and are thus not sampled. The likelihood corresponds to assumed form for the distribution of a given quantity, with the goal to be reproduced by the MCMC procedure.

Quantities computed or sampled	Likelihood (assumed form)	Priors
l_i, f_j, $\tilde{s}_L$, s_{ij}	$Y_{ij} \sim \mathcal{N}(f_j + l_i, s_{ij}^2)$	$L_i \sim \mathcal{N}(0, \tilde{s}_L^2)$ $F_j \sim \mathcal{N}(m_j, s_{F,j}^2)$ $\tilde{S}_L \sim \text{Exp}(1/\text{MAD}(y_{ij}))^\dagger$ $S \sim \text{Exp}(1/\tilde{s}_{ij})^\dagger$

† Exp(λ) corresponds to an exponential distribution with a rate parameter λ and is chosen as a maximum entropy prior for only having a point estimate. MAD denotes the median absolute deviation.

Table 2. Variances used in the work, their computation, and their corresponding symbol.

Uncertainty component	Estimation procedure	Symbol
Within-laboratory	MCMC, direct sampling	s_{ij}
Laboratory-reported uncertainty	Reported by laboratories (average of variance) †	$\tilde{s}_{ij}$, $(\tilde{s})$ †
Repeatability (approx.)	Average s_{ij}^2 across laboratories	s_r
Between-laboratory	MCMC, standard deviation of l_i	s_L
Prior for between-laboratory	MCMC, direct sampling	$\tilde{s}_L$
Reproducibility	$s_R^2 = s_L^2 + s_r^2$	s_R
Reproducibility for a single laboratory	$s_{R,ij}^2 = s_L^2 + s_{ij}^2$	$s_{R,ij}$
Between-filter	MCMC, standard deviation of f_j	s_F
Total	$s_{TOT}^2 = s_F^2 + s_R^2$	s_{TOT}

†Value in brackets is the average laboratory-reported uncertainty, computed by averaging the variance from each laboratory and filter.

3 Results

3.1 Statistical analysis of EC, OC, and TC

Figure 4 shows a sample of the results for EC for 5 of the 20 filters, corresponding to the 100 µg/m^3 mass concentration case. In Figure 4, laboratories are ordered according to the median EC measured over all 20 filters. This order is roughly respected across all of the filters. Results are generally consistent with the remaining 15 filters, not shown, though filter-to-filter differences in some cases exceeded the range shown in this subset of the filters. Uncertainties did not exhibit a trend with mass concentration or the measured value for EC, OC, or TC. Nominal or consensus values were determined by taking the mean of the filter effects as determined by the MCMC procedure.

Uncertainties, alongside their decomposition into their respective components, are reported in Table 3**Error! Reference source not found.**. While the quantities in Table 2 are expressed as variances, for the subsequent discussion uncertainties are expressed as coefficients of variation (or relative standard errors), where the coefficient of variation is the square root of the corresponding variance divided by the nominal value of EC, OC, and TC, as appropriate. Further, *expanded* uncertainties, defined as an interval about the result of a measurement that may be expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand (JCGM, 2008), are used. The expanded uncertainties are defined by a coverage factor, k, defined as a numerical factor used as a multiplier of the combined standard uncertainty in order to obtain an expanded uncertainty. Following convention, $k = 2$ is used for the expanded uncertainty throughout this work, which is roughly equivalent to a 95% confidence interval.

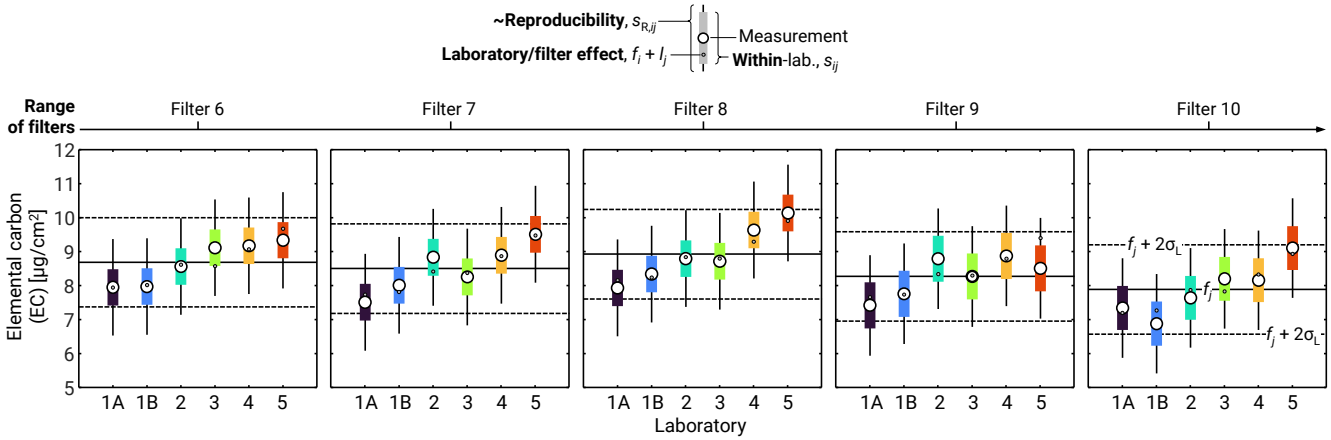

Figure 4. Sample results for the MCMC uncertainty procedure. Open circles correspond to laboratory-reported data, while small filled circles correspond to the combined filter and laboratory effects. Wide, colored bars correspond to within-laboratory uncertainties, s_{ij}, while whiskers correspond to reproducibility, including the average within- and between-laboratory uncertainties combined in quadrature as $s_{R,ij} = (s_{r,ij}^2 + s_L^2)^{1/2}$. Horizontal, solid lines correspond to the realized filter effect, while black dashed lines correspond to expanded reproducibility intervals (k = 2). Results shown are for elemental carbon and the 100 µg/m^3 case. Vertical axes are identical across all of the panels.

Table 3. Breakdown of uncertainties in the TOA measurements, stated as expanded coefficients of variation (or relative standard errors, that is the expanded standard error divided by the nominal value of EC, OC, and TC) for a coverage factor of k = 2. Reproducibility variance, s_R^2, is a combination of the within- and between-laboratory variances (that is, $s_R^2 = s_{ij}^2 + s_L^2$). The total row corresponds to a combination of reproducibility and between-filter uncertainties, again by summing the variances.

		Expanded coefficient of variation (k = 2) [%]*				
Uncertainty component	Symbol	EC	OC	TC	EC/OC	EC/TC
Within-lab. (repeatability)	s	6.8	12.5	4.7	3.8	2.7
Between-lab.	s_L	16.0	7.3	11.7	17.8	6.6
Reproducibility	s_R	16.5	8.0	12.1	18.6	6.8
Between-filter	s_F	21.1	8.6	14.8	23.1	9.1
Total	s_{TOT}	26.8	11.8	19.1	29.7	11.3
Laboratory-reported	$\tilde{s}$	12.4	14.3	13.1	-	-

† Within-filter uncertainties are taken as average values of the variance over all of the filters. This same averaging is then propagated to the total uncertainties. *Coefficient of variation are stated using the nominal value of EC, OC, and TC measurements of 8.1, 4.7, and 12.9 µg/cm^2 and EC/OC and EC/TC ratios of 1.74 and 0.63.

Figure 5 shows a decomposition of the two kinds of within-filter uncertainties (within- and between-laboratory), averaged over all of the filters and laboratories and presented as a proportion of the observed variance. For EC and TC, uncertainties are dominated by between-laboratory contributions, making up roughly 85% of the observed variance in both cases. This is a clear

indication that repeatability is a poor measure of the overall uncertainties in these measurements, and that there is indeed a requirement for larger uncertainties to account for true variability in the TOA method.

Figure 6 complements Figure 5 with a plot of the measurements for each laboratory across the filters and normalized by the filter effects, f_j. Filters are sorted in ascending order of the nominal value for each of EC, OC, and TC, such that the filter order differs between the panels. Trends in Figure 6 for EC and TC were similar, given that EC concentrations were typically double OC concentrations. We note that the uncertainty in TC is smaller than the uncertainty in either OC and EC, because OC and EC are calculated by splitting TC into two parts. Additional uncertainty arises due to this split, which is determined from the laser transmission through the filter, a software algorithm, and the estimation of how much OC had charred to form EC. Results for EC and TC exhibit a consistent bias (or systematic error (JCGM, 2008)) across the different filters; that is, a laboratory that reported a single above-average value generally did so for all of its reported values. For example, Laboratory 5 produced EC and TC values consistently above the other laboratories, while Laboratory 1 (in both the 1A and 1B samples) produced EC and TC values consistently below the other laboratories. Consistency within each laboratory drives smaller within-laboratory contributions (as each laboratory was consistent with itself), with a corresponding expansion of the between-laboratory contributions to account for the remaining spread. The observed biases in the data may also give insight into the physical causes of these uncertainties. For example, minor biases in calibration would lead to the observed systematic errors, while random operator error would not. Other potential sources of error (e.g., in terms of FID response) have been discussed in detail elsewhere (Boparai et al., 2008).

Laboratory-reported uncertainties always exceeded the repeatability computed by the MCMC procedure. The reason for this becomes apparent from the data. Laboratory-reported uncertainties, also shown in Figure 5, appear to accommodate all of the within-laboratory contributions as well as some of the between-laboratory contributions. We denote the discrepancy between the reproducibility and the laboratory-reported uncertainties as *dark* (Thompson and Ellison, 2011) contributions, given that such contributions would be hidden outside of an interlaboratory study and so as to distinguish them from the more precise between-laboratory contributions determined by the MCMC procedure. While the laboratory-reported variances are just slightly below the combined MCMC between-laboratory variance and within-laboratory variance for TC and for OC, the laboratory-reported variance grossly underestimates the combined MCMC variances for EC, requiring a further 93% of the laboratory-reported variance to match the MCMC combined variances.

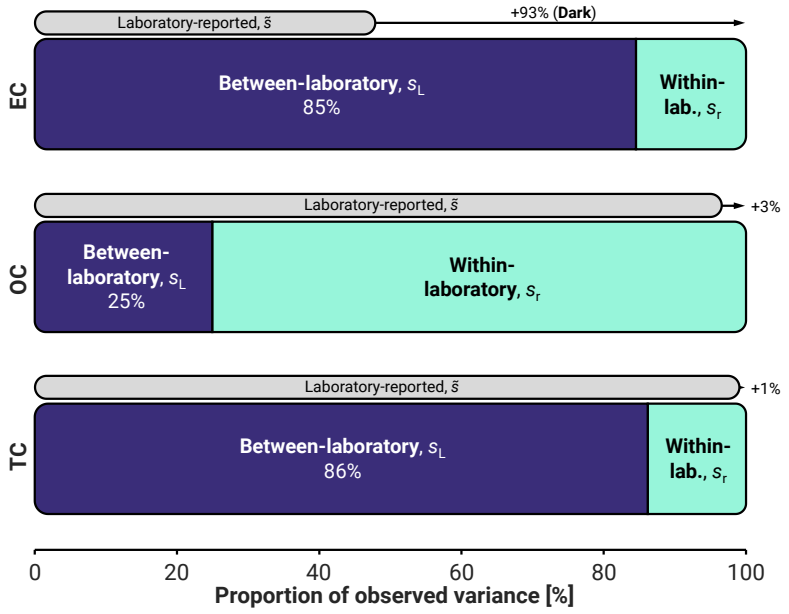

Figure 5. Breakdown of the within-filter variance in the TOA measurements into the between-laboratory variance, s_L^2, and within-laboratory variance, s_r^2, stated as a proportion of the overall within-filter variance, such that all span 0–100 %. Also shown is the corresponding average variance reported by the laboratory, $\tilde{s}^2$. See Table 3 Error! Reference source not found. **for the numerical values of the uncertainties. Percentages in the dark bar correspond to the required increase in uncertainties over the within-laboratory values.**

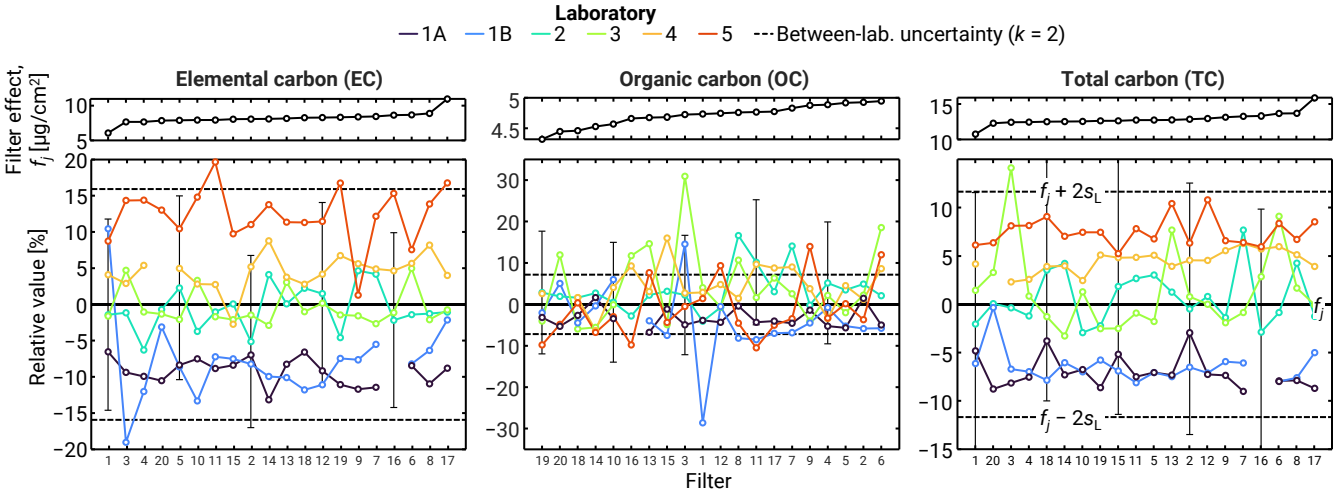

Figure 6. Laboratory measurements across the different filters, showing the associated effects/biases in the data. In each case, filters are re-sorted such that the consensus values for the filters are monotonically increasing for each of the measurands. As such, the order of filters is not the same across the panels. Upper panels show the consensus values for the different filters. Bottom panels show measurements from each laboratory normalized by those consensus values. Breaks in lines correspond to results that were not available. Error bars in the lower panels correspond to expanded (k = 2) uncertainties reported by the laboratories and, while only included for select points, were similar in magnitude across all of the data.

Unlike EC and TC, Figure 5 indicates that OC uncertainties are driven by within-laboratory contributions. This is reflected in the fact that consistent biases were less evident for OC in Figure 6. In other words, repeatability within a laboratory is of the same order-of-magnitude as the overall reproducibility for OC. Again, laboratory-reported uncertainties seem to account for the overall reproducibility in OC, in this case even accommodating the between-filter variability.

For all three of EC, OC, and TC, reproducibility reported here is smaller than that reported by Panteliadis et al. (2015), who provided values equivalent to expanded (k = 2) uncertainties of 40–50 % and 24–30 % for EC and TC, respectively, using the ISO 5725-2 method (ISO, 2019). This is most likely due to a greater variability between the atmospheric samples measured by Panteliadis et al., relative to the single aerosol source and volatile removal device in our study. Our within-laboratory relative standard errors are also smaller than those measured by a single laboratory in Conrad and Johnson (2019). Those authors provided expanded uncertainties of 20 %, 44 %, and 17 % for EC, OC, and TC (from Table 2 in that work) relative to the 6.8 %, 13 %, and 4.7 % observed in the present work. Conrad and Johnson also determined that TC is the most repeatable, while OC is the least repeatable, again consistent with the current observations. The relative breakdown of within- and between-laboratory contributions to the uncertainties for TC here are also similar to the relative contributions observed by Schmid et al. (2001), though the uncertainties here are again smaller (expanded between-laboratory uncertainties of 18 % in Schmid et al. versus 12 % in this work). These collective observations indicate that the current measurements share many of the same trends as previous works but uncertainties in this work are consistently smaller. We hypothesize that our smaller uncertainties are primarily due to the removal of volatile organics with a catalytic stripper, as organics are subject to transformation and mass loss during handling and storage. Since we observed lower uncertainties for TC than other studies, our lower uncertainties for EC and OC cannot be attributed to the EC/OC split alone. This is further supported by the lack of negative correlation between EC and OC (see Sec. 3.2), indicating that the split point was determined reliably. Further, it is likely that the use of a single particle source, a single thermal protocol, a single instrument model, and a common version of software all contribute to the smaller uncertainties observed in this study.

Overall, for our samples, expanded (k = 2) uncertainties for reproducibility in EC, OC, and TC for a given filter are 17 %, 8.0 %, and 12 % of the nominal values, respectively.

3.2 Statistical analysis of EC/OC and EC/TC ratios

Little to no correlation, R, was observed between EC and OC measured by the different laboratories (R_{ECOC} = 0.11), while TC was dominated by, and thus highly correlated with, the EC contributions (R_{ECTC} = 0.94). Combining this with the fact that the measured EC showed consistent biases across the laboratories, it is logical that this is equally reflected in the TC results. OC and TC were poorly correlated (R_{OCTC} = 0.43), given that the TC incorporates but is not dominated by OC. The low level of correlation between EC and OC indicates that the split point is unlikely to be the leading driver of variabilities in the results,

as this would result in a negative correlation between EC and OC, where more of the total carbon is attributed to one of the components at the cost of the other.

Unlike the absolute values for EC, OC, and TC, the EC/OC ratio is expected to be similar across all of the filters, regardless of loading and is a widely used quantity for characterizing the particles emitted. For the EC/OC ratio, simple propagation of errors yields (Sipkens et al., 2023; JCGM, 2008):

$$\mathrm{var}\left(\mathrm{EC}/\mathrm{OC}\right)=\left(\frac{\mathrm{EC}}{\mathrm{OC}}\right)^2\left[\frac{1}{\left(\mathrm{EC}\right)^2}\mathrm{var}\left(\mathrm{EC}\right)+\frac{1}{\left(\mathrm{OC}\right)^2}\mathrm{var}\left(\mathrm{OC}\right)-\frac{2}{\left(\mathrm{EC}\right)\left(\mathrm{OC}\right)}\mathrm{cov}\left(\mathrm{EC},\mathrm{OC}\right)\right]. \tag{5}$$

As noted above, EC and OC are not significantly correlated for these measurements, such that the covariance term can be neglected. Overall, the EC/OC ratio is 1.74 ± 0.52 (k = 2) for the full set of measurements. The expanded uncertainties for the EC/OC ratio, including the contributions from the different sources of variability, are also reported in Table 3. Expanded uncertainties are in terms of the coefficient of variation (or relative standard error), with a coverage factor of k = 2. This produces a relatively uniform estimate of the expanded uncertainty for reproducibility in the EC/OC ratio across all of the measurements, at 19 % of the nominal value. This is comparable to the laboratory-reported uncertainties. Between-filter variability was significant, increasing the expanded uncertainties to 30 % of the nominal value. Note that this is larger than the uncertainties in the individual EC and OC measurements, as it incorporates uncertainties in both EC and OC at the same time.

In this data, Laboratory 5 produced an EC/OC value consistently above the other laboratories, a consequence of measuring higher than average EC in combination with a generally lower than average OC. There was also some trend in EC/OC with mass concentration and sampling time, for all laboratories, indicated by the generally increasing filter number on the x-axis of Figure 7. This results from a similar slight increase in EC and a slight decrease in OC as the sampling period decreases. Since this effect was minor, our preceding discussion summarized the data using the means of EC, OC, TC, and EC/OC ratio.

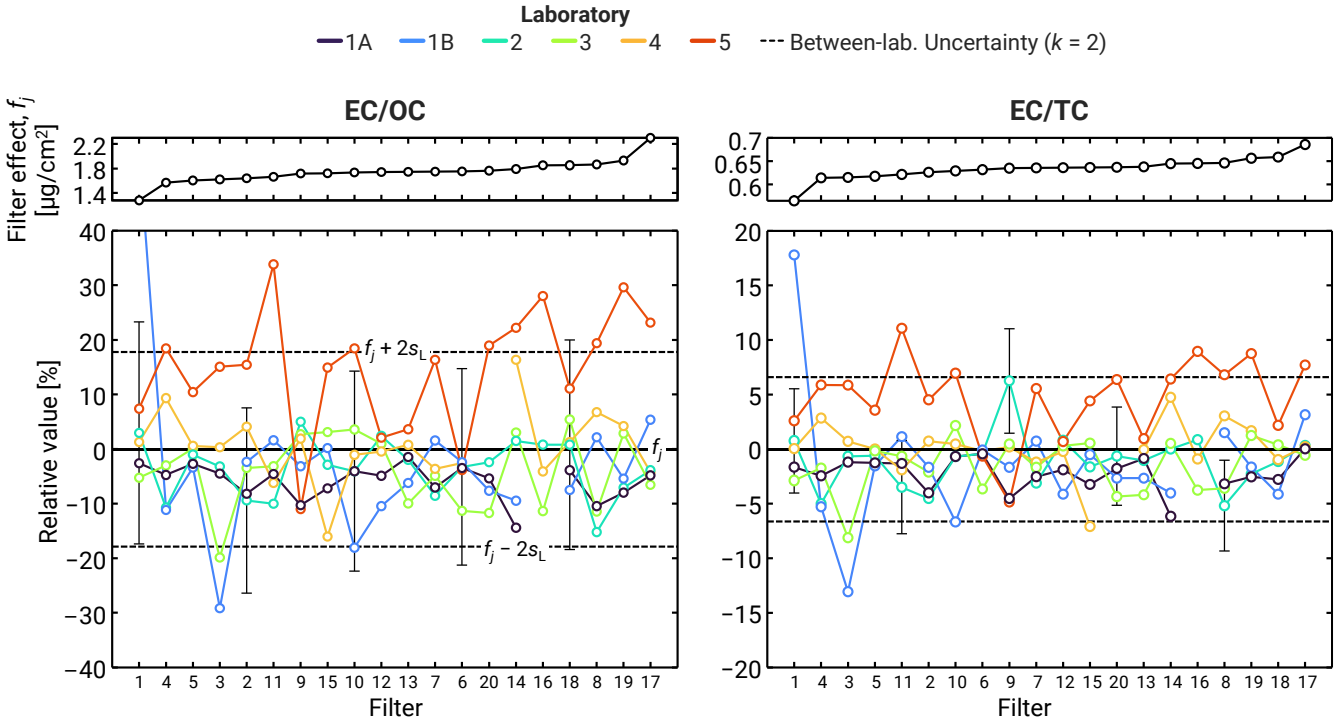

Figure 7. Variability in the EC/OC and EC/TC ratios over the measurements. Uncertainty intervals correspond to the between-laboratory expanded uncertainties at a level of k = 2. Upper panel shows consensus EC/OC values against which the filters are sorted. Bottom panel shows measurements from each laboratory normalized by those consensus values. As with Figure 6, error bars in the lower panel correspond to expanded (k = 2) uncertainties reported by the laboratories and, while only included for select points, were similar across all of the data.

Similar principles can be applied to the EC/TC ratio, where

$$\mathrm{var}\left(\mathrm{EC/TC}\right) = \left(\frac{\mathrm{EC}}{\mathrm{TC}}\right)^2 \left[\frac{1}{\left(\mathrm{EC}\right)^2}\mathrm{var}\left(\mathrm{EC}\right) + \frac{1}{\left(\mathrm{TC}\right)^2}\mathrm{var}\left(\mathrm{TC}\right) - \frac{2}{\left(\mathrm{EC}\right)\left(\mathrm{TC}\right)}\mathrm{cov}\left(\mathrm{EC,TC}\right)\right]. \quad (6)$$

This time, the covariance term will necessarily be significant, as TC is largely composed of EC. The present analysis uses a correlation of R_{ECTC} = 0.94, as noted above, and rephrases Eq. (6) in terms of the correlation:

$$\mathrm{var}\left(\mathrm{EC/TC}\right) = \left(\frac{\mathrm{EC}}{\mathrm{TC}}\right)^2 \left[\frac{1}{\left(\mathrm{EC}\right)^2}\mathrm{var}\left(\mathrm{EC}\right) + \frac{1}{\left(\mathrm{TC}\right)^2}\mathrm{var}\left(\mathrm{TC}\right) - \frac{2R_{\mathrm{ECTC}}\left[\mathrm{var}\left(\mathrm{EC}\right)\mathrm{var}\left(\mathrm{TC}\right)\right]^{1/2}}{\left(\mathrm{EC}\right)\left(\mathrm{TC}\right)}\right]. \quad (7)$$

The resultant uncertainties are quite small, due to the high degree of correlation, amounting to expanded (k = 2) uncertainties of 6.8 % within a given filter and 11 % when adding between-filter variability. A majority of this variability stems from between-laboratory variability, consistent with the observations for EC and TC.

4 Conclusions

This work investigated the between-laboratory uncertainties associated with thermal-optical analysis (TOA) applied to aircraft engine particulate emissions. These conditions represent optimal samples for TOA, in that they are primarily composed of combustion particles that are stripped of their volatile components. Uncertainties are not expected to be related to the split point, due to a lack of correlation between EC and OC (where a reduction in OC results in an increase in EC).

EC and TC measurements are highly correlated with the laboratory (i.e., reflected by a fixed bias/systematic error), with some laboratories measuring consistently above or below the average. These laboratory biases suggest a potential link to laboratory-specific calibration that affects the EC (and, by extension, the TC) measurement. This results in EC and TC uncertainties being dominated by between-laboratory contributions (~ 85 % of the variance). Further, replicates, that is repeat measurements by a single laboratory, are unlikely to properly capture these uncertainties. For data sets comparable to ours (i.e., PM dominated by soot, treated to remove volatile organic carbon, and containing negligible elemental impurities), net expanded (k = 2) relative standard errors of 17 % for EC, 8.0 % for OC, 12 % for TC, 19 % for the EC/OC ratio, and 6.8 % for EC/TC ratio are expected and account for reproducibility. These values correspond to a lower limit on the uncertainties for EC, OC, and TC, given the use of a single particle source, a single thermal protocol, a single instrument model, and a catalytic stripper to remove volatile organics. This expanded uncertainty should be used in future measurements with this test method. For application to the calibration of instruments to measure the mass concentration of nvPM emissions from aircraft engines, the expanded uncertainties for EC, 17 %, and for the EC/TC ratio, 6.8 %, are the most significant quantities.

The authors see limited scope for reducing the uncertainties of TOA through refinements to the calibration procedures and quality controls. While promising alternatives to TOA are emerging for calibration of instruments, such as the CPMA-Electrometer Reference Mass Standard (CERMS) (Titosky et al., 2019; Corbin et al., 2020), the corresponding interlaboratory variability of these alternatives have yet to be validated and should be a topic of future work.

The treatment in this work does not directly address the interpretation of OC and EC concentrations reported by TOA, nor does this work evaluate the accuracy of the TOA TC concentration (e.g., by indicating traceability to an SI unit). Rather, this work addresses metrological reproducibility of the TOA method by comparing results from the same sample, measured by different laboratories and analysts.

Data availability

A simplified form of the raw data, including the laboratory-reported measurements and uncertainties, has been included in the Supplemental Information as CSV files. One file is provided for each of EC, OC, and TC. The first columns contain information about the laboratory and whether or not the row corresponds to a measurement (y) or laboratory-reported uncertainty (std). Each column contains the results for a different filter.

Author contribution

Timothy A. Sipkens: Data Curation, Formal analysis, Methodology, Software, Writing - Original draft, Writing - Review & Editing, Visualization. **Joel C. Corbin**: Investigation, Writing - Review & Editing. **Brett Smith**: Investigation, Writing - Review & Editing. **Stephanie Gagné**: Investigation, Writing - Review & Editing. **Prem Lobo**: Supervision, Project administration, Writing - Review & Editing. **B.T. Brem**: Investigation, Writing - Review & Editing. **A. Fischer**: Investigation, Writing - Review & Editing. **M. Johnson**: Investigation, Writing - Review & Editing. **Gregory J. Smallwood**: Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing - Review & Editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We are grateful to the Rolls Royce Team for their support in obtaining these samples. This work was supported by Transport Canada, Civil Aviation, Environmental Protection and Standards (A1-023090). Filter analysis was supported in part by the Swiss Federal Office of Civil Aviation Project "EMPAIREX" SFLV- 2015-113. We also thank Jason Olfert for useful discussions of the statistical modeling presented in this work.

References

Bae, M.-S., Schauer, J. J., and Turner, J. R.: Estimation of the monthly average ratios of organic mass to organic carbon for fine particulate matter at an urban site, Aerosol Science and Technology, 40, 1123-1139, 2006.

Bautista, A. T., Pabroa, P. C. B., Santos, F. L., Quirit, L. L., Asis, J. L. B., Dy, M. A. K., and Martinez, J. P. G.: Intercomparison between NIOSH, IMPROVE_A, and EUSAAR_2 protocols: Finding an optimal thermal–optical protocol for Philippines OC/EC samples, Atmos. Pollut. Res., 6, 334–342, 10.5094/apr.2015.037, 2015.

Birch, M. and Cary, R.: Elemental carbon-based method for monitoring occupational exposures to particulate diesel exhaust, Aerosol Science and Technology, 25, 221-241, 1996.

Boparai, P., Lee, J., and Bond, T. C.: Revisiting thermal-optical analyses of carbonaceous aerosol using a physical model, Aerosol Science and Technology, 42, 930-948, 2008.

Brown, R. J. C., Beccaceci, S., Butterfield, D. M., Quincey, P. G., Harris, P. M., Maggos, T., Panteliadis, P., John, A., Jedynska, A., Kuhlbusch, T. A. J., Putaud, J.-P., and Karanasiou, A.: 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, Environ. Sci. Processes Impacts, 19, 1249–1259, 10.1039/c7em00261k, 2017.

Cavalli, F., Viana, M., Yttri, K. E., Genberg, J., and Putaud, J. P.: Toward a standardised thermal-optical protocol for measuring atmospheric organic and elemental carbon: the {EUSAAR} protocol, Atmos. Meas. Tech., 3, 79-89, 10.5194/amt-3-79-2010, 2010.

Cheng, Y., Duan, F.-k., He, K.-b., Du, Z.-y., Zheng, M., and Ma, Y.-l.: Intercomparison of thermal-optical method with different temperature protocols: Implications from source samples and solvent extraction, Atmos. Environ., 61, 453–462, 10.1016/j.atmosenv.2012.07.066, 2012.

Cheng, Y., He, K. B., Duan, F. K., Zheng, M., Ma, Y. L., Tan, J. H., and Du, Z. Y.: Improved measurement of carbonaceous aerosol: evaluation of the sampling artifacts and inter-comparison of the thermal-optical analysis methods, Atmos. Chem. Phys., 10, 8533–8548, 10.5194/acp-10-8533-2010, 2010.

Conrad, B. M. and Johnson, M. R.: Split point analysis and uncertainty quantification of thermal-optical organic/elemental carbon measurements, JoVE (Journal of Visualized Experiments), e59742, 2019.

Corbin, J. C., Moallemi, A., Liu, F., Gagné, S., Olfert, J. S., Smallwood, G. J., and Lobo, P.: Closure between particulate matter concentrations measured ex situ by thermal–optical analysis and in situ by the CPMA–electrometer reference mass system, Aerosol Science and Technology, 54, 1293-1309, 2020.

Figueiredo, J. L., Pereira, M., Freitas, M., and Orfao, J.: Modification of the surface chemistry of activated carbons, carbon, 37, 1379-1389, 1999.

Giannoni, M., Calzolai, G., Chiari, M., Cincinelli, A., Lucarelli, F., Martellini, T., and Nava, S.: A comparison between thermal-optical transmittance elemental carbon measured by different protocols in PM2.5 samples, Sci. Total Environ., 571, 195–205, 10.1016/j.scitotenv.2016.07.128, 2016.

Hornik, K., Leisch, F., Zeileis, A., and Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, Proceedings of DSC,

ISO: ISO 5725-2:2019: 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, International Standards Organization, 2019.

JCGM: Evaluation of measurement data — Guide to the expression of uncertainty in measurement, BIPMJCGM 100:2008, 2008.

Lack, D. A., Moosmüller, H., McMeeking, G. R., Chakrabarty, R. K., and Baumgardner, D.: Characterizing elemental, equivalent black, and refractory black carbon aerosol particles: a review of techniques, their limitations and uncertainties, Anal Bioanal Chem, 406, 99-122, 2014.

Lobo, P., Hagen, D. E., Whitefield, P. D., and Raper, D.: PM emissions measurements of in-service commercial aircraft engines during the Delta-Atlanta Hartsfield Study, Atmospheric Environment, 104, 237-245, 10.1016/j.atmosenv.2015.01.020, 2015a.

Lobo, P., Durdina, L., Brem, B. T., Crayford, A. P., Johnson, M. P., Smallwood, G. J., Siegerist, F., Williams, P. I., Black, E. A., and Llamedo, A.: Comparison of standardized sampling and measurement reference systems for aircraft engine non-volatile particulate matter emissions, Journal of Aerosol Science, 145, 105557, 2020.

Lobo, P., Durdina, L., Smallwood, G. J., Rindlisbacher, T., Siegerist, F., Black, E. A., Yu, Z., Mensah, A. A., Hagen, D. E., and Miake-Lye, R. C.: Measurement of aircraft engine non-volatile PM emissions: Results of the aviation-particle regulatory instrumentation demonstration experiment (A-PRIDE) 4 campaign, Aerosol Science and Technology, 49, 472-484, 2015b.

Melanson, J. E., Thibeault, M.-P., Stocks, B. B., Leek, D. M., McRae, G., and Meija, J.: Purity assignment for peptide certified reference materials by combining qNMR and LC-MS/MS amino acid analysis results: application to angiotensin II, Anal Bioanal Chem, 410, 6719-6731, 2018.

Olfert, J. S., Dickau, M., Momenimovahed, A., Saffaripour, M., Thomson, K., Smallwood, G., Stettler, M. E. J., Boies, A., Sevcenco, Y., Crayford, A., and Johnson, M.: Effective density and volatility of particles sampled from a helicopter gas turbine engine, Aerosol Science and Technology, 51, 704-714, 10.1080/02786826.2017.1292346, 2017.

Panteliadis, P., Hafkenscheid, T., Cary, B., Diapouli, E., Fischer, A., Favez, O., Quincey, P., Viana, M., Hitzenberger, R., Vecchi, R., Saraga, D., Sciare, J., Jaffrezo, J. L., John, A., Schwarz, J., Giannoni, M., Novak, J., Karanasiou, A., Fermo, P., and Maenhaut, W.: ECOC comparison exercise with identical thermal protocols after temperature offset correction – instrument diagnostics by in-depth evaluation of operational parameters, Atmos. Meas. Tech., 8, 779–792, 2015.

Righi, M., Hendricks, J., and Beer, C. G.: Exploring the uncertainties in the aviation soot–cirrus effect, Atmospheric Chemistry and Physics, 21, 17267-17289, 2021.

SAE: Procedure for the Continuous Sampling and Measurement of Non-Volatile Particulate Matter Emissions from Aircraft Turbine Engines (SAE ARP6320), Warrendale, PA, USA, 10.4271/ARP6320., 2018.

Saffaripour, M., Thomson, K. A., Smallwood, G. J., and Lobo, P.: A review on the morphological properties of non-volatile particulate matter emissions from aircraft turbine engines, Journal of Aerosol Science, 139, 105467, 2020.

Schauer, J. J., Mader, B., Deminter, J., Heidemann, G., Bae, M., Seinfeld, J. H., Flagan, R., Cary, R., Smith, D., and Huebert, B.: ACE-Asia intercomparison of a thermal-optical method for the determination of particle-phase organic and elemental carbon, Environmental science & technology, 37, 993-1001, 2003.

Schmid, H., Laskus, L., Abraham, H. J., Baltensperger, U., Lavanchy, V., Bizjak, M., Burba, P., Cachier, H., Crow, D., and Chow, J.: Results of the "carbon conference" international aerosol carbon round robin test stage I, Atmospheric environment, 35, 2111-2121, 2001.

Singh, M. and Vander Wal, R. L.: The role of fuel chemistry in dictating nanostructure evolution of soot toward source identification, Aerosol Science and Technology, 54, 66-78, 2020.

Sipkens, T. A., Corbin, J. C., Grauer, S. J., and Smallwood, G. J.: Tutorial: Guide to error propagation for particle counting measurements, Journal of Aerosol Science, 106091, 2023.

Ten Brink, H., Maenhaut, W., Hitzenberger, R., Gnauk, T., Spindler, G., Even, A., Chi, X., Bauer, H., Puxbaum, H., and Putaud, J.-P.: INTERCOMP2000: the comparability of methods in use in Europe for measuring the carbon content of aerosol, Atmospheric Environment, 38, 6507-6519, 2004.

Thompson, M. and Ellison, S. L.: Dark uncertainty, Accreditation and Quality Assurance, 16, 483-487, 2011.

Titosky, J., Momenimovahed, A., Corbin, J., Thomson, K., Smallwood, G., and Olfert, J. S.: Repeatability and intermediate precision of a mass concentration calibration system, Aerosol Science and Technology, 53, 701-711, 2019.

Turpin, B. J. and Lim, H.-J.: Species contributions to PM2. 5 mass concentrations: Revisiting common assumptions for estimating organic mass, Aerosol Science & Technology, 35, 602-610, 2001.

Watson, J. G., Chow, J. C., and Chen, L.-W. A.: Summary of organic and elemental carbon/black carbon analysis methods and intercomparisons, Aerosol and Air Quality Research, 5, 65-102, 2005.

Wu, C., Huang, X. H. H., Ng, W. M., Griffith, S. M., and Yu, J. Z.: Inter-comparison of NIOSH and IMPROVE protocols for OC and EC determination, Atmos. Meas. Tech., 9, 4547–4560, 10.5194/amt-9-4547-2016, 2016.

495